Forward genetics analysis in *Physcomitrella patens* identifies a novel ABA regulator.

The *ABA NON-RESPONSIVE* locus encoding the trimodular MAP3 kinase *PpANR* unique to basal land plants.

Sean Ross Stevenson

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of Leeds
Centre for Plant Sciences

September, 2015

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© 2015 The University of Leeds and Sean Ross Stevenson

Acknowledgements

I would like to thank my supervisor Dr. Andrew Cuming for his guidance and enthusiasm throughout this project and for making the project a pleasure to work and learn on. I am very grateful to him, and those that helped him out, for all the work carried out to get the project set up including growing up all the segregants and extracting DNA. Special thanks also to Dr. Yasuko Kamisugi who has always been on hand for advice and help in the lab and for her expertise in moss tissue culture work.

The protein work presented in this thesis would not have been possible without the help and guidance of Prof. Thomas Edwards who accepted me as one of his own for my brief but fruitful time in his lab. Thanks to his other lab members, Jenny, Hannah, Rachel and Phil for being on hand for all my trivial questions and problems. A big thank you must also go to Dr. Chi Trinh who gave so much of his time and expertise carrying out the crystallography. I now understand some of this ‘dark art’ because of his thorough explanations and diagrams. A thank you also to Dr. James Ault and Dr. Iain Mansfield who helped with the design and interpretation of the mass spectrometry experiments.

Abstract

Land plants evolved from a group of aquatic algae known as charophytes and molecular evidence suggests that they were pre-adaptated to life on land. Early land plants necessarily required mechanisms to survive dehydration and the plant hormone abscisic acid (ABA) is known to play a vital role in this conferring desiccation tolerance in all land plants. The basal non-vascular land plants, made up of the liverworts, hornworts and mosses, rely heavily on ABA-mediated vegetative dehydration/desiccation tolerance (D/DT) as they lack anatomical adaptations to retain water and this trait remains a developmentally regulated feature of the angiosperm seed. ABA non-responsive (*anr*) mutants were identified in the model bryophyte *Physcomitrella patens* and genotyping of segregating populations enabled the mapping of the *PpANR* locus. This locus encodes a trimodular MAP3 kinase comprising an N-terminal PAS domain, a central EDR domain and a C-terminal MAPKKK-like domain ("PEK" structure). Mutants of PpANR showed dehydration hypersensitivity and an inability to respond to exogenous ABA demonstrating the vital role of PpANR in the ABA-dependent osmotic stress responses. RNA-seq analysis of wild-type and *anr* mutant plants also revealed potential components of the wild-type ABA-dependent osmotic stress response not yet characterised in bryophytes. Phylogenetic analysis reveals PpANR to be part of a basal plant-specific subfamily of MAP3Ks closely related and possibly ancestral to the "EK" structured negative ethylene regulator CTR1 and the "PK" structured positive ABA regulators Raf10/11. The establishment of these subfamilies in the charophytes suggests them as potential vital components of ancestral water stress responses. The PAS domain likely originated from a domain swap from histidine kinases in the green algae and the solving of the crystal structure of this domain reveals it to form a homodimer with each domain taking the canonical PAS fold structure. A model is suggested for a key role of PpANR in an ancestral ABA-dependent osmotic stress signalling pathway.

Abbreviations

TF	Transcription factor
ABI	ABA insensitive
bZIP	basic leucine zipper domain
NGS	next generation sequencing
RNAseq	RNA sequencing
ABA1	ABA deficient 1
ZEP	zeaxanthin epoxidase
NCED	9-cis-epoxycarotenoid dioxygenase
GO	gene ontology
LEA	late embryogenesis abundant
SNP	single nucleotide polymorphisms
ANR	ABA non-responsive
PTC	premature stop codon
PAS	Per-ARNT-Sim
bHLH	basic helix-loop-helix
FMN	flavin mononucleotide
FAD	flavin adenine dinucleotide
MS-ES	mass spectrometry electrospray
MS-MS	tandem mass spectrometry
RMSD	root-mean-square-deviation
D/DT	dehydration/desiccation tolerance
ROS	reactive oxygen species
CE	coupling element
AP2/ERF	apetala 2/ethylene response factor
MAPK	Mitogen-activated protein kinases
EDR	Enhanced Disease Resistance
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
OST1-like	Open stomata 1-like
PP2C	Protein phosphatase 2C superfamily
SnRK2	sucrose non-fermenting-1-related protein kinase 2
PYL	Pyrabactin resistant-like
FDR	False discovery rate
KO	Knockout
WT	Wild-type

Table of contents

Table of contents 1
List of Tables 6
List of Figures 7
Chapter 1 – General Introduction **10**
1.1 The evolution of land plants - charophytes and preadaptation. 10
1.2 What is ABA and what is its role? 16
1.2.1 ABA and desiccation tolerance 16
1.3 ABA biosynthesis 20
1.4 ABA-dependent signalling 22
1.5 Crosstalk with other ABA-dependent and independent stress response pathways 27
1.6 MAP kinase signalling 28
1.7 PAS domains 32
1.8 *Physcomitrella patens* life cycle 35
1.9 Summary of project and aims 35
Chapter 2 – Forward genetics in *Physcomitrella patens* identifies the novel ABA regulator PpANR **39**
2.1 Introduction 39
2.1.1 Why forward genetics? 39
2.1.2 SNP marker assisted mapping 40
2.2 Materials and Methods 41
2.2.1 PCR and cloning protocols 41
2.2.2 Creation and screening of *anr* mutants by UV mutagenesis 42
2.2.3 Genotyping of the *anr* loci 42
2.2.4 Identification of the causal mutation 43
2.2.5 Creation of *Ppanr* mutant lines 44
2.2.6 Growth assays 44
2.3 Results 45
2.3.1 Identification of ABA non-responsive mutants displaying strong ABA resistance. 45
2.3.2 Mapping of causal loci. 47
2.3.3 Identification of the *PpANR* locus. 49
2.3.4 Knockout of the *PpANR* locus induces ABA non-responsiveness 54
2.3.5 An otherwise isogenic line containing the *anr4* PTC, *Ppanr4PTC*, also shows the *anr* phenotype 54

2.3.6 *Ppanr* mutants fail to show ABA-inducible gene expression 57
2.3.7 *Ppanr* mutants fail to respond to ABA pre-treatment 57
2.3.8 *PpanrKO* shows slower, less vigorous growth 60
2.3.9 *PpanrKO* protonemal cells are longer than wild-type 60
2.4 Conclusions 64
2.4.1 Forward genetics now a tool in bryophyte genomics 64
2.4.2 The *anr* mutants 65
2.4.3 Getting lucky? 65
2.4.4 Effects of *Ppanr* mutants 66
2.4.5 Interactions of ABA and ethylene 67
Chapter 3 – RNA-seq analysis of the PpANR null mutant, *PpanrKO*, reveals a vital role in ABA-dependent osmotic stress signalling. **69**
3.1 Introduction 69
3.2 Methods 71
3.2.1 Treatments and RNA extractions 71
3.2.2 RNA-seq processing and mapping 71
3.2.3 Differential gene expression analysis 71
3.2.4 Correlation analysis 72
3.2.5 GO term analysis 72
3.3 Results 73
3.3.1 RNA library quality 73
3.3.2 Library relationships 75
3.3.3 *PpANR* regulation and effect on core components 79
3.3.4 Effect on ABA biosynthesis 81
3.3.5 GO term analysis of significantly regulated transcripts 84
3.3.6 ABA/stress regulated genes and co-regulation – cluster analysis 87
3.4 Discussion 98
3.4.1 Validity of differential gene expression findings 98
3.4.2 The wild-type ABA and stress responses 99
3.4.3 The role of endogenous ABA and signalling reinforcement 100
3.4.4 Role of PpANR 101
3.4.5 Relationship of PpANR and the core signalling pathway 102
Chapter 4 – Phylogenetic analysis of PpANR and related MAP3 kinases reveals insights into the molecular mechanisms in the conquest of land. **105**
4.1 Introduction 105
4.1.1 PAS containing kinases 105
4.1.2 “EK” kinases 107

4.1.3 The serine/threonine kinases 107
4.2 Materials and methods 108
4.2.1 PAS, EDR and kinase trees 108
4.2.2 Global *Physcomitrella* MAP kinase analysis 109
4.3 Results 110
4.3.1 Raf-like serine/threonine kinase domain analysis 110
4.3.2 EDR domain analysis 115
4.3.3 PAS domain 120
4.3.4 Comparison with original B1, B2 and B3 phylogenetic definitions 125
4.3.5 Wider analysis of MAPK-related kinases in *Physcomitrella* 127
4.3.5.1 Raf-like MAP3 kinases 127
4.3.5.2 MEKK-like MAP3 kinases and related MAP kinases 130
4.3.5.3 MAP kinases 133
4.4 Discussion and conclusions 133
4.4.1 The MAP kinase protein families in *Physcomitrella* 133
4.4.2 Phylogenetic robustness 134
4.4.3 PpANR's place among the Raf-like MAP3 kinases 134
4.4.4 A histidine kinase origin for the PpANR PAS domain? 136
4.4.5 PpANR dual function a special case? 137
Chapter 5 – Structural analysis of the N-terminal PAS domain in PpANR 140
5.1 Introduction 140
5.1.1 The PAS domain superfamily 140
5.1.2 bHLH-PAS family 142
5.1.2.1 The AHR-ARNT interaction 143
5.1.2.2 The HIFα-ARNT interaction 144
5.1.3 PAS role in PpANR and aims of study 145
5.2 Methods and Materials 146
5.2.1 Bacterial strains and vector 146
5.2.2 Protein Expression 146
5.2.3 OD600 measurement 146
5.2.4 Nickel affinity chromatography 146
5.2.5 Size exclusion chromatography 147
5.2.6 Protein concentration 147
5.2.7 SDS PAGE gel electrophoresis 147
5.2.8 ABA and FAD binding 148
5.2.9 Circular Dichroism (CD) Spectroscopy 148
5.2.10 Crystallography 148

5.2.10.1 Factorials 148
5.2.10.2 Optimisation 149
5.2.10.3 Cryo cooling 149
5.2.10.4 PpANR PAS domain crystallography 149
5.2.11 Metabolite binding screen 149
5.2.12 Structural Relatedness Analysis 150
5.3 Results 150
5.3.1 PAS expression construct prediction and synthesis 150
5.3.2 Expression of recombinant PAS domain 153
5.3.3 Metabolite interaction screen with PAS 158
5.3.4 Circular Dichroism and ABA binding 164
5.3.5 FAD binding 164
5.3.6 Crystal screens and optimisation 169
5.3.7 X-ray crystallography 169
5.3.7.1 General structure features 169
5.3.7.2 Dimerisation interfaces 174
5.3.7.3 Highly conserved residues 176
5.3.8 Structural Comparisons 176
5.3.8.1 Structures with the highest pairwise similarity to the PpANR PAS domain 181
5.3.9 PAS domain ligand binding mechanisms 182
5.3.9.1 The PpANR PAS domain binding cavity 182
5.3.9.2 FAD binding 185
5.3.9.3 Haem binding 187
5.3.9.4 Indole binding 187
5.4 Discussion 187
5.4.1 Possibility of PpANR PAS ligand binding 190
5.4.2 Dimerisation 192
5.4.3 General PAS domain mechanisms of signal transduction 193
Chapter 6 – General discussion **196**
6.1 Summary 196
6.2 A model for the evolution of ABA signalling in dehydration and desiccation tolerance. 197
6.3 A role for PpANR and the ancestral DDT pathway 198
6.3.1 A histidine kinase origin of D/DT signalling? 198
6.3.2 The PAS domain 199
6.4 PpANR dual signalling role: an exception or the rule? 200

6.4.1 How might a dual role be accomplished? 201
6.5 Overview of the ABA-dependent signalling mediated by PpANR 203
6.5.1 The activator module 203
6.5.2 The 'brake' module 205
6.5.3 The problem: How does ABA alone exert regulation? 205
6.6 Evidence for the ancestral D/DT osmotic stress response 206
6.7 Main mechanisms requiring answers 207
References 208
Supplementary data 224
Appendices 224

List of Tables

Table 2.1. List of primers used in the identification, creation and genotyping of *Ppanr* mutants.
Table 3.1. Summary of RNA-seq mapping statistics.
Table 3.2. Fold changes of established and putative 'core' ABA signalling components in *Physcomitrella*.
Table 3.3. Fold changes of established and putative ABA metabolism and catabolism components in *Physcomitrella*.
Table 3.4. Summary of gene expression clusters from wild-type tissue.
Table 3.5. Identification of putative signalling components strongly implicated in ABA-dependent osmotic stress signalling.
Table 5.1. Mixtures used for SDS-PAGE gels.
Table 5.2. Data collection, processing and refinement statistics for the PpANR PAS domain.
Table 5.3. Key statistics and features of PAS domains showing the highest pairwise structural similarity to the PpANR PAS domain.

List of Figures

Figure 1.1. Consensus phylogenetic relationship between green algae and land plants.
Figure 1.2. ABA biosynthetic pathway in *Arabidopsis*.
Figure 1.3. Simplified 'core' ABA signalling pathway.
Figure 1.4. Trans-acting factors and *cis*-acting elements regulating ABA-dependent and ABA-independent stress signalling.
Figure 1.5. Some of the known plant MAP kinase cascades in stress signalling.
Figure 1.6. Diversity of PAS domain structures.
Figure 1.7. Life cycle of the moss *Physcomitrella patens*.
Figure 2.1. Wild type response of developing protonemal tissue to ABA.
Figure 2.2. Identification and phenotype of UV mutagenized *anr* mutants.
Figure 2.3. Identification of overlapping regions on chromosome 12 co-segregating with the phenotype in both A4V and A3V segregants.
Figure 2.4. Confirmation and visualization of mapped regions for *anr* mutants.
Figure 2.5. The *anr4* mutation induces a premature stop codon upstream of the kinase domain in PpANR.
Figure 2.6. Constructs used in the creation of *Ppanr* mutants.
Figure 2.7. Genotyping of *Ppanr* mutant transformants.
Figure 2.8. Growth assay of *PpanrKO* mutant confirms *anr* phenotype.
Figure 2.9. Screening and isolation of *Ppanr4PTC* transformants showing the *anr* phenotype.
Figure 2.10. *Ppanr* mutants fail to induce ABA and stress-responsive gene expression.
Figure 2.11. *Ppanr* mutants are not desiccation tolerant following ABA pre-treatment.
Figure 2.12. *PpanrKO* plants have distinct filament architecture as compared to wild-type.
Figure 2.13. *PpanrKO* plants show distinct and reduced growth on well provisioned, thick media
Figure 2.14. *PpanrKO* chloronemal cells are significantly longer than wild-type.
Figure 3.1. RNA-seq analysis revealed a significant set of ABA and stress regulated genes had their regulation abolished in *PpanrKO*.
Figure 3.2. Relationships in gene expression between libraries.
Figure 3.3. Strong overlap in gene expression between up-regulated genes in wild-type tissue is not mirrored in down-regulated genes.
Figure 3.4. GO terms featuring in the up-regulated genes in wild-type tissue.
Figure 3.5. GO terms featuring in the down-regulated genes in wild-type tissue.
Figure 3.6. Five clusters containing genes strongly up-regulated by all treatments in wild-type tissue.
Figure 3.7. Four clusters containing genes up-regulated by ABA and dehydration in the wild-type.
Figure 3.8. Four clusters containing genes significantly up-regulated by mannitol in the wild-type.
Figure 3.9. Three clusters containing genes strongly down-regulated by all treatments in wild-type tissue.
Figure 3.10. Four clusters containing genes strongly down-regulated by mannitol treatment in wild-type tissue.
Figure 3.11. Six clusters containing genes strongly down-regulated by dehydration in the wild-type.
Figure 4.1. Alignment of kinase domains from MAP3Ks used in phylogenetic analysis.
Figure 4.2. Protein trees of the kinase domains from PpANR and related MAP3 kinases.
Figure 4.3. Alignment of EDR domains from MAP3Ks used in phylogenetic analysis.
Figure 4.4. Protein trees of the EDR domains from PpANR and other EDR-containing kinase including related MAP3 kinases.
Figure 4.5. Alignment of PAS domain protein sequences used in phylogenetic analyses.

Figure 4.6. Protein trees of the PAS domains from PpANR, related MAP3Ks and other PAS-containing histidine kinases.
Figure 4.7. Re-analysis of original B group subfamily relationships confirms these were incorrect.
Figure 4.8. Large scale analysis of all *Physcomitrella* kinases showing significant similarity to *Arabidopsis* MAP3 kinases.
Figure 4.9. Subtree of the Raf-like MAP3Ks.
Figure 4.10. Subtree of the canonical MAP kinases.
Figure 4.11. Tree showing the MAPKs and related kinases.
Figure 5.1. Schematic representation of a PAS fold structure as defined by Hefti *et al*., 2004.
Figure 5.2. I-TASSER structure prediction of PAS fold and flanking 15a.a.
Figure 5.3. I-TASSER structure prediction of PAS fold sequence used in expression construct.
Figure 5.4. SDS-PAGE gel of initial PpANR PAS protein expression conditions.
Figure 5.5. SDS-PAGE gel of fractions from first nickel column affinity purification of PpANR PAS protein.
Figure 5.6. Fractions from second nickel column affinity purification yielded high purify of the PpANR PAS protein.
Figure 5.7. Size exclusion chromatography shows single peak of the PpANR PAS protein.
Figure 5.8. Mass-spectrometry of ABA standards.
Figure 5.9. Metabolite profile of protonemal tissue from two extraction methods.
Figure 5.10. Enrichment of metabolites by protein affinity.
Figure 5.11. METLIN (Scripps) database search results with query of 148.04m/z and tolerance of +-30ppm.
Figure 5.12. METLIN (Scripps) database search results with query of 112.98m/z and tolerance of +-300ppm.
Figure 5.13. Purified PpANR PAS protein is structured.
Figure 5.14. Thermal melt analysis reveals no PpANR PAS::ABA interaction.
Figure 5.15. Size exclusion chromatography of the PpANR PAS protein incubated with FAD.
Figure 5.16. Spectrophotometry reveals no PpANR PAS::FAD interaction.
Figure 5.17. Crystal screening identifies favourable conditions.
Figure 5.18. Crystallography reveals the PpANR PAS domain to be a homodimer with canonical PAS fold structure.
Figure 5.19. Asymmetries are found between the two chains.
Figure 5.20. Analysis of dimerisation interfaces.
Figure 5.21. Analysis of highly conserved and structurally important residues.
Figure 5.22. Structural similarity analysis of PAS structures reveals clusters of phylogenetically and functionally related structures.
Figure 5.23. PAS structures as monomers or dimers have distinct structures.
Figure 5.24. Both chains of the PpANR PAS domain crystal structure have internal cavities.
Figure 5.25. Ligand binding leads to a measurable increase in cavity size in HIF2α.
Figure 5.26. Conserved mechanisms of FAD binding not found in the PpANR PAS domain.
Figure 5.27. Conserved mechanisms of haem binding not found in the PpANR PAS domain.
Figure 5.28. Structural analysis of indole binding in bacterial HTR-like PAS domain (3BWL).
Figure 6.1. Overview of ABA-dependent osmotic stress signalling model incorporating PpANR.

CHAPTER 1

Chapter 1 – General Introduction

This thesis presents the work involved in the identification and characterisation of a novel MAP3 kinase ABA regulator in *Physcomitrella patens*. This introduction seeks to place this gene and the study organism in evolutionary context by describing some of the latest theories on the origins of land plants. As an ABA regulator, an overview of ABA signalling is given with an emphasis on ancestral mechanisms. The wider stress signalling pathways are also introduced, which include the MAP kinases. These constitute ancient and vital eukaryote signal transduction pathways and give context to the gene family to which PpANR belongs. Finally, some background on the N-terminal PAS domain is given in an effort to understand what possible roles those found in the MAP3 kinases may have.

1.1 The evolution of land plants - charophytes and preadaptation.

Using the model bryophyte *Physcomitrella patens* allows comparative genomics to infer evolutionary processes that allowed land plants to proliferate and become so successful. Bryophytes represent the earliest branching land plant lineages and so they represent the closest extant plants to the ancestors of all land plants and the limited fossil record of the earliest land plants supports this (*e.g.* Kenrick and Crane, 1997). The conquest of land by plants was a significant event in the Earth's natural history, with a transformative effect on the terrestrial habitat. Consequently, understanding the evolutionary processes that allowed the genesis of the first land plants is necessary to fully understand how the barren land became colonised and transformed. It is widely accepted (see Leliaert *et al.,* 2012 for review) that land plants evolved from aquatic green algae, being most closely related to a group known as the charophytes - which together with the embryophytes (all land plants) make up the streptophytes. Recent advances in molecular phylogenetics coupled with ever-improving genetic resources for basal plants, such as the onekp consortium (www.onekp.com), have revealed some important evolutionary relationships among the charophytes and basal land plants (Figure 1.1). Two major areas of contention have only recently begun to reach consensus: First, the order and relationship of the bryophyte lineages to each other and the rest of the land plants and the timeline of terrestrial plant evolution and second, which of the charophyte groups is sister to all land plants. Liverworts have frequently been suggested as the earliest branching lineage and sister to all other land plants (*e.g.* Qiu *et al.,* 1998 based on mitochondrial sequences), whereas others have suggested the three bryophyte lineages, liverworts, hornworts and mosses, were monophyletic (*e.g.* Nishiyama *et al.,* 2004 based on chloroplast sequences), with conflicts arising between these views depending on the different methods and taxa used. A recent large-scale analysis has still failed to resolve these conflicts but

provides good support for mosses and liverworts forming a single clade with the position of hornworts uncertain but the bryophytes may be monophyletic: the hornworts have been largely ignored, and are in need of greater sampling (Wickett *et al.,* 2014). Various studies are in greater agreement over the position of the algal group, the zygnematophyceae, as sister to embryophytes (Wickett *et al.,* 2014; Wodniok *et al.,* 2011; Timme *et al.,* 2012; Civan *et al.,* 2014; Turmel *et al.,* 2006). This is a group of anatomically simple single celled or filamentous green algae that include two distinct orders: the desmidiales and the zygnematales which are characterised by their ability to reproduce by conjugation which is possibly a secondary adaptation following the loss of motile sperm (Wodnoik *et al.,* 2011). There are far more anatomically complex charophytes such as the charophyceae which have a multicellular body plan superficially similar to land plants or the coleochaetophyceae which can form multicellular discoids. It is possible that these lineages developed these complexities independently or it has been suggested that morphological complexity was a trait of this group (including zygnematophyceae and embryophytes) but that the zygnematophyceae had a secondary loss of complexity (Wodnoik *et al.,* 2011).

Since the earliest land plants lacked anatomical complexity, the key to the conquest of the land was likely dependent on molecular adaptations. These would have primarily required changes and developments in stress signalling to enable plants to withstand the variable environment on land – leading to exposure to dehydration, increased radiation and greater and more rapid fluctuations in temperature. Land plants share a number of features with charophytes that show their close relatedness. These include multicellularity (as seen in the charophyceae, coleochaetophyceae and the filamentous zygnematales), plasmodesmata, asymmetrical cell division, zygote retention, branching and cellulosic cell walls among others (reviewed in Graham *et al.,* 2000). In order for these aquatic (likely fresh water) land plant ancestors to make the move onto land they necessarily required some preadaption to the stresses encountered - variable water availability as well as greater temperature fluctuation, elevated UV radiation and growth in the non-buoyant air – as these would have been too severe to allow gradual changes. A plausible model therefore may involve these aquatic ancestors inhabiting the waters edge where they would periodically be exposed to these stresses as water levels rose and fell. The development of molecular adaptations to the stresses and problems described would have enabled survival of ever increasing periods of exposure offering a selective advantage and eventual complete terrestrialisation. Preadaptation is therefore importantly different to so called ‘exaptations’ (discussed below) in which selective pressures for traits existed before the genesis of land plants, *senso stricto*. This model is conjecture as direct evidence is unlikely to be come by,

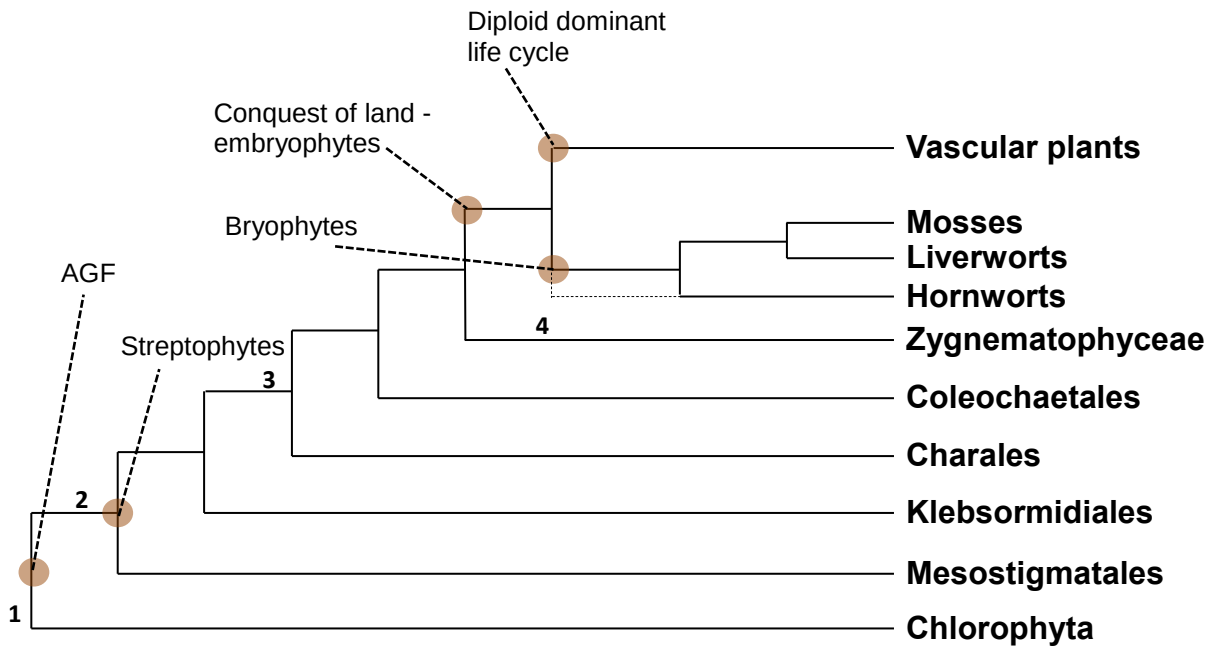


Figure 1.1. Consensus phylogenetic relationship between green algae and land plants. Not all charophyte lineages are shown but zygnematophyceae are recovered as sister to land plants. Bryophytes are recovered as monophyletic and are the earliest branching land plants anatomically similar to the earliest land plants, however the position of hornworts is the least certain possibly representing a distinct land plant lineage (represented by dashed line). Origin of key groupings are shown including the ancestral green flagellate (AGF). Streptophytes include the charophyte green algae and land plants. The development of vascularization and the ever increasing anatomical sophistication is associated with the switch to a diploid dominant life cycle. 1 – Evidence for presence of all 5 major plant hormones (ABA, ethylene, auxin, gibberellin and cytokinin) in micro algal. 2 – origins of charophytes accompanied by number of traits shared with land plants. 3 – evolution of morphological complexity. 4 – Secondary loss of morphological complexity in these unicellular or filamentous algae and loss of motile gametes resulted in conjugation. Relationships and key events based on Wickett *et al*., 2014 and references therein.

however, we can see what molecular components and signalling mechanisms are shared among the streptophytes and infer how these were brought about in the 'aquatic' plants.

Plant hormones play vital roles in regulating plant growth and development as well as coordinating responses to external stimuli including stresses. The presence of the stress hormone abscisic acid (ABA) in almost all algae analysed (Hirsch *et al.,* 1989) is suggestive of a biosynthetic capacity, however the mechanisms remains unclear. Many studies have shown the presence of phytohormones in other photosynthetic lineages with ABA being detected in cyanobacteria, diatoms and brown and red algae (see Lu and Xu, 2015 for review). Importantly, these authors also highlight the presence of all five major phytohormones in green algae: auxin, gibberellin, cytokinin, ABA and ethylene. Of these 5 hormones, ABA and ethylene are generally considered water stress hormones although they are implicated in other stresses such as pathogen and cold stress. Some very interesting recent studies have begun to show an important role for these hormones in algal stress responses. The terrestrial algae *Klebsormidium crenulatum* has been shown to be desiccation tolerant and shows 'land plant-like' molecular responses to desiccation (Holzinger *et al.,* 2014). These authors found transcripts encoding the desiccation-protectant LEA proteins - intrinsically unstructured hydrophilic proteins hypothesised to have multiple roles in desiccation tolerance and which are discussed in detail below - as well as those involved in potential osmolyte (raffinose) and reactive oxidative species (ROS) metabolism accompanied by elevated glutathione levels. This landmark study revealed that the potential for dehydration tolerance may be present in many charophytes; a trait found in many non-plant organisms too such as nematodes where LEA proteins are also found. The important question still remains however: how did these ancient tolerance mechanisms become tightly regulated, largely by ABA, to allow the land plant ancestor to colonise the land? As more genetic resources become available for charophytes, the extent to which components of both the stress responses and their regulation are conserved will become apparent. The study on *Klebsormidium*, for example, was able to detect likely homologues for all components of the 'core' ABA signalling pathway except for the PYR/PYL/RCAR (pyrabactin resistance/pyrabactin resistance-like/regulatory component of ABA receptor) receptors (Holzinger *et al.,* 2014). Interestingly, searches of the onekp algal sequence data reveals a likely PYR/PYL/RCAR homologue with partial conservation of key ABA binding residues in the zygnematophyte *Mesotaenium endlicherianum* (but not in any other lineages) suggesting that these may represent one of the latest adaptations to occur prior to the conquest of land; this requires a more in depth analysis however. In more distant algal species possible, although often

contradictory, roles for ABA have been found: for example in conferring oxidative stress tolerance to *Chlamydomonas reinhardtii* by boosting the activity of catalases (Yoshida *et al.,* 2003). It is thought that ABA produced by many other basal algae and even other organisms such as lichens is secreted with possible signalling roles but with little effect on the actual organism itself (see Hartung, 2010 for review).

It has long been hypothesised that many hormone biosynthetic pathway components were established in charophytes prior to the conquest of land (*e.g.* Kenrick and Crane, 1997) although no conclusive findings have been found. It is worth mentioning here that caution must be maintained in drawing conclusions about putative functional homologues. For example, one recent article claimed clear evidence for auxin biosynthetic origins in the charophytes based on relatively low protein homology scores and questionable phylogenetics (Wang *et al.,* 2014a) that have been rightly questioned by others (Huang *et al.,* 2014a). Interestingly, it is thought that while charophytes contain auxin, the ability to regulate cellular concentrations through auxin conjugation was a land plant specific adaptation (Cooke *et al.,* 2002; Sztein *et al.,* 1995) which would have enabled its function as a crucial growth and developmental regulator. The development of charophyte model systems enabling reverse and forward genetic analysis will be vital in addressing these important and interesting questions.

A recent study taking advantage of the available genetic tools in a developing charophyte model showed that the actual mechanisms behind hormone signalling, in this case ethylene, were functionally conserved (Ju *et al.,* 2015). These authors showed that the zygnematophyte *Spirogyra pratensis* not only showed a typical ethylene response – marked by cell elongation – but that this was likely mediated by an interaction between a homologue of the negative regulator constitutive triple response 1 (CTR1) and an ethylene receptor 1 (ETR)-like ethylene receptor as found in *Arabidopsis* (Kieber *et al.,* 1993). They also found that transgenic expression of the *Spirogyra* receptor was able to partially recover an *Arabidopsis etr* mutant and that overexpression of the SpEIN3 (ethylene insensitive 3) transcription factor (TF) in the *Arabidopsis* wild-type background conferred ethylene hypersensitivity. Another important mechanism of ethylene signalling involves the cleavage and translocation of the EIN2 (ethylene insensitive 2) protein from the endoplasmic reticulum (ER) to the nucleus and this too was observed in *Spirogyra*. In short, an almost complete canonical ‘higher’ plant phytohormone signalling pathway has been shown to be functionally conserved in a charophyte; and importantly a zygnematohpyte.

Outside of hormone-mediated stress signalling, other biochemical changes were required to help protect cells from elevated UV and the lack of buoyancy. Important

compounds include phenolics which have a range of protective roles including from UV radiation (Aigner *et al.,* 2013) and are found in charophytes, albeit with a lower complexity, where they may have represented ‘exaptations’ (pre-adaptations which were likely important in other functions) (Graham *et al.,* 1996). Such compounds likely had their complexities increased, through growing enzyme families, in stages through land plant evolution but the basic building blocks in the biosynthetic pathways were already present in the charophytes. For example it is known that lignin-like polymers are present in bryophytes with the key evolutionary changes allowing lignin synthesis a later adaptation crucial in vascular plants (*e.g.* Ligrone *et al.,* 2008). Charophytes have also long been recognised as having resistant cell walls, as measured by resistance to high temperature acid treatment, likely important in preadaptation (Kroken *et al.,* 1996). Such ‘tough’ cell walls were likely important in the extra structural support required on land. Many charophytes were found to induce formation of these resistant cell walls upon desiccation stress (Kroken *et al.,* 1996) which is very much in support of fresh water algae having adaptations to periodic dehydration through variable water availability, which became constitutive in the first land plants. A recent study taking advantage of increased genetic resources for green algae has proposed that cell wall polysaccharide complexity greatly increased in charophytes prior to the conquest of land (Mikkelsen *et al.,* 2014). These authors found the number of glycosyltransferase families and their members were comparable between charophytes and land plants showing a dramatic increase as compared to prasinophyte algae; a group of early diverging green algae in the chlorophyta. These enzymes play important roles in polysaccharide biosynthesis with greater family numbers and family sizes associated with the ability to synthesize a greater complexity of products. One of the key roles this possible exaptation had was to facilitate subsequent tissue specialisation such as the vascular tissues so important for the proliferation of vascular plants through subsequent selective pressures.

The work presented in this thesis supports the general view that charophytes had many molecular features shared with land plants that made the move onto land possible. The subsequent genome duplications and gene family expansions seen in many land plant lineages (*e.g.* Lang *et al.,* 2008; Rensing *et al.,* 2007) were likely secondary processes that built up the complexity of these functions through gene duplication and subfunctionalisation enabling further anatomic and molecular adaptations that allowed plants to dominate the land (*e.g.* Ohno, 1970; Zhang, 2003; Flagel and Wendel, 2009). It is also interesting to see what components of stress signalling are lost during plant evolution, another important theme of this thesis, such as the loss of many Ca^{2+} signalling mechanisms (Wheeler and Brownlee, 2008) as well as the loss of a number

of histidine kinase families present in the green algae but absent in land plants (Chapter 4).

1.2 What is ABA and what is its role?

We have seen that while the presence of the carotenoid-derived phytohormone ABA in basal plants has been confirmed, its role is far more enigmatic. In land plants, ABA is well characterised with a role in a range of processes. ABA is one of 5 central plant hormones (phytohormones), the others being auxin, gibberellin, cytokinin and the gaseous ethylene, which together enable the regulation of almost all developmental processes as well as coordinating responses to external stimuli. Other phytohormones such as jasmonic acid (abiotic stress responses), salicylic acid (pathogen responses) and the brassinosteroids (growth regulators) were later land plant developments. While ABA is principally involved in abiotic stress responses, and as such is referred to as a stress hormone, it also has roles in regulating development. In the angiosperms for example, ABA plays a central role in seed maturation and in inhibiting germination. This process additionally involves ABA conferring desiccation tolerance to seeds which enables future generations to remain quiescent until favourable conditions are encountered; sometimes remaining dormant and viable for thousands of years (Sallon *et al.,* 2008). In angiosperms, and other stomata bearing vascular plants, the role of ABA has been very well characterised in ABA-dependent stomatal closure in times of reduced water availability, where it is synthesised in the vascular tissue (Koiwai *et al.,* 2004) before acting on stomata to reduce transpiration, thereby minimising water loss long before the shoots experience any osmotic stress (Blackman and Davies, 1985). ABA has been most extensively studied in angiosperms, notably the model *Arabidopsis thaliana*. These relatively sophisticated plants have many anatomical adaptations that enable the retention of water, meaning that vegetative desiccation tolerance has largely been lost, with the exception of a few so-called resurrection plants. These adaptations include extensive root systems to collect water and complex vascularisation allowing its transport throughout a well waterproofed plant reinforced by lignin and coated by cuticular wax and suberin. As a consequence, the role of ABA in these plants has become highly restricted to the seeds, bud formation and the stomata; the latter simply needing to be closed to minimise the loss of water when required. If simply closing stomata is not sufficient, ABA also has a ‘higher’ plant-specific role in leaf abscission, the shedding of leaves in certain species to save water, the process in which it was originally discovered and from which it takes its name.

1.2.1 ABA and desiccation tolerance

The basal land plants lack anatomical sophistication. The earliest branching extant land plants are the bryophytes among which a large number of species are poikilohydric,

meaning they are unable to retain water and readily equilibrate with surrounding water potentials. As a consequence, tolerance to dehydration at a minimum, and often to desiccation, is required throughout the plant to ensure its survival under conditions of water limitation. Dehydration and desiccation tolerance can be thought of as essentially the same process operating on a continuum (differing in the amount of water loss tolerated) but essentially involving the maintenance of ionic balances, minimising and repairing cellular damage and regulating growth and development (Ingram and Bartels, 1996). The process can be broken into three basic, but overlapping, parts (Bewley, 1979), the balance of which can affect how tolerant the organism can be: **i)** the constitutive level of protection, **ii)** the speed and strength of stress-induced responses and **iii)** the readiness for repair mechanisms upon rehydration. During periods of dehydration, plants become largely quiescent and the level of tolerance amounts to the speed with which normal metabolic processes can be resumed, which is a function of the balance and strength of these parts. In contrast, dehydration intolerant organisms show irreversible damage because they lack some of these processes. Desiccation associated damage can result in irreversible protein misfolding, overproduction of reactive oxygen species and the physical damage to cell walls and membranes, including thylakoid membranes. All the proteins and transcripts that have roles in ameliorating one or more of these processes are heavily reliant on ABA-dependent gene expression mediated by key transcription factors (TFs). Many bryophytes that display desiccation tolerance have a high level of constitutive protection, such as *Tortula ruralis* (also known as *Syntrichia ruralis*) (Oliver *et al.,* 2004), so that less damage is incurred both during the desiccating and rehydrating stages and normal metabolic processes resume relatively quickly. While ABA acts at specific sites in angiosperms, it is required to act in all cells in the bryophytes to confer vegetative dehydration/desiccation tolerance (D/DT).

One of the important stages of tolerance is osmoregulation in which the plant tries to reduce the loss of cellular water to the drier environment by lowering the water potential in the cells. A short term advantage can be conferred if the cells are able to maintain normal metabolic activity in the presence of osmotically active solutes. Potentially, many solutes could accumulate to regulate water potential, but in practice the accumulation of, for example charged ions could be toxic, and so there is a preference for the accumulation of components less likely to exert detrimental effects on the cellular environment. These are referred to as ‘compatible osmolytes’ and include a range of mostly disaccharide sugars, specific amino acids and their derivatives such as glycine-betaine (see Krasensky and Jonak, 2012 for recent review). In *Physcomitrella* the dominant osmoltye is sucrose (Oldenhof *et al.,* 2006) which has

long been known to have protective properties on enzymes (Hanafusa, 1969) and is commonly used by plants. Other sugars are also used as compatible osmolytes such as trehalose (*e.g.* Bianchi *et al.,* 1993) which too can stabilise proteins and membranes (Paul *et al.,* 2008). Another group of sugars often implicated is the raffinose family which accumulate in a wide range of plant seeds (Peterbauer and Richter, 2001) but which are also specifically implicated in reactive oxygen species (ROS) scavenging (Nishizawa *et al.,* 2008). The amino acid proline is another major osmolyte active in a wide range of organisms including outside the plants which is able, through overexpression of its key biosynthetic enzyme, to confer osmotic stress tolerance to bacteria (Mahan and Csonka, 1983). Again, like the raffinoses, proline is also found to have a role in ROS scavenging (reviewed by Verbruggen and Hermans, 2008). As we can see from these few examples, while the prolonging of normal cellular turgor and water availability is clearly advantageous, many osmolytes have secondary roles with protective functions. Some sugars are able to confer protection by forming a glass-like state at sufficient concentrations which helps stabilise proteins and membranes in the absence of water (Alpert and Oliver, 2002) and have been theorised to effectively replace water by the interaction with polar (hydrophilic) regions of proteins (Crowe and Crowe, 1984).

Some osmolytes have been implicated in ROS scavenging but this aspect of the stress responses in dehydration tolerance involves many more components. The onset of dehydration has severe consequences for photosynthesis, respiration and indeed most metabolic processes. A reduction in water availability reduces the electron source in photosynthesis and the process is largely stopped, however excitation energy from light is still being absorbed by chlorophyll. This can lead to photoreduction of oxygen to produce superoxide, which can secondarily form hydrogen peroxide (H_2O_2) and hydroxyl radicals, and singlet oxygen (reviewed by Smirnoff, 1993); ROS are capable of reacting with and often damaging (or inactivating in the case of enzymes) almost all proteins. Responses to elevated levels of ROS involve both enzymatic and molecular antioxidants. Some of the major enzymes include the peroxidase family and catalases which both decompose H_2O_2. Being enzymes, these can only function while cellular conditions allow and so other molecular ROS scavengers are required when dehydration becomes more severe and these include glutathione, ascorbate and phenolics among others (reviewed by Apel and Hirt, 2004). These become oxidised following ROS scavenging and require reductase enzymes to recycle them as ROS scavenging sources.

As stated, dehydration tolerance is achieved by the efficient resumption of normal metabolic processes following rehydration. There are a number of stress-associated

processes specific to this stage which includes membrane leakage which is a side effect of phase transitions in the membranes due to repartitioning and restructuring. In this process, the uptake of water and movement of various amphiphilic molecules from the membranes back to the cytoplasm coupled with membrane phase changes cause holes to form and leakage to occur (Hoekstra *et al.,* 1999). In some desiccation tolerant bryophytes such as *T. ruralis*, an extensive molecular response is induced during rehydration resulting in the synthesis of 'rehydrins' (reviewed in Oliver *et al.,* 2005); not to be confused as a gene family but is a collective term. An interesting feature of this process in *T. ruralis* is that it sequesters a pool of rehydrin mRNA in messenger ribonucleoprotein (mRNP) particles during dehydration which are then released and translated following rehydration to confer a rapid response (Wood and Oliver, 1999). Among these rehydrins are LEA and ELIP proteins which are common molecular components of stress responses (Oliver *et al.,* 2004). Rehydration is thought to enable photosynthesis to restart rapidly, however ROS scavengers are required to deal with its resumption if this occurs in the light. The ELIPs are thought to have photoprotective roles being related to the chlorophyll a/b-binding (CAB) proteins and their likely function is to dissipate excess light-induced energy minimising the production of ROS (reviewed by Adamska, 1997).

The other major class of rehydrins are the LEA proteins which are also expressed during the dehydration process and are ubiquitous components of dehydration tolerance (Cuming, 1999). These are a diverse superfamily of proteins that are characterised by being highly hydrophilic and rich in glycine but with no consensus on their mode of action. Due to their perceived plasticity they may well have a wide range of functions. These include functions in protein stabilisation preventing aggregation and misfolding, as 'space-fillers' preventing cellular collapse, as antioxidants, in ion sequestration, in membrane stabilisation and even as hydration buffers given their hydrophilicity (reviewed by Tunnacliffe and Wise, 2007). It has also been proposed that they may be primarily involved in conferring protection during the rehydration stage (Oliver *et al.,* 2004) through "managing" the rate of water uptake, although it is likely that they have roles during dehydration too; as the dehydrin (group 2 LEAs) name suggests and most of the listed possible roles are likely to have relevance during both stages. Despite the functional uncertainty, these are often used as molecular markers for the ABA-dependent dehydration response (*e.g.* Kamisugi and Cuming, 2005).

Vegetative D/DT would have been vital for the earliest land plants, which evolved from fresh water algae and were anatomically similar to bryophytes (*e.g.* Edwards *et al.,* 1995; Kenrick *et al.,* 2012). Whilst today's freshwater algae appear to lack the genetic capacity for ABA-signalling as understood in land plants, including the bryophytes, the

evolution of these mechanisms in conferring D/DT were likely to have been a crucial component in the conquest of land; a process that changed the planet (Kenrick *et al.,* 2012). By studying ABA signalling in bryophytes we look through a window back in time to understand both how this transition was accomplished and to better understand how and why it functions as it does in the vascular plants, setting the stage for the incredible proliferation of these 'higher' plants we see dominating the landscape today. The study of the ancestral role of ABA in conferring D/DT is the focus of this thesis.

1.3 ABA biosynthesis

While ABA is detected in charophytes, and indeed many eukaryote lineages, its biosynthetic pathway has not been established in these algae. However, the mechanisms behind it are a crucial first step in ABA signalling. ABA signalling relies not only on its biosynthesis but its translocation in vascular plants to the sites of action. Here is a clear difference between bryophytes and vascular plants where biosynthesis itself is clearly a vital part of the signalling process as each cell necessarily must respond to changes in endogenous levels given the lack of any means of long distance transportation; although plasmodesmata may facilitate cell-to-cell transport. In vascular plants, ABA biosynthesis has been shown to be strongly up-regulated in response to abiotic stresses such as dehydration (reviewed by Xiong and Zhu, 2003) besides during developmentally controlled phases such as seed development. It is derived from carotenoid precursors in the chloroplast before becoming active in the cytoplasm or being actively transported, and involves 5 steps encoded by the following enzyme families (and cofactor) in order (Figure 1.2): zeaxanthin epoxidases (ZEP), 9-cis-epoxycarotenoid dioxygenases (NCED), short-chain dehydrogenase/reductases (SDR), aldehyde oxidases (AAO) and the molybdenum cofactor (MoCo) which requires its own biosynthesis. All the components have been characterised in *Arabidopsis* and recently the PpZEP enzyme has been shown to be crucial for ABA biosynthesis in the moss *Physcomitrella patens* (Takezawa *et al.,* 2015) which along with the presence of putative homologues for all remaining steps (except SDRs) strongly suggests this pathway is largely conserved among land plants (reviewed by Sakata *et al.,* 2014 and Takezawa *et al.,* 2011; see Chapter 3). An important part of ABA biosynthetic up-regulation is the dehydration stress-mediated up-regulation of genes encoding the key enzymes, with NCEDs being the primary target as they are thought to catalyse the rate-limiting step (Qin and Zeevaart, 1999; Audran *et al.,* 1998). The application of exogenous ABA itself is also known to increase endogenous ABA biosynthesis (reviewed by Xiong and Zhu, 2003) allowing positive feedback. Vascular plants require not only transportation of ABA but its active efflux into the vascular system and uptake into target cells which has been shown to rely on ATP-binding cassette (ABC)

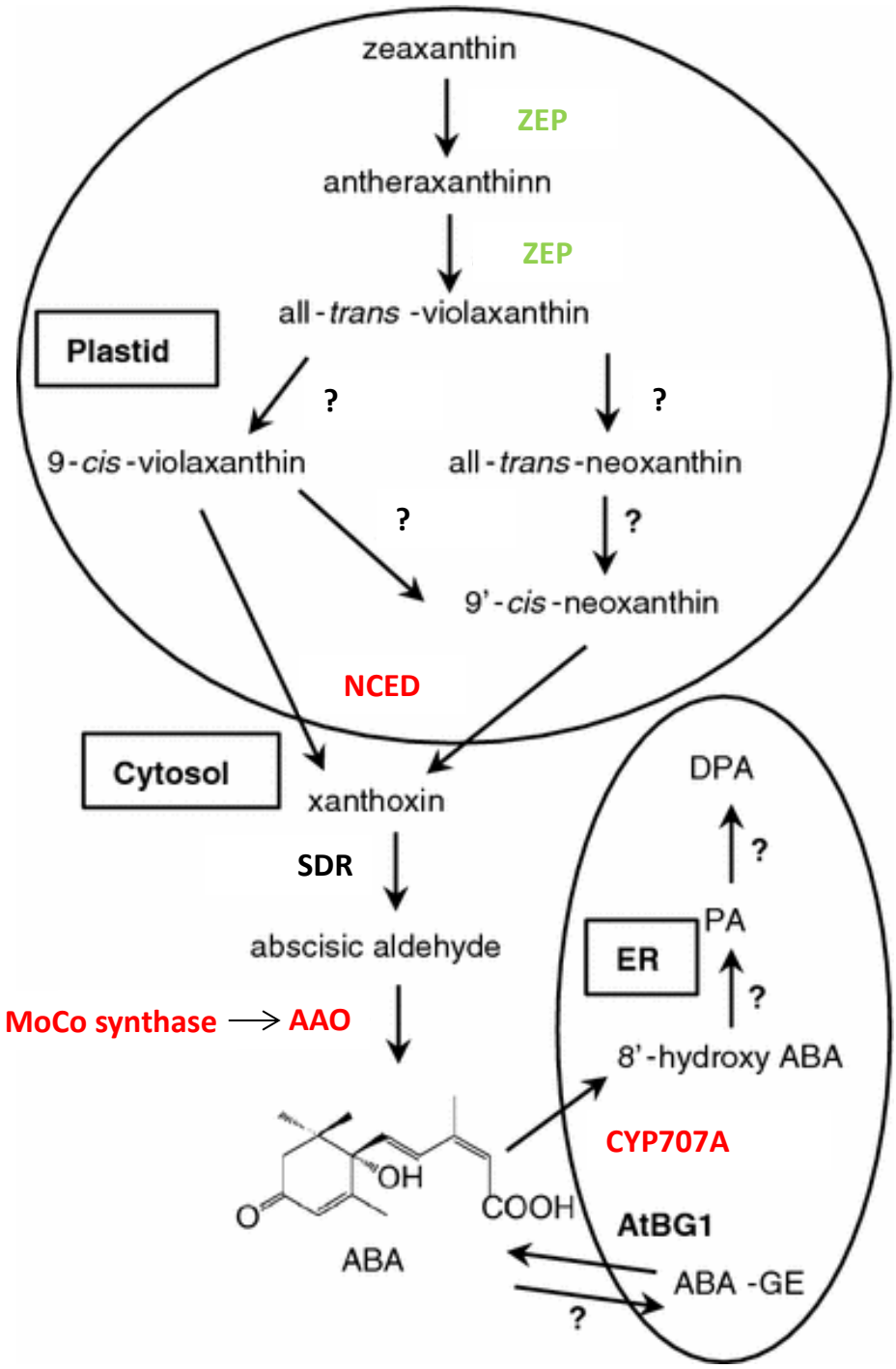


Figure 1.2. ABA biosynthetic pathway in *Arabidopsis*. Zeaxanthin precursor is derived from carotenoids synthesized in the plastid. Zeaxanthin epoxidase (ZEP) is responsible for the two-step epoxidation of zeaxanthin. Some intermediate steps possibly carried out by AtABA4 (not shown), an uncharacterized protein found in mutant screens. Key step in ABA biosynthesis is carried out by the multigene family of NCED enzymes. This conversion to xanthoxin is though to be the rate-limiting step. Conversion to ABA aldehyde carried out by SDR1 (also known as ABA2). Final oxidation by AAO enzymes requires the MoCo cofactor. This is synthesized by the MoCo synthase. ABA catabolism is primarily carried out by CYP707A family at the endoplasmic reticulum (ER) which results in production of phaseic acid (PA). Other catabolites are ABA glucosyl ester (ABA-GE) and dihydro phaseic acid (DPA). Enzyme in green is functionally conserved in *Physcomitrella*, enzymes in red have putative homologues in *Physcomitrella*. SDR has no clear homologue. Figure adapted from Seo and Koshiba, 2011.

transporters (Kuromori *et al.,* 2010 and Kang *et al.,* 2010) as well as members of the NRT1/PTR family (Kanno *et al*., 2012). The regulation of ABA levels in plants is the first stage of the signal transduction pathway which translates this signal into molecular changes affecting the range of processes already mentioned.

1.4 ABA-dependent signalling

While a number of ABA receptors have been proposed (reviewed in Cutler *et al.,* 2010), this thesis will only discuss those that appear to be part of the 'core' ABA signal transduction pathway conserved in *Physcomitrella* and all land plants (Takezawa *et al.,* 2011), and for which convincing evidence is available. This pathway has been well characterised in *Arabidopsis* where it is headed by a 14 member subfamily of START proteins known as PYR/PYL/RCAR proteins (Park *et al.,* 2009; Ma *et al.,* 2009). The ABA response is achieved through the highly conserved core signalling pathway (Figure 1.3), in which the receptors bind ABA, thereby potentiating a molecular interaction with group A protein phosphatase 2C superfamily (PP2C) phosphatases typified by the products of the *ABA INSENSITIVE ABI1* and *ABI2* genes. Sequestration of these phosphatases enables the phosphorylation of SnRK2 serine/threonine protein kinases, typified by OST1 ("Open Stomata 1"), that in turn initiate a cascade of protein phosphorylation that activate transcription factors (*e.g.* the ABI5/ABA-response element binding factors (ABF) bZip and ABI3 B3 transcription factors) that trigger ABA-mediated transcription of the genes encoding dehydration tolerance-associated proteins. These TFs themselves are also found to be inhibited by the ABI1/2 PP2Cs and so these are too released from inhibition (Yuan *et al*., 2013). These targets are most commonly characterised by the presence of ABA responsive binding elements (ABREs) recognised by the ABI5-related TFs which are often found in conjunction with other *cis*-elements such as the Ry/Sph element recognised by the B3 domain TFs (such as ABI3), the dehydration responsive elements (DREs) recognised by the ABI4 and DRE binding (DREB) AP2/ERF TFs and NAC-recognition sites (important for salt responsive gene expression) (Figure 1.4) making the regulation of transcription an important stage for crosstalk (reviewed by Roychoudhury *et al.,* 2013). Functional conservation in *Physcomitrella* and other bryophytes has been demonstrated for the ABI1-related protein phosphatases (Komatsu *et al.,* 2009; Komatsu *et al*., 2013), the OST1 kinase (Chater *et al.,* 2011) and the key transcription factors ABI3A/B/C (Khandelwal *et al.,* 2010) which together regulate a similar set of genes as those found in seeds (Kamisugi and Cuming, 2005; Cuming *et al.,* 2007). This signalling pathway is thought to function in both the cytoplasm and nucleus where it can regulate specific targets by phosphorylation such as the ion channel slow anion channel-associated 1 (SLAC1) in angiosperm stomata and ABA-regulatory TFs in most cell types – a

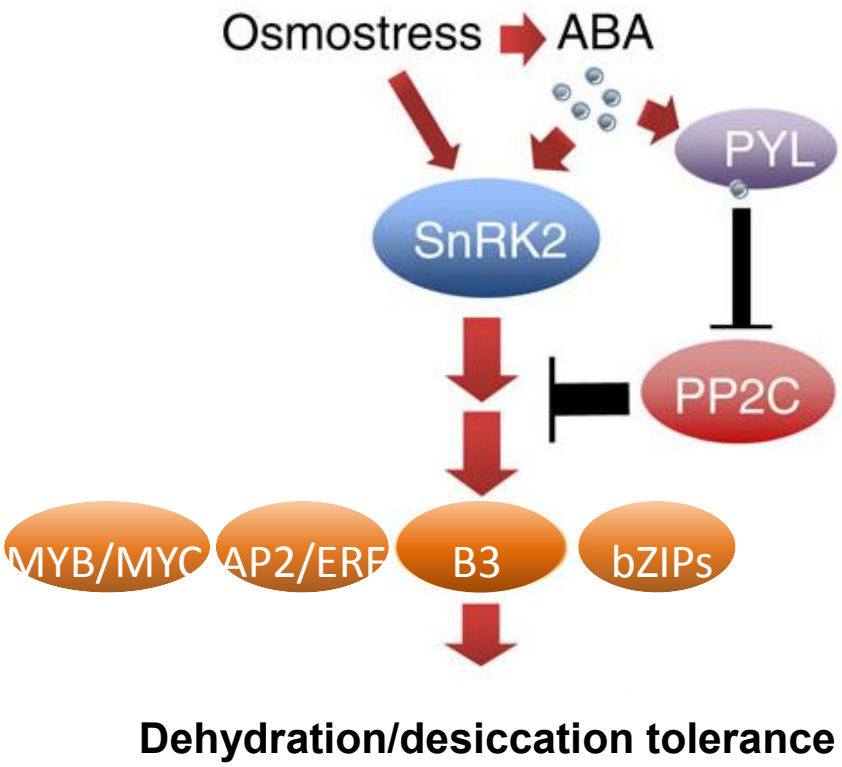


Figure 1.3. Simplified 'core' ABA signalling pathway. Increased ABA biosynthesis, following osmostress, leads to activation of the PYR/PYL/RCAR receptors through binding of ABA. ABA binding induces the formation of an interface on receptors that binds PP2C phosphatases and so sequesters them from inhibition of the SnRK2 kinases. These kinases may self activate through autophosphorylation in *Arabidopsis* but in more basal plants this may be also under the control of upstream phosphorelay activation; possibly from signal transduction pathways sensing stresses directly. The transcription factor (TF) targets for SnRK2 kinases include some B3 domain, bZIP, AP2/ERF, MYB and MYC TFs. bZIP TFs are direct targets for SnRK2s (*e.g. AREB1*) in *Arabidopsis* where they recognize the ABA-responsive element (ABRE) which has not yet been shown in *Physcomitrella* where instead ABI3 homologues appear to be the main ABA-dependent TF components although ABRE remains vital. TFs interact in the promoter region of ABA-responsive genes in variety of ways. The genes regulated by these TFs encode proteins involved in conferring desiccation tolerance in angiosperm seeds and bryophyte vegetative tissue. Figure adapted from Komatsu *et al*., 2013.

response more prominent in nuclei of the bryophyte dominant gametophyte generation to confer gene expression mediated dehydration tolerance.

While there are other ABA-dependent signalling components, this 4 component core pathway has been demonstrated as sufficient to induce ABA-dependent gene expression in a reporter system in a triple sugar non-fermenting 1 (SNF1)-related SnRK2/OST1-like mutant; and therefore ABA-non responsive background (Fujii *et al.,* 2009). This important experiment demonstrated that the target TFs required phosphorylation by the OST1 kinase and that OST1 itself was phosphorylated by default in the absence of the PP2C phosphatases. When ABI1 PP2C was included it acted as a strong OST1 inhibitor via dephosphorylation, indicating crucial roles for phosphorelays in ABA-dependent signalling. Key structural work has demonstrated the mechanisms behind the ABA-induced interaction of the PYR/PYL/RCAR receptors and PP2Cs (Park *et al.,* 2009; Ma *et al.,* 2009). Briefly, binding of ABA in the receptor pocket resulted in conformational change of a loop region that locked ABA inside while creating a new interacting interface with PP2C (Santiago *et al.,* 2009; Melcher *et al.,* 2009; Miyazono *et al.,* 2009; Yin *et al.,* 2009). This complex is how the PP2Cs are sequestered from SnRK2 inhibition, by competitive binding of the PP2C active site: and without this mechanism, plants fail to respond to ABA. This well characterised and relatively simple signal transduction pathway requires interactions with other pathways to regulate the subtle and specific responses to ABA and particular stresses through crosstalk. For example, besides regulation of phosphorylation, the ubiquitin-proteasome system (UPS)-mediated protein degradation pathways are also important posttranslational regulatory mechanisms. Many components of the ABA and stress response pathways are ubiquitin tagged, a process involving three components (called E1, E2 and E3) to tag lysine residues, for degradation by the 26S proteasome (reviewed by Stone, 2014). The E3 ubiquitin ligase components are substrate specific and recently the colourfully named SALT- AND DROUGHT-INDUCED REALLY INTERESTING NEW GENE FINGER1 (SDIR1) E3 ligase, part of the RING-type E3 family, was found to tag its target SDIR interacting protein 1 (SDIRIP1) (Zhang *et al.,* 2015). These authors found that while SDIR1 acts as a positive regulator of various ABA-responsive bZIP TFs, SDIRIP1 was found to specifically regulate ABI5.

As stated, the canonical ABA-responsive promoter motif is known as the ABA-responsive binding element (ABRE) and is recognised by members of the bZIP family (such as ABI5a and AREBs) (Figure 1.4). In the well characterised *Arabidopsis*, this bZIP subfamily of ABA-dependent TFs (known as group A) contains around 12 members and includes other ABA responsive element binding factors besides ABI5 (AREBs, ABFs and DPBFs). The two with major roles in seed maturation are AtABI5

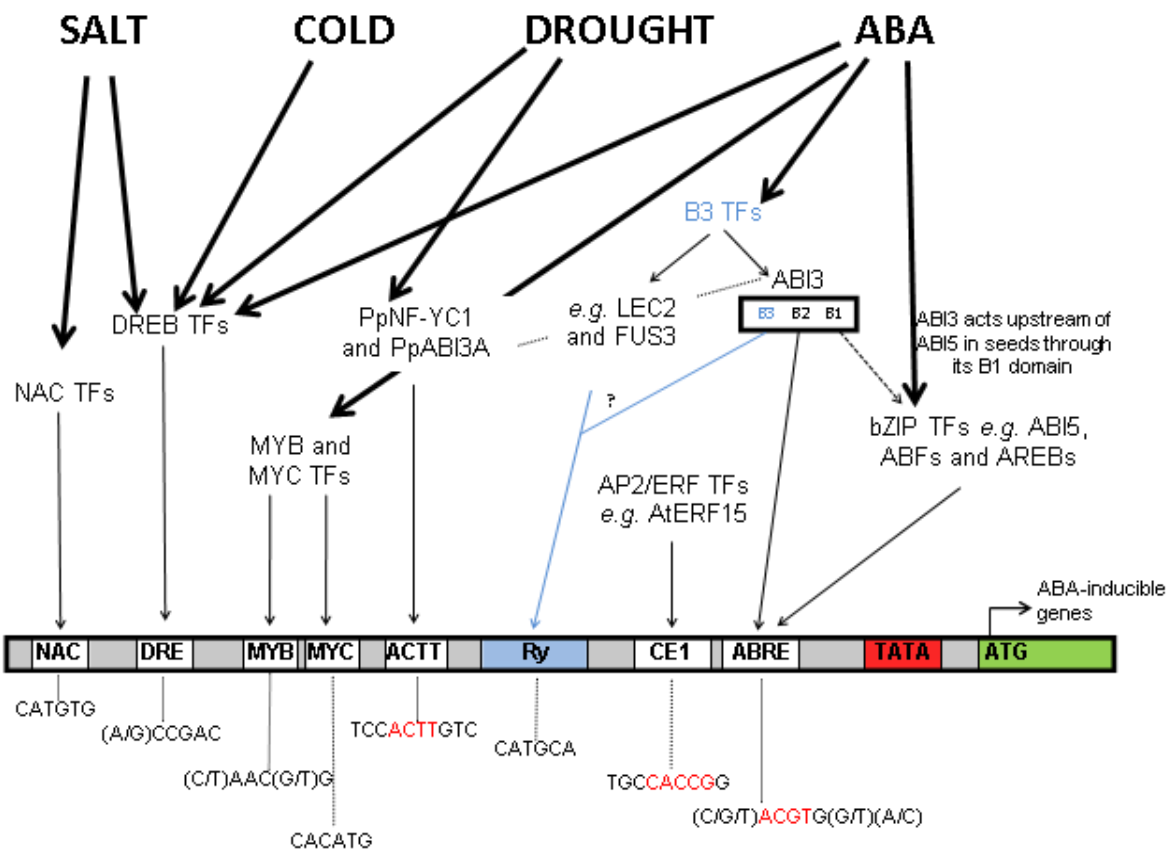


Figure 1.4. Trans-acting factors and *cis*-acting elements regulating ABA-dependent and ABA-independent stress signalling. bZIP transcription factors (TFs) (ABI5/AREB/ABFs) are primary components of ABA-responsive gene expression through recognition of the ABA-responsive element (ABRE). The B3 domain TFs such as LEC2 and FUS3 bind the Ry element which are often clustered together with ABREs (part of the G-box elements)where they are referred to as the Ry/G-box. ABI3, composed of the B1, B2 and B3 domains, may interact with the Ry elements directly or through interaction with other B3 TFs. ABI3 does act upstream of ABI5 through a physical interaction with the B1 domain. The B2 domain may also interact with the ABRE directly. ABI3 is also found to interact with the nuclear factor Y C1 (NF-YC1) subunit in *Physcomitrella* mediating binding to the ACTT-core element. AP2/ERF TFs recognize coupling element 1 (CE1) which is often 'coupled' with ABRE and required for ABA-responsive expression. Both MYB and MYC TFs are ABA responsive recognizing their own *cis*-elements. Drought responsive element (DRE) are bound by DREBs which are also involved in a range of stress responses. NAC TFs are involved in salt responses recognizing the NAC element. Thick black arrows represent the stress stimuli; solid thin arrows represent likely direct binding to *cis*-elements, dashed arrows indicate likely TF complexes.

and AtEEL (for “Enhanced Em level” – Em being a group 1 LEA protein) which have overlapping expression that allows fine-tuning. Many other group A members have weak loss of function phenotypes showing redundancy but this is likely part of a general ability to fine tune responses for example by forming heterodimers, through their leucine repeat region, and through crosstalk with ABA-independent pathways (*e.g.* Bensmihen *et al.,* 2002). The four characterised ABFs (ABF1-4) show differential induction following cold, drought or salt treatment (Choi *et al.,* 2000), likely due to different upstream phosphorylation pathways, which demonstrates how stress specific responses can be fine-tuned. Of course, if all these related TFs recognise the same *cis*-element then such fine-tuning is impossible unless further interactions are involved. In fact the ABRE elements are rarely found alone and are often found with coupling elements (CEs) (Figure 1.4). These are so called because of their often coupled function with ABRE in that some ABRE containing genes require the CE for full ABA-induced gene expression (Shen and Ho, 1995). A suite of AP2/ERF (Apetala2/Ethylene Response Factor) TFs were found to bind to CE1 in yeast one-hybrid screens (Lee *et al.,* 2010) and one of these has since been functionally demonstrated as being a positive ABA regulator (Lee *et al.,* 2015a). There is another group of MYB/MYC-related ABA-dependent TFs which recognise MYC- and MYB-binding sites respectively (Figure 1.4) and these can function alongside the bZIP-mediated ABA signalling, although these TFs themselves are likely induced by ABA signalling, making them part of the ABA-dependent secondary response (Iwasaki *et al.,* 1995). Two characterised members, AtMYB2 and AtMYC2, showed additive roles in ABA-induced gene expression of a range of transcripts which generally contained both MYC and MYB *cis*-elements (Abe *et al*., 1997; Abe *et al.,* 2003). While distinct TFs, they are often treated together given their strong overlap in target gene regulation.

Regulation of ABA-dependent gene expression is largely mediated by the presence of *cis*-element complexes which enable both direct physical and indirect molecular interactions to occur between ABA-dependent and ABA-independent pathways which are discussed below. In the angiosperm seed, ABI5 is found to act downstream of ABI3, a B3 domain TF, with ABI3 inducing the expression of ABI5 itself (Lopez-Molina *et al.,* 2002). This has also been shown to involve a physical interaction between the B1 domain of ABI3 and ABI5 with only ABI5 binding the ABRE element of target genes (Nakamura *et al.,* 2001). This interaction is also found to increase the binding of bZIP TFs to DNA through the action of the ABI3 B2 domain (Hill *et al.,* 1996). This level of physical interaction goes further with the finding that PpABI3A can act synergistically with the nuclear factor Y (NF-Y) complex to bind an ACTT element in *Physcomitrella* (Figure 1.4) to regulate ABI3-mediated gene expression (Yotsui *et al.,* 2013). The B

subunit in this NF-Y complex is homologous with AtLEC1 which has long been known to be involved in embryo development (Meinke *et al.,* 1994). Other B3 TFs such as LEC2 and FUS3 are found to directly bind the Ry/G-box elements in seed maturation genes suggesting that not all members of the highly-expanded B3 TF family function the in the same way (Kroj *et al.,* 2003). These studies revealed a hierarchical regulatory system that regulates the group A bZIP TFs even before ABA-dependent SnRK2 phosphorylation has occurred.

1.5 Crosstalk with other ABA-dependent and independent stress response pathways

Molecular crosstalk involves the interaction between various components in pathways such that one exerts an effect on the other becoming extra sources of input besides the 'primary' stimulus. This is a very important part of plant, and indeed all life forms, signalling enabling fine tuning of responses to enable greater efficiency and appropriateness of responses. Of the 'core' components already mentioned, there are three SnRK2/OST1-like kinases in *Arabidopsis* that appear to mediate the ABA-dependent increase in phosphorylation of 166 peptides and a decrease in phosphorylation of 117 peptides (Wang *et al.,* 2013). Many of these are likely to be targets of secondary signalling following ABA-induced gene expression but these authors found 58 proteins to be possible substrates. The OST1-like kinases likely represent a key nexus in ABA and stress signalling with far more roles than in simply activating the target ABA-dependent TFs. Of the 58 substrates, 18 were likely further signal transduction components which included typical TF targets but also included other kinases, including a MAP kinase (MPK6), a uridine kinase (TTM1) and another serine/threonine kinase SIK1 (Wang *et al.,* 2013). What these represent are most likely crosstalk partners that are integral parts of other pathways that are modulated by the ABA-signalling OST1-like kinases. For example, MPK6 is known to be involved in a blue light sensing pathway (MKK3-MPK6-MYC2) (Sethi *et al.,* 2014), phosphate acquisition (MKK9-MPK6/MPK3) (Lei *et al.,* 2014) and ethylene signalling (ETR-CTR1-SIMKK-SIMK/MMK3/MPK6-EIN2-EIN3) (Ouaked *et al.,* 2003). In fact, MPK6 has been found to potentially have 39 substrates (Feilner *et al.,* 2005). This brief example serves to demonstrate the complexities of signalling pathways and serves as a particularly relevant example for this thesis.

While a myriad of interactions occur between kinases and phosphatases in stress responses (and especially MAP kinases which are discussed in more detail below) many are beyond the scope of this thesis. However, one level of crosstalk already indicated, and more tractable to analysis, is the complex of TF/*cis*-element interactions found in ABA/stress responsive genes. The main TF families involved in stress

responses are the B3-domain, bZIP/ABF/AREBs, MYB, MYC, NAC, AP2/ERF and DREB/C-repeat binding factor (CBF) families. The bZIP/ABR/AREBs, MYBs, MYCs and DREB/CBFs are involved in ABA-dependent gene expression while NACs (salt), ERFs (ethylene) and MYC-like ICE TFs (cold stress) are more involved in ABA-independent stress responses. Their roles are not rigid, however, as the DREBs are also heavily involved in ABA-independent drought and salt responses and NACs can be activated by ABA-dependent pathways. The multi-gene families of TFs enable non-redundant functions to evolve through neofunctionalisation which can be as simple as evolving a novel interacting partner. A general trend of plant evolution facilitating this process has been an increase in the size of many gene families involved in stress signalling through whole genome and partial duplications creating ever more paralogs (*e.g.* Takezawa *et al.,* 2011). Many stress inducible genes also have multiple *cis*-elements allowing co-regulation by a range of TFs acting at the terminus of many stress signalling pathways. For example, *Arabidopsis* was found to have over 2000 genes with both DRE and ABRE motifs of which 72% were within 400bp of each other, suggesting direct interactions between the ABF and DREBs to be important (Mishra *et al.,* 2014) or simply that the two TF families share strong co-regulation of gene targets (see chapter 3). The interaction between DREB/CBFs and ABFs in ABA-independent and -dependent crosstalk is well studied and even demonstrates crosstalk mechanisms whereby the expression levels of one are regulated by the other such as in the case of DREB2 which is up-regulated by ABFs (reviewed by Yoshida *et al.,* 2014). The picture from studies of crosstalk between the various stress responses shows that there is a great deal of overlap with no pathway acting in true isolation. Studies of the ‘core’ ABA signalling pathway demonstrated that only four components are necessary and sufficient for ABA-induced gene expression (Fuji *et al.,* 2009) but by coordinating interactions between a myriad of such modules, plants enable the fine tuned responses required to thrive in almost all habitats.

1.6 MAP kinase signalling

As touched upon, stress signalling is a highly complex process involving many levels of interaction. One aspect of plant stress signalling receiving ever more attention is the role of MAP kinase cascades (see Figure 1.5 and Sinha *et al*., 2011). MAPK cascades are ubiquitous signal transduction pathways found in all eukaryotes including fungi and animals in which entire modules are conserved (*e.g.* Widmann *et al.,* 1999). The canonical signal transduction pathway involves detection of external stimuli by a receptor which interacts with and activates a MAP3K (although some MAP4K modules have been identified (*e.g.* Champion *et al.,* 2004a)). The MAP3K (or MAP4Ks; although not all plant MAP4Ks appear to act on MAP3K targets in plants (Champion *et al*.,

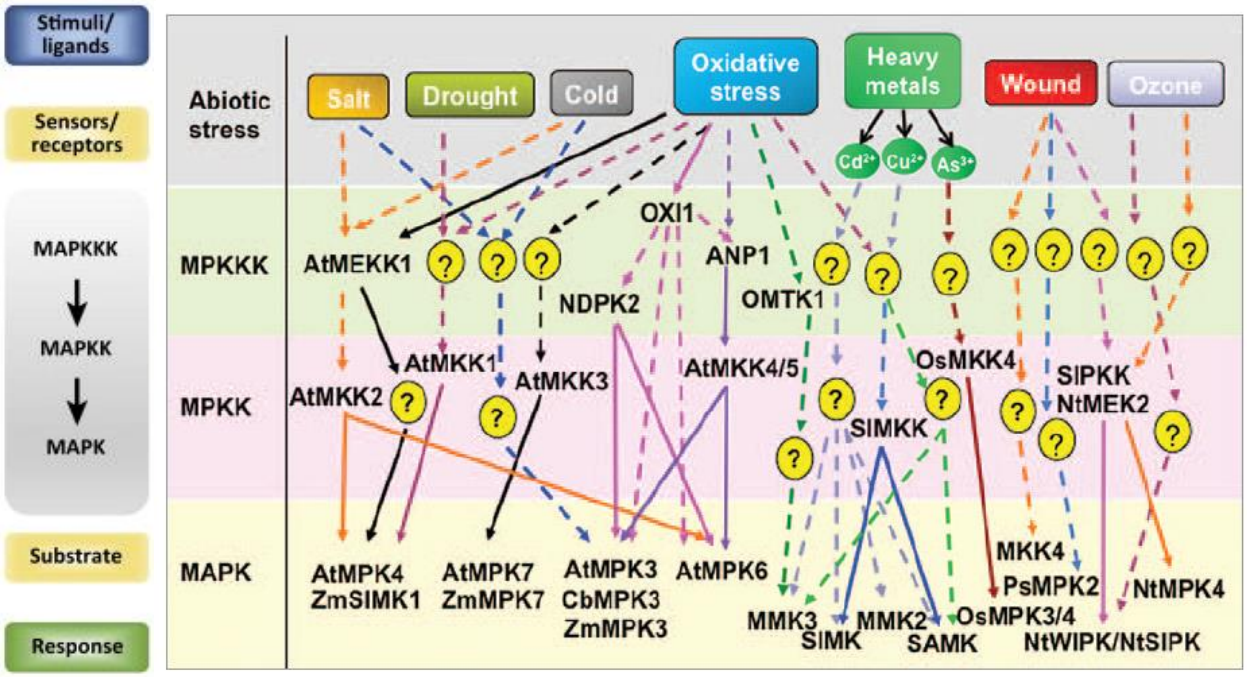


Figure 1.5. Some of the known plant MAP kinase cascades in stress signalling. The canonical MAP kinase signal transduction pathway is shown on left that links sensory components (*e.g.* receptors) with responses primarily mediated by the MAPK substrate which is often a transcription factor resulting in gene expression. Many MAP kinase phosphorylation cascades have been elucidated in plants which respond to environmental stress. The number of MAP kinase components found to be involved in stress signalling is growing and many are involved in crosstalk with other pathways through phosphorelays. Note the involvement of MPK6 in multiple pathways. Figure adapted from Sinha *et al*., 2011.

2004b)) then initiates a sequential phosphorylation cascade through MAP2Ks and MAPKs which then act on a target substrate with each interaction relying on targets having specific phosphorylation sites (Figure 1.5). Given these simple building blocks, many MAP kinase pathways have been modified and rearranged but the essential processes are conserved; namely in phosphorylating a target. The first global analysis of plant MAP kinases was first carried out in *Arabidopsis* by the MAPK group in 2002 (Ichimura *et al.,* 2002) that set up the initial definitions for the subfamilies and showed their relationships. They revealed *Arabidopsis* to have 60 MAP3Ks (since updated to 80 (Moustafa *et al.,* 2014)), 10 MAP2Ks and 20 MAPKs (Ichimura *et al.,* 2002) which was later updated with the identification of 10 MAP4Ks (Champion *et al*., 2004b). These studies revealed the MAP3Ks to be split into two phylogenetically distinct groups; the MEKK-like family and the Raf-like family. The MEKK-like family are closely related to the MAP2K and MAPK families and more closely still to the MAP4K family (Champion *et al*., 2004b; Chapter 4) and as such appear to derive from a canonical MAP kinase ancestral sequence. The Raf-like MAP3Ks share more in common with animal Raf-kinases than the MEKK-like kinases and indeed all other MAP kinase families (Ichimura *et al.,* 2002; Champion *et al*., 2004b). As such they can be thought of as non-canonical MAP3Ks. These Raf-like kinases are closely related to a very large sister group of plant receptor-like kinases (RLK) (Champion *et al*., 2004b; Chapter 4). This thesis has a special interest in these Raf-like MAP3Ks but the general role of MAP kinase cascades in stress signalling will be discussed.

MAP kinases have been heavily implicated in stress signalling. Plant MAP3 kinases are divided into the MEKK-like (Group A) and Raf-like (Groups B and C) which are phylogenetically distinct and for example, just within the B group a number have been implicated in: ABA and drought stress responses (Lee *et al.,* 2015b; Sasayama *et al.,* 2011; Ning *et al.,* 2010), ethylene (Kieber *et al.,* 1993), salt and sugar (Gao and Xiang, 2008; Huang *et al.,* 2014b), pathogen responses (Frye *et al.,* 2001; Tang *et al.,* 2005) and most recently, and relevantly, a dual role in ABA and ethylene signalling (Yasumura *et al.,* 2015). The two MAP3Ks involved in ethylene and pathogen responses, AtCTR1 (Keiber *et al.,* 1993) and AtEDR1 (Frye *et al.,* 2001), as well as the sugar insensitive kinase AtSIS8 (Huang *et al.,* 2014b) all contain N-terminal regulatory EDR domains which are among the best characterised in their interaction with the ETR receptors (*e.g.* Clark *et al*., 1998). The N-terminal regulatory domain for those kinases involved as drought and ABA regulators, Raf10/11 (Lee *et al.,* 2015b) and MAP3Kδ4 (Sasayama *et al.,* 2011), have an N-terminal Per-ARNT-Sim (PAS) domain which are discussed in more detail. Interestingly, MAP3Kδ4 was also found to be auxin responsive (Sasayama *et al.,* 2011). The rice DSM1 (drought-hypersensitive mutant1)

MAP3K is most closely related to the EDR1 kinase in *Arabidopsis* but is found to be involved in drought responses possibly regulating ROS responses (Ning *et al*., 2010).

Many other groups of MAP3 kinases are modular in nature such that N-terminal domains act as input domains allowing regulation of the kinase domain activity and subsequent downstream signal transduction. Other subfamilies of Raf-like MAP3Ks with defining regulatory domains include the C2 subfamily which contain N-terminal ACT domains (***A***spartate kinase, ***C***horismate mutase and ***T***yrA ligand-binding fold) and the C1 subfamily which contains N-terminal ankyrin domains. Our knowledge of these group of kinases is expanding however and the discovery of AtVIK, a C1 subfamily member, as a stress regulator of vascular glucose uptake (Wingenter *et al.,* 2011) may indicate a role for ankyrin mediated protein-protein interaction in this function (Li *et al.,* 2006). Another Raf-like MAP3K, called Raf43, from the C5 subfamily has been found to have a role in a range of abiotic stresses including drought, mannitol and fungal attack (Virk *et al*., 2015).

The large number of MAP3Ks, relative to MAP2Ks and MAPKs, suggests that there has been a selective advantage in increasing the possible number of these interactions with signal sensor components or other inputs, thereby increasing the specificity of responses. In the hierarchies beneath the MAP3Ks, many MAP2Ks and MAPKs have also been implicated in stress signalling demonstrating that MAP kinase cascades are likely to be an important building block of stress responses. For example, water stress in *Arabidopsis* is able to induce the expression of MPK2-5/12, while salt stress induced the expression of MPK9-11/17/18 (Moustafa *et al.,* 2008). These 10 MAPKs represent half of the full *Arabidopsis* MAPK complement, indicating a major role in stress signalling. Similarly, half of rice MAP2Ks have been shown to be up-regulated by drought (Kumar *et al.,* 2008). The 5 drought- and 5 salt-induced MAPKs (Moustafa *et al*., 2008) also demonstrate that redundancy in responses is likely but that, while drought and salt stresses have physiological consequences in common, specificity in the responses is still maintained through the signal transduction pathways likely emanating from a specific MAP3K mediated interaction. Overexpression of the ABA-inducible rice orthologue of MPK3 (OsMAPK5) has been shown to enhance drought tolerance (Xiong and Yang, 2003) showing that these components are important for leading to adaptive molecular and physiological changes that help plants survive and cope with abiotic stresses. This ABA-induction indicates not only the role of MAP kinases in stress signalling but also the clear role of ABA. Many of the MAPKs identified in stress responses have been shown to have probable functional roles in the nucleus (*e.g.* Shi *et al.,* 2011); suggesting that activation of transcription factors is often the end point of MAP kinase signal transduction. A direct interaction between MPK6,

previously seen to be involved in multiple pathways, and the TF MYB41 in *Arabidopsis* has been demonstrated enabling salt stress responses to confer salt tolerance (Hoang *et al.,* 2012). MPK6 was also shown to be regulated by phosphatidic acid, known to be a signalling component of salt stress (Yu *et al.,* 2010). MAPKs have also been implicated in heat stress responses (Suri and Dhindsa, 2008) reflecting their role in signal transduction of a wide range of abiotic (and biotic *e.g.* AtEDR1 in powdery mildew responses (Frye and Innes, 1998)) stress responses. The finding that many MAP kinases are themselves up-regulated by stress (and in some cases by ABA itself) shows that positive feedback is an important feature of stress signalling.

1.7 PAS domains

As seen, a number of PAS-containing MAP3Ks are involved in ABA and stress signalling, in which these N-terminal domains are thought to act as regulatory domains for the kinase activity, linking environmental stimulus sensing with MAP kinase phosphorylation cascades. These are not the only PAS-containing proteins found in plants; two other important families of regulatory protein that contain PAS domains are the light sensing phytochromes and phototropins. The PAS domain superfamily is a group of structurally similar protein sensor modules found in all kingdoms of life. They are principally involved in signal transduction (Taylor and Zhulin, 1999) - enabling the detection of and response to a range of environmental stimuli. Despite low sequence homology across the sub-families the 100-120 amino acid PAS domain core forms a canonical PAS fold including a 5 antiparallel strand β-sheet and (usually) 4 α-helices (A_β, B_β, C_α, D_α, E_α, F_α, G_β, H_β, I_β) (Taylor and Zhulin, 1999; Hefti *et al.,* 2004) however, there is some variation in the number and size of the α-helices (Möglich *et al.,* 2009a) (Figure 1.6). This PAS fold is often flanked by further α-helices which show considerable structural variation between the PAS subsets but that are thought to be important principally by interacting with the β-sheet in various ways (Möglich *et al.,* 2009a) affecting both quaternary structure (Ayers and Moffat, 2008) and by propagating signals from the PAS fold as coiled-coil linkers. This relatively simple module is highly versatile enabling the PAS domain to:

- Sense light by associating with flavin cofactors (flavin mononucleotide or flavin adenine dinucleotide) as well as sensing redox potential (see Becker *et al.,* 2011 for recent review).
- Sense oxygen using a haem cofactor (see Gilles-Gonzalez and Gonzalez, 2004 for review).
- Bind a range of small ligands (see Henry and Crosson, 2011 for bacterial review).

- Interact with transcriptional co-effectors as well as other PAS domains to enable modulation of its signal output (see Partch and Gardner, 2010 and Bersten *et al.,* 2013 for reviews in bHLH-PAS transcription factors).

PAS domains have been best functionally characterised in bacteria where they are principally found in two-component regulatory systems; of which the ethylene receptors are plant-specific descendants. These are single or paired membrane associated proteins in which the (often) modular sensory component directly or indirectly senses changes in the environment upon which conformational changes allow or block autophosphorylation of a histidine kinase effector domain which transduces the signal into the cell, typically resulting in the phosphorylation of TFs, mediating a response (reviewed by Mascher *et al.,* 2006 and Szurmant *et al.,* 2007). Also found in bacteria are the two-component related PAS-containing histidine kinases known as bacteriophytochromes which are ancient light sensors and are the forerunners of the red/far red-light sensing phytochromes in plants. Phytochromes too are histidine kinase-related proteins that typically have 3 PAS domains alongside the distantly related GAF and PHY domains (Sharrock, 2008). The N-terminal PAS domain (often termed PAS-like domain (PLD)) has no sequence homology with the MAPK PAS domains (Chapter 4) however the more distal pair show some (albeit low) homology, with PASb showing the most. The other family of PAS-containing plant photoreceptors are the blue-light receptors: the phototropins. These contain PAS subfamily variants known as light-oxygen-voltage (LOV) domains that are usually found as a tandem repeat and are frequently associated with a serine/threonine kinase domain (Christie, 2007). These LOV domains bind flavin mononucleotide (FMN) cofactors that enable light excitation to be transduced into perception and responses via kinase activation. However these share no significant sequence homology with MAP3 kinases; in either the PAS or kinase domains (Chapter 4).

The small PAS sensor module is rarely found alone but is found in a range of contexts befitting its range of functions including modular kinases (including the histidine kinase bacterial ‘two-component’ regulatory systems), transcription factors and circadian clocks. PAS domains are sometimes found as part of a signal sensor module complex with other sensor domains and are also frequently found with other PAS domains; occasionally ones with different origins and functions. PAS domains are almost always found intracellularly (but see Chang *et al.,* 2010 for an extracytoplasmic example), although they are occasionally membrane associated, despite their occasional sensing of external stimuli (such as light or toxins) meaning the cues must necessarily cross into the cell. More often than not however, proxies of environmental conditions such as changes to redox potentials or intracellular oxygen concentrations are sensed.

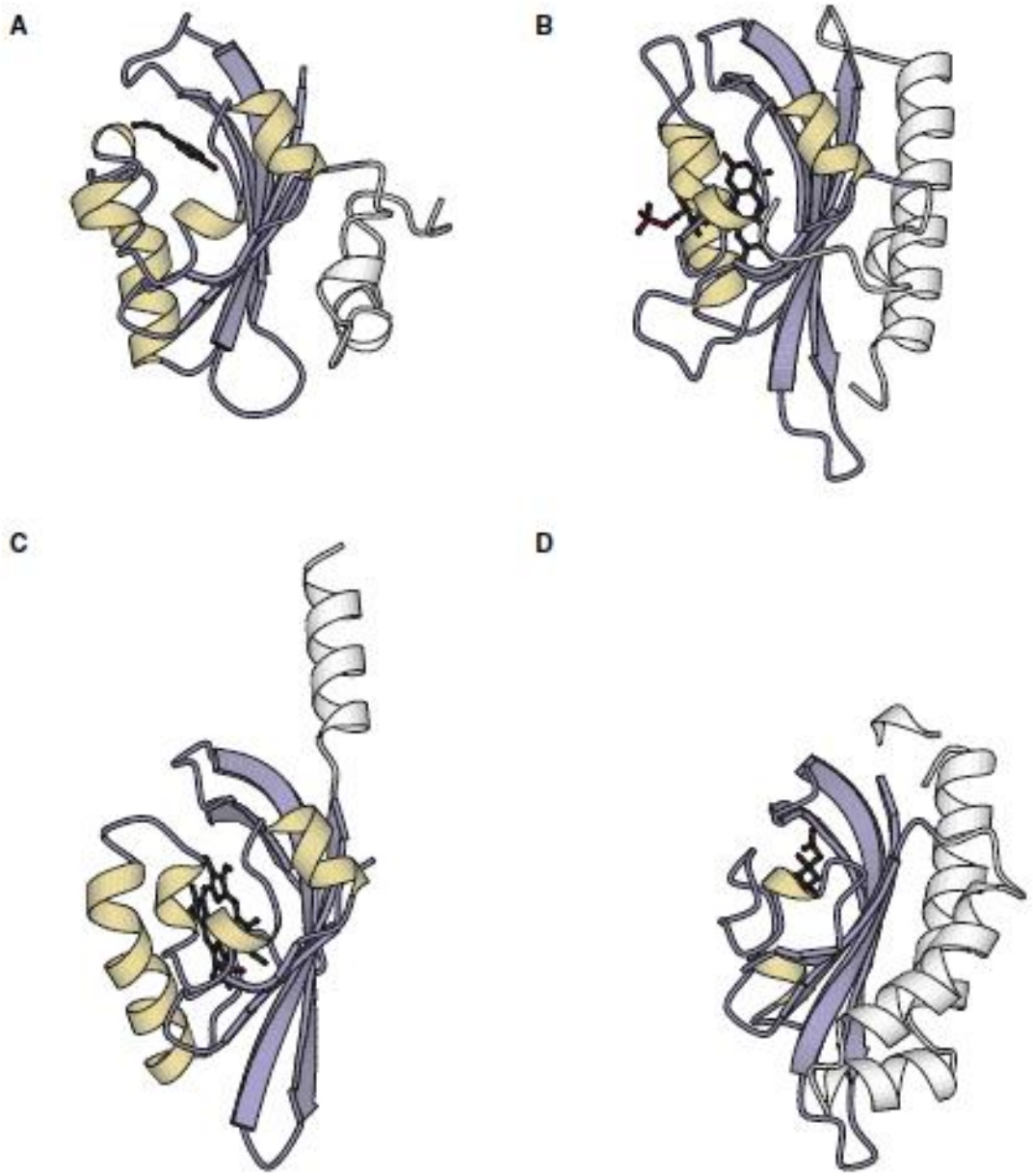


Figure 1.6. Diversity of PAS domain structures. A) *p*-Coumeric acid binding in the photoactive yellow protein (PYP) photosensor. **B)** Flavin mononucleotide (FMN) binding LOV domain of a phototropin. **C)** Bacterial haem binding redox sensor FixL. **D)** Citrate binding of the CitA citrate sensor. All PAS domains form a typical 'PAS fold' conformation often resulting in the formation of an internal cavity often adjacent to the large F_{α} helix (pale yellow). Interactions with other PAS domains and signal transduction often conferred by conformational changes in the β-sheet (blue). The 'PAS fold' is often flanked by further helices (white) which can act as 'molecular rods' enabling signal transduction to the effector domain, often a kinase. As can be seen between A and C, these helices can take very different conformations either folding back to interact with the PAS core or pointing away and acting more as a molecular 'rod'. Figure taken from Möglich *et al*., 2009a.

The focus of this thesis is on the PAS containing MAP3K, PpANR, and in the context of MAP kinase stress signalling, the role of its 'modulator' domains is of paramount importance. The question of whether it may be involved in ligand binding or in protein-protein interactions must be resolved to understand the mechanism that links environmental stimuli with MAP kinase signal transduction. These questions are addressed in Chapter 5 which has focused on a structural approach to understand how these so far uncharacterised PAS domains may function.

1.8 *Physcomitrella patens* life cycle

While the focus of this introduction has been on stress signalling mechanisms and plant evolution, little has been said about the model organism upon which this thesis's findings are based. The identification of aberrant phenotypes in this model moss requires some knowledge of what constitutes 'normal' growth and development. As in all plants, phytohormones play crucial roles in moss growth and development (reviewed by Decker *et al.,* 2006). As a bryophyte, *Physcomitrella* is anatomically simple but still has morphologically distinct stages and tissue (Figure 1.7). Moss has a haploid dominant life cycle with the leafy gametophyte bearing the diploid sporophytes. The gametophyte develops mitotically from a spore released from the sporophyte, first going through the juvenile filamentous stage in which chloronema and caulonema can be distinguished. While ABA can stimulate the characteristic ABA growth response at this stage, which includes retarded growth including blocking gametophore production and induction of 'brood' cell formation, the positive growth regulators auxin and cytokinin drive on growth and eventual bud formation (reviewed by Ashton *et al.,* 1979). These buds become the leafy gametophores which form rhizoids at their base and the sex organs at their top. Self- or cross-fertilisation results in the fusion of gametes to form the diploid sporophyte, which bears ABA regulated stomata (Chater *et al.,* 2011). The terminal spore capsules undergo meiosis resulting in the formation of haploid spores which are released and dispersed.

1.9 Summary of project and aims

Abiotic stress signalling is heavily reliant on ABA signalling and this likely played a vital role in the conquest of land by plants. Functional conservation of the 'core' signalling components has been demonstrated across 500 million years of land plant evolution, principally through using reverse genetic approaches in the model moss *Physcomitrella* taking advantage of excellent homologous recombination-mediated gene targeting (Schaefer and Zryd, 1997 Kamisugi *et al.,* 2005 and Kamisugi *et al.,* 2006). This has been a fruitful approach which has helped build our understanding of the signalling pathways likely to be important in anatomically simple early land plants. However, this "reverse genetics" approach is necessarily limited to testing hypotheses based on *a*

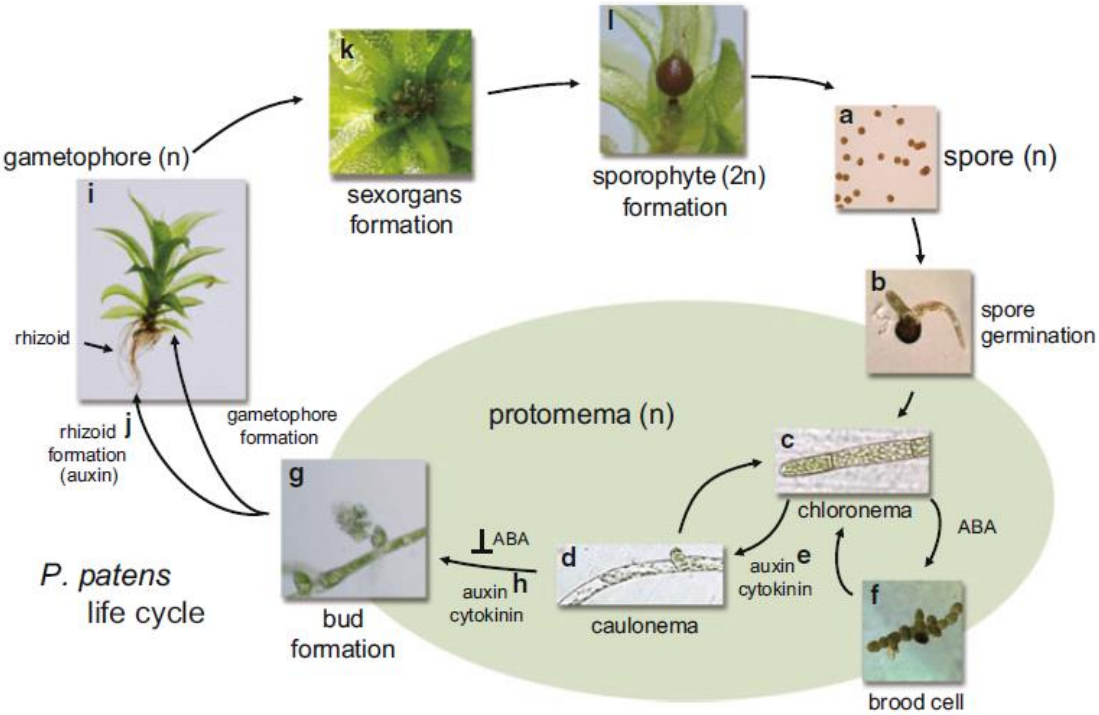


Figure 1.7. Life cycle of the moss *Physcomitrella patens*. a) spores, **b)** spore germinating starts the protonemal stage, **c)** heavily chloroplasted chloronema cells, **d)** adventurous caulonema cells, **e)** protonema differentiation regulated by auxin and cytokinin, **f)** brood cell formation induced by ABA, **g)** bud formation, **h)** bud formation regulated by auxin and cytokinin and repressed by ABA, **i)** leafy gametophore with rhizoids (root-like), **j)** auxin induces rhizoid formation, **k)** short day and cold treatment (15°C) induce sex organ formation, **l)** following fertilisation, the egg cell develops into the diploid sporophyte within which meiosis leads to haploid spore production. Figure and text taken and adapted from Sakata *et al*., 2014.

priori assumptions gleaned from angiosperms. The progress made in developing the genetic resources available for *Physcomitrella* (http://www.phytozome.net/physcomitrella_er.php, Rensing *et al.,* 2008; Zimmer *et al.,* 2013; Kamisugi *et al.,* 2008) has now meant that forward genetics can be employed to identify novel genes involved in ABA-dependent stress signalling. This approach benefits from a good genome assembly with rich genetic marker data to allow mapping of causal loci. Before genome assemblies became available, forward genetics was still possible but extremely laborious often involving map based cloning and chromosome 'walking'. Using the latest genetic resources which have enabled the generation of chromosome scale scaffolds, this project aimed to identify the underlying genotypes in ABA non-responsive (*anr*) mutants.

This approach has resulted in the identification of a novel ABA regulator, PpANR, which is shown to be a crucial component of ABA-dependent stress signalling. This thesis presents functional data showing the phenotypic effects of *Ppanr* mutants on growth and on the ABA responses both at the developmental and molecular level (Chapters 2 and 3). The finding that no clear orthologue of the ANR gene exists in higher plants not only explains why PpANR had not been previously discovered but also begged the questions of its origin. A thorough phylogenetic analysis is also presented which shows PpANR to be part of a basal plant specific subfamily of Raf-like MAP3 kinases closely related to other hormone and stress signalling regulators (Chapter 4). The presence of a PAS domain at the N-terminus of PpANR suggests it to be a potential regulatory domain mediating interactions with upstream signalling components. In order to uncover the function of the PAS domain, various binding interactions were tested, including with ABA, accompanied by the structural analysis of the crystal structure of the PpANR PAS domain (Chapter 5). Together this thesis presents novel data on an ancient, essential but previously unsuspected component of MAP kinase stress signalling. These MAP kinases were established prior to the evolution of the land plants where they were very possibly functional (*e.g.* ethylene signalling - Lu *et al.,* 2015) in the ancestors of land plants suggesting they played a vital role in the conquest of land.

CHAPTER 2

Chapter 2 – Forward genetics in *Physcomitrella patens* identifies the novel ABA regulator PpANR

2.1 Introduction

2.1.1 Why forward genetics?

The use of *Physcomitrella* in demonstrating the functional conservation of many key ABA regulators has been very successful and only the PYR/PYL/RCAR receptors remain uncharacterised of the 'core' pathway. This process has relied on the excellent gene targeting capabilities mediated by homologous recombination (Kamisugi *et al*., 2005; Schaefer *et al*., 1997). The great drawback to this approach is that it is based on *a priori* knowledge, namely what the functions are of ABA regulators in *Arabidopsis*. This largely closes the door on new discoveries unless by chance the presumed homologues in fact have distinct functions. The position of *Physcomitrella*, as a bryophyte, at the base of the land plant lineage means it is well placed for comparative genomics which can help understand the evolutionary processes that led to the origins and proliferation of land plants. This 'comparative' approach should be considered in two directions however, and there are likely features unique to bryophytes that proved surplus to the requirements of subsequent land plants. Equally likely are lineage specific features which indicate the various routes evolution can take and that can often prove interesting and unexpected. For example, the discovery of neochromes in ferns has spawned an interesting and unexpected avenue of research which has shed light on evolutionary processes mediated by novel gene formation. The neochromes are chimeric proteins in which the N-terminal chromophore binding domains from the far-red/red sensing phytochromes are fused with the blue-light sensing phototropins (Kawai *et al*., 2003). These neochromes apparently arose in the hornworts and are thought to have been passed to ferns by horizontal gene transfer (Li *et al*., 2014). Similar neochromes have also been discovered in the charophyte *Mougeotia scalaris* in which they are thought to have arisen independently (Suetsugu *et al*., 2005). Regardless of their interesting origins, these neochromes have been proposed to confer a selective advantage to ferns in low light habitats (Li *et al*., 2014) where they are frequently found beneath the taller 'higher' plants; the gymnosperms and angiosperms. The discovery of neochromes, initially through the analysis of phototropic mutants in *Adiantum*, demonstrates the advantage of forward genetics in discovering the unexpected and novel and the development of this approach in the model *Physcomitrella* will be a powerful tool in understanding some of the early events in land plant evolution.

2.1.2 SNP marker assisted mapping

The forward genetic approach presented here has relied on the use of single nucleotide polymorphisms (SNPs) as genetic markers. These are an abundant form of genomic variation, accounting for around 90% of genomic variation in humans (Collins *et al*., 1998), that can be used as markers. In human genetics, a SNP is defined as having to occur in at least 1% of a population, however in the context of mapping SNPs in crosses where two distinct ecotypes are used the SNP genotyping allows the parental origin of regions of the genome to be identified. SNPs can, in theory, occur anywhere in genomes including in coding regions of genes where they can be associated with natural variation in gene functions and often disease in humans. In *Physcomitrella*, the forward genetic screen employed has taken advantage of the identification of thousands of SNPs between the Gransden and Villersexel K3 laboratory strains. These were used to develop an Illumina genotyping platform able to analyse large numbers of individuals: an important part of mapping where the greater the number of individuals analysed, the greater the possible number of recombination events which define the resolution of mapping. The details of the approach used in this thesis are described in a following section (Chapter 2.2.3); however the basic principles and theory behind SNP based mapping will be described. This approach deals with bi-allelic SNPs (X↔Y) in which allele X occurs in genotype A (Gransden ecotype) and allele Y occurs in genotype B (Villersexel ecotype). The crossing of these genetically distinct ecotypes results in a diploid sporophyte genome containing SNPs from both parents which become shuffled following recombination during meiosis. In mosses, the gametophytes produce the sperm (antheridium) and egg (archegonium) producing organs, and the gametes are produced by mitosis. Following fertilisation the diploid sporophytes form and grow from the archegonium. These form diploid sporangia at their tips which then produce spores through meiosis. The spore-derived progeny are haploids and are equivalent to recombinant inbred lines in diploid species (Figure 1.6). Using tools to implement linkage disequilibrium statistics, the positional relationships between SNPs can be worked out. This process relies on the basic principle that SNPs closer together are less likely to be separated by recombination events occurring between them, which approaches 0% when they are adjacent, while SNPs on different chromosomes (linkage groups), or sufficiently separated on the same chromosomes (*i.e.* at opposite ends) will show no association (50% assortment in progeny).

By implementing this basic mapping process on ABA non-responsive (*anr*) mutants in *Physcomitrella patens*, we hoped to identify novel regulators. The typical wild-type response to exogenous ABA in developing moss tissue is characterised by slow growth and the formation of brood cells (Figure 1.7) (reviewed by Decker *et al*., 2006) which

ultimately blocks the progression of bud formation and gametophore growth. Any mutant in which this response no longer occurs results in a positive mutant phenotype of normally growing plants which, all else being equal, will progress through normal growth and development without the inhibitory effects of ABA being manifest. Any gene that mediates this response is likely to be an ABA regulator that is no longer functional.

2.2 Materials and Methods

2.2.1 PCR and cloning protocols

20µl standard PCR reactions were carried out using Taq polymerase in a 'hot start' protocol in which the Taq mixture (0.5µl Taq, 0.5µl 10x buffer, 4µl H_2O) was added after an initial denaturation step at 95°C for 2mins. Standard mixtures for PCR reactions/sample were as follows: 6.1µl H_2O, 1.5µl 10x buffer, 2µl dNTP (1µM each), 2µl $MgCl_2$ (optimised but usually 15mM) 1.2µl each primer (5µM), 1µl DNA/cDNA template. Templates were generally used at 1ng/µl for cDNA and 5ng/µl for gDNA. Cycles involved 95°C for 2mins, 80°C to load Taq mixture then 35x[95°C for 20s, 58°C for 20s, 70°C for Xs] where X is 1min/1kbp fragment.

For colony PCR reactions, the bacterial colonies were resuspended in 15ul H_2O from which 7.1 µl was used per 20 µl reaction. These reactions were then run as for standard PCRs without the addition of further template.

The proof reading KOD polymerase (Takara) reactions were carried out in 50µl reactions: 27µl H_2O, 5µl 10x buffer, 5µl 2mM dNTP, 3µl 25mM $MgSO_4$, 3µl each primer (5µM), 3µl template, 1µl KOD. 95°C for 2min then 35x[95°C for 2s, 60°C for Xs, 70°C for 2s] where X is 20s/1kbp product. Templates were generally used at 1ng/µl for cDNA and 5ng/µl for gDNA.

Products for cloning were concentrated by ethanol precipitation and phosphorylated by incubating a 25µl mixture (Xµl DNA, 5µl 5x DNA ligase buffer, up to 21µl H_2O, 1µl polynucleotide kinase (PNK)) for 1hr at 65°C where X is variable depending on the concentration of DNA and the required amount to use a 3:1 molar ratio for insert:vector ligation step.

Ligations were carried out in 15µl reactions (10µl PNK mixture, 2µl digested vector (20ng/µl), 1µl 10mM ATP, 1µl 5x ligase buffer, 1µl 1unit/µl T4 ligase (Invitrogen)) incubated over night at 15°C. Ligation reactions were used to transform competent DH5a *E. coli* cells using electroporation.

2.2.2 Creation and screening of *anr* mutants by UV mutagenesis

Protoplasts were isolated as described (Hohe *et al*., 2003; Kamisugi *et al*., 2005) and embedded in "PRM-T" agar medium (BCDAT containing 6% (w/v) mannitol, 10mM $CaCl_2$ and 0.4% agar) on cellophane overlaying the same medium containing 0.55% (w/v) agar in Petri dishes. Approximately 50,000 protoplasts were suspended in 3ml PRM-T for each 9cm Petri dish. These were exposed to 25,000µJ.cm^{-2} ultraviolet radiation (280nm) in a UV Stratalinker 2400, and incubated in darkness for 24 hrs. The plates were then transferred to the growth room (25°C, continuous illumination), and allowed to regenerate. This treatment resulted in *ca.* 10-20% of the irradiated protoplasts surviving and regenerating as protonemata. After 2 days to permit cell wall regeneration, the cellophanes bearing the embedded regenerants were transferred to plates containing standard BCDAT-agar medium (1mM $CaCl_2$, no mannitol) supplemented with 10^{-5}M ABA for 13 days, by which time *anr* mutants were distinguishable and could be routinely subcultured. These steps were carried out by numerous other people under the supervision of Dr. A. Cuming.

2.2.3 Genotyping of the *anr* loci

Isolated mutants were crossed with the Villersexel K3 ('Vx') wild-type by inoculating explants adjacent to each other on BCD agar medium (lacking ammonium tartrate) as described previously (Kamisugi *et al*., 2008). After several weeks of growth, cultures were transferred to inductive conditions (15°C, 8h-light, 16h dark (Hohe *et al*. 2003)) until antheridial development was apparent in the 'Vx' plants. At this point, plants were irrigated with sterile water. The Gransden strain exhibits low levels of male fertility, and consequently the appearance of developing sporophytes on this strain was generally indicative of cross-fertilisation by the 'Vx' parent. Mature spore capsules on the Gransden parent were harvested and were surface-sterilised before releasing the spores by crushing in sterile water. Spores were germinated and the progeny were replica-picked onto medium with and without 10^{-5}M ABA. These steps were carried out by other people under the supervision of Dr. A. Cuming.

DNA was extracted from individual segregants using a CTAB protocol as previously described (Knight *et al*., 2002) and 96 segregants (48 *anr* : 48 w.t.) from the anr4xVx cross and 82 segregants from the anr3xVx cross (45 *anr* : 37 w.t) were genotyped on a custom SNP marker array (GoldenGate, Illumina), comprising 4309 loci, at the Joint Genome Institute Oak Ridge National Laboratory. DNA was extracted by other people under the supervision of Dr. A. Cuming and the genotyping was developed and carried out by Gerald Tuskan and Wellington Muchero at the Joint Genome Institute. Data were analysed by a custom Python script (Appendix 2.1) which essentially calculated the genotype ratio for each SNP marker in the *anr* and wild-type segregants

respectively. The same process allowed chromosome scale mapping using a separate mapping population in order to work out the relative order of SNP markers. The analysis allowed the identification of Gransden-specific SNPs co-segregating with the *anr* progeny and the reciprocal Vx-specific SNPs co-segregating with the wild-type progeny (Table 2.1). These SNPs were then located on the *Physcomitrella* genome assembly to define the physical limits of the genetic interval. SNP specific primers from within this region were used to confirm the genotyping.

Table 2.1. List of primers used in the identification, creation and genotyping of *Ppanr* mutants.

Primer name	Sequence	Use
462_10F	TAGCTCCAGGTCACACCCAGAT	*anr4* sequencing; transformant genotyping
462_10AR	TGGGGACGCAGTTCACTTAC	*anr4* sequencing; PpanrKO construct
462_10BF	GAACACGGTGGTGACACAGAAG	*anr4* sequencing
462_10BR	ATGTGGCTTGGGGAGACATAGA	*anr4* sequencing
462_10CF	GAAATGGAGATCCACACCTTCG	*anr4* sequencing; *Ppanr4PTC* construct
462_10CR	TAAAGGTGGAGGCATTGGAAGA	*anr4* sequencing; *Ppanr4PTC* construct
462_10DF	TTTGCGAGGGAGTTTGAGATTC	*anr4* sequencing
462_10R	AGAGCGTAGCAATACAGGGTGG	*anr4* sequencing; transformant genotyping
E6_F	TGGTGGGAGCAGTGTTTCAGTA	*anr4* PTC sequencing
E6_R	TGGAAAAGGAGAACAGCTCACC	*anr4* PTC sequencing
ANR_KOS	GGAAAGATTTCGTGGGATTGG	*PpanrKO* construct
ANR_3S	CCATCGTCATGCTGACTTCAAT	*PpanrKO* construct
ANR_3A	TGGCCTCACAATACAACTACGC	transformant genotyping
462_57138FG	ACTTTCGATCTAAATCTGAGTGGAG	SNP genotyping of A4V segregants - Gr specific
462_57138FV	ACTTTCGATCTAAATCTGAGTGGAA	SNP genotyping of A4V segregants - Vx specific
462_57138R	GGCAACGTACAAAAGGATTGA	SNP genotyping of A4V segregants

2.2.4 Identification of the causal mutation

Sequencing of candidate genes was carried out by amplification of fragments from a single *anr* segregant using KOD polymerase (Takara) to amplify overlapping segments which were cloned in pBluescriptKS- and sequenced (Source Bioscience, UK) using both universal and custom primers (Table 2.1). Identification of a CAG>TAG nonsense mutation in the gene Phpat.012G009800 (Pp1s462_10V6; Phypa_30352;

Pp3c12_3550) was confirmed by amplification and sequencing of a 350bp fragment containing this site in multiple *anr* and wild-type segregants by direct PCR product sequencing in which PCR products were recovered from agarose gels by a freeze-squeeze method. This involved excising the gel fragment, freezing it in liquid N_2 before squeezing out the eluate between Para-film and collecting this with a pipette.

2.2.5 Creation of *Ppanr* mutant lines

The candidate *anr4* locus was functionally confirmed by (i) targeted deletion of the protein-coding sequence and (ii) targeted point mutagenesis of the mutated base to generate a CAG>TAG mutant in isogenic Gransden genetic background. For targeted deletion, a strain '*PpanrKO*' was created by transforming wild-type 'Gransden 2004' protoplasts with a DNA fragment containing a central CaMV35S-nptII-CaMVter selection cassette flanked by regions of homology corresponding to the 5'- and 3'-ends of the Phpat.012G009800 locus. The targeting sequences comprised 1,116bp (chromosome 12 sequence coordinates 2,809,033-2,810,198, comprising the first two exons of the gene) at the 5'-end and 878bp (coordinates 2,817,434-2,818,311 comprising the last two exons of the gene) (Table 2.1). The transgene was amplified from a pBluescriptKS- based clone and used to transform protoplasts (Hohe *et al*., 2003; Kamisugi *et al*., 2005). Targeted point mutagenesis was undertaken by marker-free transformation of wild-type protoplasts with a 1,246bp PCR amplicon derived from the *anr4* mutant (sequence coordinates 2,813,530-2,814,775) containing the C>T mutation at position 2,814,230 approximately centrally located (Table 2.1). This resulted in the creation of a number of *Ppanr4PTC* lines showing the *anr* phenotype which were isolated for further analysis. Transformants containing single copy inserts, no wild-type Phpat.012G009800 sequence and no ectopic insertions were identified by the standard methodology using gene- and transgene-specific PCR and Southern blot analysis (Kamisugi *et al*., 2005). The entire ANR locus in the *Ppanr4PTC* mutant lines were additionally sequenced to verify that no unintended mutations had been introduced (Table 2.1).

2.2.6 Growth assays

The *anr* response was assayed by measuring the growth of explants, small tweezer-picked fragments of protonemal tissue, inoculated on both normal (BCDAT) and ABA (10^{-5}M) supplemented media. Two replica plates were set up in which each was split into three equal sections within which 8 explants from each line were inoculated with even spacing. These lines included the wild-type Gransden strain, *PpanrKO*, *Ppanr4PTC* and the ABI3 triple knockout *Ppabi3* (Khandelwal *et al*., 2010). Digital photographs were taken every 3-4 days and analysed using ImageJ to calculate mean

colony size over time (Kamisugi *et al*., 2012). Data processing included removing anomalous explants such that the minimum sample size analysed per line was n=13.

For the ABA growth response assay, small explants (n=8) of protonemal tissue were inoculated on plates with or without 10^{-5}M ABA. Plates included spot inocula from the *PpanrKO* mutant alongside 'Gr' WT and the triple ABI3 knockout line *Ppabi3ko* (Khandelwal *et al*., 2010) as controls. Digital photographs were taken every 3-4 days and analysed using ImageJ to calculate mean colony size over time (Kamisugi *et al*., 2012). For long term growth, spot inocula were grown in glass jars containing 10ml BCDAT + agar medium and kept well hydrated for 3 months.

2.3 Results

2.3.1 Identification of ABA non-responsive mutants displaying strong ABA resistance.

Protoplasts in the Gransden background were irradiated with UV-C to induce mutations. Mutations affecting ABA growth response were identified by generating the protoplasts on an ABA containing medium. If wild-type protoplasts or protonemal explants are regenerated in the presence of ABA, the cells exhibit a characteristic change in growth habit characterised by cellular differentiation producing brachcytes ('brood cells') – small, round, thick-walled non-vacuolate cells in which vacuolation and cell expansion is suppressed (Figure 2.1) – interspersed with tmema cells (empty, thin-walled cells that enable dispersal of dehydrated brood cells that can act as "vegetative spores") (Decker *et al.*, 2006). The consequence of this cellular change is a highly reduced growth rate in which regenerating plants have a dwarf, twisted protonemal morphology.

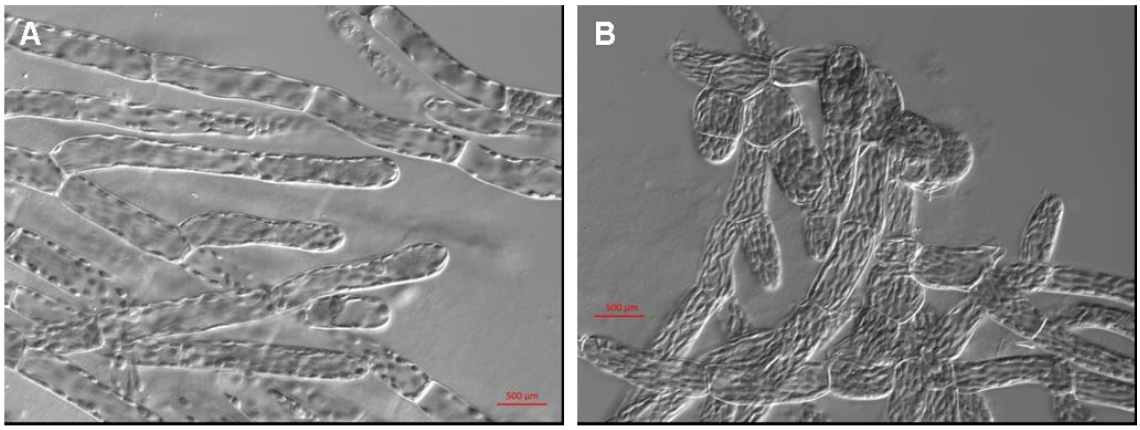


Figure 2.1. Wild type response of developing protonemal tissue to ABA. **A)** Filaments are elongated in normal conditions but **B)** with the addition of ABA, brood cells form which are rounder thicker celled cells more able to survive dehydration. Credit Y. Kamisugi.

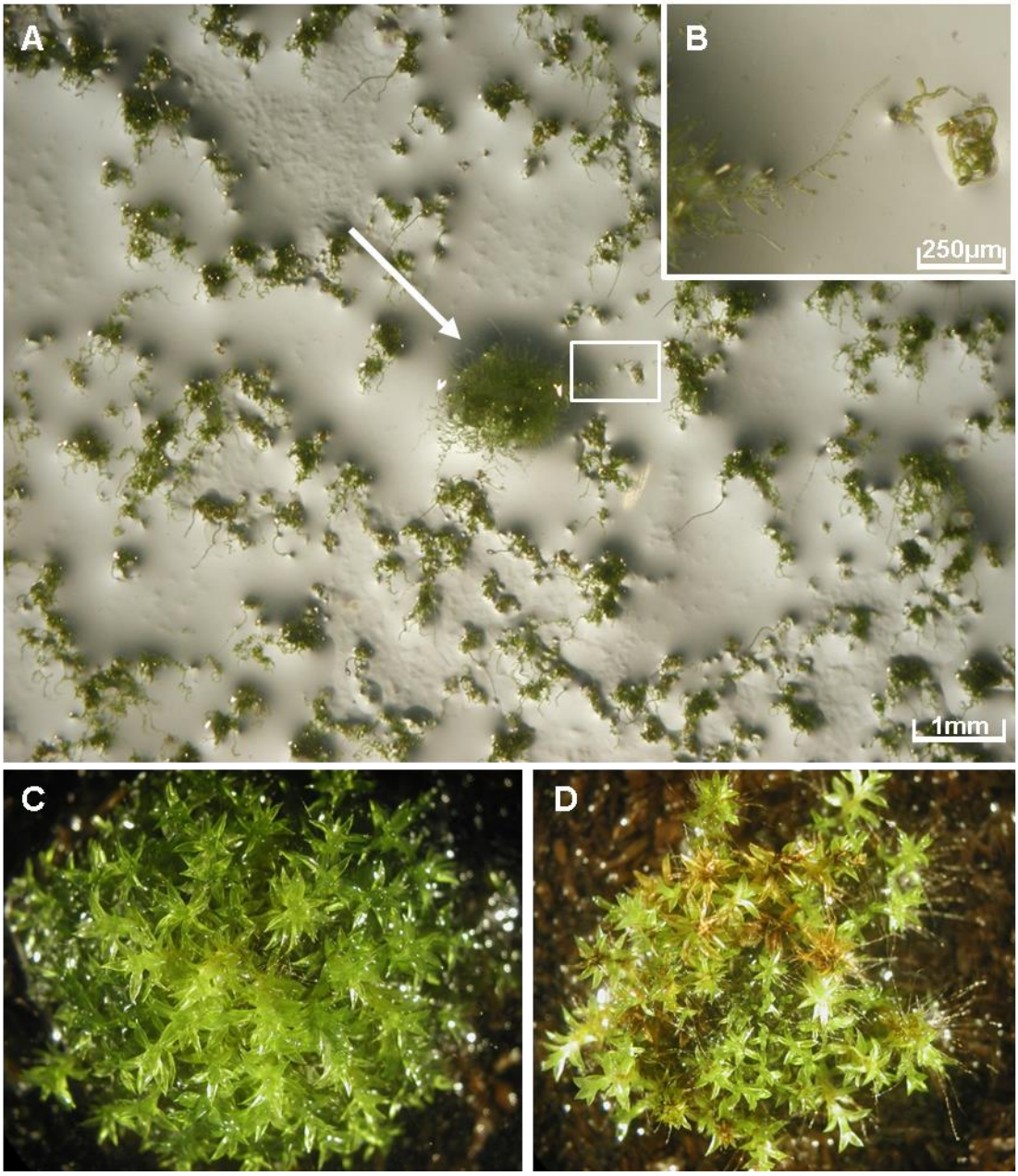


Figure 2.2. Identification and phenotype of UV mutagenized *anr* mutants. A) Identification of *anr* mutants (arrow) was clear as UV irradiated protoplasts regenerated 'normally' in the presence of ABA. **B)** Close up of *anr* mutant and wild-type neighbour (white box) shows normal growth habit with extensive filamentous growth as though ABA was not present which is contrasted with the neighboring wild-type colony displaying the typical ABA growth response in which growth is severely retarded accompanied by brood cell formation. **C)** Wild-type colonies develop into lush green plants in normal conditions but **D)** *anr* mutants were dehydration hypersensitive showing premature senescence of gametophytes unless continuously irrigated. Credit A. Cuming

In this way, mutants in the ABA response were identified as fast-growing plants, normal in appearance despite the presence of ABA, which clearly contrasts with the wild-type response (Figure 2.2A, B). These mutants also exhibited drought hypersensitivity, the aerial gametophore tissue undergoing premature senescence unless continually irrigated (Figure 2.2C, D). A number of independent mutants showing this ABA non-responsive (*anr1-7*) phenotype were isolated. These lines were created and analysed by other people under the supervision of Dr. A. Cuming.

2.3.2 Mapping of causal loci.

The independent mutants, designated *aba non-responsive* (*anr1-7*), were cross-fertilised by co-culture with plants of the genetically divergent 'Villersexel K3' strain (Vx) in order to generate segregating populations (Kamisugi *et al*., 2008). Hybrid diploid sporogonia recovered from the mutant (maternal; Gr) plants were identified by spore germination and growth testing of the sporelings on ABA-supplemented medium: hybrid sporogonia yielded spores that segregated 1:1 for the *anr* and wild-type phenotypes. In *Physcomitrella*, the gametophyte produces gametes by mitosis that fuse to generate the diploid sporophyte. This in turn generates haploid spores by meiosis, so the spores obtained from hybrid sporogonia represent a segregating population containing recombinant chromosomes – the equivalent of recombinant inbred lines in diploid species. Ninety-six segregants (48 mutant and 48 wild-type) from the *anr4* x Vx cross (A4V) were genotyped at 4309 SNP markers, using an Illumina GoldenGate genotyping platform, to identify a genetic interval containing the mutant locus. This was the same platform that was used to genotype a mapping population of the same cross (Gr v Vx) (Kamisugi *et al.*, 2008) to guide the recent chromosome scale genome assemblies for *Physcomitrella patens* (V3.0 and 3.1). Plotting the SNP genotype ratio of Gr alleles for the 48 A4V mutants across each chromosome in the *Physcomitrella* V3.0 sequence assembly revealed a single peak on chromosome 12. The converse relationship was observed between the wild-type phenotype and the Gr alleles (Figure 2.3A). PCR based genotyping using selected SNP-specific primers confirmed the mapped region as segregating with the phenotype in the A4V segregants (Figure 2.4A). All *anr* segregants displayed the Gr genotype while all wild-type segregants displayed the Vx genotype as expected. The genetic interval contained 3 mapped SNPs, encompassing a region of 88,366bp containing 8 gene models (Figure 2.4B). The same process was carried out for the *anr3* x *Vx* (A3V) cross which resulted in the identification of a larger but overlapping region as to that for *anr4*. The minimum interval for *anr3* was found to be 206,036bp and the genotype ratio of Gr alleles for 45 *anr* mutants and 37 wild-type segregants across chromosome 12 is shown (Figure 2.3B). Unlike for the *anr4* mapping, the minimum interval was obtained from

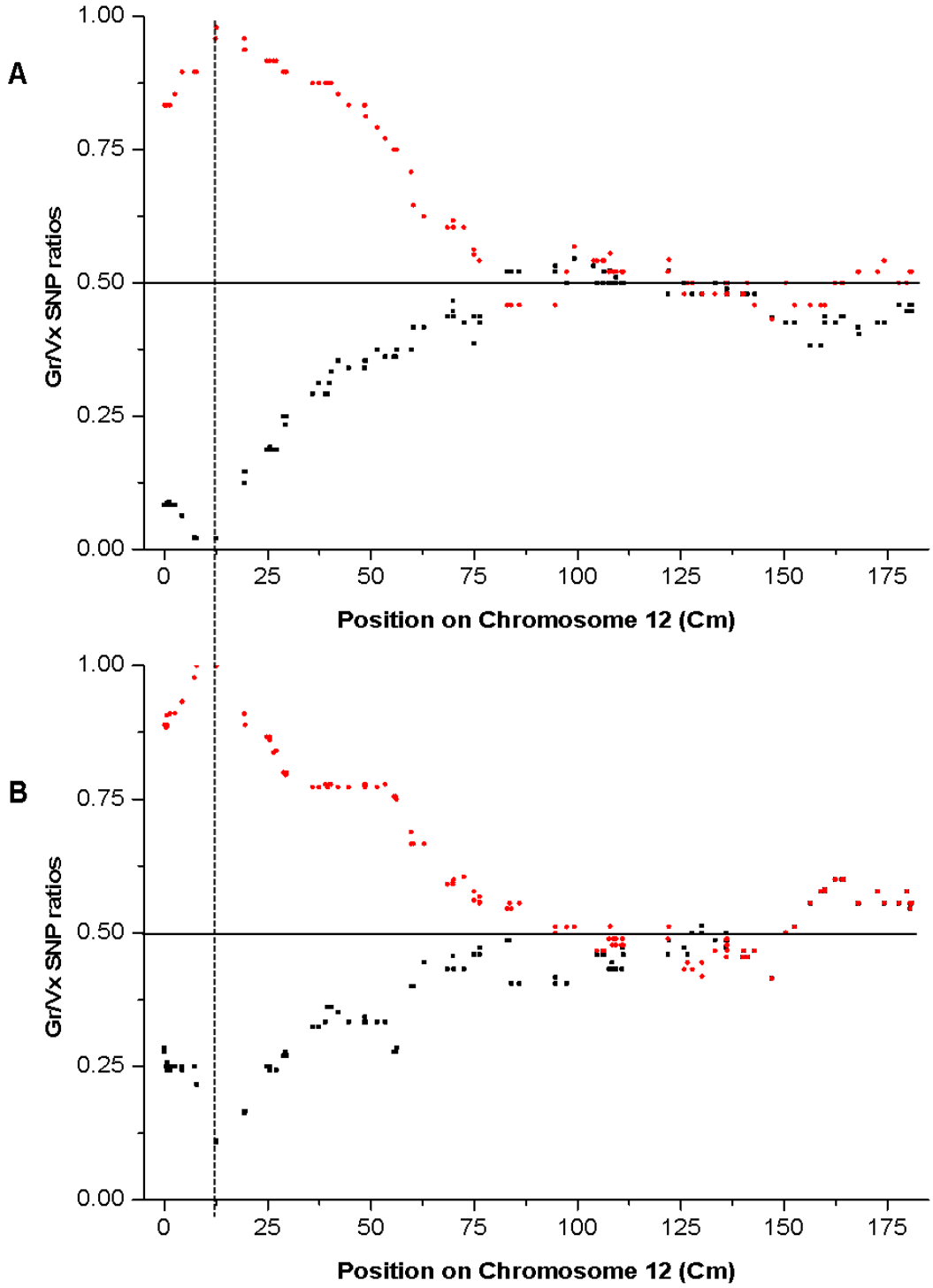


Figure 2.3. Identification of overlapping regions on chromosome 12 co-segregating with the phenotype in both A4V and A3V segregants. Plotting of the genotype ratios for SNP markers in either the *anr* (red circles) or wild-type (black squares) segregant progeny from **A)** the A4V (*anr4* x Vx) and **B)** A3V (*anr3* x Vx) crosses. SNP ratios of 0.5 represent regions that show no co-segregation (random assortment) with the phenotype whereas values approaching 1 represent regions of the genome in almost all *anr* mutants that originated exclusively from the Gr mother while values approaching 0 represent regions in the wild-type progeny originating exclusively from the Vx father. These regions are most likely to harbor the causal mutation underling the *anr4* and *anr3* phenotypes respectively. These mutations were presumed to be allelic as they map to overlapping regions.

recombination events in the wild-type segregants. This region, while over twice as big as that defined by the *anr4* mapping, still only contains 12 gene models (Figure 2.4B). This suggested these mutations were allelic and candidate genes were sequenced in both mutant lines.

2.3.3 Identification of the *PpANR* locus.

In order to identify the causal mutation, candidate genes within the mapped region required sequencing. Sequencing of overlapping PCR fragments was carried out on mutant (*anr*) segregants from both the A4V and A3V crosses. Of the 8 gene models within the shared mapped region of both mutants, the smaller ~88kbp region defined in the *anr4* mapping, a number were prioritised based on functional annotations. One was predicted to encode a subunit of CCAAT-box binding nuclear transcription factor Y (NF-Y) (specifically subunit A) and the other a trimodular MAP3-like kinase. NF-Y is a heterotrimeric complex involved in transcription and these have been implicated in ABA signalling through interaction with ABI5-like bZIP TFs in *Arabidopsis* (Yamamoto *et al.,* 2009) and with ABI3 in *Physcomitrella* (Yotsui *et al.,* 2013). Sequencing of the NF-YA gene, however, found no likely mutations in the coding regions of either the A3V or A4V mutants. Sequencing of the MAP3K-like kinase found no mutations in the coding region of A3V mutants but did reveal a point mutation predicted to introduce a premature stop codon (PTC) (CAG→TAG: Q^{712}Ter) upstream of the kinase domain in the A4V mutants (Figure 2.5A). This mutation was confirmed by sequencing of this region in 5 further mutant and 6 wild-type A4V segregants in which the PTC was found exclusively in the *anr* mutants (Figure 2.5B).

This gene is referred to as *PpANR* (also known as Pp1s462_10V6; Phypa_30352; Phpat.012G009800.1; Pp3c12_3550 in the V1.6, 1.2, 3.0 and 3.1 assemblies, respectively (http://www.cosmoss.org; http://phytozome.jgi.doe.gov/pz/portal.html#)) and was recently shown to have a dual role in ABA and ethylene signalling (Yasumura *et al.,* 2015); making it a regulator of water responses. The wild-type *PpANR* gene encodes a MAP3 kinase-like protein containing two additional conserved domains: a central “EDR” domain, found in the *Arabidopsis* Raf-like MAP3 kinase ethylene and pathogen response regulators CTR1 and EDR1 respectively (Kieber *et al*., 1993; Frye *et al*., 1998), and an N-terminal PAS domain: a small and ancient sensor domain best described in bacterial ‘two-component’ systems but also found in a range of eukaryotic response regulators (Henry and Crosson, 2011).

What are the extra four genes found in the *anr3* mapped region as compared to the *anr4* mapping region? One encodes a putative histone acetylation protein (Phpat.012G009700), another encodes a putative RNA binding FET family protein

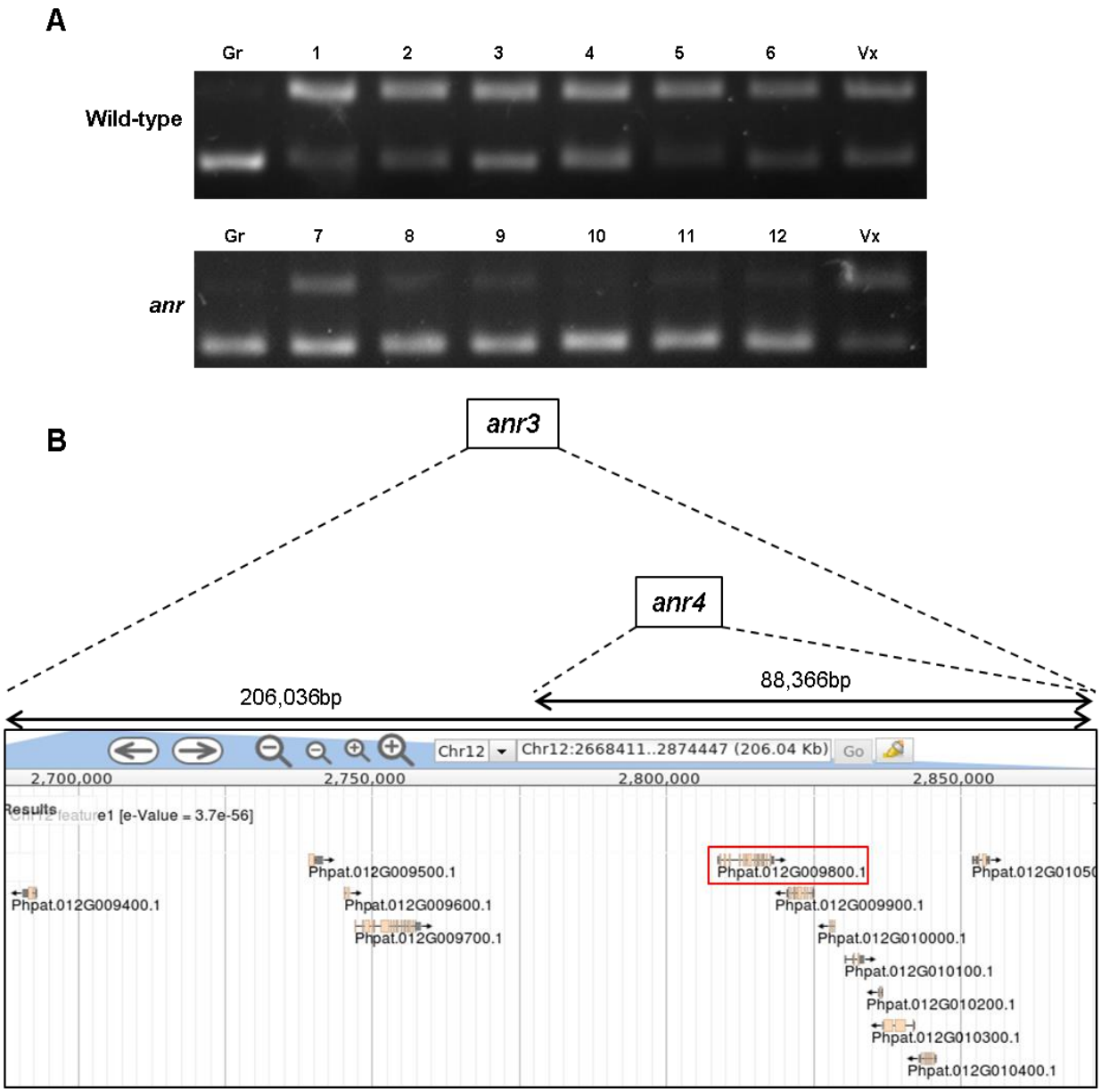


Figure 2.4. Confirmation and visualization of mapped regions for *anr* mutants. A) SNP specific primers (for a SNP at position 2823714 on chromosome 12) were used to confirm the GoldenGate genotyping in A4V segregants. All 6 wild-type segregants (1-6) showed the Vx genotype while the 6 *anr* mutants (7-12) showed the Gr genotype. **B)** The overlapping mapped regions for *anr3* and *anr4* viewed on the V3.0 genome browser. The *anr3* mutation mapped to a region of 206,036bp containing 12 gene models. The *anr4* mutation mapped to a region of 88,366bp containing 8 gene models. The *PpANR* locus is highlighted.

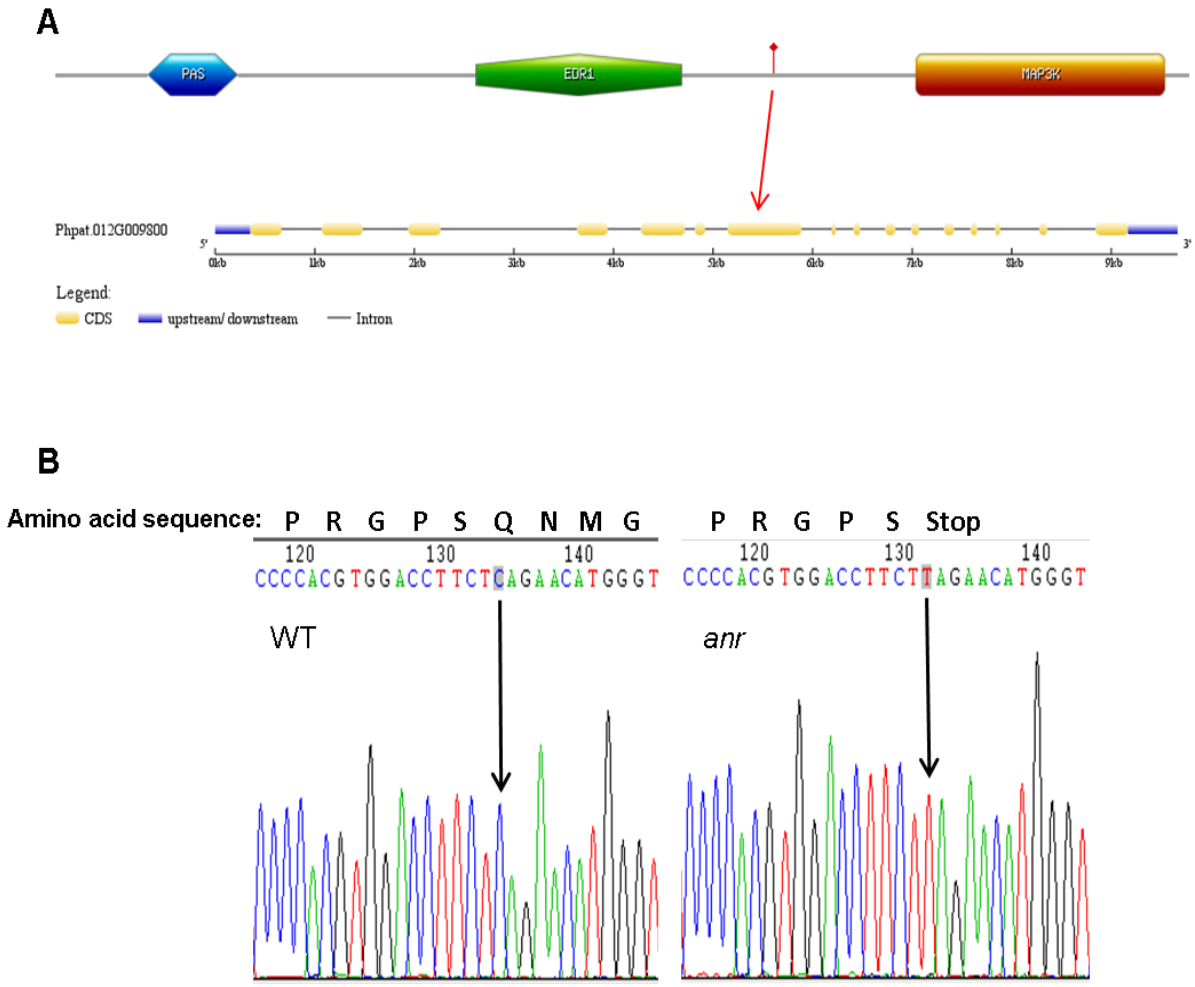


Figure 2.5. The *anr4* mutation induces a premature stop codon upstream of the kinase domain in PpANR. A) Schematic of *PpANR* locus made with Prosite showing the trimodular structure: an N-terminal PAS domain, a central EDR domain and a C-terminal MAP3K-like kinase domain. The location of the predicted premature stop codon is indicated by the red dot. This lies in exon 7 at residue 712 (CAG→TAG; Q^{712}Ter) which is upstream of the kinase domain predicted to result in a truncated and non-functional protein. **B)** Following identification of C to T mutation in a A4V mutant, sequencing around this position in a number of further A4V mutants and wild-type segregants showed this premature stop codon (PTC) to co-segregate with phenotype with unambiguous traces from Sanger sequencing. This strongly suggested it as the causal mutation.

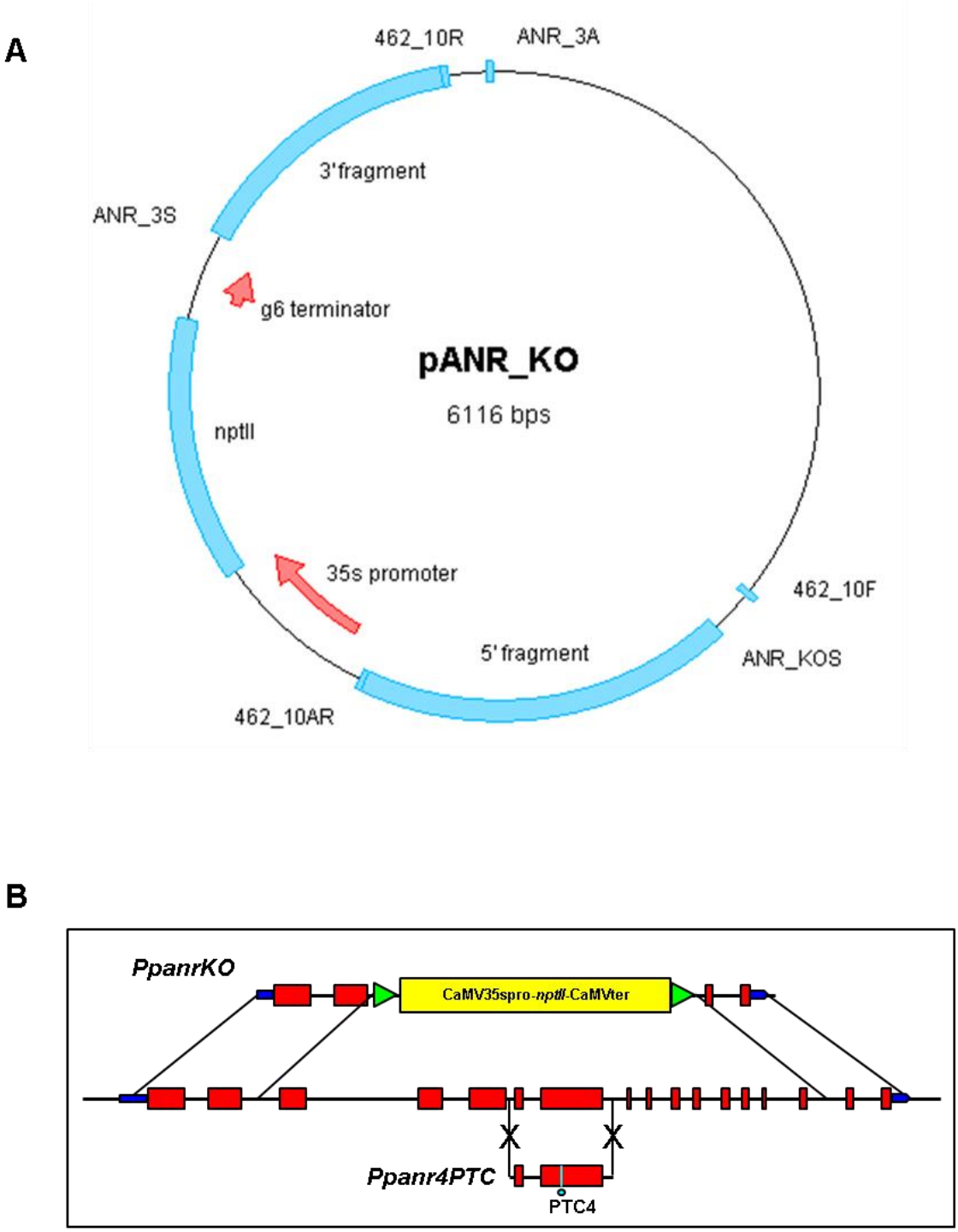


Figure 2.6. Constructs used in the creation of *Ppanr* mutants. A) Schematic of the clone used in *PpanrKO* construction. Two ~1kbp regions of homology flank a central neomycin phosphotransferase (*nptII*) selection cassette. Various primers used are detailed in Table 2.1. **B)** Schematic of the transgenes delivered to wild-type protoplasts to make the *Ppanr* mutant lines. In the null mutant, *PpanrKO*, most of the coding region was deleted and replaced with the *nptII* selection cassette enabling marker assisted transformant selection. The marker-less *Panr4PTC* lines relied on the mutation conferring the *anr* growth phenotype to enable screening of transformants.

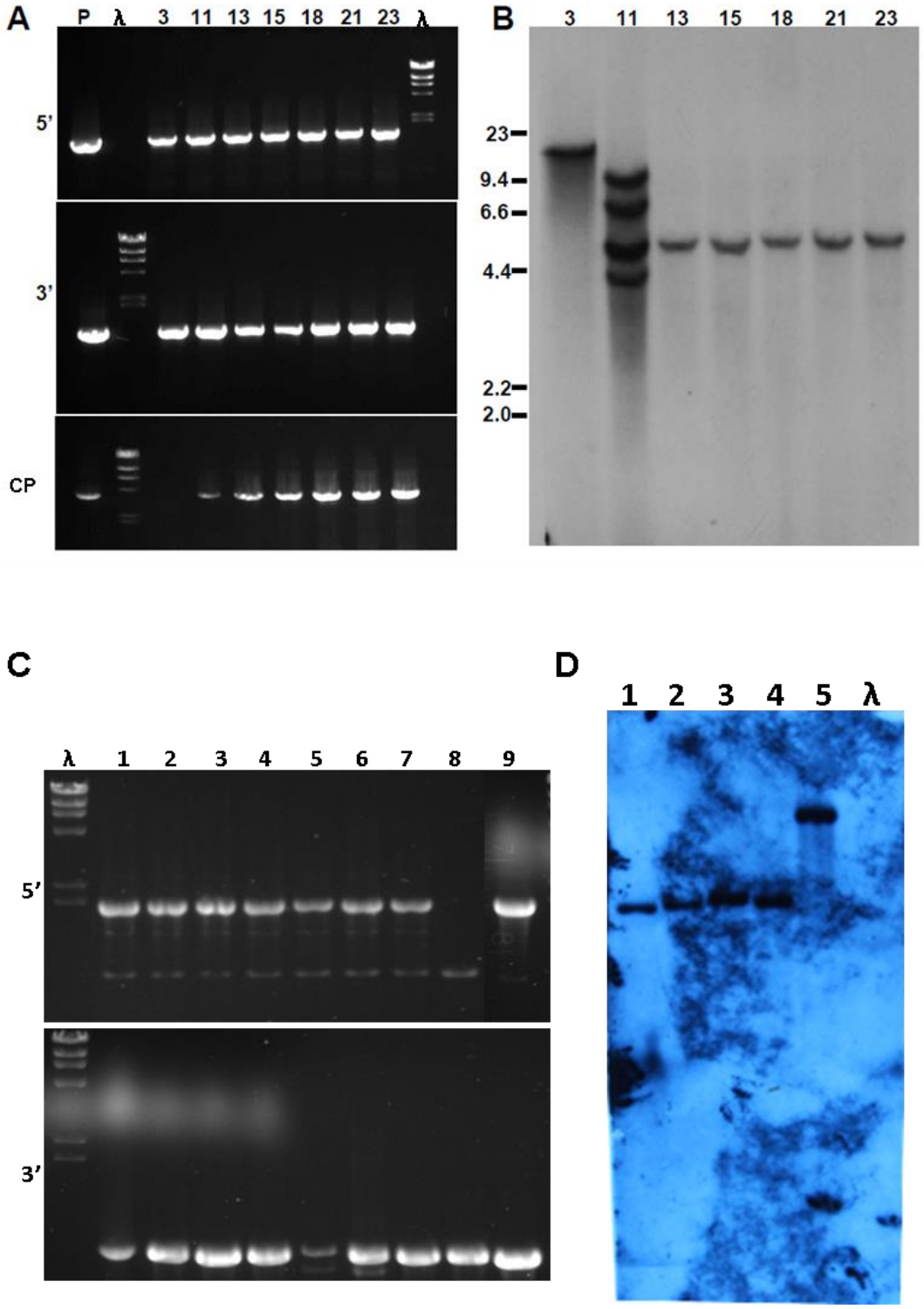


Figure 2.7. Genotyping of *Ppanr* mutant transformants. A) PCR checks for targeting of the *PpanrKO* transgene to the *ANR* locus using both 5' and 3' specific primers. Insert copy number with external primers shows line 3 to have a concatemer. **B)** Southern blot analysis confirms line 3 as having a concatemer while line 11 shows ectopic insertions (Credit A. Cuming). Lines 13, 15, 18, 21 and 23 are single insertion single copy transformants. **C)** PCR targetting checks for *Ppanr4PTC* transformants. **D)** lines 1-4 were single copy single insert transformants while line 5 had a concatemer. Lambda *Hind*III molecular marker was used (λ) and plasmid control when shown (P).

(Phpat.012G009500) and one encodes a putative lipase (Phpat.012G009400) with one likely pseudogene. The failure to find a mutation in the coding region of *PpANR* in the *anr3* mutants may be due to the mutation lying in more distant *cis*-regulatory regions or coincidentally in a neighbouring gene. These 4 'extra' genes do not have annotations that strongly suggest they could be harbouring the mutation but, if prioritising of these for sequencing was to be done, the putative RNA binding and histone acetyltransferase are the most likely candidates for a signalling role. Sequencing in the 5' UTR and flanking genomic region of *PpANR* in *anr3* mutants was not thoroughly pursued due to time constraints but sequencing of the first 2000bp 5' flanking region was almost complete (missing the central 353bp) revealing a single base pair mutation at position 2,808,172; a G→T substitution. However, this potential mutation is not predicted to affect any predicted *cis*-elements. Further analysis, including possible fine-mapping of the *anr3* mutation, is necessary to identify the causal mutation.

2.3.4 Knockout of the *PpANR* locus induces ABA non-responsiveness

In order to confirm the mapping and identification of the PTC upstream of the kinase domain as the causal mutation behind the *anr* phenotype in *anr4*, targeted gene knockout of the *PpANR* locus, in which all but the first and last two exons were replaced with a selection cassette, was first carried out to produce a line known as *PpanrKO* (Figures 2.6 and 2.7A, B). The *PpanrKO* mutant demonstrated a clear *anr* phenotype by growing uninhibitedly in the presence of ABA (Figure 2.8). The uninhibited growth in the presence of ABA was comparable to that seen in the triple knockout mutant of the *PpABI3 genes*: *abi3ko* (Khandelwal *et al.,* 2010). The growth curves in normal conditions (no ABA) suggest that *PpanrKO* plants grew significantly slower than either wild-type or the *Ppabi3* mutant. While this may represent a genuine feature of this mutant it is worth noting that the condition of *Physcomitrella* cultures is dependent on the frequency of sub-culturing as tissue 'fatigues' during long term storage (at 4°C). Cultures from long term storage do not grow as fast as those from a subculture of a continuously growing culture. While efforts were made to keep all lines in similar states, *Ppanr* mutants may be more susceptible to this effect given its dramatic effect on ABA responses (Chapter 3); an important hormone required for normal growth and development in land plants, and for tolerance to chilling (Minami *et al*., 2003).

2.3.5 An otherwise isogenic line containing the *anr4* PTC, *Ppanr4PTC*, also shows the *anr* phenotype

While the deletion of the *PpANR* coding region confirmed it as an ABA regulator, additional evidence for the role of the PTC identified in *anr4* was sought. We generated a targeted point mutation in which the wild-type exon 7 "CAG" in *PpANR* was converted

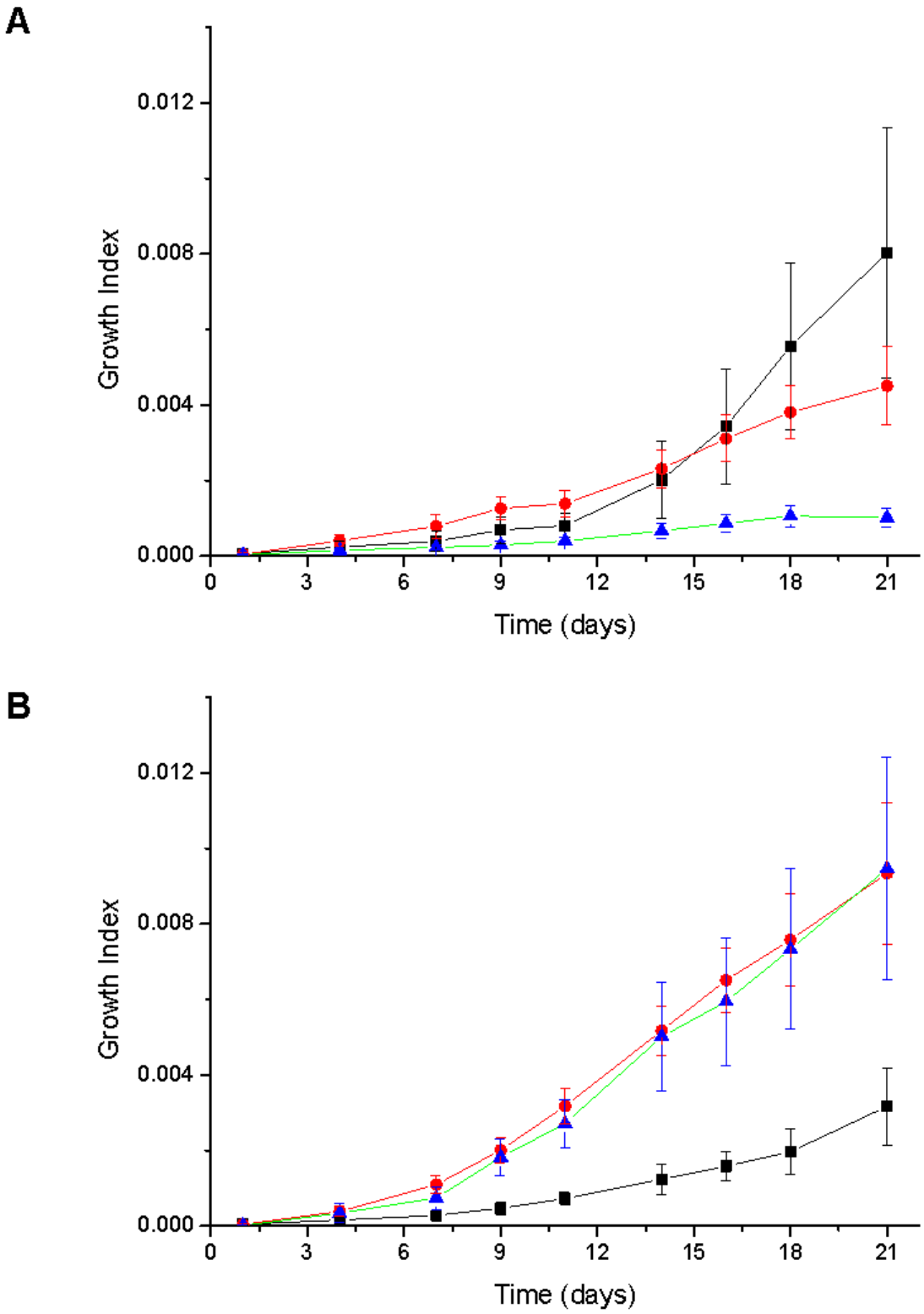


Figure 2.8. Growth assay of *PpanrKO* mutant confirms *anr* phenotype. Growth assay comparing the growth rates of explants from wild-type (blue triangle with green line), a triple ABI3 knockout (*Ppabi3*) (red circles with red line) and *PpanrKO* (black squares with black line) lines on **(B)** normal BCDAT media and on **(A)** media supplemented with 10^{-5} M ABA. *PpanrKO* grows slower in normal conditions than the wild-type and *Ppabi3* lines but clearly shows an ABA non-responsive growth phenotype in the presence of ABA; comparable to that of *Ppabi3*. The typical wild-type response in the presence of ABA is to dramatically reduce growth instead entering a state of quiescence. The growth index is a mean colony size as measured by ImageJ.

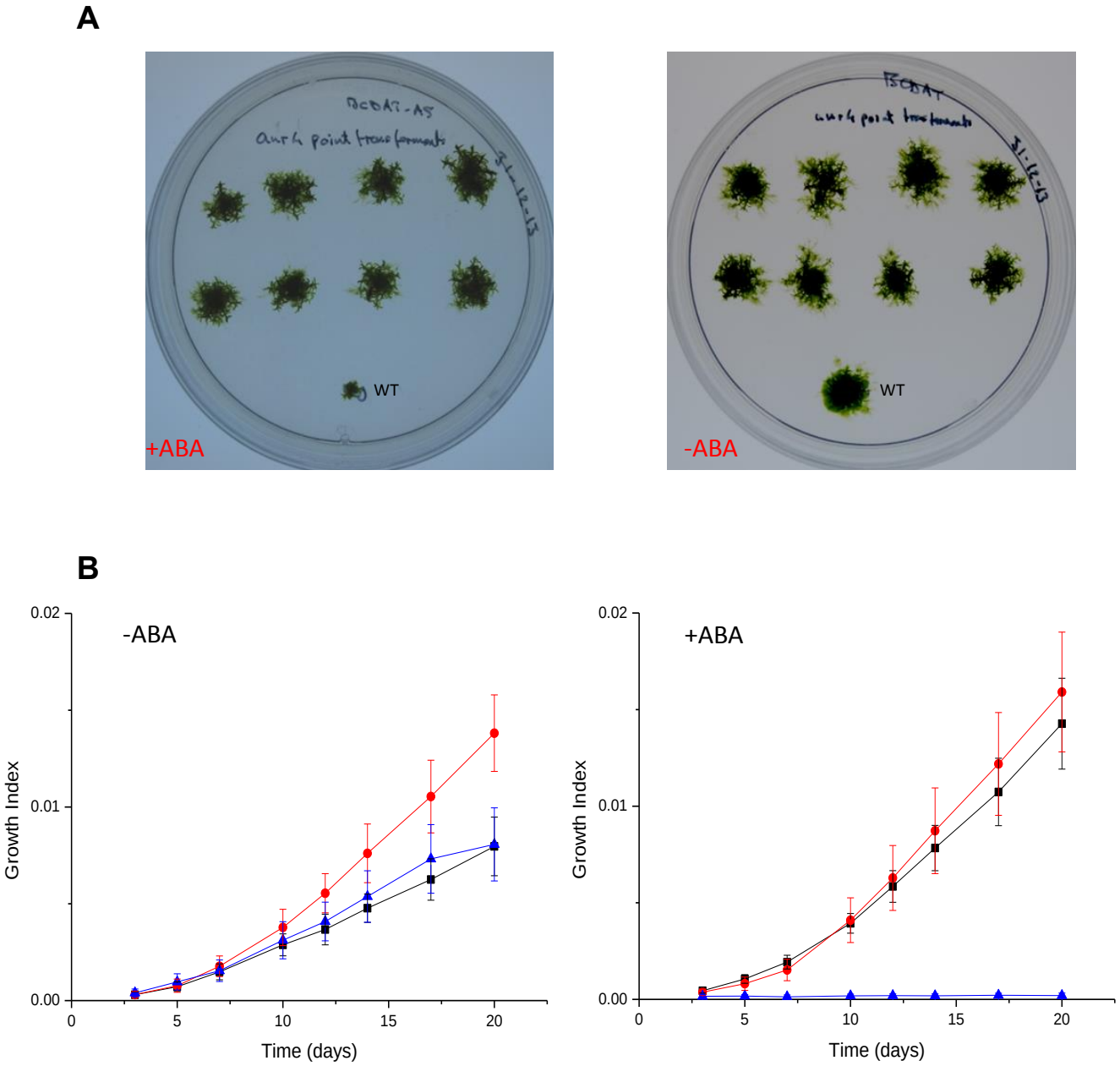


Figure 2.9. Screening and isolation of *Ppanr4PTC* transformants showing the *anr* phenotype. A) 8 transformants were identified by their uninhibited growth on ABA-containing medium following transformation of protoplasts with DNA containing the *anr4* premature termination codon (PTC). Isolated mutants were confirmed as ABA non-responsive by growing explants for 24 days with or without ABA (10^{-5} M) alongside a 'Gr' WT colony. **B)** Selected single copy and single insertion *Ppanr4PTC* lines (black squares) were re-analysed alongside the *PpanrKO* (red circles) and wild-type (blue triangles) lines. The two *Ppanr* mutant lines showed strong ABA non-responsiveness confirming the PTC identified in the *anr4* mutant as sufficient to confer ABA non-responsiveness. The growth index is a mean colony size as measured by ImageJ.

to “TAG” thus generating an *anr4* point mutant line (*Ppanr4PTC*) otherwise isogenic with the Gr background (Figures 2.6 and 2.7C, D). This similarly displayed an *anr* growth phenotype (Figure 2.9) confirming the identification of the point mutation as sufficient to induce the *anr* phenotype which showed comparable growth kinetics as that seen in *PpanrKO*. While the growth assay which compared *PpanrKO* and *Ppabi3* showed clear differences between growth of *PpanrKO* tissue between the control and ABA conditions (Figure 2.8), this was not the case in the comparison with *Ppanr4PTC*. It is worth noting too, that the explants grew to a greater size (the growth index values are normalised against the size of the petri dishes so comparisons are possible) in this assay which may reflect differences in the conditions between the plates used; older plates may be drier for example. *PpanrKO*, given its lack of a key hormone regulator, may be disproportionately affected. This assay did, however, demonstrate that the kinase domain is vital for function of PpANR as an ABA regulator, suggesting a role in ABA-dependent phosphorylation of downstream targets.

2.3.6 *Ppanr* mutants fail to show ABA-inducible gene expression

In order to confirm the role of PpANR as an ABA regulator at the molecular level, semi quantitative PCR analysis was carried out using a number of known ABA responsive genes as identified from a previous study (Cuming *et al*., 2007). In wild-type tissue, clear induction of ABA-responsive genes was found which was largely abolished in the *Ppanr* mutants (Figure 2.10). In the wild-type, ABA-induction of transcripts was largely correlated with mannitol and dehydration treatments with the notable exceptions of “*Up1*” and *PpLea1*. Interestingly, ABA/stress-induction appears to be partially retained in *Ppanr4PTC* for some genes. This may reflect that the truncated protein, lacking the kinase domain, has a partial function through interactions of the PAS and/or EDR domains. This would require further analysis however as the power of this technique, especially when dealing with low expression, is not sufficient to make firm conclusions. Both the deletion of the *ANR* gene and the truncation before its kinase domain do however clearly result in a dramatic loss of ABA-dependent gene expression.

2.3.7 *Ppanr* mutants fail to respond to ABA pre-treatment

We have seen that *Ppanr* mutants fail to respond to ABA both at the phenotypic and molecular level. To see whether this translated into a direct loss of ABA-mediated stress tolerance an assay based on Khandelwal *et al*., 2010 was carried out. *Physcomitrella patens* is defined as dehydration but not desiccation tolerant but is able to increase this tolerance through slow drying. This process induces ABA-dependent gene expression encoding protective proteins which enable survival and the pre-treatment with exogenous ABA can mimic and greatly strengthen this effect. In wild-type tissue, pre-treatment for 24hrs with a range of ABA concentrations (10^{-6}, 10^{-5} and

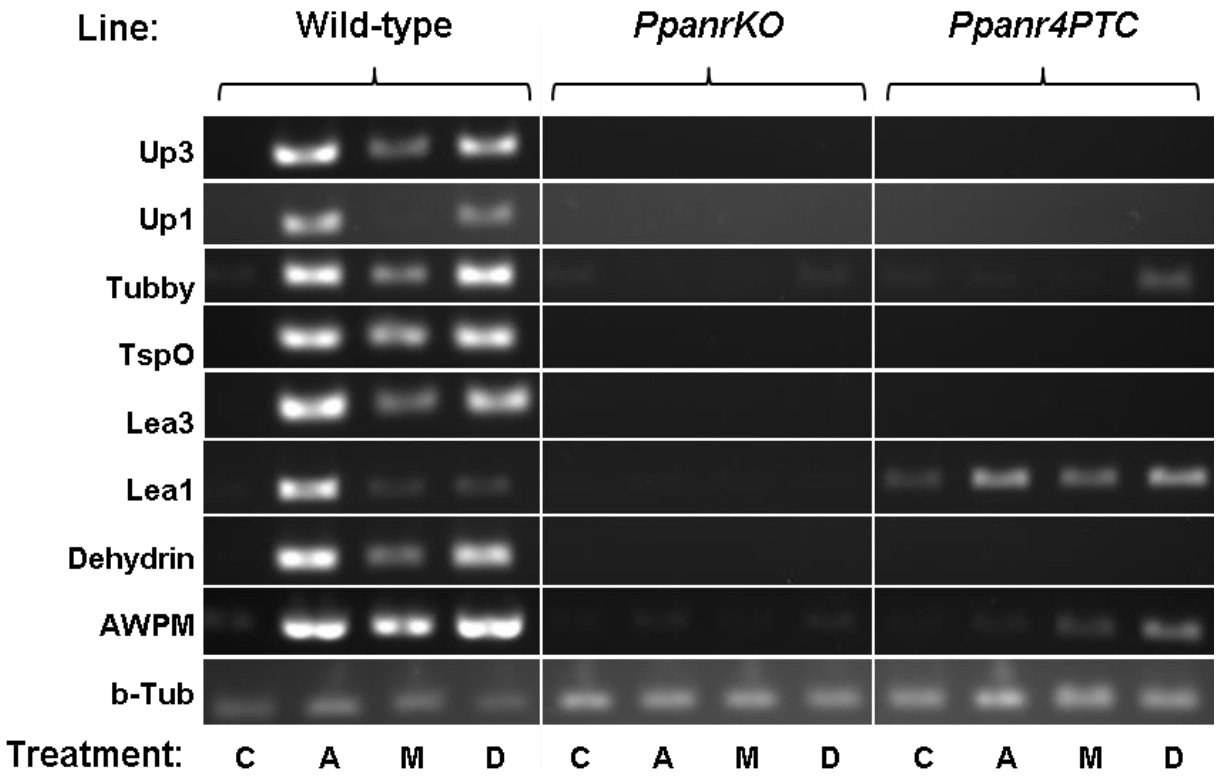


Figure 2.10. *Ppanr* mutants fail to induce ABA and stress-responsive gene expression. Semi quantitative PCR of 8 ABA/stress-inducible genes in wild-type, *PpanrKO* and *Ppanr4PTC* lines. Treatments were control (C), 10^{-5} M ABA (A), 10% mannitol w/v (M) and dehydration to 70% fresh weight (D). All genes showed low or undetectable levels of expression in control conditions and were strongly unregulated by ABA and in most case by mannitol and dehydration treatments too. *PpanrKO* showed almost complete abolition of ABA/stress responses in all genes. *Ppanr4PTC* also showed a dramatic reduction of ABA/stress-induced gene expression, however, some detectable signal following treatments was apparent for Tubby and AMPW. This may indicate that *Ppanr4PTC* retains partial functionality or it may be an artefact. The V3.0 gene IDs for the genes analysed are: Phpat.004G114600 (AWPM), Phpat.012G092400 (Tubby), Phpat.012G077600+Phpat.012G077500 (Lea3), Phpat.020G034600 (Up1 but gene model incorrect), Phpat.010G025800 (TspO), Phpat.005G044800 (Dehydrin), Phpat.008G073700 (Up3), Phpat.005G076000 (b-Tub). Up1 was known as Phypa_66163 but has no gene model in V3.0.

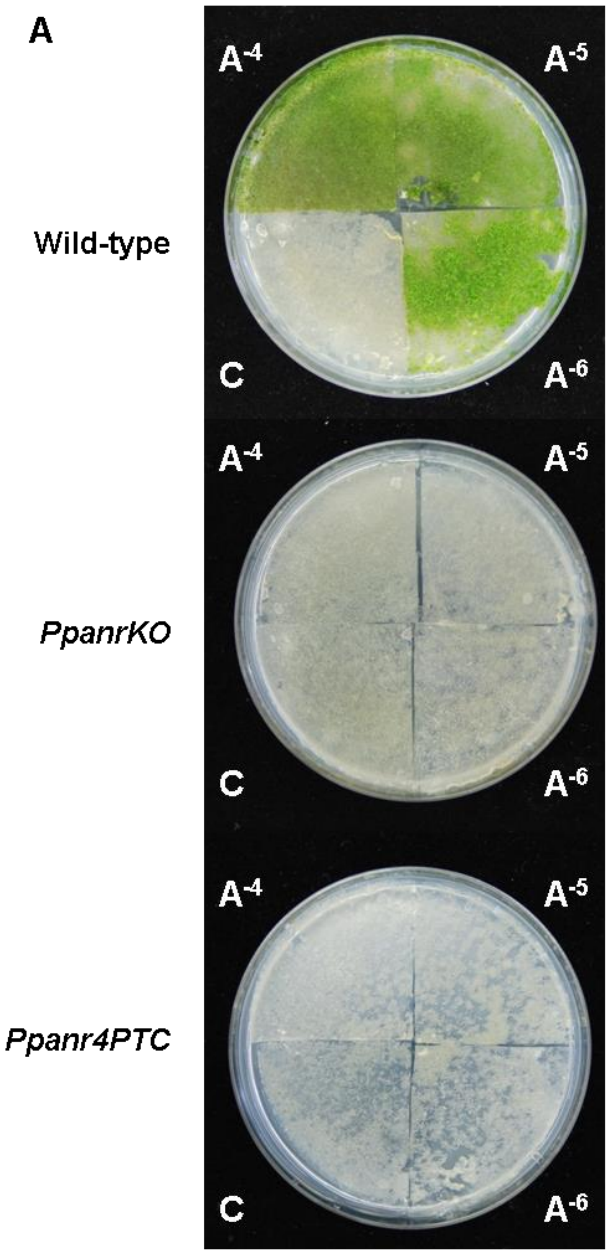
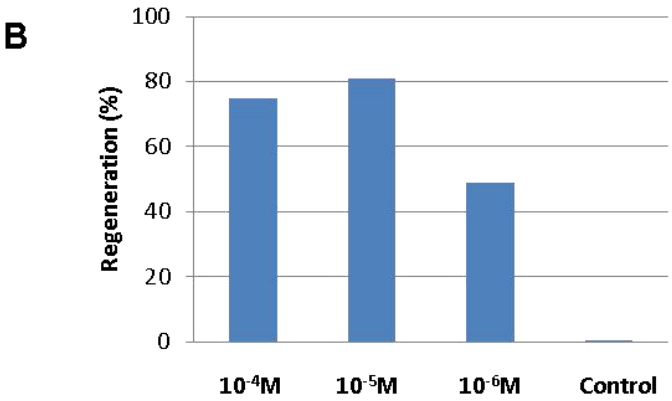


Figure 2.11. *Ppanr* mutants are not desiccation tolerant following ABA pre-treatment. A) Regeneration of rehydrated tissue following complete desiccation was dependent on ABA pre-treatment in wild-type tissue but this response is lost in the *Ppanr* mutants; *PpanrKO* and *Ppanr4PTC*. Each slice represents the tissue pre-treatment conditions (*i.e.* A^{-4} is ABA 10^{-4}M) for 24hrs prior to severe desiccation. Desiccation was carried out in a lamina flow hood for 12hrs. **B)** The level of protection conferred by ABA pre-treatment in the wild-type was partially concentration dependent with an increase in regeneration seen between 10^{-6} and 10^{-5}M. This concentration appears to saturate the response however with no increase seen between 10^{-5} and 10^{-4}M. Regeneration (%) was calculated as the proportion of the whole plate that regenerated as 'green' tissue as compared to the area of 'green' tissue prior to desiccation using ImageJ.



10^{-4}M) was able to confer complete desiccation tolerance (Figure 2.11). Tissue was completely desiccated in a laminar flow hood until no further weight loss occurred and then rehydrated and allowed to regenerate on normal medium. Tissue not pre-treated with ABA failed to regenerate while all three concentrations of ABA pre-treatment enabled vigorous regeneration in a partially concentration dependent manner (Figure 2.10B). Quantification of regeneration measured the area of green tissue after regeneration as a fraction of the area prior to desiccation. An increase is seen between the 10^{-6} and 10^{-5}M pre-treatments but not between 10^{-5} and 10^{-4}M suggesting saturation of the response was reached. Considering this pre-treatment was for 24hrs and many ABA-inducible genes are induced within minutes of exposure (Cuming *et al*., 2007) this might be expected. The *Ppanr* mutants failed to induce desiccation tolerance and so demonstrated the necessity of *PpANR* in regulating ABA-mediated stress tolerance.

2.3.8 *PpanrKO* shows slower, less vigorous growth

The distinct plant architecture phenotype of the *Ppanr* mutants, including less aerial growth and smaller gametophores, observed from sub-culturing and growth assays, was analysed in more detail. First, plants were allowed to grow over 4 weeks on a thin film of agar medium (surrounded by thicker medium which acted as a reservoir). This allowed the architecture of *PpanrKO* to be compared to wild-type. The *PpanrKO* mutant grew more slowly, and to a lower density and with less regularity (Figure 2.12). This phenotype is comparable to that seen following both ethylene and submergence treatments (Yasumura *et al.,* 2012; Yasumura *et al.,* 2015) of wild-type moss plants. It is worth noting that the thin film on which these colonies were grown may limit nutrient uptake and the phenotype of *PpanrKO* may reflect a disproportionate effect of this and not a direct effect of altered development and growth due to perturbed hormone signalling. In order to ensure that nutrient availability was not a major contributor to this phenotype, colonies were allowed to develop on well provisioned (thicker) media for longer periods. After several weeks, *PpanrKO* plants still displayed slower growth accompanied with a reduced plant ‘bulk’ as compared to the wild-type plant (Figure 2.13). While the wild-type plants develop into thick ‘pin cushion-like’ structures, *PpanrKO* gametophytes remain smaller and underdeveloped. This is reflected in the length of the gametophores which are shorter in *PpanrKO* while the ‘leaves’ too are shorter and reduced (Figure 2.13B, C). *PpanrKO* also shows premature senescence as observed in the original *anr* mutants (Figure 2.2).

2.3.9 *PpanrKO* protonemal cells are longer than wild-type

While reduced colony sizes, most notable after the onset of gametophore development, appear to be partly explained by reduced gametophore size including reduced leaves

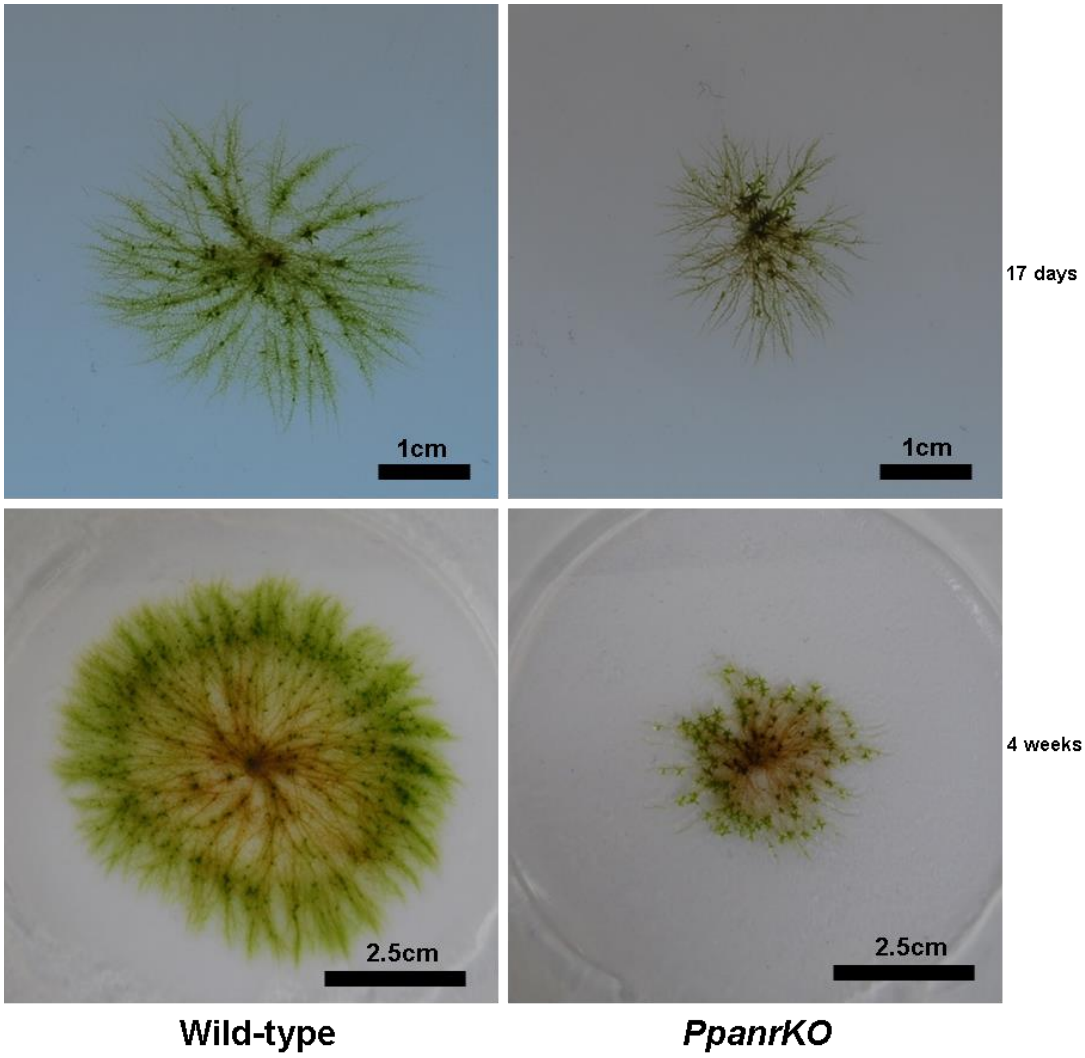


Figure 2.12. PpanrKO plants have distinct filament architecture as compared to wild-type. Explants were allowed to regenerate on a thin film of media surrounding by a reservoir of thicker media. Images were taken after 17 days (top panels) and after 4 weeks (bottom panels) for wild-type (left panels) and *PpanrKO* (right panels). After 17 days, the colonies look similar although *PpanrKO* is growing slower. After 4 weeks the differences become clear with wild-type colonies showing a uniform growth pattern with dense filamentous growth while *PpanrKO* appears less uniform and is far reduced in size.

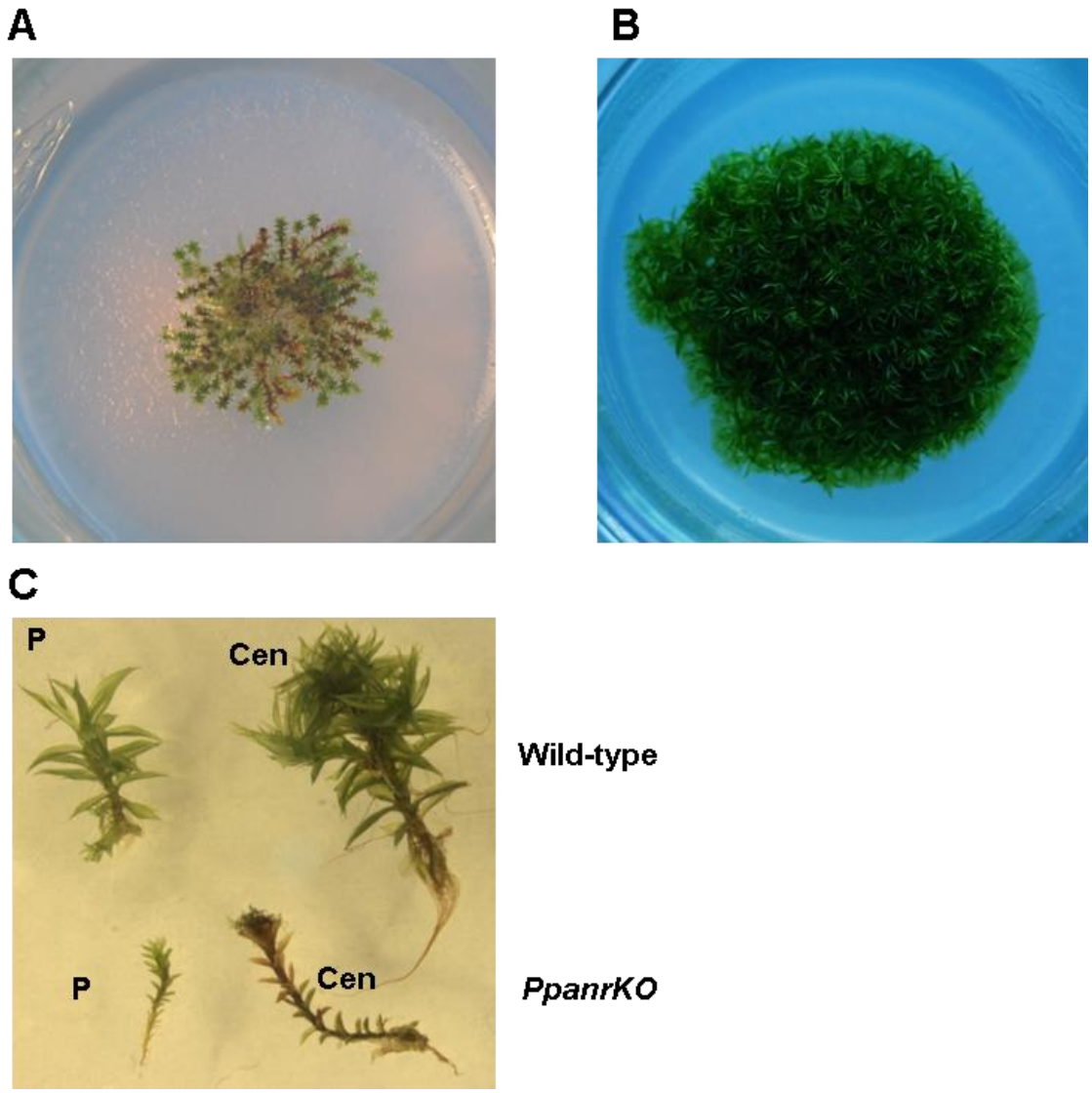
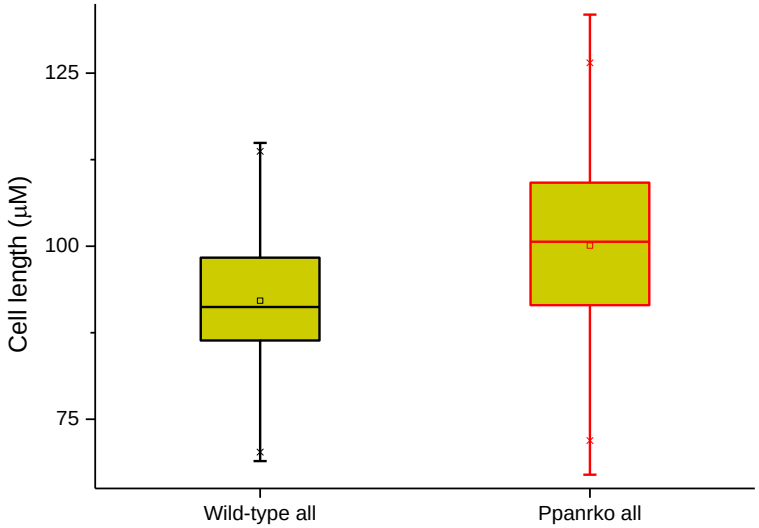


Figure 2.13. *PpanrKO* plants show distinct and reduced growth on well provisioned, thick media. A) Plant structure of *PpanrKO* after 3 months growth on BCDAT medium. This is dramatically different to the wild-type plant structure **B)** which has denser foliage coverage and more vigorous growth resembling a 'pin-cushion'. **C)** The reduction in plant size and foliage coverage can partly be explained by the reduction in individual gametophores. The gametophores shown are taken from the centre (Cen) and periphery (P) of the plants which reflect different ages as the plant grows out from centre.

A



B

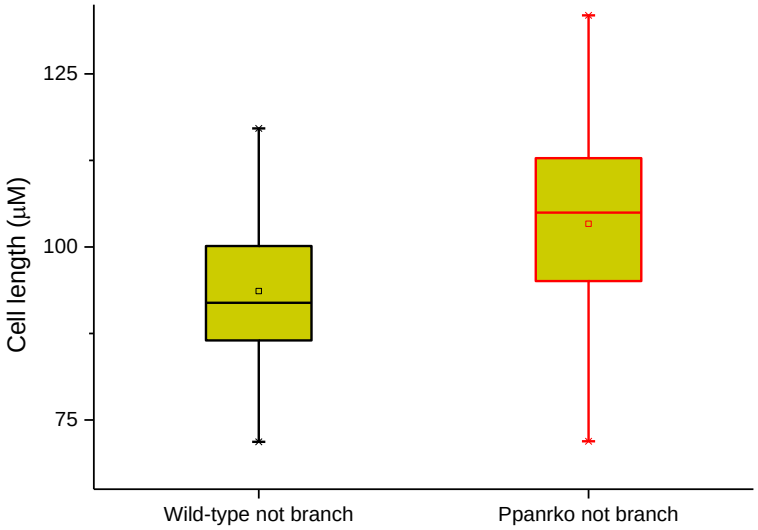


Figure 2.14. *PpanrKO* chloronemal cells are significantly longer than wild-type. A) Distribution of chloronemal cell lengths in wild-type and *PpanrKO*. *PpanrKO* has significantly (2-tailed t-test P-value<0.001) longer cells (mean=100.1µm) than wild-type (mean=92.1µm). **B)** The same analysis was carried out excluding branching chloronemal cells which slightly reduced overall cell lengths but again the same pattern was found. Elongation of cells is consistent with a role for PpANR as a negative regulator of ethylene signalling. Distribution shows box with 75 and 25% of mean (black or red squares) with whiskers showing 99 and 1% of range following outlier removal.

further analysis was carried out on the cellular level. Given the role of PpANR as a negative regulator of ethylene responses (Yasumura *et al*., 2015), the observed phenotype may represent a constitutive ethylene-stimulated growth phenotype. Further analysis of the cell morphology was analysed by measuring cell lengths of *PpanrKO* and wild-type chloronema cells; assimilatory filamentous protonemal cells (identified as chloroplast-rich protonemal cells with perpendicular cross-walls). This revealed that the *PpanrKO* mutant had cells significantly (2-tailed t-test P-value<0.001) longer (mean=100.1µm, SD=14.1µm, n=109) than wild-type (mean=92.1µm, SD=9.8µm, n=105) (Figure 2.14A). One possible source of bias in this measurement was the difference in frequency of branches between the lines. Branching occurs at the proximal end of a sub-apical cell in which the daughter cell is the result of reorientation of the mitotic spindle to produce a projection from the parent cell and this process can affect the cell's length. A comparison was therefore carried out (Figure 2.14B) excluding any cells in which branching occurred. This did not greatly affect the overall mean values obtained (wild-type mean=93.6µm, SD=10.0µm, n=91; *PpanrKO* mean=103.6µm, SD=13.5µm, n=86). This analysis showed that the slower growing, less dense colonies were not due to shorter cells. The interplay of plant hormones in regulating proper growth and development interacts with regulation of the cell cycle, division and cell expansion and clearly the deletion of *PpanrKO* affects this process by perhaps through failing to execute ABA mediated inhibition of cell elongation as well as by conferring constitutive ethylene signalling.

2.4 Conclusions

2.4.1 Forward genetics now a tool in bryophyte genomics

The implementation of forward genetics is now clearly possible in the model bryophyte *Physcomitrella patens*. This study represents the first known example of this approach being successfully carried out taking advantage of advances in the genetic resources available. The latest genome assembly of *Physcomitrella* (V3.1; Rensing *et al*., 2015 in preparation) now has chromosome scale linkage maps that represent one of the final stages in genome assembly, barring the ever improvement of gene models. With rich marker information and ever more affordable high throughput techniques, this makes *Physcomitrella* an excellent system for gene discovery. While mapping of the *ANR* locus relied on an Illumina GoldenGate array, developed specifically for the generation of the chromosome scale assembly, other sequence-based techniques are likely to be more applicable and cost-effective in future forward genetic approaches. Bulked segregant RNA-seq analysis, in which separately bulked mutants and wild-type segregants can be compared, can combine knowledge of SNP markers with *de novo*

identification of likely causal mutations negating the need for candidate gene sequencing and even combine expression analysis. This is likely to be a widely used process as NGS gets cheaper. One caveat is that this approach relies on the mutations lying in transcripts and not in *cis*-elements for example. However, a similar approach based on whole genome sequencing of bulked segregant pools is also feasible. The reliable identification of mutants is a crucial part of any forward genetic approach and the screening of positive phenotypes such as the faster growing *anr* mutants in the presence of ABA clearly has its advantages. This suggests the analysis of further *anr* mutants should be successfully pursued in *Physcomitrella* especially given that novel mechanisms are clearly involved in bryophyte ABA signalling. This proof of concept should now allow the interrogation of *Physcomitrella* for novel discoveries.

2.4.2 The *anr* mutants

The identification of the *ANR* locus resulted from unambiguous mapping data that identified a single region associating with the phenotype. The finding that the *anr3* and *anr4* mutations mapped to overlapping regions, in which the *anr3* region only had 4 extra gene models, suggested they were allelic. However, no mutations were found in the coding region of A3V mutants and sequencing of other candidate genes in the mapped region are has yet to reveal a candidate. Sequencing of potential regulatory regions for *PpANR* in A3V mutants may reveal the causal mutation although none have been found to date in the promoter region. One of the interesting differences between *anr3* and *anr4* not presented here is that *anr3* still retains ABA-responsive gene expression while clearly showing an *anr* growth phenotype. This data is preliminary and *anr3* will be the subject of fine mapping in order to identify the mutation. It is however, an interesting mutation to try identify and characterise as it appears that not only is it not in a predictable location, based on current knowledge, but this mutation somehow dissects the growth and molecular ABA-responses. Another mutant not presented here was *anr7* which was found to have a PTC between the PAS and EDR domains so predicted to encode an even more truncated protein than that in *anr4*. While more *anr* mutants remain to be analysed, a fascinating and unexpected story has already been started by the implementation of forward genetics in *Physcomitrella* and the identification of such a vital and unexpected ABA regulator.

2.4.3 Getting lucky?

The mapping of a relatively small region (88,366bp) for *anr4* greatly aided its eventual identification within exon 7 of the *ANR* gene but just how likely was this to have happened? Using a formula from Durrett *et al*., 2002, the probability of mapping to such a small region using 80 segregants, which are equivalent to 80 F_2 meiotic gametes, was around 18% which suggests we were among the ‘lucky’ studies. This formula does

rely on an average, not local, recombination frequency; in this case calculated for the whole of chromosome 12. Using their formula to predict the number of segregants that would have been required to have a probability of 0.95 of mapping to this relatively small region, we get 495 suggesting future mapping approaches use a greater number of segregants if mapping to small (100kbp) regions is desired. This again highlights the advantages that RNA-seq technologies would offer as traditional mapping simply serves to identify a small enough region within which to sequence candidates. With RNA-seq approaches, assuming the mutation lies within mRNA, coarse mapping or bulked segregant analysis may be sufficient to narrow the window for *in silico* mutation identification based on the segregation of novel mutations with the phenotype. This is again where a good reference genome allows the identification of these 'novel' mutations induced as part of the forward genetic approach.

2.4.4 Effects of *Ppanr* mutants

PpANR is clearly involved in ABA signalling which has effects at the phenotypic and molecular level. Growth assays confirm that *Ppanr* mutants fail to respond to the presence of ABA in the medium which in wild-type tissue greatly slows growth as protenemal development enters a stress adaptive state characterised by formation of 'brood cells' (Decker *et al*., 2006). Both *PpanrKO* and *Ppanr4PTC* show similar levels of ABA non-responsiveness but they appeared to have subtly different phenotypes with *Ppanr4PTC* having a more intermediate phenotype characterised by more gametophyte growth. This observation was not properly quantified; however gene expression analysis suggested that the truncated ANR protein, lacking the kinase domain, may have partial functionality. While an interesting possibility, this remains speculative and elucidating the roles of the remaining PAS and EDR domains would enable a mechanism for this to be found. For example, the recent finding of PpANR as a dual ethylene and ABA regulator (Yasumura *et al*., 2015) showed that PpANR was able to interact with the PpETR7 ethylene receptor, albeit with a much reduced affinity than its *Arabidopsis* CTR1/ETR1 functional counterparts, and this interaction, mediated by the EDR domain, may in itself have an effect. What was clear however was that both *Ppanr* mutants had severely attenuated ABA-induced gene expression. This was evident for a number of genes previously identified as strongly ABA up-regulated (Cuming *et al*., 2007) including LEA/dehydrin encoding transcripts. This lack of molecular responses to ABA showed clear mechanisms behind both the dehydration hypersensitivity of mutants and the inability for mutants to induce ABA-mediated desiccation tolerance. These processes are crucial to anatomically simple plants, such as bryophytes as well as the ancestral land plants, enabling molecular changes to

confer protection to cells. The full scope of this molecular response process is explored in Chapter 3 along with the extent to which these responses are abolished in *PpanrKO*.

2.4.5 Interactions of ABA and ethylene

The growth phenotype of *PpanrKO* was also analysed under normal conditions. This revealed that not only is ANR required for ABA-dependent stress responses but for regulating normal growth. Mutants appeared to grow slower, most notably after the onset of gametophore growth, than wild type colonies. As cell length is clearly longer in *PpanrKO*, the smaller colonies are likely the result of more complex effects on development such as the much reduced gametophores. While not quantified, the consistently smaller, less dense *PpanrKO* plants were likely due to the reduced gametophores which also had much reduced leaves. Wild-type plants develop thick canopies likely important in maximising photosynthesis and therefore likely to be reduced in *PpanrKO*. Analysis of the plant architecture revealed that *PpanrKO* mutants grew with far less uniformity and with a phenotype superficially similar to that observed in the ethylene/submergence response characterised by reduced filamentous growth in the colony centre and greater filament growth at the peripheries (Yasumura *et al*., 2012; Yasumura *et al*., 2015). ANR also acts as a negative regulator of ethylene (Yasumura *et al*., 2015) which means *PpanrKO* mutants have some level of constitutive ethylene signalling and this is likely to explain these similarities. While not quantified in these previous studies into ethylene and submergence responses, one feature of the response is the promotion of elongation which facilitates the escaping of submergence. This ‘escape’ process is an attempt by plants to escape the detrimental effects that submergence has on gas diffusion whereby CO_2 uptake and O_2 release is severely inhibited (reviewed by Jackson, 2008). One of the key mechanisms behind this elongation is the increase in final cell length rather than an increase in cell number (*e.g.* Musgrave *et al*., 1972). *PpanrKO* chloronemal cells were found to be significantly longer than wild-type which is in strong accord with this process suggesting that basal land plants have conserved submergence mechanisms likely under the same hormonal regulation as those in higher plants. Functional characterisation and phenotypic analysis suggests that not only is ANR involved in ABA and ethylene responses but that, as a nexus of hormone signalling, it is important for maintaining normal growth. Crosstalk between all the major plant hormones is a crucial part of maintaining a plants normal growth and development and when one, or in the case of *Ppanr* mutants two, of these is affected, adverse effects can become apparent.

CHAPTER 3

Chapter 3 – RNA-seq analysis of the PpANR null mutant, *PpanrKO*, reveals a vital role in ABA-dependent osmotic stress signalling.

3.1 Introduction

ABA signalling is known to induce global gene expression changes in embryophytes in response to a range of abiotic stresses, primarily through the induction of ABA biosynthesis which ultimately acts on gene targets containing the ABA responsive element (ABRE) among others. These are closely integrated with ABA-independent pathways such as those that act on, for example, the drought responsive element (DRE) or the NAC-recognition site (salt response) through which much crosstalk can be mediated through combinations of these *cis*-regulatory factors in ABA/stress responsive genes (reviewed by Roychoudhury *et al.*, 2013 and Lindemose *et al.*, 2013). Many analyses have been carried out in angiosperms that have delineated these various pathways to identify the specific responses they regulate. As discussed before (Chapter 1) these mechanisms appear to be largely shared across the embryophytes although further work is needed to fully understand the extent of conservation and differences given the relatively little, but ever increasing, attention the basal bryophytes have received. Transcriptomic analyses aim to uncover global effects of treatments or mutations on tissues or whole plants. Our extensive understanding of abiotic stress responses in the angiosperms has largely been driven by a desire to understand what the key components are that confer stress-tolerance with an eye to improving crop traits. More recently, the bryophytes have received attention as attractive model systems since they frequently exhibit high tolerances to a range of abiotic stresses. Unsurprisingly, given the wealth of molecular resources available, the most recent focus has been on *Physcomitrella patens*, which has a well-characterised genome sequence (Rensing *et al.*, 2008) and unique tools for genetic manipulation.

To date, relatively few transcriptomic datasets have been produced for *Physcomitrella* and indeed for bryophytes in general. An early EST analysis profiled the transcriptional changes that occurred in the fully desiccation tolerant moss *Tortula ruralis* (also known as *Syntrichia ruralis*) during both dehydration and rehydration (Oliver *et al*., 2004). This revealed an emphasis on maintaining ionic balance and protecting membranes during the dehydration phase while transcripts encoding proteins involved in oxidative stress responses were more involved during rehydration. Importantly, these authors also identified LEA proteins as a crucial component of both phases. *T. ruralis*, as a desiccation tolerant species, has relatively high levels of constitutive expression of transcripts normally thought of as ABA/stress-inducible, such that it is more protected

at the onset of stresses and so better able to survive (Wood *et al.*, 1999); a pattern also observed in a salt tolerant Arabidopsis-related species (*e.g.* Taji *et al.*, 2004). More recently, a very similar experiment was carried out on the wide ranging, and occasionally desert living, desiccation tolerant *Bryum argenteum* in which it showed a strong recovery of photosynthetic genes following rehydration indicative of strong protective measures during dehydration as in *T. ruralis* (Gao *et al.*, 2015) while analysis of the cold response of another wide ranging moss, *Pohlia nutans*, capable of surviving in artic habitats, further characterised the responses of these hardy bryophytes (Liu *et al.*, 2013). These experiments demonstrated the ability for RNA-seq analyses to simultaneously generate *de novo* transcriptome assemblies for non-model species alongside gene expression analysis and to provide resources to compare differences within the bryophytes. This greatly contributes to the burgeoning genetic resources available for these plants, further aided with the release of the draft transcriptomes of the moss *Ceratodon pupureus* (Szövényi *et al.*, 2015) and the liverwort *Marchantia polymorpha* (Sharma *et al.*, 2014). *Physcomitrella* is considered a drought tolerant but desiccation sensitive species, although able to survive dehydration to 91% relative humidity which generally equates to 0.32g H_2O/g dry matter (Koster *et al.*, 2010). There have been a number of microarray studies carried out on *Physcomitrella patens* which confirmed the importance of ABA as the stress signalling hormone in conferring this tolerance (Cuming *et al.*, 2007; Richardt *et al.*, 2010; Beike *et al.*, 2015) with many similarities found between the responses of bryophyte vegetative tissue and of maturing angiosperm seeds (Kamisugi and Cuming, 2005). A further large scale microarray analysis produced a database of transcript profiles not only to stresses but to developmental stages and tissues (Hiss *et al.*, 2014) and the subsequent microarray analyses have increased the number of target genes being probed. However, not many studies in *Physcomitrella* have taken advantage of next generation sequencing (NGS) technologies except for a few investigating alternate splicing (Wu *et al.*, 2014; Chang *et al.*, 2014).

This study is primarily interested in characterising the global effects of *PpANR* in regulating responses not only to ABA but also to osmotic and dehydration stresses and this requires the comparison with wild-type responses. Thus an RNA-seq transcriptomic analysis data of the *Physcomitrella* response to ABA, dehydration and mannitol was carried out. This is also the first study in *Physcomitrella* that has employed RNA-seq to assess the extent to which a single gene acts, enabled by the high sensitivity of NGS (in which gene expression can be simultaneously monitored across tens of thousands of genes, expressed over a range of five orders of magnitude) at relatively low cost; a trend likely to continue. The discovery of *PpANR* as

a novel ABA regulator not part of the current model of 'core' ABA signalling necessitated a more detailed analysis of its effects on global gene expression. The aim of the RNA-seq analysis was to not only confirm that *PpANR* was a vital component of ABA signalling but to try uncover what effects it may have on targets or pathways outside of ABA signalling. The expression data produced should also be an important resource in future efforts to form a model of how *PpANR* and its signal transduction pathway operates in ABA/stress responses.

3.2 Methods

3.2.1 Treatments and RNA extractions

The treatments and RNA extractions were as described in chapter 2.2. Three biological replicas per sample were created and RNA was extracted before bulking in equal quantities for cDNA synthesis and sequencing.

3.2.2 RNA-seq processing and mapping

Complementary DNA synthesised from 0.5 µg RNA, bulked from three biological replicas per sample, was run on an Illumina HiSeq 2500 platform to produce at least 30million 50bp single end reads per library (GATC Biotech., Konstanz). Raw reads of individual libraries were first viewed with fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to identify any biases and issues to guide pre-processing with trimmomatic (http://www.usadellab.org/cms/?page=trimmomatic) using the following command:

```
$ java -jar ~/bin/Trimmomatic-0.32/trimmomatic-0.32.jar SE input.fastq output _trimmed.fastq
ILLUMINACLIP:TruSeq3-SE.fa:2:30:10 HEADCROP:11 LEADING:30 TRAILING:30
SLIDINGWINDOW:10:30 MINLEN:20
```

Trimmed reads were aligned to the *Physcomitrella* V3.0 genome and transcriptome assembly (http://phytozome.jgi.doe.gov/pz/portal.html#!bulk?org=Org_Ppatens) using tophat2 with the following command:

```
$ tophat2 –o tophat_out --b2-very-sensitive --transcriptome-
index=./annotation_index/Ppatens_251_v3.0.gene V3genome output_trimmed.fastq
```

3.2.3 Differential gene expression analysis

Tophat_out files in .bam format were processed to obtain count values for each gene locus first using samtools (http://samtools.sourceforge.net/):

```
$ samtools_0.1.18 sort accepted_hits.bam accepted_hits_sorted

$ samtools_0.1.18 index accepted_hits_sorted.bam
```

The 8 processed, aligned and formatted libraries then had count data extracted using bedtools (http://bedtools.readthedocs.org/en/latest/):

```
$ bedtools multicov -bams <libA libB …..libH> -bed Ppjustgenes.gff3 > counts.txt
```

Tabular output files were trimmed leaving only 9 columns containing the gene name and its count values for each of the 8 libraries which could be used as input for the Trinity gene expression package (http://trinityRNA-seq.github.io/) which wraps the EdgeR Bioconductor package. Initial EdgeR analysis done on non-normalised data:

```
$ <path_to_trinity>/Analysis/DifferentialExpression/run_DE_analysis.pl --matrix counts.txt --method edgeR
```

Counts were normalized to library size (TMM scaling):

```
$ <path_to_trinity>/Analysis/DifferentialExpression/run_TMM_normalization_write_FPKM_matrix.pl --matrix count.txt --just_do_TMM_scaling
```

Normalized data analysed for differential gene expression using default cutoff settings (2-fold change and FDR P-value<0.001):

```
$ <path_to_trinity>/Analysis/DifferentialExpression/analyze_diff_expr.pl –matrix count_TMM_scaled.txt
```

Further cluster analysis was carried out by manually defining clusters using the Trinity package wrapping the statistical program R as described to obtain biologically meaningful clusters.

3.2.4 Correlation analysis

For correlation analysis, the count data were trimmed so that only transcripts with >50 sequence reads in either of the pairwise comparisons were included, *i.e.* first must be over 50 in WT_C or WT_ABA, then must be over 50 in WT_C or KO_C. This ensures expression change values are less likely to derive from stochastic error likely to arise when only a handful of reads are obtained from genes expressed at very low levels. Thus the transcripts analysed at least show robust expression values in one of the libraries. Values of 0 were converted to 1.

3.2.5 GO term analysis

All significantly regulated genes were run through the Blast2Go (https://www.blast2go.com/blast2go-pro/download-b2g) annotation pipeline following recommended procedures. The outputs for GO terms were manually analysed for interesting and meaningful patterns and the data presented.

3.3 Results

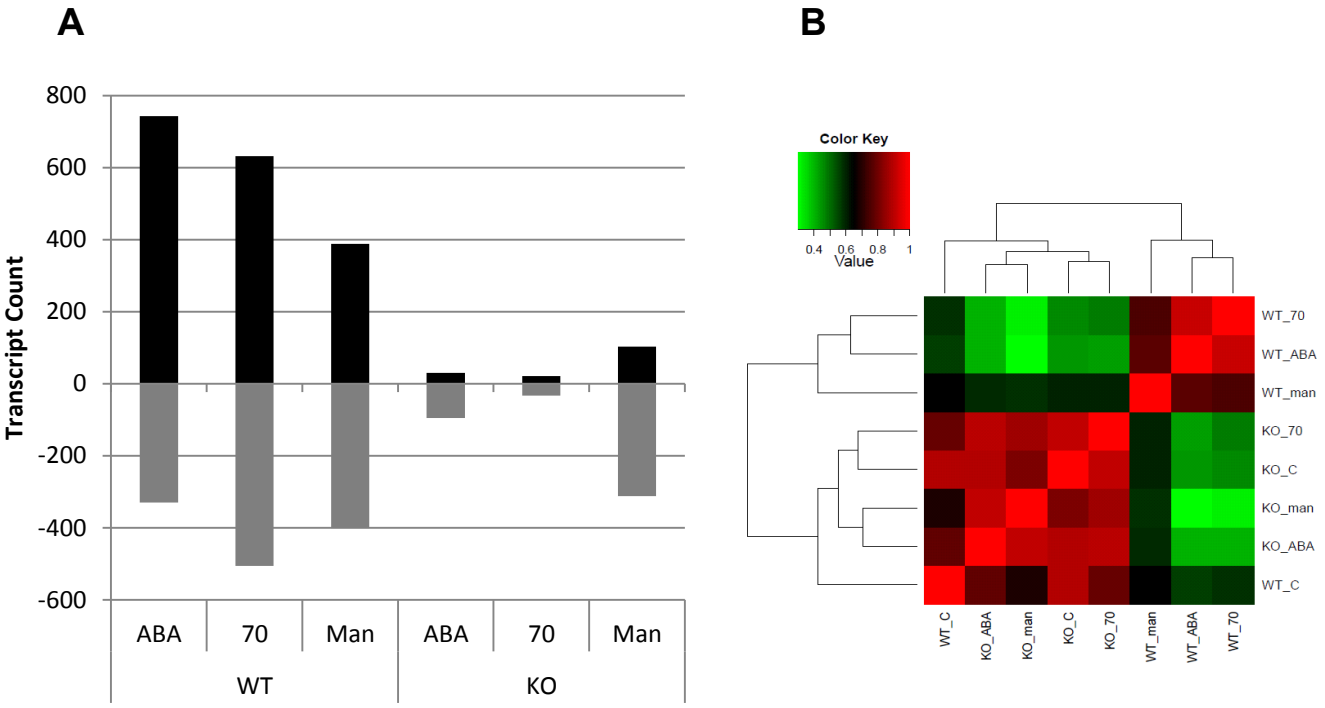
3.3.1 RNA library quality

To analyse the transcriptomic effects of *Ppanrko*, 8 libraries were sequenced on the Illumina HiSeq 2500 platform. Four libraries were prepared from wild-type tissue and 4 from *Ppanrko* which included a control and three treatments: ABA (10^{-5}M), mannitol (10% w/v) and dehydration (to 70% hydrated weight loss). The 8 libraries were successfully sequenced with all 8 yielding at least 26 million mappable reads (Table 3.1).

Table 3.1. Summary of RNA-seq mapping statistics.

Line	Library	Raw reads	Trimmed %	Tophat2 input	Mapped (% of Tophat2 input)	Models not matched (26610 total)
wild-type	**Control**	31753175	5.4	30036710	26101901 (86.9%)	3365 (12.6%)
	ABA	39352684	7.6	36372909	31498939 (86.6%)	3946 (14.8%)
	Mannitol	33360756	5.7	31456179	30481037 (96.9%)	3152 (11.8%)
	Dehydration	31480291	8.8	28711648	25007845 (87.1%)	4130 (15.5%)
Ppanrko	**Control**	38310341	7.7	35366412	30839511 (87.2%)	3608 (13.6%)
	ABA	46356009	3.1	44929329	39133446 (87.1%)	3841 (14.4%)
	Mannitol	34533871	5.6	32608685	31369555 (96.2%)	3355 (12.6%)
	Dehydration	32544262	2.4	31753631	27593905 (86.9%)	4055 (15.2%)

Trimming involved removing low quality base calls as well as the first 10bp from each read as these contained nucleotide biases as flagged by the fastqc program. Only a small proportion of reads from each library were removed at this stage (2-9%). Each library had reads that mapped to between 84.5 and 88.2% of the V3.0 gene models. For each library, around 80% of the original raw reads were mapped: a value higher than that obtained in previous studies (Wu *et al.*, 2014; Chang *et al.*, 2014). A small number of gene models (1739; 6.5%) failed to have any reads mapped from any library representing a suite of candidate pseudogenes, silenced or highly specialized genes (*e.g.* gametophore or sporophyte-specific). Using a cutoff of 20 reads from all libraries, 5358 (20%) gene models could be considered as having very low expression. For wild-



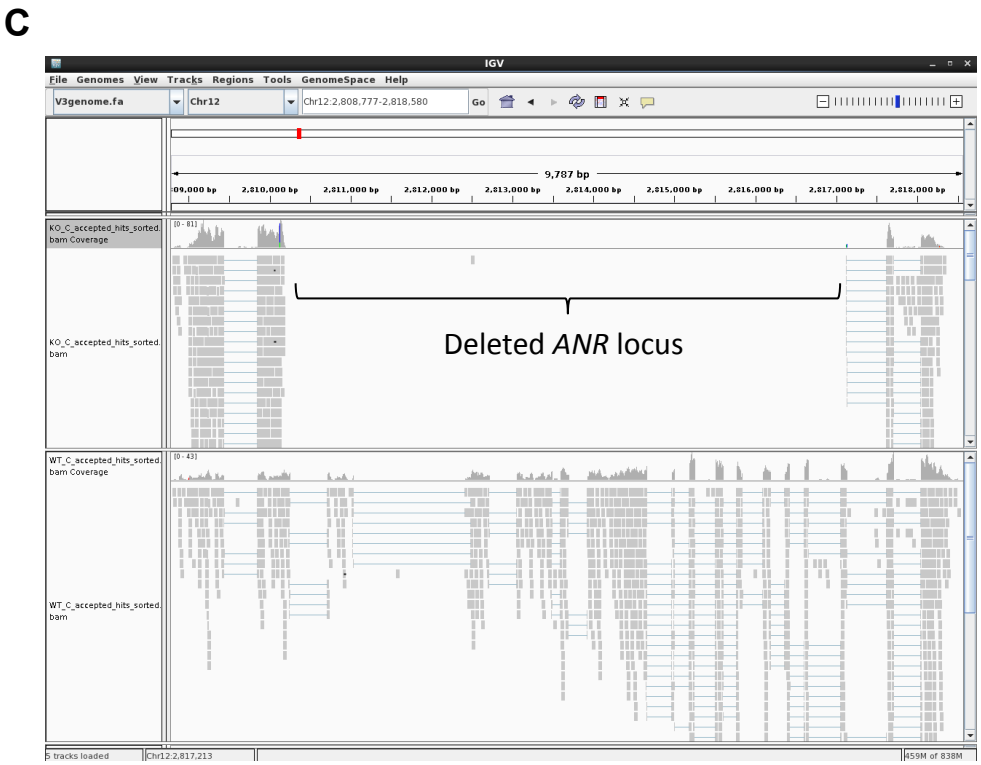


Figure 3.1. RNA-seq analysis revealed a significant set of ABA and stress regulated genes had their regulation abolished in *PpanrKO*. A) Total transcripts found to be significantly (>2-fold, FDR P-value< 0.001) up- and down-regulated in the 3 treatments (see below) for both wild-type (WT) and *Ppanrko* (KO) tissue. **B)** Correlation matrix showing overall relationship and similarity in global expression patterns of transcripts between all 8 libraries. ABA= 1hr 10^{-5} M ABA treatment, 70= dehydration to 70% weight loss over approx. 12hrs, man= 1hr mannitol treatment and C= control. Note the strong relationships between the WT_C and all KO libraries reflecting a loss of transcriptional response. **C)** Screenshot of *ANR* locus on IGV genome viewer shows clear deletion of *ANR* coding region in *PpanrKO* (top panel) while wild-type (bottom panel) tissue shows good coverage with thousands of reads mapped.

type moss, analysis of the significantly regulated genes (>2 fold up-regulation and FDR P-value<0.001) revealed that the treatments had significantly modified gene expression with 743, 632 and 388 genes showing up-regulation by ABA, dehydration and mannitol respectively and 330, 506 and 401 genes showing down-regulation in response to ABA, dehydration and mannitol respectively (Figure 3.1A and 3.3A). There was considerable overlap in the genes differentially expressed between treatments in wild-type tissue so that the differentially expressed genes comprised a total of 911 and 903 unique genes significantly up-regulated and down-regulated respectively (Figure 3.1A and 3.3A). The number of genes showing significant differential regulation in the *PpanrKO* mutant revealed that responses to ABA and dehydration were largely abolished, with only 30 and 20 genes respectively up-regulated by these treatments and only 95 and 33 respectively down-regulated (Figure 3.1A and 3.3B). Interestingly, the mannitol treated *PpanrKO* tissue appeared less affected with 103 (27% of wild-type) and 312 (78% of wild-type) genes up- and down-regulated respectively. The correlation matrix of library similarities demonstrated that the abolition of the response in the *PpanrKO* mutant resulted in all four of its libraries remaining highly similar to that for the wild-type control condition (Figure 3.1B). Analysis of the mapping files additionally confirmed the deletion of the *PpANR* gene (Figure 3.1C) which was in confirmation with the genetic analysis (see Chapter 2).

3.3.2 Library relationships

Having successfully sequenced and mapped a large number of reads to produce robust differential expression data, comparisons between the treatments could be carried out. In bryophytes, a central role of ABA in stress signalling has been established (*e.g.* Cuming *et al*., 2007; Oliver *et al.*, 2004) and a clear positive correlation can be found for gene expression (log fold change) under ABA and dehydration (Pearson's two-tailed correlation=0.887, P<0.001, N=17719) and for ABA and mannitol (Pearson's two-tailed correlation=0.682, P<0.001, N=17743) (Figures 3.2A and 3.2B). This is in accordance with the overlap in the identity of the genes significantly regulated by ABA, dehydration and mannitol in wild-type tissue (Figure 3.3A) as ABA and dehydration have a greater number of co-regulated genes (551/743) than ABA and mannitol (286/743). To better understand the global effects of *PpANR*, the relationship between the expression of genes in *PpanrKO* and wild-type ABA/stress treatments was analyzed. A scatter plot of expression changes due to ABA treatment in wild type and *PpanrKO* shows no clear relationship (Figure 3.2D). There is, however, a significant negative correlation in those genes up-regulated by ABA (Pearson's two-tailed correlation=-0.222, P<0.001, N=10516). This suggests that there is a measureable level of constitutive ABA-related signalling in *Physcomitrella* as these

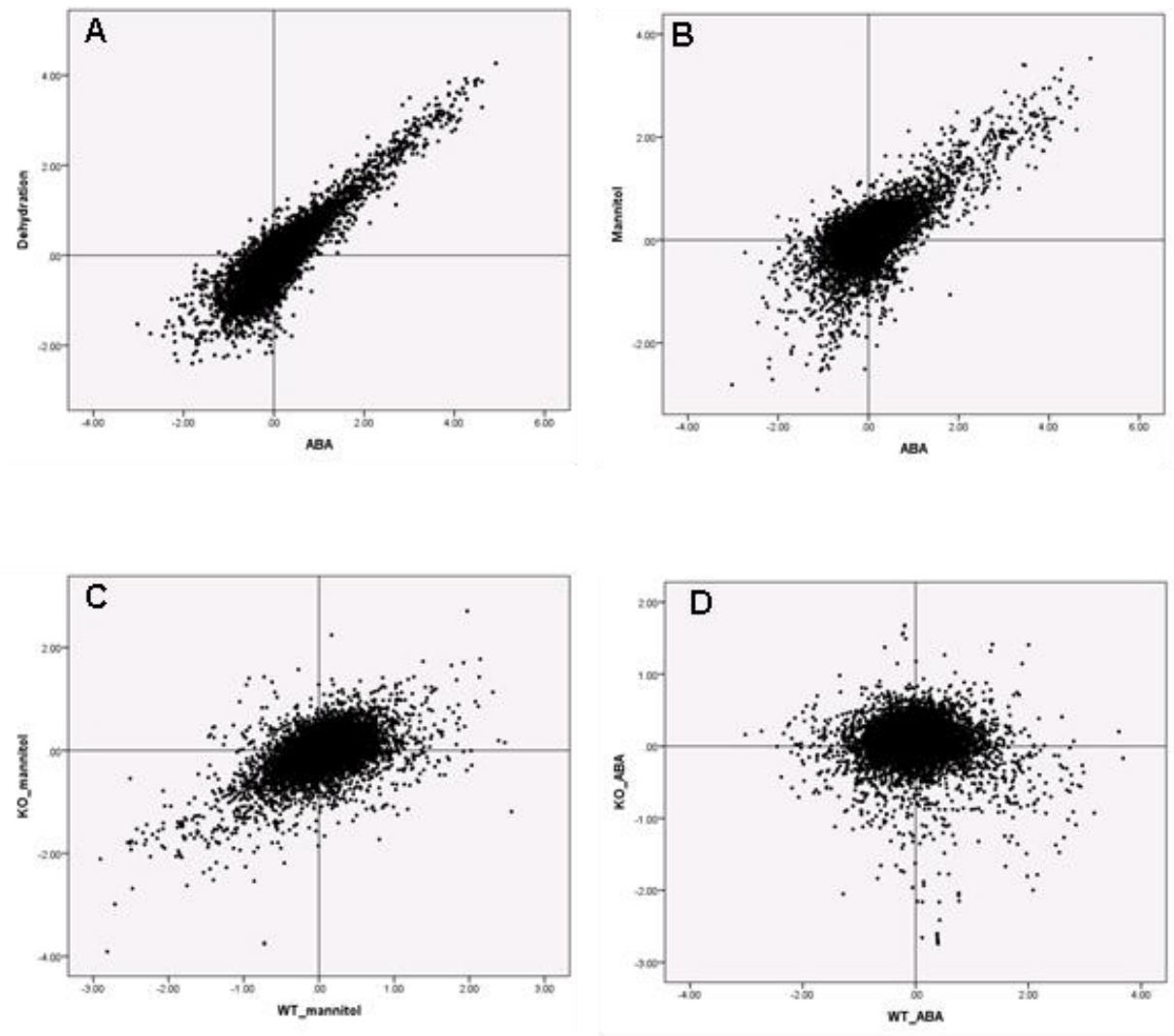


Figure 3.2. Relationships in gene expression between libraries. Scatter plots of pairwise log fold changes between **A)** ABA and dehydration in wild-type tissue, **B)** ABA and mannitol in wild-type tissue, **C)** manntiol treatments in wild-type and *PpanrKO* tissue, **D)** ABA in wild-type and *PpanrKO* tissue. Clear positive correlations observed between ABA and stress treatments in wild-type tissue and to a lesser degree between the mannitol treatments in both lines.

genes clearly require *PpANR* to maintain a baseline level of expression. This is contrasted with the observation that mannitol treatment of *PpanrKO* tissue is well correlated with mannitol treatment in wild-type tissue (Figure 3.2C) and there is a particularly strong positive correlation between those genes down-regulated by mannitol in in wild-type and those similarly down-regulated in *Ppanr*KO (Pearson's two-tailed correlation=0.629, P<0.001, N=4289). This indicates the likely presence of an ABA-independent mannitol/osmotic stress response pathway that does not require *PpANR*. These relationships highlight some of the stress specific responses seen in *Physcomitrella* which have been observed before (*e.g.* Beike *et al.*, 2015 in the cold response) and also serve to underscore the role of *PpANR* in the ABA-dependent signalling pathways.

The relationships between ABA and stress treatments in wild-type tissues represent the relationship between ABA-dependent and ABA-independent pathways. Much is known about the various TF families in angiosperms, including WRKY, bZIP (including ABI5), bHLH, AP2/ERF (including ABI4) and ABI3, and their roles in both ABA-dependent and ABA-independent pathways with multiple levels of crosstalk (see reviews by Lindemose *et al.*, 2013 and Roychoudhury *et al*., 2013) and similar processes are likely to occur in *Physcomitrella* (*e.g.* Khandelwal *et al.*, 2010). For example, there is clear overlap between the genes up-regulated by ABA and dehydration, with the 551 shared genes representing a substantial proportion of dehydration responsive genes mediated by ABA-signalling (Figure 3.3A). The 81 genes (12.8%) that are up-regulated by dehydration but not ABA presumably represent those under the control of an ABA-independent pathway (for example through possible DRE element regulation). While the correlation between ABA- and dehydration- regulation of genes is particularly high (Figure 3.2A) it is less so between ABA and mannitol (Figure 3.2B). 96 genes, representing 24.7% of those up-regulated by mannitol, are not up-regulated by ABA which indicates that a higher proportion of the genes up-regulated by mannitol are regulated through ABA-dependent pathways as compared to dehydration. 332 genes (65.6%) are down-regulated by dehydration and not by ABA which indicates the down-regulation of genes by dehydration is possibly less ABA-dependent than for those that are up-regulated. It is worth noting that the RNA-seq analysis provides a proxy measurement of gene regulation but is in fact a measurement of the steady-state abundance of individual transcripts, which comprises a balance between transcript synthesis and degradation. It is likely that the latter process is more significant in the dehydrated tissue since the tissue experienced a longer period of physical stress (overnight as compared to 1hr ABA and mannitol exposure). The disparity between up- and down-regulated transcripts is seen to an even greater extent with mannitol

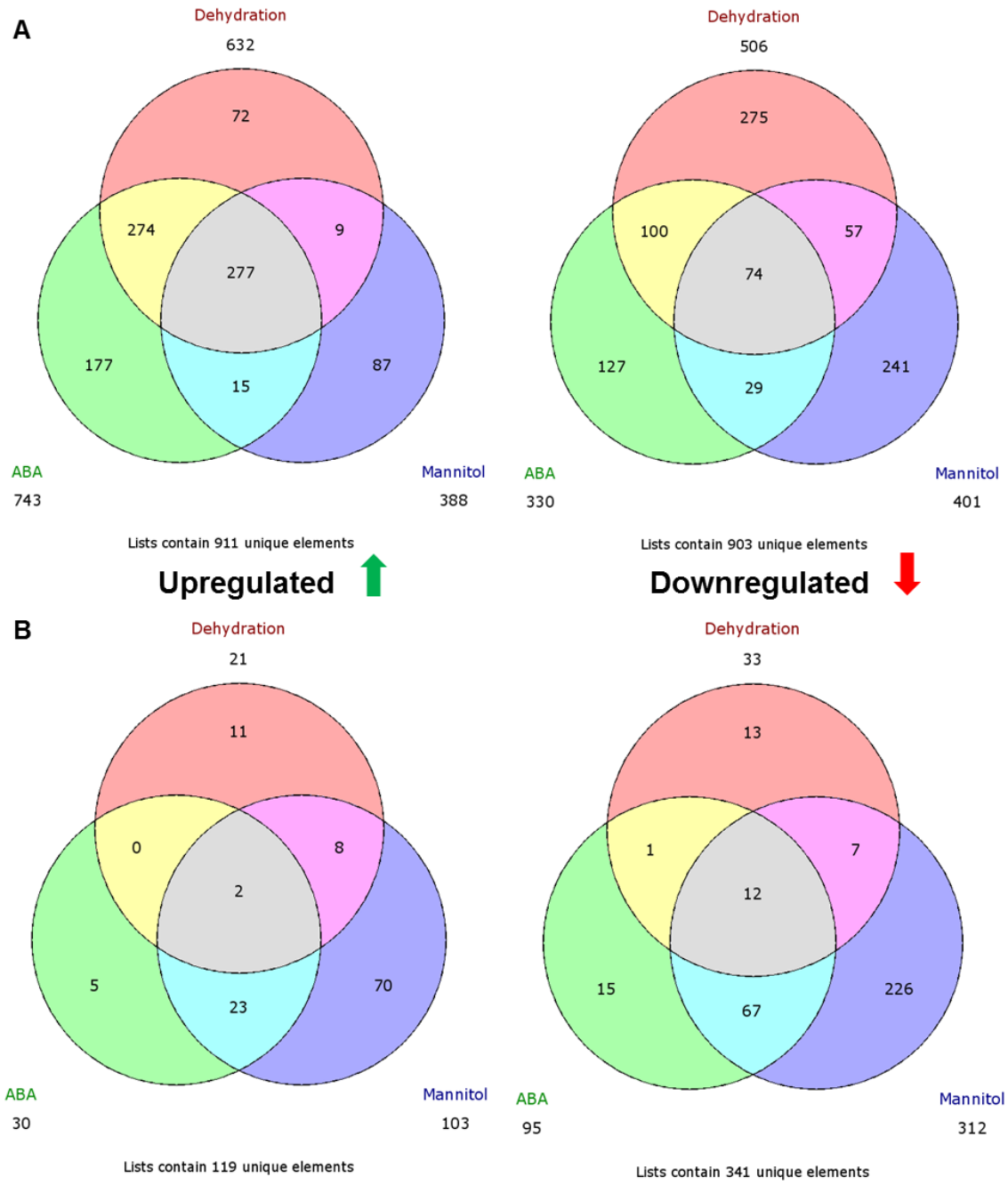


Figure 3.3. Strong overlap in gene expression between up-regulated genes in wild-type tissue is not mirrored in down-regulated genes. Venn diagram showing the overlaps in regulation for significantly (>2-fold, FDR P-value<0.001) up- (left panels, green arrow) and down-regulated (right panels, red arrow) transcripts in wild-type tissue **(A)** and *PpanrKO* **(B)**. In the wild-type response there is a strong overlap in those transcripts significantly up-regulated by ABA/stress treatments. In *PpanrkO*, the total number of up-regulated transcripts is significantly reduced with only 30 transcripts significantly up-regulated by ABA. The reduction in the number of down-regulated transcripts is less dramatic with the majority being down-regulated by mannitol treatment.

treatment where 298 (74.3%) of osmotically down-regulated genes are not significantly down-regulated by ABA. This suggests that in general, those genes down-regulated by stresses show greater treatment-dependent specificity than those up-regulated. A greater number of up-regulated genes are likely to encode components required in greater quantities than during normal conditions, including those involved in more general osmotic stress protection such as cellular stabilisation (LEA proteins for example) and oxidative stress responses. On the other hand, those genes down-regulated may be more likely to be involved in repressing specific metabolic or developmental processes such as cell wall modifications with greater stress specificity.

3.3.3 *PpANR* regulation and effect on core components

The 'core' ABA signalling pathway has been well described in Arabidopsis (Cutler *et al.*, 2010) and is thought to be largely functionally conserved in the bryophytes (*e.g.* Hanada *et al.*, 2011). *PpANR* is not currently known to be part of this pathway and as a single gene is clearly not functionally redundant, given the global abolition of ABA responses in *PpanrKO*. However, its effect on the transcription of the genes that encode the components of the 'core' response pathway is an important aspect of understanding its role. The 'core' pathway is made up of 5 principal gene families: PYR/PYL/RCAR receptors, the PP2C phosphatases, the SnRK2 kinases, the B3 domain (*e.g. ABI3*) and the bZIP (*e.g. ABI5*, *AREB1*) TFs (see Figure 1.3). The expression profiles of these members in all libraries have been analyzed (Table 3.2). Of the four putative PYR/PYL/RCAR receptors, 2 show up-regulation by ABA, with one showing up-regulation by all treatments in wild-type tissue (Phpat.009G076800/*PYL5*). Both PP2Cs in *Physcomitrella* (Sakata *et al.*, 2009; Komatsu *et al.*, 2009) show very strong ABA and stress induction (Phpat.007G024200/*ABI1* and Phpat.011G067300/*ABI2*). Three of the 4 putative SnRK2 kinases (*PpOST1-1/2/3/4*) show strong ABA-mediated up-regulation with *PpOST1-1* (Phpat.005G078000) showing the strongest up-regulation by ABA/stress of any of the components (>20-fold in all treatments). Three of the ABI3s show similar responses to ABA (4-fold up-regulation) but subtly different responses to dehydration and mannitol. For example, *PpABI3B* (Phpat.017G059900) shows particularly strong up-regulation by mannitol (7-fold) suggesting ABA-independent pathways could modulate their expression in response to specific stresses. The fourth putative ABI3-like TF, *PpABI3D* (Phpat.013G044400), shows around a 2-fold up-regulation for all treatments. A further group of basic leucine zipper (bZIP) TFs is known to regulate ABA-responsive gene expression in *Arabidopsis.* These include ABI5, the ABA response element binding protein (AREB3) and the Enhanced Em Level (EEL) transcription factor. Three putative homologues of these bZIP ABA regulators have been identified in Physcomitrella

known as bZIP49 and bZIP9a/b. This nomenclature derived from their location on the "version 1" sequence scaffolds 49 and 9, respectively. The *bZIP9a* and *bZIP9b* genes are identical, occurring as an inverted ("head-to-head") repeat on chromosome 23. These genes are incorrectly annotated in the various iterations of the genome browser, but have been correctly identified in our laboratory (Y. Kamisugi, unpublished data). RNA-seq reads from

Table 3.2. Fold changes of established and putative 'core' ABA signalling components in *Physcomitrella*. Fold changes are conditionally formatted such that darker colours represent stronger regulation (green= up-regulation; pink= down-regulation). Wild-type (WT); *PpanrKO* (KO). Fold changes between -1.49 and 1.49 are displayed as '-'.

	ABA-fold		Dehydration-fold		Mannitol-fold	
	WT	KO	WT	KO	WT	KO
PpANR	2.0	-	1.5	-	2.0	1.8
PYL4-1	-1.6	1.5	-2.2	-	-1.2	-
PYL4-2	2.4	-	1.8	-	1.6	-1.7
PYL4-3	-	-	-1.6	-	-	-
PYL5	7.2	1.8	4.0	1.6	4.6	1.5
ABI1	10.8	-	10.9	-	9.2	-
ABI2	16.7	-	8.8	-	8.6	-2.1
OST1-1	27.6	2.6	21.7	-	29.7	-
OST1-2	7.1	-	5.8	-	2.7	-1.6
OST1-3	-	-	-	1.5	1.5	-
OST1-4	4.8	-	2.5	-1.8	2.0	-1.5
ABI3A	4.1	-	1.8	-	4.3	-
ABI3B	4.3	1.5	4.3	-	7.3	1.7
ABI3C	4.0	1.6	2.6	-	1.7	-1.9
ABI3D	2.8	1.5	2.1	-	2.5	-
bZIP9	10.0	-	6.1	-1.7	2.5	-2.3
bZIP49	18.8	-	16.9	-1.6	5.6	-2.0

both have been mapped to just one of the tandem repeats (confirmed by inspecting the mapping bam file) so the expression changes calculated are a likely average of both. The data suggests that these bZIP TFs are strongly ABA/dehydration up-regulated with bZIP49 showing around 17-fold up-regulation to ABA while bZIP9a/b shows around 7-fold up-regulation. Their up-regulation to mannitol is less than for the other treatments but still significant. As for *PpANR* itself, it is found to be mildly ABA/stress responsive showing around 2-fold up-regulation by ABA and mannitol but only 1.4-fold by dehydration suggesting posttranslational regulation may be more important. There is

also some evidence that this gene may be subject to alternative splicing offering another possible level of regulation (D. Lang, University of Freiburg, Pers. Comm.)

3.3.4 Effect on ABA biosynthesis

Many *Arabidopsis* mutants were identified as ABA deficient (known as ABA1 through 4) and these were associated with mutations in genes encoding enzymes in the ABA biosynthetic pathway. Many of the key ABA-biosynthetic enzymes are up-regulated in response to stress leading to an increase in endogenous ABA and the associated ABA-dependent stress responses.

Table 3.3. Fold changes of established and putative ABA metabolism and catabolism components in *Physcomitrella*. Fold changes are conditionally formatted such that darker colours represent stronger regulation (green= up-regulation; pink/red= down-regulation). Fold changes between -1.49 and 1.49 are displayed as '-'.

	ABA-fold		Dehydration-fold		Mannitol-fold	
	WT	KO	WT	KO	WT	KO
ZEP/PpABA1	-	-	-1.9	-	1.6	-
NCED1	22.3	-	14.9	-3.1	13.2	-
NCED2	24.5	-3.1	16.1	-	14.8	-
NCED3	2.9	-	2.4	-	2.3	-
NCED4	-	-1.7	-4.0	-1.8	2.1	-
NCED5	-	-	-	-1.6	2.1	-
AAO1	2.1	1.6	-	-	1.6	-1.5
AAO2	-	-	-	-1.5	-	-1.7
AAO3	2.1	1.6	-	-	1.6	-1.5
AAO4	2.2	-	-	-	2.8	-
AAO5	2.8	-	1.9	-	-	-1.8
MoCo sulphurase	-	2.1	-	-	-	-
CYP707A1	-	-	-	-	1.7	-
CYP707A2	-5.3	-3.5	-38.4	-13.3	-93.3	-12.1
CYP707A3	-	-2.5	-47.4	-10.0	-	-12.7

Physcomitrella is thought to possess the canonical ABA biosynthetic machinery (Table 3.3) and recently *PpABA1* (ABA deficient 1), which encodes the single copy zeaxanthin epoxidase (ZEP) enzyme that catalyzes the first step in ABA biosynthesis, has been demonstrated to be functionally conserved among land plants (Takezawa *et al.*, 2015). *PpABA1* shows no ABA/stress induced up-regulation, in support with previous findings,

as it is not thought to be a rate limiting step (Audran *et al.*, 1998). Our RNA-seq analysis reveals that two of the five putative NCED (9-cis-epoxycarotenoid dioxygenase) genes (Phpat.025G017600/*NCED1* and Phpat.016G070400/*NCED2*) showed very strong up-regulation by both ABA and stress treatments (>20-fold to ABA) in support with previous findings (see Xiong and Zhu, 2003 and references therein), an effect completely abolished and even reversed in the *PpanrKO* mutant (Table 3.3). A third member of this family (Phpat.022G023300/*NCED3*) shows much higher levels of constitutive expression than these two genes and yet is also moderately up-regulated (between 2- and 3-fold) by ABA and stress treatments. The transcriptional regulation of these biosynthetic enzymes has long been known to be a crucial component of stress-induced ABA biosynthesis (reviewed by Xiong and Zhu, 2003). The cleavage step carried out by the NCED enzymes is the first committed step in ABA biosynthesis and so the strong transcriptional up-regulation of these components is thought to greatly boost the endogenous ABA levels in part by preventing other pathways scavenging 9-cis-neoxanthin. This is considered the primary rate-limiting step in angiosperms (Tan *et al.*, 1997; Zeevaart and Creelman, 1988) and the strong ABA/stress-up-regulation of NCED transcripts is indicative that this is likely the case in *Physcomitrella* too. We also see a moderate up-regulation of 3 of the 4 putative AAO (ABA-aldehyde oxidase) transcripts, by around 2–fold, but only in response to ABA. The response to ABA alone demonstrates the positive feedback involved in ABA biosynthesis with exogenous ABA able to increase endogenous ABA biosynthesis so boosting the signal. This also demonstrates the differential regulation of these multigene family members. The single MoCo sulphurase transcript, responsible for the crucial sulphation of MoCo (molybdenum cofactor), shows no significant change in wild-type tissue; however, interestingly it does show a 2-fold increase in response to ABA in *PpanrKO*.

CYP707A enzymes are the primary enzymes involved in ABA catabolism. *Physcomitrella* has 3 putative members of this family and they show strong differential regulation in wild-type tissue. *CYP707A1* (Phpat.023G012500) is expressed at high constitutive levels and shows little response to ABA/stress suggesting the role of its gene product is to continuously provide negative feedback on ABA levels by maintaining ABA catabolism. Interestingly, it shows constitutively higher up-regulation in *PpanrKO* control tissue as compared to wild-type control tissue suggesting, again, that *PpANR* is required to maintain a baseline level of ABA signalling; in this case through gene repression. The other two CYP707A transcripts (Phpat.015G014800/*CYP707A2* and Phpat.009G031600/*CYP707A3*) show lower expression levels but are very strongly down-regulated in response to stress

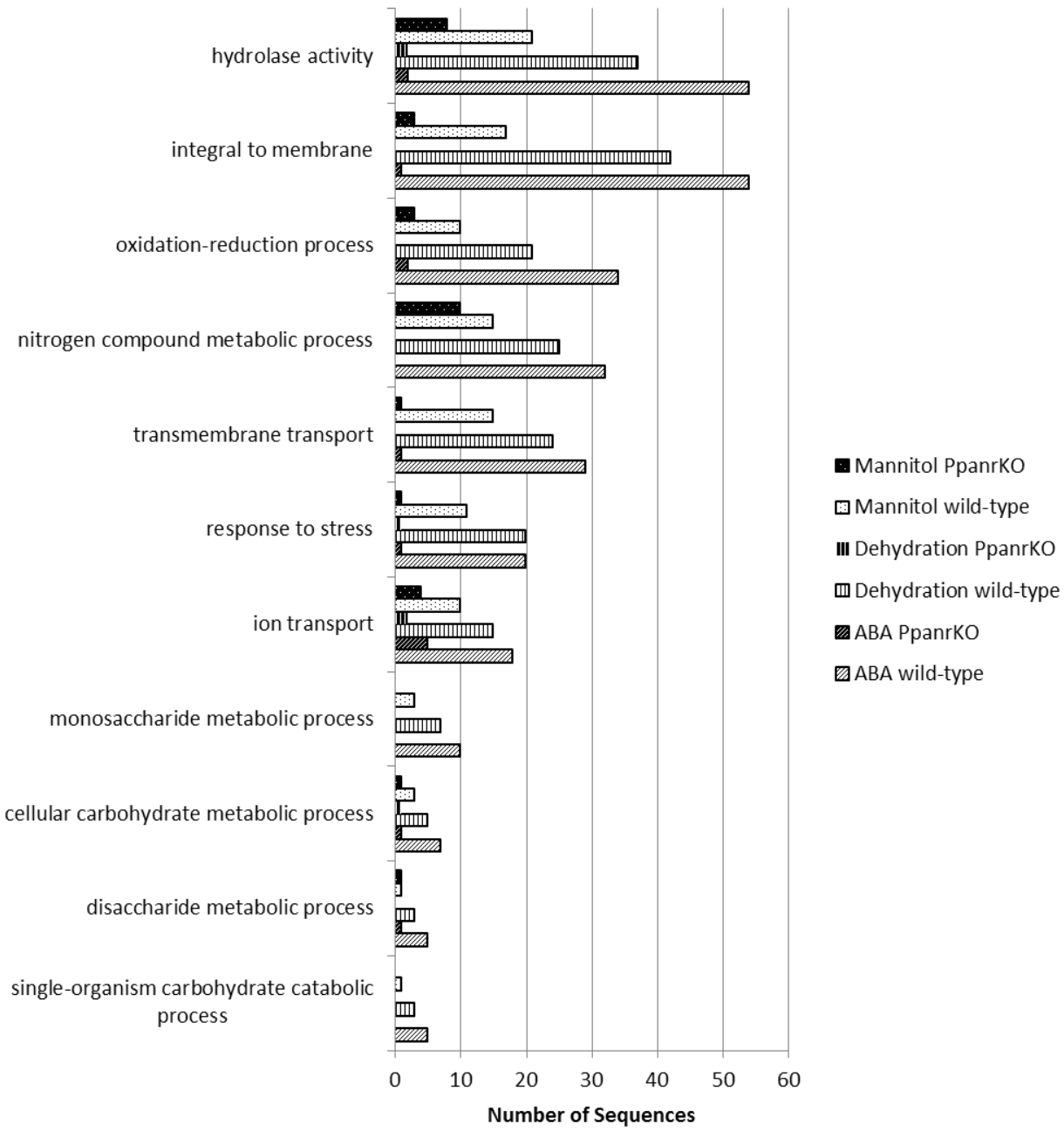


Figure 3.4. GO terms featuring in the up-regulated genes in wild-type tissue. Note the low number of transcripts matching these GO terms in *PpanrKO* (white patterns on black background) as compared to in wild-type tissue (black patterns on white background) highlighting the general abolition of ABA/stress responsive gene expression; transcripts involved in key molecular ABA/stress response such as ion balance and transportation as well as sugar metabolism. These GO terms represent molecular functions, biological processes and cellular components. The GO terms are as follows from top to bottom: GO:0016787, GO:0016021, GO:0055114, GO:0006807, GO:0055085, GO:0006950, GO:0006811, GO:0005996, GO:0044262, GO:0005984, GO:0044724.

treatments, and only *CYP707A2* is also down-regulated by ABA, demonstrating another positive feedback mechanism in ABA signalling.

3.3.5 GO term analysis of significantly regulated transcripts

Looking at the expression changes of transcripts involved in ABA signalling and biosynthesis we can see that of the many genes that are up-regulated by ABA and stress treatments in wild type tissue, almost all of them show abolition of the response in the *PpanrKO* mutant. Indeed, very few genes show any ABA-mediated regulation in the *PpanrKO* mutant. This section will look at some of the key transcripts involved in the ABA/stress response. It is generally found that plant tissues treated with ABA alone will share many molecular responses with a range of abiotic stresses such as dehydration and this is clearly seen in *Physcomitrella* wild-type tissue too (Figure 3.3). It is also known that stress responses in plants do not depend entirely on ABA mediated signalling and that there are additional ABA-independent pathways. In order to understand some of the responses regulated by these pathways, gene ontology ("GO") term annotation was carried out using the Blast2Go tool to associate differentially regulated transcripts with functions and processes. An important caveat about using this type of analysis is that many sequences have no known functional annotations and of those that do, many are incorrect as functional annotation based on sequence homology cannot for example, account for sub- or neofunctionalisation of proteins. This approach does however allow descriptive, rather than quantitative, overviews of the kinds of molecular processes, functions and locations that are affected by treatments. For example, of the 1806 significantly up- or down-regulated transcripts from all three wild-type treatments, 169 failed to match any other sequence in the database by BLASTP and 769 were not annotated with any GO terms meaning they were not part of the analysis. Another feature of GO term analysis is that because of the hierarchical nature of its structure, a single transcript can be associated with more than one GO term when one may be the 'parent' or 'child' term of another.

To get an overview of the kinds of biological and metabolic processes that ABA and stress treatments have in both wild-type and *PpanrKO* tissue, a number of terms were selected that had functional relevance. These included terms known to be involved in general stress responses (*e.g.* 'response to stress') as well as being highly represented in the GO annotation (Figures 3.4 and 3.5). As part of the up-regulated transcripts, we see a number of terms involved in membrane transport processes, such as 'ion transportation'. Adjusting ion balances is an important part of responding to decreasing water availabilities to maintain cellular homeostasis, as seen in *T. ruralis* (Oliver *et al*., 2004). Even without exposure to salt stress, a decrease in cellular water will increase the concentration of various ions that need to be moved; either by efflux or by

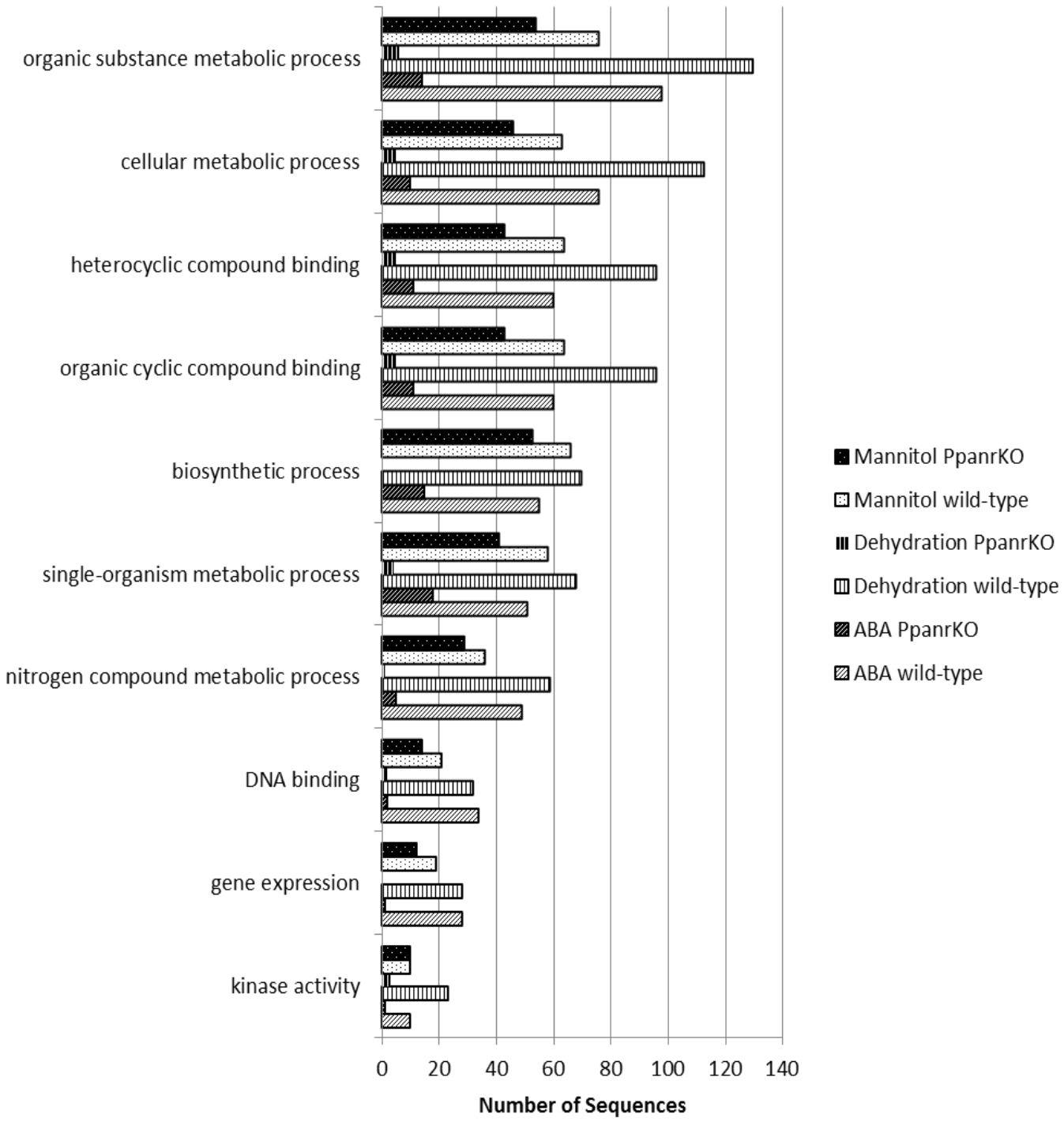


Figure 3.5. **GO terms featuring in the down-regulated genes in wild-type tissue.** The abolition in responses for *PpanrKO* (white patterns on black background) compared to wild-type (black patterns on white background) is not as severe in these down-regulated genes as compared to the up-regulated genes. This is most notable in the mannitol treatments where there is little difference between the two lines. Most of the GO terms are related to metabolism which is generally repressed during stressful conditions especially during dehydration where most proteins become nonfunctional as cellular water levels decrease. These GO terms represent molecular functions and biological processes and are as follows from top to bottom: GO:0071704, GO:0044237, GO:1901363, GO:0097159, GO:0009058, GO:0044710, GO:0006807, GO:0003677, GO:0010467, GO:0016301.

compartmentation (see Osakabe *et al.*, 2014 for recent review in angiosperms). For example, iron homeostasis is crucial for maintaining proper metabolic processes such as photosynthesis and respiration. One such transcript encodes for a putative MATE family iron transporter (Phpat.010G068000) which shows dramatic up-regulation by ABA and dehydration, and related proteins are known to have important roles in osmotic stress responses (Seo *et al.*, 2012). Two putative ferritin encoding transcripts (Phpat.007G031700 and Phpat.010G014500) which are involved in iron storage also show a similarly strong ABA up-regulation. These two represent a typical response to osmotic stresses as ions, such as iron, become more concentrated which is frequently associated with toxic side effects and they need to be sequestered to prevent this.

Another GO term that is very prominent is ‘hydrolase activity’ which is a very general term for any enzyme involved in the hydrolysis of a chemical bond. By inspecting many of the daughter terms that are involved in ‘hydrolase activity’ two of the most common terms are ‘hydrolase activity, acting on ester bonds’ and ‘hydrolase activity, acting on glycosyl bonds’. One such transcript with ‘hydrolase activity’ (Phpat.010G022400), encoding a putative GDSL hydrolase superfamily protein which include lipases and esterases, is very strongly up-regulated by ABA and stress. Those transcripts annotated as being involved in glycosyl hydrolase are likely involved in complex sugar degradation such as those involved in mobilizing storage sugars (reviewed in Minic, 2008). Other transcripts encode putative enzymes involved in cell wall related complex sugar modification. For example, one such transcript encodes a putative alpha-xylosidase (Phpat.026G032800) whose most homologous protein in *Arabidopsis* (*AtXYL1*) is known to be involved in the degradation of xyloglucan in cell walls (Gunl and Pauly, 2011). Reorganizing the cellular make up of complex and simple sugars is also seen in the terms such as ‘disaccharide metabolic process’, ‘disaccharide metabolic process’ and ‘cellular carbohydrate metabolic process’. The up-regulation of this process likely enables the production of osmolytes among other things - in *Physcomitrella* the key osmolyte is thought to be sucrose (Oldenhof *et al*., 2006) - that protect cells in the absence of water through a variety of potential mechanisms (see chapter 1.2).

Another prominent GO term in the annotation analysis is ‘nitrogen compound metabolic process’; which also features in down-regulated transcripts showing the limitations of this approach without further in-depth analysis. In general, as seen in the GO terms for many down-regulated transcripts, metabolic processes feature strongly and yet a number of transcripts, such as a putative glutamine synthase (Phpat.026G041000) and a putative glutamate synthase (Phpat.016G078900) shows strong ABA and stress up-regulation. This may be explained by glutamine being readily converted into glutamate

which is an important amino acid from which many others are readily metabolized such as proline and arginine (see Forde and Lea, 2007 for review). Proline is an osmolyte that is known to be synthesized from glutamate (Strizhov *et al*., 1997) and so this may provide a mechanism for protecting cells during dehydration. Also, by synthesizing such an amino acid central to the synthesis of many others, the resumption of protein synthesis may be more effectively started when normal conditions resume. Another aspect is that glutathione - which is a crucial component of the oxidative damage response in plants - is synthesized from glutamate (along with cysteine and glycine). This indicates that a number of specific yet energetically expensive metabolic processes may still be induced by ABA and stress if they are part of the protective processes that also have roles in the recovery phase.

Generally, though, annotations for transcripts down-regulated by ABA and stress are dominated by metabolic and biosynthetic processes. This is indicative of cellular conditions entering a metabolically dormant period, as seen in the clear ABA-responsive phenotype in which growth is arrested (Figure 1.6), as water availability decreases which negatively affects both photosynthesis and respiration so reducing energy availability. As indicated above, this includes ‘nitrogen compound metabolic process’ which is likely linked to general protein synthesis which is a major component of energy expenditure in plants. Besides most metabolic and biosynthetic processes being down-regulated, we also see a number of transcription regulation and signalling-related terms among the down-regulated gene set, such as ‘DNA binding’, ‘gene expression’ and ‘kinase activity’. A number of the transcripts annotated with ‘kinase activity’ have no clear functional role such as Phpat.005G016700, Phpat.014G068600, Phpat.001G087600, Phpat.002G000600, Phpat.005G067400, and Phpat.007G085400. Two other interesting transcripts that likely have kinase activity are Phpat.021G064400 which has high similarity to *AtHT1* involved in stomatal movements (Hashimoto *et al.*, 2006) and Phpat.001G045000 which is a MEK-like MAP3K most similar to the Arabidopsis *MAP3K3*. The pathways these putative kinases are involved in are therefore likely to be repressed by ABA-dependent stress signalling.

3.3.6 ABA/stress regulated genes and co-regulation – cluster analysis

Looking at individual genes that are regulated by ABA or stress and at the general functions these are known to or are likely to encode tells us a lot about what the patterns and features of the responses are. Cluster analysis provides a means by which we may understand another level of detail about gene co-regulation. This process statistically clusters genes by their shared expression levels and changes between libraries from which one can better understand the groups of genes that are likely to be co-regulated; by the sharing of *cis*-regulatory elements for example.

Analysis was carried out by manually defining 44 clusters from all genes that showed significant differential regulation in any one of the treatments from wild-type or *PpanrKO* tissue. By carrying out this analysis only on significantly regulated genes, which takes into account the false discovery rate (FDR) score so generally excluding low abundance genes, the differential expression detection is more robust.

Table 3.4. Summary of gene expression clusters from wild-type tissue. Number of transcripts signifincalty regulated by ABA, dehydration (DH) and mannitol (Man).

Clusters down-regulated	ABA	DH	Man	Total in cluster	Clusters up-regulated	ABA	DH	Man	Total in cluster
36	45	81	116	140	**1**	119	119	107	119
38	10	16	93	163	**10**	83	83	69	83
15	60	38	12	74	**14**	97	70	53	99
17	73	24	11	134	**5**	89	83	14	275
25	13	60	0	170	**13**	75	74	23	76
40	16	46	6	91	**7**	82	77	3	97
16	15	42	8	133	**8**	77	54	1	104
39	9	8	45	50	**20**	12	14	35	117
26	19	38	0	88	**4**	35	4	0	260
23	1	55	0	103	**11**	13	10	12	14
24	0	45	0	168	**6**	13	16	2	133
22	6	7	19	107	**43**	13	6	10	142
37	6	0	24	86	**9**	17	6	0	91
3	0	0	26	68	**12**	10	7	0	60
27	24	2	0	101	**19**	0	0	17	86
33	9	5	7	100	**34**	0	0	16	22
21	1	0	18	50	**41**	0	0	16	66
42	0	16	2	64	**2**	4	2	1	59
44	3	8	0	116	**3**	0	5	0	68
19	8	2	0	86	**32**	4	0	1	52
18	5	3	1	33	**18**	0	2	1	33
35	3	3	3	3	**30**	0	0	3	15
2	0	0	7	59	**31**	0	0	3	12
34	0	5	0	22	**26**	0	0	1	88
29	2	1	0	21	**Total**	743	632	388	2171
28	2	0	0	96					
43	0	0	2	142					
9	0	0	1	91					
32	0	1	0	52					
Total	330	506	401	2611					

In wild type-tissue, there are 24 clusters to which all the up-regulated genes from all three treatments belong (Table 3.4). There are five clusters which include genes that are strongly up-regulated in all treatments and all show a very similar expression profile (Figure 3.6) and as such can be treated as a 'super-cluster'. Four of the 5 are made up almost entirely (>98%) of genes showing significant up-regulation by ABA. These genes represent the core ABA-dependent stress responsive genes and include those encoding LEA proteins, transmembrane transporters, catalases, ELIPS and enzymes involved in sugar metabolism (see Supplementary data). These ABA-induced stress responsive genes all show abolition of the response in *PpanrKO* tissue with many even showing a measurable down-regulation in *PpanrKO* under control conditions. Within these ABA-dependent stress responsive genes are a number encoding putative signalling components (Table 3.5) including *PpOST1-1*; known to be part of the 'core' ABA signalling pathway and important in ABA-induced stomatal closure in the moss sporophyte (Chater *et al*., 2011) (Table 3.2).

Table 3.5. Identification of putative signalling components strongly implicated in ABA-dependent osmotic stress signalling. Fold changes are shown in response to ABA, dehydration (DH) and mannitol (Man) treatment.

V3.0 ID	Protein Family	Putative function	Fold change (ABA, DH, Man)	Cluster
Phpat.005G078000	SnRK2 kinase	*OST1-1* ABA signal transduction	28, 22, 30	14
Phpat.010G014900	Two-component system-like histidine kinase	Likely orthologue of *AtHK1* a putative osmosensor	27, 34, 39	14
Phpat.005G075000	PP2C; C subfamily	Likely general stress signalling	96, 79, 23	10
Phpat.017G037200	PP2C; C subfamily	Likely general stress signalling	34, 19, 26	14
Phpat.003G031200	General Regulatory Factor (14-3-3)	Large regulatory family involved in many signalling processes	129, 70, 25	10
Phpat.009G024400	CBF/NF-Y	Subunit C4 of CCAAT binding TF	184, 89, 38	10
Phpat.010G046900	DREB; subfamily A-2 protein containing AP2 domain	TF likely involved in osmotic stress responses	142, 120, 33	14
Phpat.008G027400	DREB; subfamily A-4 protein containing AP2 domain	TF likely involved in osmotic stress responses	17, 27, 18	14

These signalling components are all very strongly up-regulated by ABA, dehydration and mannitol and constitute all the elements of a signal transduction pathway, although this need not imply they function as a linear pathway. These include a homologue of the osmosensor *AtHK1*, a pair of PP2C phosphatases, the aforementioned *PpOST1-1*

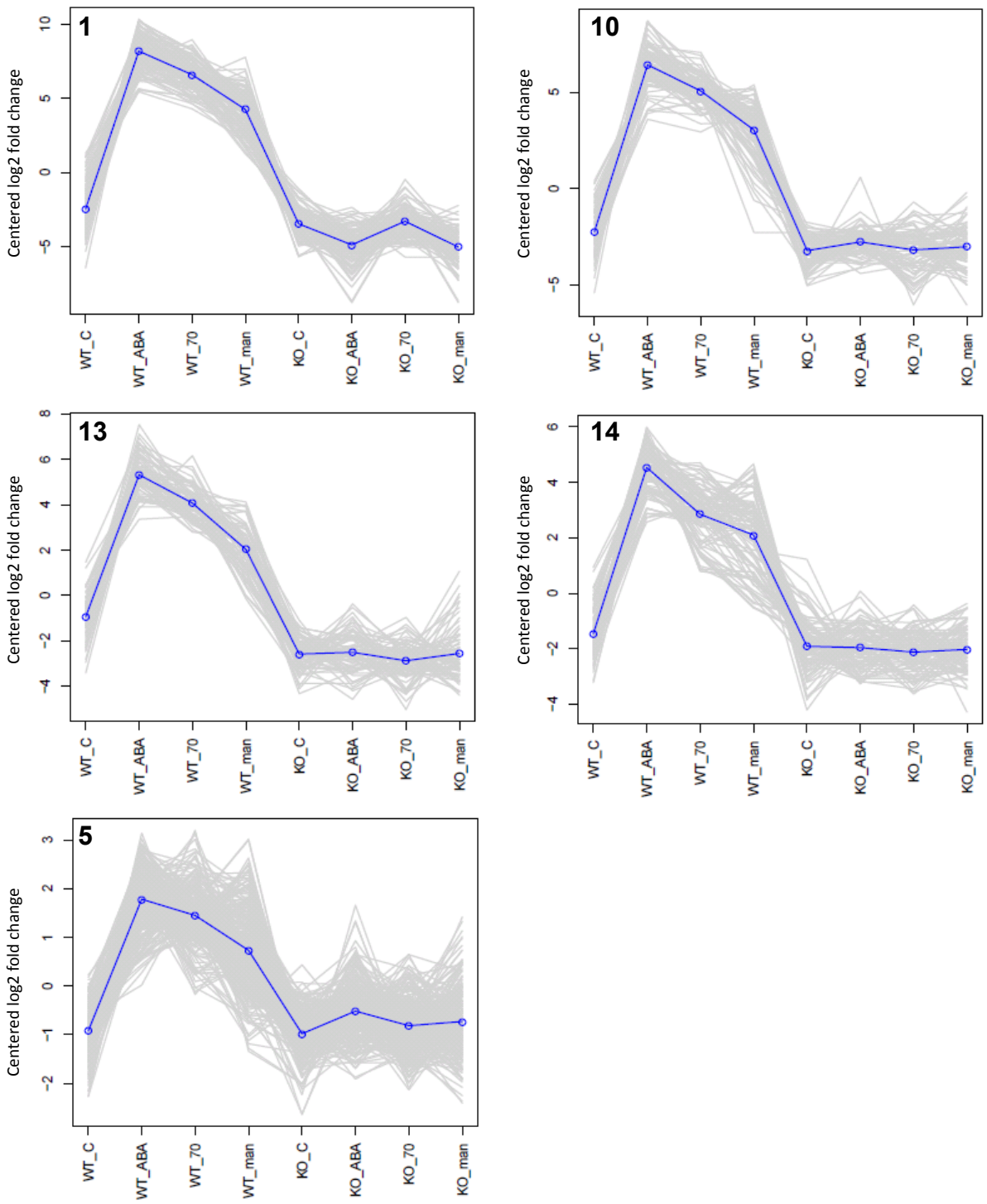


Figure 3.6. Five clusters containing genes strongly up-regulated by all treatments in wild-type tissue. All five could be considered a 'super-cluster' containing transcripts that represent the 'core' ABA-dependent stress response. Each transcripts expression is plotted as a centered log2 value which 'centers' all expression levels to a global mean so making the relative changes relevant. The mean expression for each cluster is shown (blue line) with each clusters number shown.

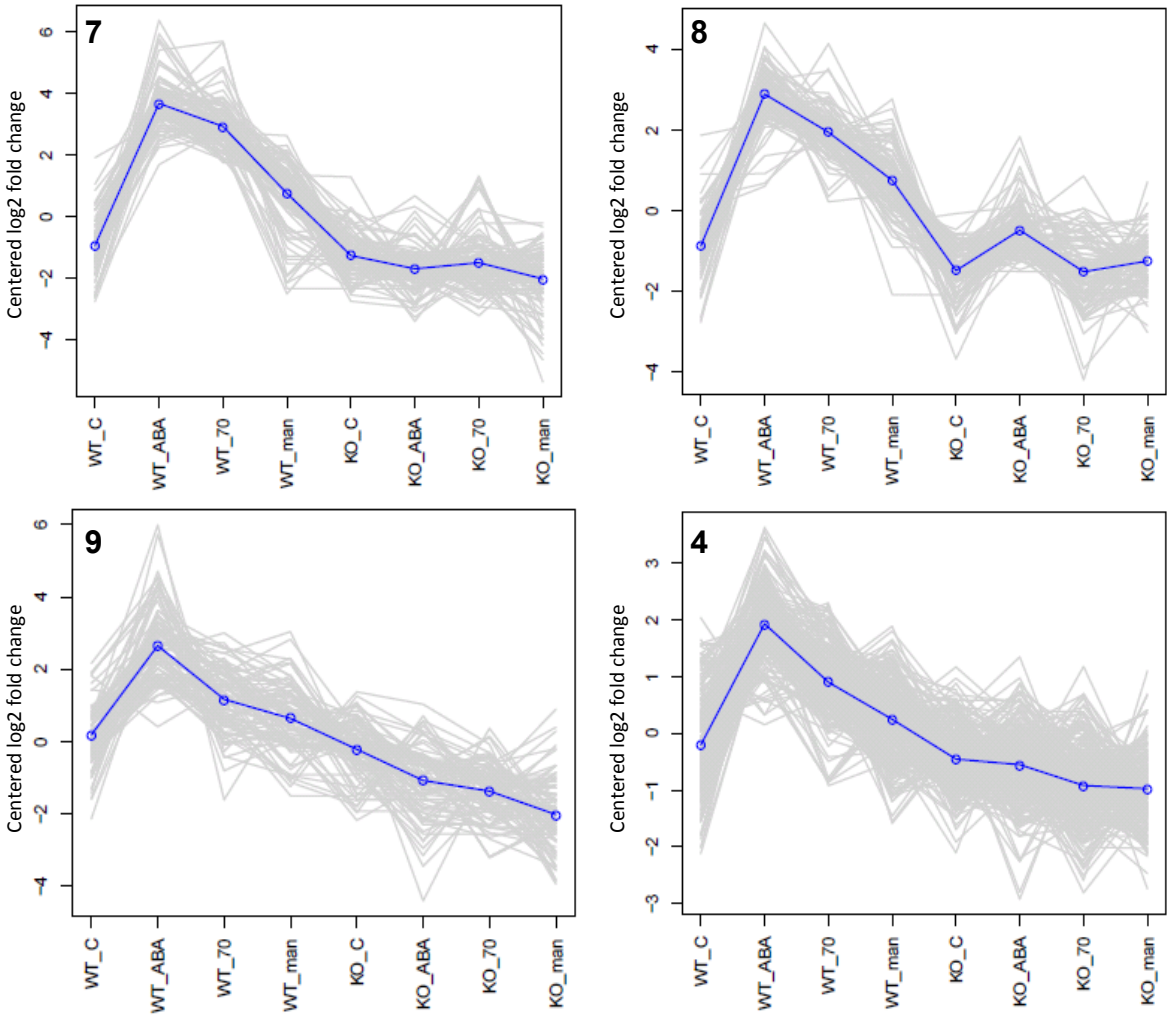


Figure 3.7. Four clusters containing genes up-regulated by ABA and dehydration in the wild-type. As can be seen, the general patterns show that these transcripts are generally up-regulated only slightly by mannitol. Each transcripts expression is plotted as a centered log2 value which 'centers' all expression levels to a global mean so making the relative changes relevant. The mean expression for each cluster is shown (blue line) with each clusters number shown.

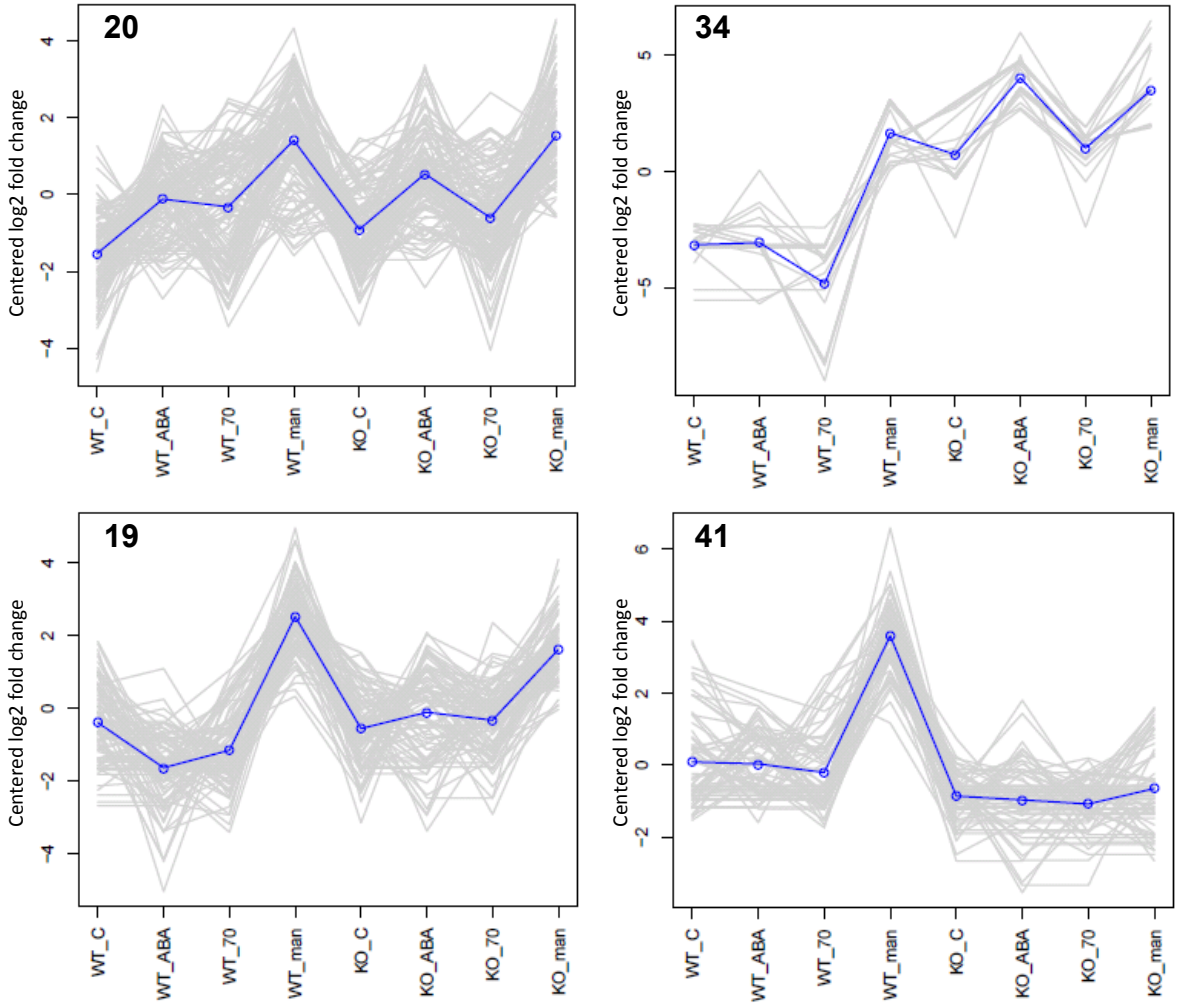


Figure 3.8. Four clusters containing genes significantly up-regulated by mannitol in the wild-type. Two of these clusters, 20 and 19, show this response to be largely maintained in *PpanrKO*. 41 shows transcripts up-regulated by mannitol in which the response is lost in *PpanrKO*. Cluster 34 contains few transcripts which are strongly up-regulated by mannitol in wild-type and appear to require *PpANR* for repression in ABA and dehydration treatments as well as control conditions. Each transcripts expression is plotted as a centered log2 value which 'centers' all expression levels to a global mean so making the relative changes relevant. The mean expression for each cluster is shown (blue line) with each clusters number shown.

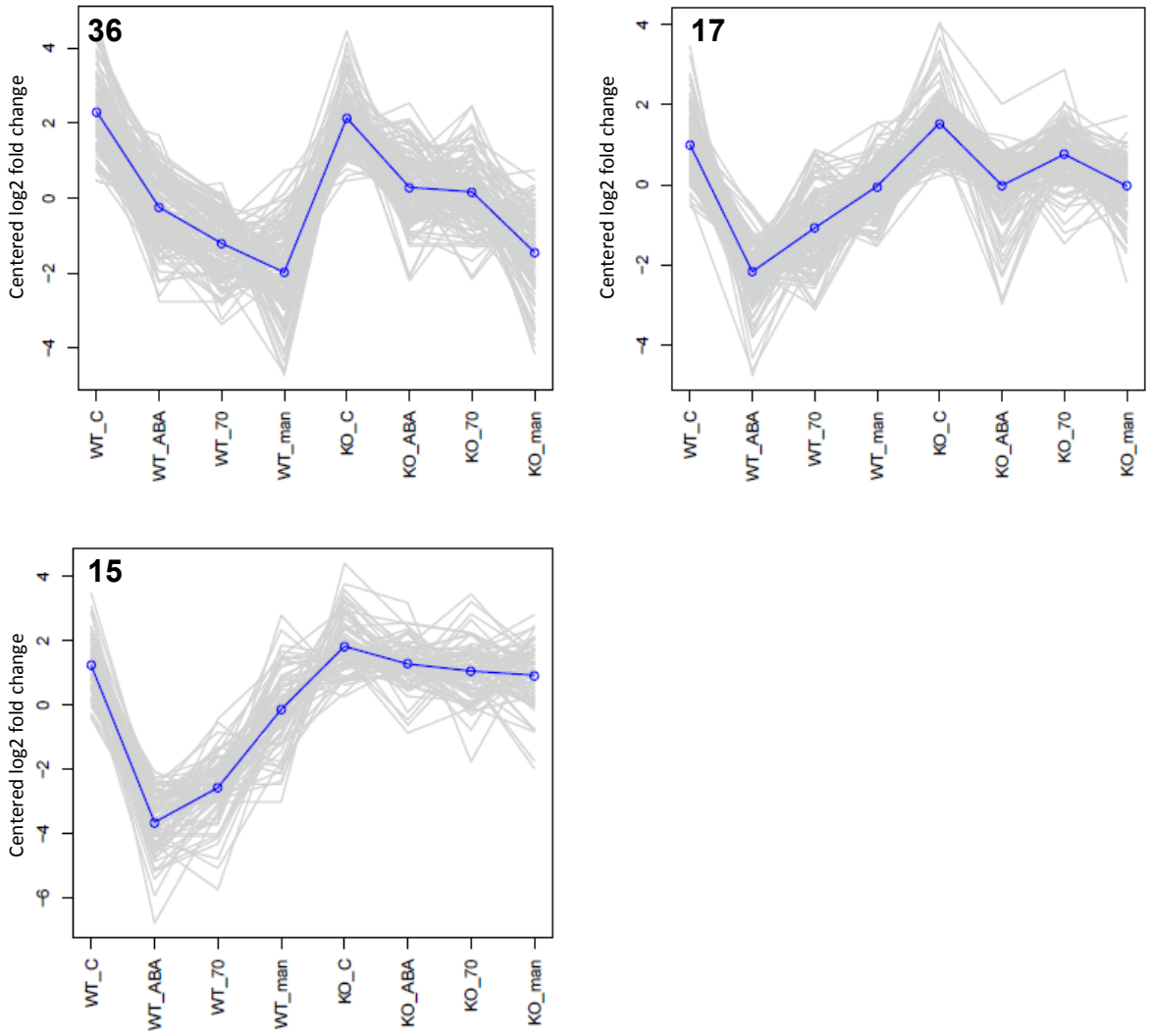


Figure 3.9. Three clusters containing genes strongly down-regulated by all treatments in wild-type tissue. Cluster 36 has a strikingly different expression profile which shows the down-regulation of these transcripts is maintained in *PpanrKO*. Clusters 17 and 15, however, show transcripts in which the down-regulation has largely been lost in *PpanrKO*. Each transcripts expression is plotted as a centered log2 value which 'centers' all expression levels to a global mean so making the relative changes relevant. The mean expression for each cluster is shown (blue line) with each clusters number shown.

kinase, a 14-3-3 regulatory factor and three TFs whose families are implicated in stress-induced gene expression. The strong implication of these components in ABA-dependent responses shows that phosphorelay pathways, involving both kinases and phosphatases, are likely crucial. Two of the TFs are part of the DREB subfamily of AP2/ERF TFs which are strongly implicated in the dehydration response mediating gene expression through the DRE (drought responsive element) *cis*-element. As all of these components, like all genes found in these clusters, show abolition of this strong response in *PpanrKO*, *PpANR* is clearly vital for their induction and the downstream responses they regulate.

Four further clusters show similar, but weaker expression changes with most genes failing to show significant up-regulation by mannitol (Figure 3.7). Four interesting clusters are also uncovered that show genes preferentially up-regulated by mannitol (Figure 3.8). Two of these clusters (19 and 20) contain genes that are still responsive to mannitol in *PpanrKO* while cluster 41 contains genes in which this response is largely lost. These clusters highlight a subtle difference between those genes in which the mannitol response is ABA-independent (clusters 19 and 20) and those whose response is ABA- (or more strictly *PpANR*-) dependent (cluster 41). Cluster 34 contains few genes with a divergent expression profile in which ABA/*PpANR* signalling is required to repress expression while mannitol treatment can clearly override this in wild-type tissue. Without *PpANR*, and thus ABA signalling, these genes are constitutively expressed at high levels (Figure 3.8). This cluster is in fact made up of a large group of putative alpha carbonic anhydrase family members which appear to comprise two clusters of near identical genes on chromosomes 7 and 26.

In wild-type tissue there are genes from 29 clusters in which expression is down-regulated by one or more of the treatments (Table 3.4). Three of these contain genes strongly down-regulated by all three treatments but showing subtle differences (Figure 3.9). Cluster 36 for example shows that down-regulation by ABA/stress treatments in these genes is retained, albeit slightly attenuated, in *PpanrKO* and this cluster contains a large number of putative aromatic (phenyalanine or histidine) ammonia lyases which are key enzymes in phenylpropanoid biosynthesis; a secondary metabolic pathway and clearly not afforded in stressful conditions. Cluster 15 and 17 contain genes whose down-regulation is largely lost in *PpanrKO* with many putative transcription factors found in both clusters including AP2/EREBPs, bHLHs and bZIPs, as well as a number of protein kinases (see Supplementary data). One possibility is that these may be regulators of ABA-dependent down-regulation mirroring those putative signalling components involved in ABA-dependent up-regulation found in clusters 1, 5, 10, 13 and 14 (Figure 3.6).

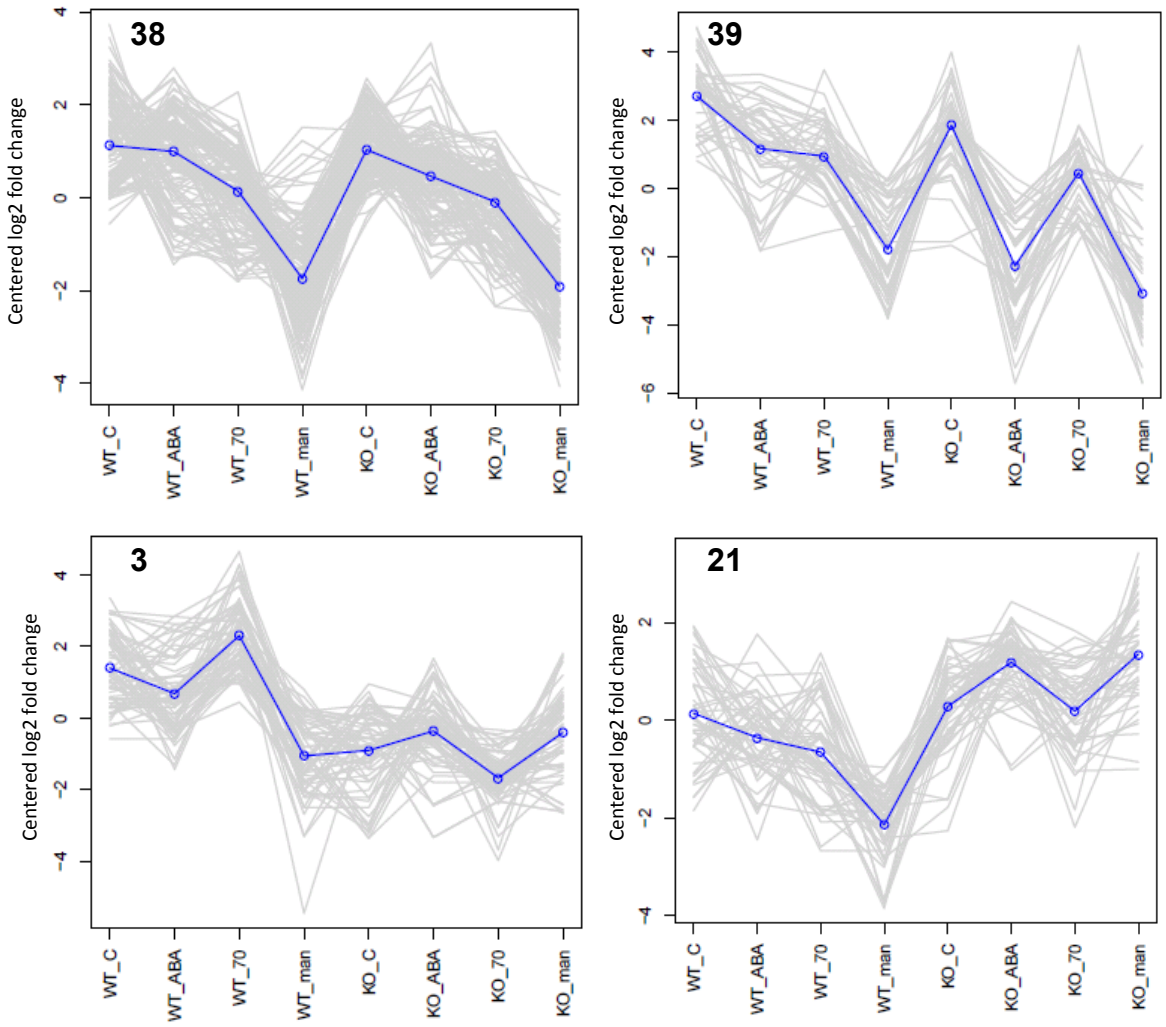


Figure 3.10. Four clusters containing genes strongly down-regulated by mannitol treatment in wild-type tissue. Cluster 38 shows the responses are largely retained in *PpanrKO*. Cluster 21 shows that mannitol has the opposite effect in either line. Each genes expression is plotted as a centered log2 value which 'centers' all expression levels to a global mean so making the relative changes relevant. The mean expression for each cluster is shown (blue line) with each clusters number shown.

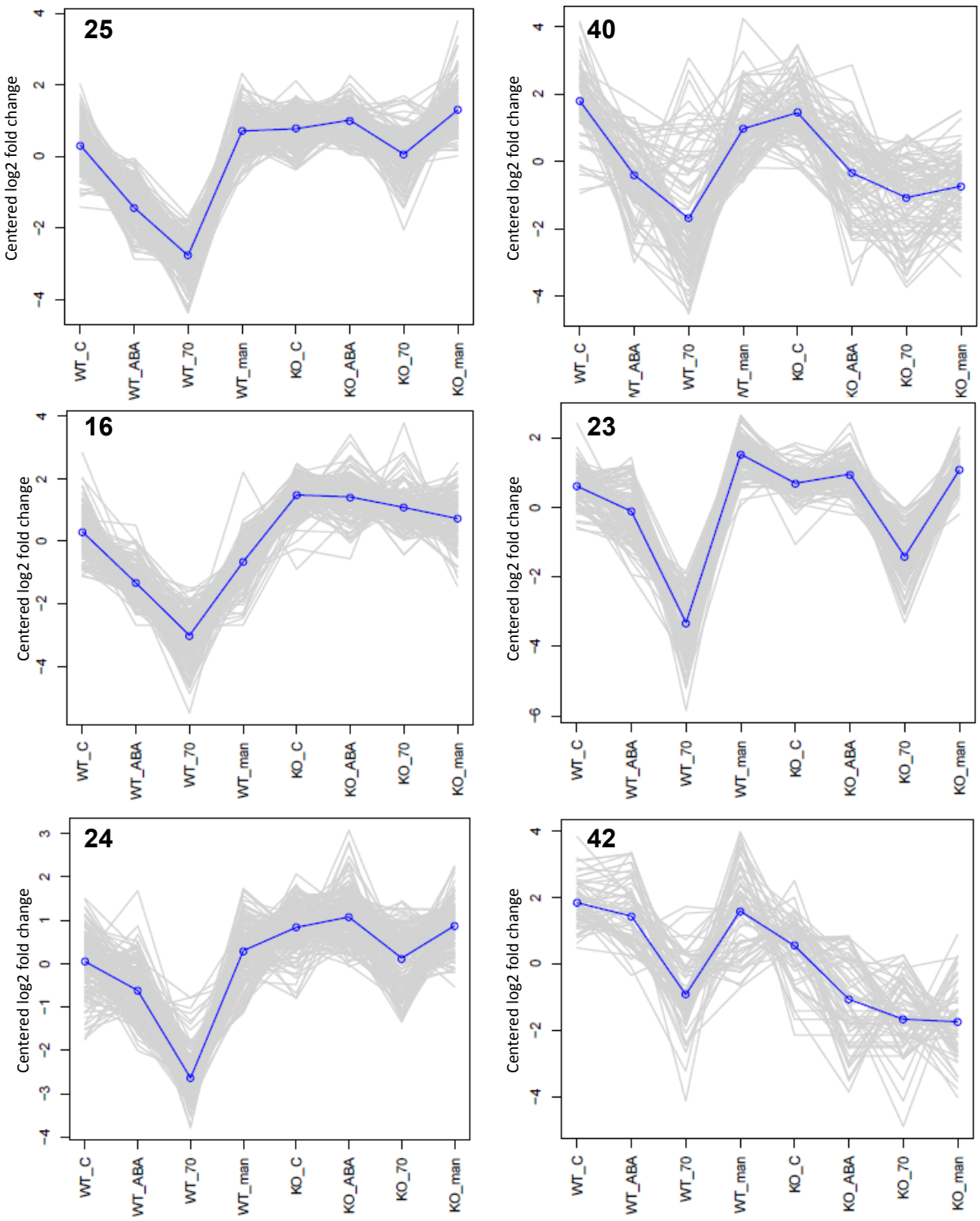


Figure 3.11. Six clusters containing genes strongly down-regulated by dehydration in the wild-type. Clusters 23, 24, 25 and 16 show very similar expression patterns with varying levels of the dehydration-induced down-regulation being lost in *PpanrKO*. Each genes expression is plotted as a centered log2 value which 'centers' all expression levels to a global mean so making the relative changes relevant. The mean expression for each cluster is shown (blue line) with each clusters number shown.

In much the same way that a suite of genes were identified with mannitol-specific up-regulation (Figure 3.8), a number of clusters contain genes specifically down-regulated by mannitol but with quite striking differences otherwise (Figure 3.10). The large cluster 38 contains genes in which mannitol-induced down-regulation is largely retained in *PpanrKO* suggesting that down-regulation of these genes is ABA-independent. Cluster 39 contains genes whose strong mannitol response is not affected in *PpanrKO* and that additionally appear down-regulated by ABA specifically in the *PpanrKO* mutant. This suggests that *PpANR* may be responsible for dampening any effect of ABA possibly by working antagonistically with some other ABA regulator. While cluster 39 contains a number of putative enzymes with hydrolase activity, it also contains a very high proportion of conserved plant proteins with unknown functions whose functions could now be associated with osmotic stress responses. Cluster 3 likely represents genes in which *PpANR* maintains constitutive expression until overridden by ABA-independent mannitol responses and when *PpANR* is deleted, the constitutive response is lost (Figure 3.10). Finally, cluster 21 contains genes whose down-regulation by mannitol is completely lost and even reversed in *PpanrKO*.

As the previous results demonstrated, there is much stronger correlation and overlap between genes that are up-regulated by ABA/stress than those that are down-regulated (Figure 3.3). In this context, there are 6 clusters which show preferentially stronger down-regulation by dehydration (Figure 3.11). Three of these clusters (23, 24 and 42) could be considered to be almost specifically dehydration-down-regulated in their response. Cluster 23 shows the maintenance of this response in *PpanrKO* and cluster 24 shows an attenuation of the response. Cluster 42, however, contains genes that likely require *PpANR* to maintain a certain level of constitutive expression which is overridden by dehydration through ABA-independent means. Clusters 16, 25 and 40 show strong dehydration- induced down-regulation and weak ABA-induced down-regulation with no response to mannitol (Figure 3.11). Cluster 40 contains genes that largely retain this down-regulation by ABA and dehydration in *PpanrKO* but, interestingly, are also down-regulated by mannitol. This cluster contains a number of predicted xyloglucan endo-transglycosylases (XET) suggesting that down-regulation of growth-related cell wall modification is an important feature of this cluster. These enzymes are involved in allowing the reconfiguration of cellulose microfibrils during growth by breakage and rejoining of hemicellulose tethers, allowing cell expansion. This is therefore a key example implicating a mechanism behind the growth arrest observed in response to ABA and stress. Clusters 16 and 25 contain a number of TFs and kinases likely indicating an important role of *PpANR* and ABA-dependent signalling in regulating secondary responses to dehydration which are lost in *PpanrKO*.

3.4 Discussion

3.4.1 Validity of differential gene expression findings

This experiment was principally aimed at characterizing the global effects of *PpANR* in ABA/stress signalling but in order to do this, a robust and accurate analysis of the wild-type response was required against which to compare the mutant. Analysis of the libraries prior to mapping and differential gene expression analysis revealed the reads to be of high quality meaning little bias was incorporated into the analysis. The general patterns of global change induced by ABA and osmotic stress treatments, discussed below, were largely consistent with previous experiments demonstrating that the RNA level changes found were likely robust proxies of molecular changes due to the treatments. There were a number of issues and caveats associated with the experiment however. One is that RNA-seq measures the sum of net RNA synthesis against degradation and the longer dehydration treatment (up to 12hrs) is likely to have resulted in greater levels of transcript degradation. It is also worth noting that ABA/stress responses have been found to be induced rapidly (*e.g.* Kamisugi and Cuming, 2005; Cuming *et al.,* 2007) and reach a plateau meaning that over the 12hrs dehydration treatment this degradation and RNA turnover will again have some bias when comparing between the other treatments. One other potential weakness, especially in the statistical analysis of differential expression, is the nature of the replication. Each library was the product of three bulked biological replicates, rather than each representing a set of individual replicates, meaning they were an average measurement of three samples; however, without the three values no true assessment of variation can be made. This analysis relied on the measurement of false discovery rate (FDR) to find significance in differential expression which enables confidence to be obtained by taking into account not only the fold change but the number of reads mapped to the gene of interest from each library which affect the influence that stochastic changes have; *i.e.* lower expressed genes have a greater amount of stochastic influence. Using an FDR cutoff value (P-value<0.001), in conjunction with good sequencing depth, ensured significance could be robustly found. In RNA-seq experiments, a judgement has to be made in which the value of additional replication is set against other parameters: which include the extent of biological variation likely, and the significant added cost associated with the analysis of individual, rather than bulked replicates. In the case of *Physcomitrella* the tissue analysed is biologically highly uniform. The wild-type and mutant lines represent, respectively, genetically uniform samples, with the *anr* mutation the only difference between the lines. The tissue in each case represents an identical stage of development: chloronemal cells derived from vegetatively subcultured tissue grown under highly controlled conditions in culture.

In short, the biological replicates that were bulked were highly uniform in nature comprising a single cellular population.

3.4.2 The wild-type ABA and stress responses

The approach taken in this experiment has enabled a number of important findings to be made about the wild-type response. The overlap in expression between ABA and the two osmotic stress treatments is clearly significant and is consistent with an important role for ABA in abiotic stress signalling. A shared and tightly co-regulated set of 'primary' ABA-dependent stress response genes have been identified by cluster analysis. These contain a large number of LEA and transmembrane proteins as well as enzymes involved in sugar metabolism. Many components of the core responses to osmotic stress have proven and hypothesized roles in protecting other proteins and membranes in the absence of water possible by acting as molecular buffers or through the formation of a glass-like state from high concentrations of sucrose (Oldenhof *et al*., 2006). LEA proteins are intrinsically disordered hydrophilic proteins thought to have a wide range of roles in osmotic stress responses even outside plants (reviewed in Battaglia *et al*., 2008 and Tunnacliffe and Wise, 2007) and quite clearly are a vital and significant component of the ABA-dependent stress response in *Physcomitrella*. The many genes that encode transmembrane proteins are predicted to be principally involved in the transport of ions to maintain homeostasis. Many ions are toxic in high concentrations and as such, as water levels decrease and their concentrations increase, they require sequestering (through efflux or compartmentalisation) with many transporters necessary for the different cargoes. There are a number of co-regulated genes that showed ABA-independent up-regulation; with specificity to either dehydration or mannitol. Down-regulated targets appeared to show less correlation between treatments possibly indicating a higher specificity in those pathways turned off during stress although, as discussed, the nature and length of treatments is likely to have an effect on this too. Many of the down-regulated genes encoded proteins likely involved in biosynthesis and metabolism which indicates that many normal processes involved in growth and development were repressed in the face of unfavourable conditions. This is in clear correlation with the observed phenotypes of the ABA/stress response in which filamentous growth is arrested (Chapter 2). One key feature of the down-regulated genes is that many genes are predicted to be involved in cell wall growth and reorganisation. XET enzymes, crucial for allowing the expansion of cell walls through the restructuring of cellulose, are a prominent component of this response and their down-regulation is likely an important part of the osmotic stress response in arresting growth. The greater specificity of those pathways down-regulated demonstrates an important feature of the ABA/stress response where the core ABA-

dependent response leads to a general protective state while fine tuning of down-regulated processes, likely through ABA-independent pathways enables more stress-appropriate responses to be mediated. Many of these may be considered secondary responses and the presence of a number of putative signalling proteins encoded by both the up- and down-regulated genes suggest likely mechanisms by which these responses are achieved as these signalling components are synthesized and activate their targets like molecular dominoes.

3.4.3 The role of endogenous ABA and signalling reinforcement

ABA signalling requires the regulation of both ABA biosynthesis, to increase endogenous levels, and signal transduction following its detection to induce molecular changes. In agreement with previous studies in angiosperms, members of the key NCED enzyme family are found to be strongly up-regulated by ABA, dehydration and mannitol although ABA-dependent cold tolerance in *Physcomitrella* has been previously shown not to be accompanied by an increase in endogenous ABA levels (Minami *et al*., 2005). This suggests that previous measurements of endogenous ABA may not have been very accurate or that only small increases are required to trigger a robust response. In angiosperms, *de novo* ABA biosynthesis is often followed by long distance transport of ABA along the vascular system to target sites, such as stomata (Koiwai *et al*., 2004), and as such transport is not present in bryophytes it seems likely that intracellular levels would be required to increase to signal the ABA-dependent response; a hypothesis backed up by this strong up-regulation of *NCED* genes. As seen by many of the reinforcing mechanisms of ABA-dependent signalling such as up-regulation of key signalling components (*e.g. OST1-1* and *PYL-5*), biosynthetic components (*e.g. NCED1* and *NCED2*) as well as likely secondary signalling components, such as those found in the core ABA-dependent response clusters, it is possible that a small spike in ABA concentrations is all that is required to activate responses. The putative *CYP707A* genes involved in ABA catabolism show strong down-regulation in response to ABA and osmotic stress indicating another shared feature with angiosperms and indicate another mechanism by which endogenous ABA levels can be increased. As mentioned, many of the ‘core’ signalling components also show ABA and stress up-regulation, in particular the OST1 family of kinases and their putative bZIP TF targets. Together, as positive regulators, these represent an important target for maximising the ABA response. Together this represents a strategy for maximising a very rapid and robust molecular response which is essential for the anatomically simple bryophytes that are more at the mercy of abiotic stresses. The activation of downstream target TFs, such as the ABI3 and bZIP-like TFs, leads to ABRE mediated ABA induced gene expression and importantly this whole process

enables positive feedback on itself through the expression of these same components likely increasing ABA sensitivity. Analysis of *cis*-element patterns and enrichments for this model would be an excellent addition to this data. This working model of how ABA-dependent molecular responses are induced is largely consistent with previous studies but the main question for this experiment was to discover the role of *PpANR* in all this.

3.4.4 Role of PpANR

One of the critical findings of this experiment was that *PpANR* is absolutely crucial for ABA-dependent molecular responses. In the *PpANR* deletion line, *PpanrKO*, the total number of genes significantly up-regulated by ABA and dehydration is almost completely abolished indicating that the protective measures that moss tissues require for dehydration tolerance are not expressed. The effect is not as dramatic for those genes regulated by mannitol treatment. The effects are manifest in the dehydration sensitivity observed in phenotypic analysis of *PpanrKO* and other *Ppanr* mutants (Chapter 2). In the absence of *PpANR*, the enhanced ABA biosynthesis and increased ABA signalling sensitivity discussed is simply lost. The up-regulation of *NCED* genes is clearly ABA and *PpANR*-dependent as no partial up-regulation is found in the dehydrated or mannitol treated *PpanrKO* tissue. This suggests that an increase in endogenous ABA levels is not induced in *PpanrKO* mutants which is also coupled with a loss of up-regulation of the ABA regulators meaning two levels of reinforcement are lost accounting for the dramatic abolition. PpANR is a MAP3 kinase and as such is likely to function through phosphorylation-mediated activation of targets and, while these targets remain unknown, the loss of such activation would clearly seem to be the vital factor in explaining the abolition of the response. As highlighted in the wild-type response, there are a number of ABA-independent pathways involved with the dehydration and mannitol responses and many of these can be seen to be still functional in *PpanrKO*. This is most clearly seen in those genes down-regulated by mannitol, where the wild-type response is largely unaffected, suggesting that the principal role of *PpANR* and ABA-dependent signalling are in gene activation, rather than gene repression in osmotic stress responses. There were a number of gene expression clusters that showed ABA-independent regulation and yet still showed perturbed responses to the respective treatments in the *PpanrKO* mutant which highlights the crosstalk between the ABA-dependent and ABA–independent pathways in coordinating complex responses. This also highlights just how far-reaching the role and function of *PpANR* is showing it to be a potential ‘master regulator’ for a number of direct or indirect targets.

3.4.5 Relationship of PpANR and the core signalling pathway

This RNA-seq experiment has been successful both in showing the wild-type responses of *Physcomitrella* to ABA and stress, and in demonstrating the dramatic affect that deletion of the *PpANR* gene has on these responses. PpANR is clearly a vitally important ABA regulator in moss. Its deletion is not functionally covered by the presence of the members of the supposed 'core' pathway or any other ABA regulators. This suggests that PpANR functions in a linear fashion within the 'core' pathway and that the way the 'core' pathway operates in angiosperms might be distinctly different to that in bryophytes; not least because there is no *PpANR* orthologue in higher plants although related MAP3Ks may be functionally equivalent (see Chapter 4). The two PP2C phosphatases, *PpABI1* and *PpABI2*, have been found to be involved in the core ABA response pathway in *Physcomitrella* (Komatsu *et al*., 2013). The deletion of these PP2Cs did not have the expected effect on what were identified as OST1-like kinases (based on predicted size in an SDS-PAGE gel) in that these proteins were not constitutively phosphorylated in the absence of PP2C-mediated dephosphorylation: their phosphorylation required ABA treatment. This led the authors to suggest that unlike in *Arabidopsis*, where the OST1-like kinases are thought to activate through autophosphorylation, the PpOST1 kinases required an ABA-dependent upstream activator. The discovery of *PpANR* as a vital ABA regulator likely acting in a linear fashion with the core signalling pathway suggests it may fulfil this missing mechanism; however, this is discussed in more detail (Chapter 6). There remain a number of problems with the conclusions reached by these authors. The first is that they did not show that the actual phosphorylated target was an OST1-like protein but rather inferred this based only on its predicted size. Secondly, they used a non-specific kinase substrate (a histone) meaning that potentially any protein kinase could have been responsible. A third issue was that, even if the activity detected was of OST1-like kinases, the expression of the OST1-like kinase genes (and therefore likely OST1-like kinase protein levels) shows strong ABA/stress induction meaning that while equal protein concentrations are loaded from each sample, those from ABA treatments are likely to contain more OST1-like kinase and therefore more phosphorylation of the histone substrate mimicking the effect of a hypothesized upstream activator. The main issue lies in how the treatment of ABA alone is able to induce the phosphorylation of these kinases without functional ABI1/2 phosphatases and this appears one of the missing pieces of the ABA-dependent stress signalling pathway. While these findings are interesting (especially since the identification of *PpANR* would be consistent with the proposed model) it is important to identify both the direct targets of PpANR activity and any potential MAP kinase cascade(s) within which it might act. From our data it appears most likely that PpANR's substrates are either the OST1-like kinases or the

downstream TFs which also require phosphorylation. If PpANR phosphorylates the OST1-like kinases, directly or indirectly, then we would expect to see neither OST1-like kinases nor its TF targets, such as the ABI5-like bZIPs, phosphorylated in the *PpanrKO* mutant following ABA treatment. If it acted between the two then we would expect to still see OST1-like phosphorylation following ABA treatment but not that of the target TFs. The elucidation of this would update our understanding of ABA signalling in the bryophytes and bring us closer to understanding the origins of ABA signalling and its role in the conquest of land. The finding of uncharacterized ABA-dependent signalling components offers further candidates for novel ABA-dependent stress signalling, possibly functioning as part of the so far unknown PpANR signalling pathway. The finding that a homologue of the osmosensor AtHK1 is strongly up-regulated by stress in an ABA-dependent manner hints at a possibility that it may be at the apex of stress signal transduction. These osmosensors appear to be functionally conserved between plants and fungi and have been suggested to function at the head of a MAP kinase cascade (Urao *et al*., 1999). The RNA-seq analysis presented here is likely to be an important resource in the future understanding of PpANR and ABA signalling in *Physcomitrella*.

CHAPTER 4

Chapter 4 – Phylogenetic analysis of PpANR and related MAP3 kinases reveals insights into the molecular mechanisms in the conquest of land.

4.1 Introduction

This chapter seeks to identify the evolutionary origins and relationships of PpANR. As a trimodular protein with a MAP3K-like serine/threonine kinase it shares homology with a number of diverse gene products in each of its domains individually and also in combinations in which either of the two N-terminal domains (the PAS (P) and EDR (E) domains) can be found with a C-terminal kinase (K) domain. However, although both “PK” and “EK” dimodular kinases have previously been identified in plants, no other similar tri-domain (“PEK”) homologues were found in any of the plant BLAST-searchable databases at the time of its discovery.

4.1.1 PAS containing kinases

The N-terminal PAS domain is a member of the highly versatile PAS superfamily which is found across all kingdoms and has roles in the detection of: blue-light (flavanoid cofactor), redox states (flavanoid cofactors), oxygen (haem cofactors) and small ligands, as well as in regulating protein-protein interactions - often forming heterodimers (see chapter 5). PAS domains have been best functionally characterised in bacteria where they are principally found as members of two-component regulatory systems. These are single or paired membrane-associated proteins in which the (often) modular sensory component directly or indirectly senses changes in the environment upon which conformational changes allow autophosphorylation of a histidine kinase effector domain which transduces the signal into the cell mediating a response (reviewed by Mascher *et al*., 2006 and Szurmant *et al*., 2007). Also found in bacteria are the two-component related PAS-containing histidine kinases known as bacteriophytochromes which are ancient light sensors and are ancestral to the red/far red-light sensing phytochromes in plants. Given their ancestry, phytochromes too are histidine kinase-related proteins that typically have 3 PAS domains with the N-terminal (PASa) found alongside the distantly related GAF and PHY domains (Sharrock, 2008). PASa (often termed PAS-like domain (PLD) given its divergence) has no sequence homology with the PpANR PAS domain (BLASTP E-value>1), however the more C-terminal pair show some, albeit low, homology with the PpANR PAS domain, with PASb being the most similar. Another class of plant photoreceptor that contains a PAS domain is the blue-light receptor family: the phototropins. These contain PAS subfamily variants known as light-oxygen-voltage (LOV) domains that are usually found as a tandem repeat and are frequently associated with a serine/threonine kinase domain

(Christie, 2007). These LOV domains bind flavin mononucleotide (FMN) cofactors that enable light excitation to be transduced into perception and responses via kinase activation. They do not, however share any sequence homology with the PpANR PAS domain (BLASTP E-value>1).

A third group of plant PAS-containing kinases are members of the B2 sub-family of Raf-like MAP3Ks identified in *Arabidopsis* (Ichimura *et al*., 2002) which have an N-terminal PAS domain and a C-terminal serine/threonine kinase domain giving them a “PK” structure. These are among the most similar proteins to PpANR (in both the PAS and kinase domains) found in plants and some of their functions are starting to be understood. Significantly, these functions appear to partially overlap with that revealed for PpANR. One B2 subfamily MAP3K (At4g23050) has been described as being involved in salt tolerance and ABA signalling by its effect on seed dormancy (Shitamichi *et al*., 2013) as well as being a more general growth regulator (Sasayama *et al*., 2011). More significantly, two functionally redundant members of the B2 subfamily (Raf10 and Raf11) were shown to additively and positively regulate seed dormancy thereby implicating them as positive regulators of the ABA response (Lee *et al*., 2015b). This has greater significance considering that the molecular responses and regulation in angiosperm seeds have the closest similarity to that in bryophytes; both conferring dehydration/desiccation tolerance at the cellular level. These authors also found molecular evidence for these MAP3 kinases being positive ABA regulators as over-expressing lines showed strong induction of ABA-responsive transcripts encoding seed-related proteins such as LEAs, oleosins, cruciferin A1 and cupins as well as the ABA-responsive TFs ABI3 and ABI5. This indicated these B2 MAP3Ks are involved in ABA-dependent signalling and the strong induction found for seed-related proteins indicates a likely role in ABA-mediated seed development with dormancy a key affected trait. Interestingly, these authors also identified mutants of Raf10/11 as osmotic stress insensitive indicating they are likely involved in mediating stress signalling responses (Lee *et al*., 2015b). Many plant MAP3 kinases are implicated in a range of stress responses (reviewed in Sinha *et al*., 2011; see also Colcombert and Hirt, 2008; see Chapter 1.6) where they act at the head of signal transduction pathways mediating responses to stimuli as diverse as pathogen attack and osmotic stress. However, the role of the PpANR PAS domain and whether it functions as a sensor module for ABA signalling remains unclear. Phylogenetic analysis of PAS domains was first carried out to uncover the relationship between the PAS domains in the MAP3 kinases; PpANR and the B2 subfamilies. This was then extended to identify some of the most similar PAS domains from other protein families including two-component-like histidine kinases, which includes phytochromes.

4.1.2 “EK” kinases

Another subfamily of Raf-like MAP3 kinases, the B3 sub-family (Ichimura *et al*., 2002-terms which are used but revised in this chapter), includes the *CONSTITUTIVE TRIPLE RESPONSE 1* (AtCTR1/At5G03730) and *ENHANCED DISEASE RESISTANCE 1* (AtEDR1/At1G08720) proteins, which each possess a conserved N-terminal EDR domain and a C-terminal serine/threonine kinase domain giving them an “EK” structure. Again, like the B2 kinases these are among the most homologous proteins to PpANR, with PpANR previously identified as a possible functional homologue of AtCTR1 (Yasumura *et al*., 2012); a finding that was recently confirmed (Yasumura *et al*., 2015) and one that is discussed in the context of the phylogenetic findings in this chapter. AtCTR1 and AtEDR1 are both negative regulators of ethylene signalling, however AtEDR1 was first characterised by its role in jasmonic acid-mediated pathogen responses (Frye and Innes, 1998) and proposed to mediate cross-talk between ethylene and pathogen responses (Tang *et al*., 2005). In line with the roles of many MAP3 kinases including the aforementioned, another “EK” member (At1g73660) of the B3 subfamily has also been linked to stress signalling having previously been described as a negative regulator of salt tolerance (Gao and Xiang, 2008). While PAS domains are found across the kingdoms, EDR domains are almost entirely limited to the Raf-like MAP3 kinases in *Arabidopsis* where they are found in all 6 of the B3 and all 3 of the B1 subfamily members and, with a far lower degree of homology, in a single Raf-related non-MAP3 kinase (AT1G04210) which does have a putative homologue in *Physcomitrella* (Phpat.008G084000) and which is characterised by N-terminal leucine-rich repeats. The EDR domain of AtCTR1 was found to interact with the histidine kinase (C-terminal) domain of AtETR1 through yeast two-hybrid assays (Clark *et al*., 1998) which has since been characterised in more detail (Huang *et al*., 2003). One such mutant, *ctr1-8*, is affected in a single amino acid (G→A) in its EDR domain that, while not affecting kinase activity, disrupts the interaction with AtETR1 and so prevents receptor mediated activation – which itself is required for negative inhibition of downstream targets. It was also shown that overexpressing an N-terminal (EDR domain-containing) fragment lead to a dominant negative phenotype through competitive binding of the AtETRs preventing AtCTR1 binding. Interestingly, this residue was also identified as invariant among all EDR containing MAP3Ks (those from the B3 and B1 subfamilies) suggesting the interaction mediated by this position is functionally important for all members (Huang *et al*., 2003).

4.1.3 The serine/threonine kinases

The angiosperm Raf-like MAP3 kinases, which contain either a PAS or EDR domain, also show very high sequence homology in their respective kinase domains with

PpANR. The kinase domains of the B3/B2 subfamilies are serine/threonine kinases but their roles as genuine MAP3 kinases are only recently being elucidated, with AtEDR1 found to regulate the MKK4/MKK5-MPK3/MPK6 kinase cascade (Zhao *et al*., 2014). To help understand how PpANR functions, its phylogenetic relationship within the green plant Raf-like MAP3 kinases needs to be resolved. This needs to be done on a domain-by-domain basis given its trimodular nature where each module may well have a very different evolutionary trajectory (analogous to horizontal gene transfer between organisms). These three independent lines of analysis should help us to understand not only the origins but also the fate of PpANR-like genes, which appear to have been lost in higher plants, and should shed light on possible mechanisms of action. In support of this study a comprehensive comparative analysis between the protein kinase C-like kinases (PKc) in the V3.0 moss genome assembly that showed significant similarity with *Arabidopsis* MAP kinases was carried out to infer not only where PpANR sits among this large kinase family but also to compare and contrast this family between *Physcomitrella* and angiosperms. This comparative genomic approach will enable the assignation of *Physcomitrella* genes to the 4 MAP kinase tiers (MAPKs, MAP2Ks, MAP3Ks and MAP4Ks) and their subfamilies aiding any future work in this field, as well as improving our evolutionary understanding of these vital signalling components.

4.2 Materials and methods

4.2.1 PAS, EDR and kinase trees

Sequences were collected (Table A1) and domains were defined by CD searches (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) with manual curation. Where the Onekp database (https://www.bioinfodata.org/Blast4OneKP/home) was used, EST sequences required processing to identify the likely protein sequence using the protein translation tool found on the ExPASy server (http://web.expasy.org/translate/). Sequences were aligned with Clustal OMEGA (http://www.ebi.ac.uk/Tools/msa/clustalo/ - using 10 iterations) and alignments were visualized on Jalview (http://www.jalview.org/download). For the PAS structural alignment, the secondary structure of the PpANR PAS domain (chain A) was used to enforce indels to only occur in loop regions by manual curation. The alignments were then run using both a Bayesian (MrBayes - http://mrbayes.sourceforge.net/download.php) and maximum likelihood (RAxML - http://sco.h-its.org/exelixis/web/software/raxml/index.html) approach for comparison. MrBayes required nexus format alignments which were converted from fasta format (http://sequenceconversion.bugaco.com/converter/biology/sequences/fasta_to_nexus.php). All runs in MrBayes were run with the default setting except for the following:

mixed model setting to automatically find the best amino acid model (prset aamodelpr=mixed) and gamma rates setting (lset rates=gamma). 2 parallel runs were run with 4 chains until the 'Average standard deviation of split frequencies' dropped below 0.01 (set using the stoprule=yes and stopval=0.01 mcmc options). Trees were generated with the default setting where the burnin phase (first 25% of sampled trees) was discarded to avoid non-stable trees being included in the consensus. RAxML was run using the combined rapid bootstrap and ML tree search method to produce the best tree with bootstrap values and was called using the following command line:

```
$ raxmlHPC-SEE3 –f a –m PROTGAMMAAUTO –p 418307 –x 418307 -# autoMRE –s input_alignment.fasta –n output.
```

The consensus or best trees were viewed in FigTree (http://tree.bio.ed.ac.uk/software/figtree/) where the EDR and kinase domain trees were rooted by the B1 clade (or by midpoint rooting for the PAS trees) and nodes ordered with branches labelled by their posterior probability (MrBayes) or bootstrap values (RAxML).

4.2.2 Global *Physcomitrella* MAP kinase analysis

Given the large numbers of genes involved, a high throughput method was developed to extract sequences and identify domains as sequence input for phylogenetic analyses. The analysis was based around *Arabidopsis* sequences which were selected from established trees of MAP kinases to give the trees a scaffold (Ichimura *et al*., 2002; Champion *et al*., 2004b). Sequences used in final analysis were chosen which sampled across all previously defined subfamilies. These sequences were used as input in BLASTP searches of the V3.0 genome assembly of *Physcomitrella patens* in Phytozome (http://www.phytozome.net/). Results from each *Arabidopsis* query were extracted in fasta format using the BioMart tool in Phytozome. All fasta sequences were collected in one file for each search group, where one file contained all sequences from *Arabidopsis* MAP3K queries and one from the *Arabidopsis* MAPK queries. These files were then processed first by stripping all new line characters to simplify future steps using the single line unix command:

```
$ awk '!/^>/ { printf "%s", $0; n = "\n" } /^>/ { print n $0; n = "" }END { printf "%s", n }' input.fasta > output.fasta
```

Duplicate entries were removed as well as all non-primary models using a custom Python script (Appendix 4.1). These fasta files were then used as input for a batch CD-search (http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). Outputs were downloaded as tabular files in concise form and limited to super-families only. A custom Python script (Appendix 4.2) was then used to extract the coordinates of the domains defined as 'PKc_like' which included the kinase domains for all the *P. patens*

sequences. The output of this was a tabular text file (named coordinates.txt) with each gene ID followed by the start and end of the 'PKc_like' domain. Genes in which more than one 'PKc_like' domain was found were manually curated to include both sequences in the analysis. The fasta file used as input for the batch CD-search was also used to make a blastable database using the makeblastdb tool with the following command:

```
$ makeblastdb -in input.fasta -parse_seqids -hash_index -out DB_name.
```

The blastdbcmd tool was then used in a single unix awk command line to extract the domain sequences only from each sequence using the coordinate.txt file as an input:

```
$ awk '{system("blastdbcmd -db DB_name -entry "$1" -range "$2"-"$3"");}' coordinates.txt >
output.fasta
```

All *A. thaliana* protein sequences used in the analysis were manually added to this file. Final fasta files containing all the 'PKc_like' domain protein sequences were aligned and analysed with MrBayes as previously described with the following change: mcmc Nchains=2 (rather than the default 4) to reduce the run times given the very large alignments. Trees were produced and visualized in the same way using FigTree.

4.3 Results

4.3.1 Raf-like serine/threonine kinase domain analysis

Phylogenetic analysis of the kinase domains of the Raf-like MAP3 kinases most similar to PpANR and its homologues (B1, B2 and B3 subfamilies) offers the best opportunity to understand the evolutionary origins and relationship of PpANR and its homologues. This is because B1 and B3 members lack PAS domains (so were not part of the PAS domain analyses) and B2 members lack EDR domains (so were not part of the EDR domain analyses), but all have kinase domains. Kinase domains are also highly conserved, as evidenced by high sequence homologies across vast evolutionary distances due to the constraints imposed by their importance as signalling molecules. Alignment of the analysed kinase domain amino acid sequences highlights a number of indels which are found to be specific to the recovered groups, such as an insertion at position 86-87 for the CTR1-like (but not in the EDR1-like clade) and B1 clades (Figure 4.1). B1 kinases are further shown to be divergent with two deletions at position 113-114 and at 221-223. The high conservation means phylogenetic models are less

Figure 4.1. Alignment of kinase domains from MAP3Ks used in phylogenetic analysis. Alignment was done using the clustal omega tool and coloured using the clustalx option in Jalview. The kinase domain is highly conserved with few features differentiating the subfamily sequences. Starred are the representative sequences for each subfamily which include *AtCTR1*, *AtEDR1*, *PpANR*, *AtRaf10*, AT2G31010 (B1 subfamily). Highlighted by black brackets (157-165) are the residues identified in both Raf-like and MEKK-like MAP3Ks as a highly conserved motif in the catalytic domain (Rodriguez *et al*., 2010).

Figure 4.1

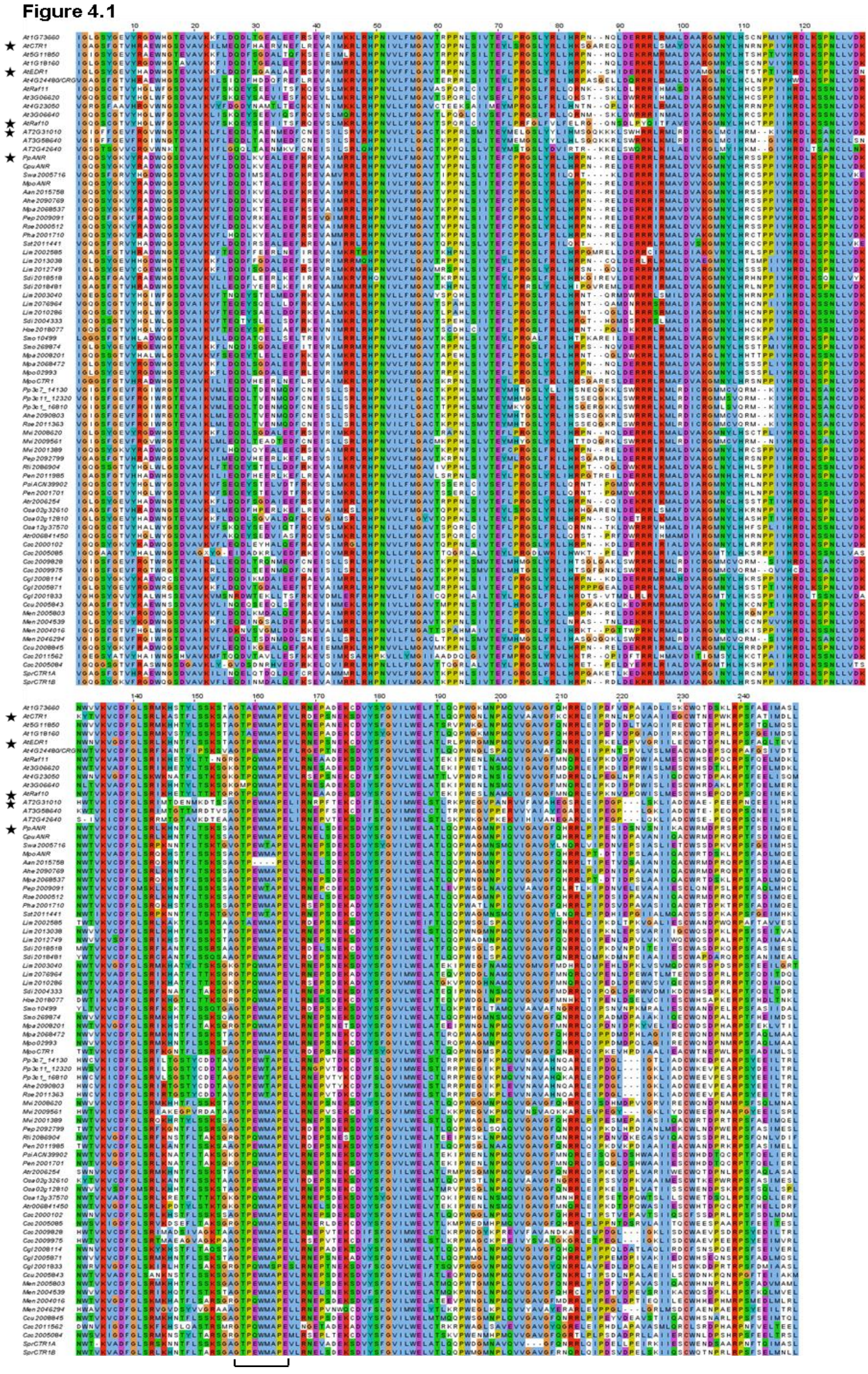

At1G73660
AtCTR1
At5G11850
At1G18160
AtEDR1
At4G24480/CRG
AtRaf11
At3G06620
At4G23050
At3G06640
AtRaf10
AT2G31010
AT3G58640
AT2G42640
PpANR
GpuANR
Swa2005716
MpoANR
Aan2015758
Ahe2090769
Mpa2068537
Pep2009091
Rse2000512
Pha2001710
Sat2011441
Lim2002585
Lim2013038
Lim2012749
Sdi2018518
Sdi2018481
Lim2003040
Lim2076964
Lim2010286
Sdi2004333
Hbe2018077
Sao10499
Sao269874
Mpa2008201
Mpa2068472
Mpo02993
MpoCTR1
Pp3c7_14130
Pp3c11_12320
Pp3c1_16810
Ahe2090803
Rse2011363
Mvi2008620
Mvi2009561
Mvi2001389
Pep2092799
Rli2086904
Pen2011985
PslACN39902
Pen2001701
Atr2006254
Osa02g32610
Osa02g12810
Osa12g37570
Atr006841450
Ceu2000102
Ceu2005085
Ceu2009828
Ceu2009975
Cgl2008114
Cgl2005871
Cgl2001833
Ceu2005843
Men2005803
Men2004539
Men2004016
Men2046294
Ceu2008845
Ceu2011562
Ceu2005084
SpCTR1A
SpCTR1B

Figure 4.2A

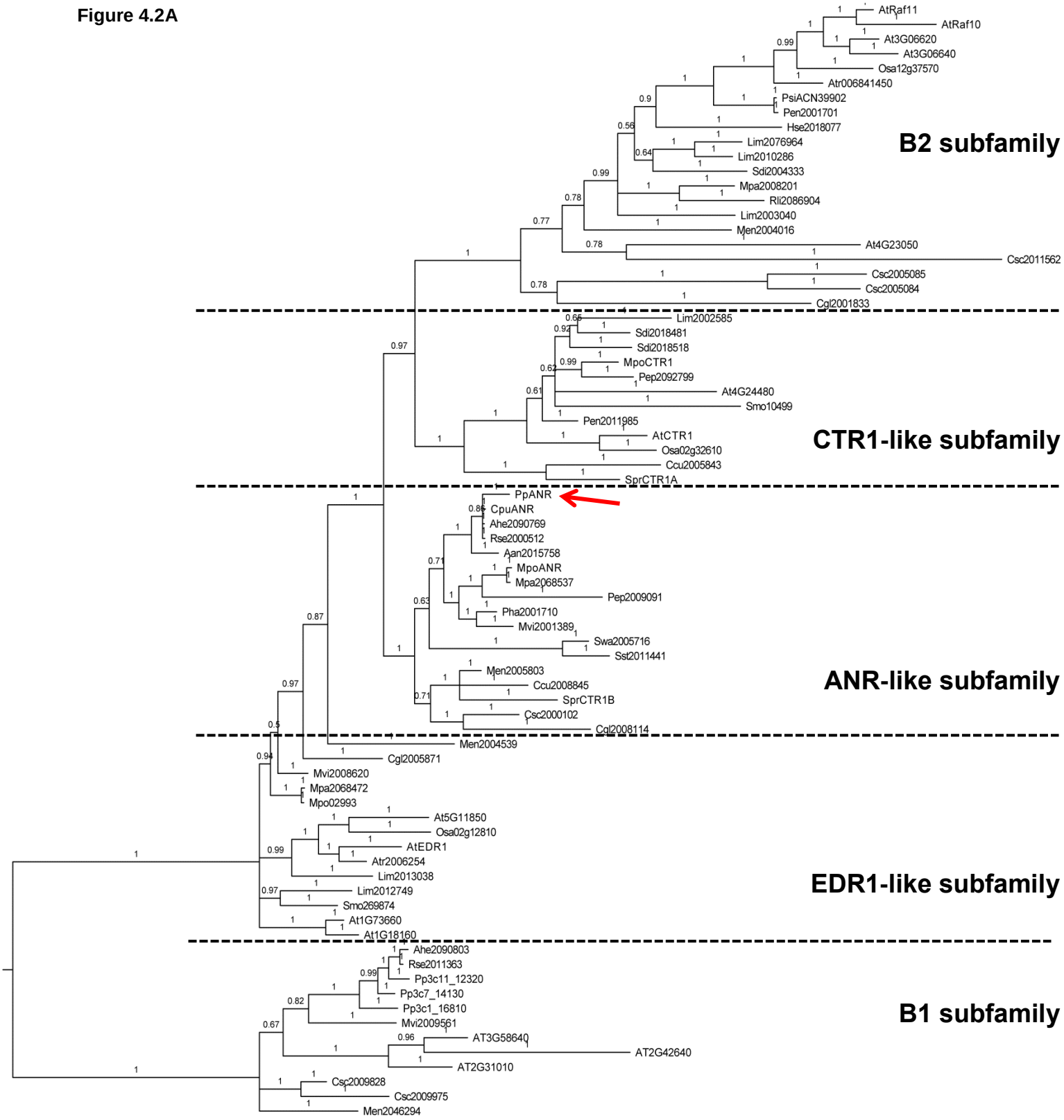

AtRaf11
AtRaf10
At3G06620
At3G06640
Osa12g37570
Atr006841450
PsiACN39902
Pen2001701
Hse2018077
Lim2076964
Lim2010286
Sdi2004333
Mpa2008201
Rli2086904
Lim2003040
Men2004016
At4G23050
Csc2011562
Csc2005085
Csc2005084
Cgl2001833
B2 subfamily
Lim2002585
Sdi2018481
Sdi2018518
MpoCTR1
Pep2092799
At4G24480
Smo10499
Pen2011985
AtCTR1
Osa02g32610
Ccu2005843
SprCTR1A
CTR1-like subfamily
PpANR
CpuANR
Ahe2090769
Rse2000512
Aan2015758
MpoANR
Mpa2068537
Pep2009091
Pha2001710
Mvi2001389
Swa2005716
Sst2011441
Men2005803
Ccu2008845
SprCTR1B
Csc2000102
Cgl2008114
ANR-like subfamily
Men2004539
Cgl2005871
Mvi2008620
Mpa2068472
Mpo02993
At5G11850
Osa02g12810
AtEDR1
Atr2006254
Lim2013038
Lim2012749
Smo269874
At1G73660
At1G18160
EDR1-like subfamily
Ahe2090803
Rse2011363
Pp3c11_12320
Pp3c7_14130
Pp3c1_16810
Mvi2009561
AT3G58640
AT2G42640
AT2G31010
Csc2009828
Csc2009975
Men2046294
B1 subfamily

Figure 4.2B

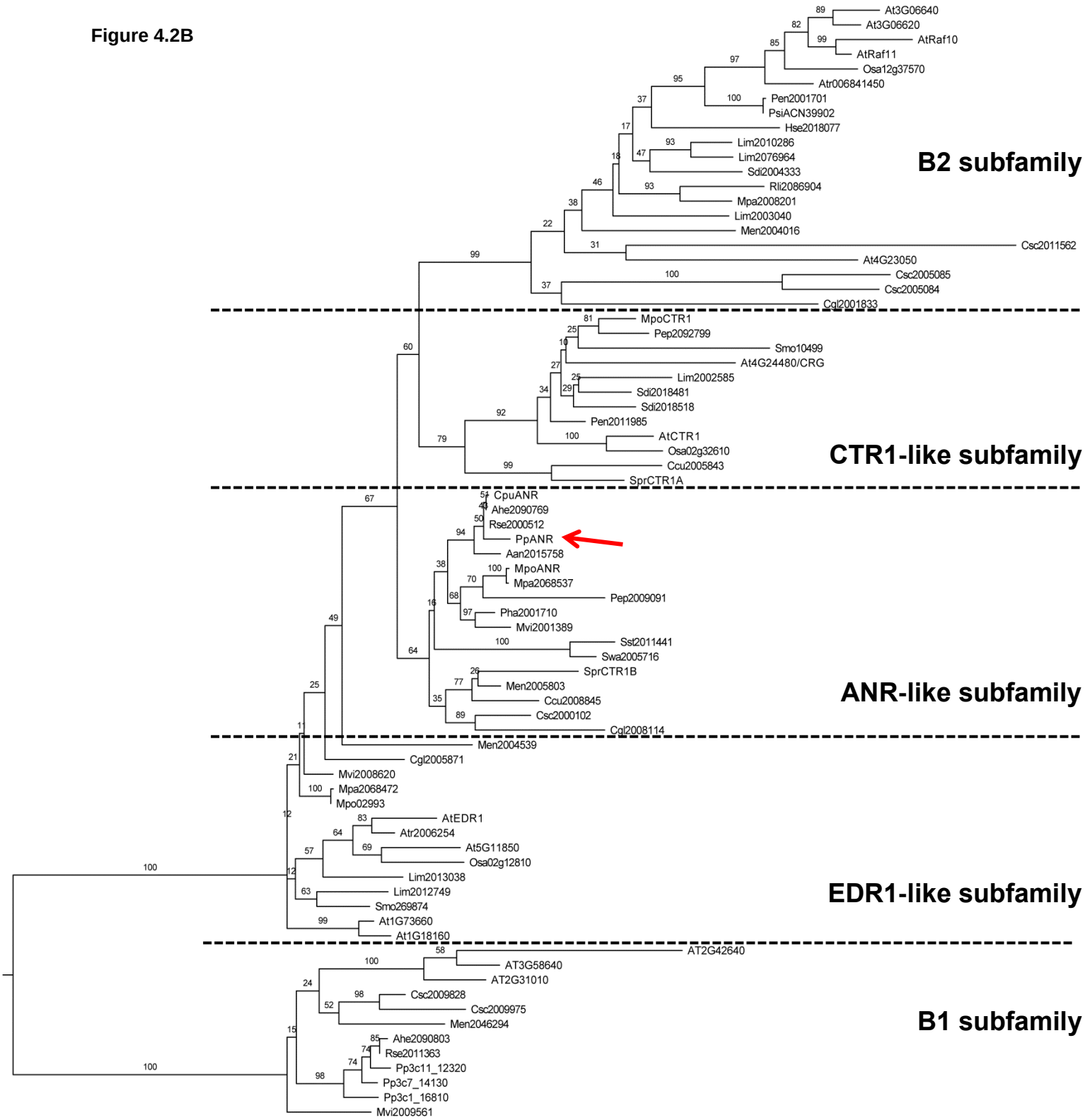


Figure 4.2. Protein trees of the kinase domains from PpANR and related MAP3 kinases. These were produced using both a Bayesian **(A)** and Maximum Likelihood (ML) **(B)** approach rooted on the B1 subfamily clade. Sequences from charophytes, bryophytes, lycophytes, ferns, gymnosperms and angiosperms were included. Both approaches produce almost identical topologies with the main difference being that the ML approach resolves the polytomies as short branches. The major subfamilies are labeled: B1 (outgroup), EDR1-like, ANR-like, CTR1-like and B2. Branches are labeled by posterior probability in the Bayesian tree **(A)** and with bootstrap values in the ML tree **(B)**. PpANR is found in a well supported moss specific sub-clade within the ANR-like clade. The EDR-like kinases assume a basal position to the ANR-like, CTR1-like and B2 subfamilies.

affected by noise caused by more rapidly diverging proteins or those with longer evolutionary separations (such as in PAS domain phylogeny). A motif thought likely to be important in kinase activity shared between all MAP3Ks is also found at positions 157-166 (Rodriguez *et al*., 2010). Within this motif at position 100 all kinases found in the B2 subfamily have a glutamine (Q) while all other members have an glutamic acid (E) suggesting this to be a subfamily specific change in a putatively functional region.

In order to help elucidate the origins of MAP3 kinases related to PpANR, charophyte sequences were also included. These were gathered from five charophyte species: The multicellular *Coleochaete scutata* and unicellular *Chaetosphaeridium globosum* (both Coleochaetophyceae), the unicellular *Mesotaenium endlicherianum* and *Cylindrocystis cushleckae* and the filamentous *Spirogyra pratensis* (all Zygnematophyceae) which represent two of the most closely related lineages to embryophytes. While no algal genomes are currently as complete or as well annotated as *Arabidopsis* or *Physcomitrella* for example, although *Klebsormidium flaccidum*, *Chara braunii* and *Chlamydomonas reinhardtii* are among the most advanced, the EST database from Onekp was mined for all putatively useable Raf-like sequences (with similarity to any of the three domains of interest) either as complete or partial transcripts from each of these species and as such the full context of sequences used is unknown.

Both Bayesian and ML approaches revealed very similar topologies by employing (Figure 4.2) different, but related, amino acid matrix replacement models (MrBayes=Jones; RAxML=LG) with the only differences being that RAxML resolved polytomies into short branches (Figure 4.2B). When the same clades are reported, the bootstrap (BS as a percentage) and posterior probabilities (PP as a proportion of 1) are presented in the form (BS/PP). The rooting of the trees on the B1 subfamily, previously found to be sister to the B2 and B3 subfamilies (Ichimura *et al*., 2002), received maximum support (100/1). The three “EK” proteins found in *Physcomitrella* are recovered in the B1 subfamily meaning that moss has no canonical ‘EK’ members truly homologous with AtCTR1 or AtEDR1. Both approaches also agree that the previously defined B3 subfamily is not in fact a subfamily but instead is split into the newly defined CTR1-like and EDR1-like subfamilies; a finding supported by other domain analyses. The grouping of ANR-like kinases with the B2 and CTR1-like clades also receives moderately good support (49/0.87) confirming that these kinases are those most closely related to PpANR. Within this group, the relationship of the B2 and CTR1-like clades as sister groups receives good support (60/0.97) suggesting that they were both derived from an ANR-like ancestor. Importantly, the resolving of an ANR-like clade (64/1) is the first clear evidence that ANR-like kinases form a distinct subfamily within

the ‘B-group’ MAP3 kinases; most closely related to the CTR1-like and B2 subfamilies. Algal sequences are resolved into all 5 subfamilies which demonstrate that these subfamilies were established in the charophytes and therefore in the divergence of the ancestral land plants. In all the subfamilies, except the EDR1-like subfamily, the majority of algal sequences occupy the earliest branch indicative of their basal position among the plants analysed; however the internal structures of each of the clades does not follow this pattern with the land plant lineages. Another important finding of this analysis is that EDR1-like kinases, while not recovered as a clade and so therefore a term used with more caution, are ancestral to the ANR-like, CTR1-like and B2 subfamilies.

4.3.2 EDR domain analysis

The EDR domain found in PpANR links it most closely with the B1 and B3 subfamilies of MAP3 kinases, to which a clear relationship is found in the kinase domain analysis. These represent the only groups of proteins containing EDR domains, with the exception of the putative orthologues AT1G04210 and Phpat.008G084000 which have low sequence homology with MAP3Ks and have N-terminal leucine rich repeats suggesting they are divergent.

The alignment of the plant EDR domains (Figure 4.3) shows that these Raf-like kinases share a high degree of homology at the amino acid level, with several sites globally conserved. The *ctr1-8* mutant affecting a residue previously shown to be highly conserved is indeed found to be invariant in all EDR domains analysed, including those from algae, suggesting a prominent role in EDR domain function as evidenced by the disruption of AtCTR1-AtETR1 binding in *ctr1-8* (Huang *et al*., 2003). A number of indels appear to correlate with the grouping found in the phylogenetic analyses such as an insertion in CTR1-like and ANR-like sequences at positions 36-42 in the alignment. This is missing in both the B1 and EDR1-like subfamilies supporting the finding that not only are EDR1-like and CTR1-like subfamilies distinct but that the B1 and EDR1-like subfamilies are basal among the analysed subfamilies.

Figure 4.3. Alignment of EDR domains from MAP3Ks and other EDR-containing kinases used in phylogenetic analysis. Alignment was done using the clustal omega tool and coloured using the clustalx option in Jalview. The non-MAP3K EDR domains (the bottom 5 taxa in alignment) contain a large insertion from position 160-229 (except Men2008640) and are the most divergent of the EDR sequences. The B1 subfamily EDR sequences have a distinctive insertion between 294-303. Both ANR-like and CTR1-like EDR sequences also have a shared insertion between 36-47 which is not shared with EDR1-like sequences. Starred are the representative sequences for each subfamily which include *AtCTR1*, *AtEDR1*, *PpANR*, *AT2G31010* (B1 subfamily) and *AT1G04210* (non MAP3K EDR-containing kinase). Highlighted with a red arrow is the invariant glycine (G) which was shown to be vital for CTR1::ETR1 binding in *Arabidopsis*.

Figure 4.3.

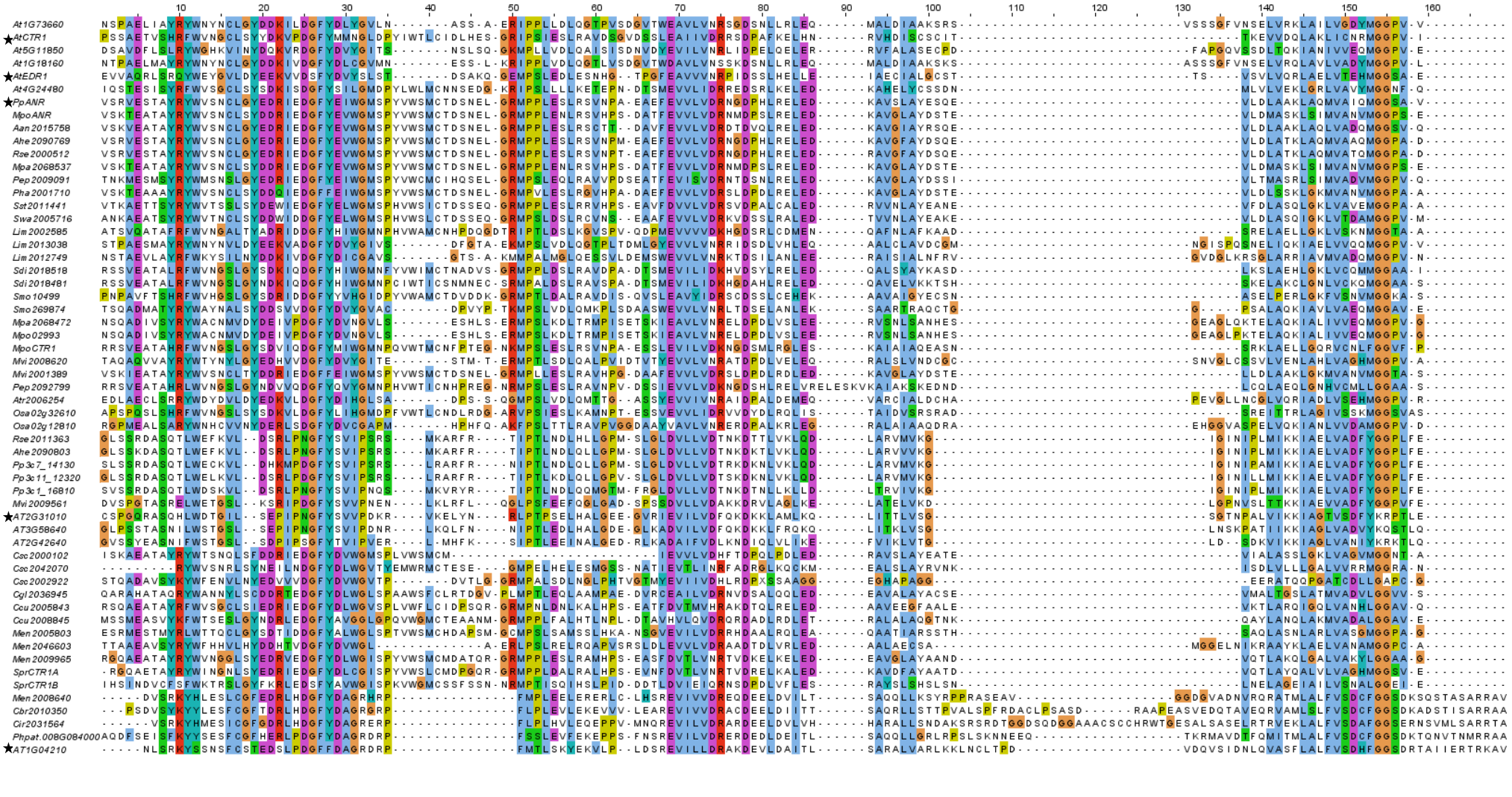

10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160
At1G73660
AtCTR1
At5G11850
At1G18160
AtEDR1
At4G24480
PpANR
MpoANR
Aan2015758
Ahe2090769
Rse2000512
Mpa2068537
Pep2009091
Pha2001710
Sst2011441
Swa2005716
Lim2002585
Lim2013038
Lim2012749
Sdi2018518
Sdi2018481
Smo10499
Smo269874
Mpa2068472
Mpo02993
MpoCTR1
Mvi2008620
Mvi2001389
Pep2092799
Atr2006254
Osa02g32610
Osa02g12810
Rse2011363
Ahe2090803
Pp3c7_14130
Pp3c11_12320
Pp3c1_16810
Mvi2009561
AT2G31010
AT3G58640
AT2G42640
Csc2000102
Csc2042070
Csc2002922
Cgl2036945
Ccu2005843
Ccu2008845
Men2005803
Men2046603
Men2009965
SprCTR1A
SprCTR1B
Men2008640
Cbr2010350
Cir2031564
Phpat.008G084000
AT1G04210

Figure 4.3 continued.

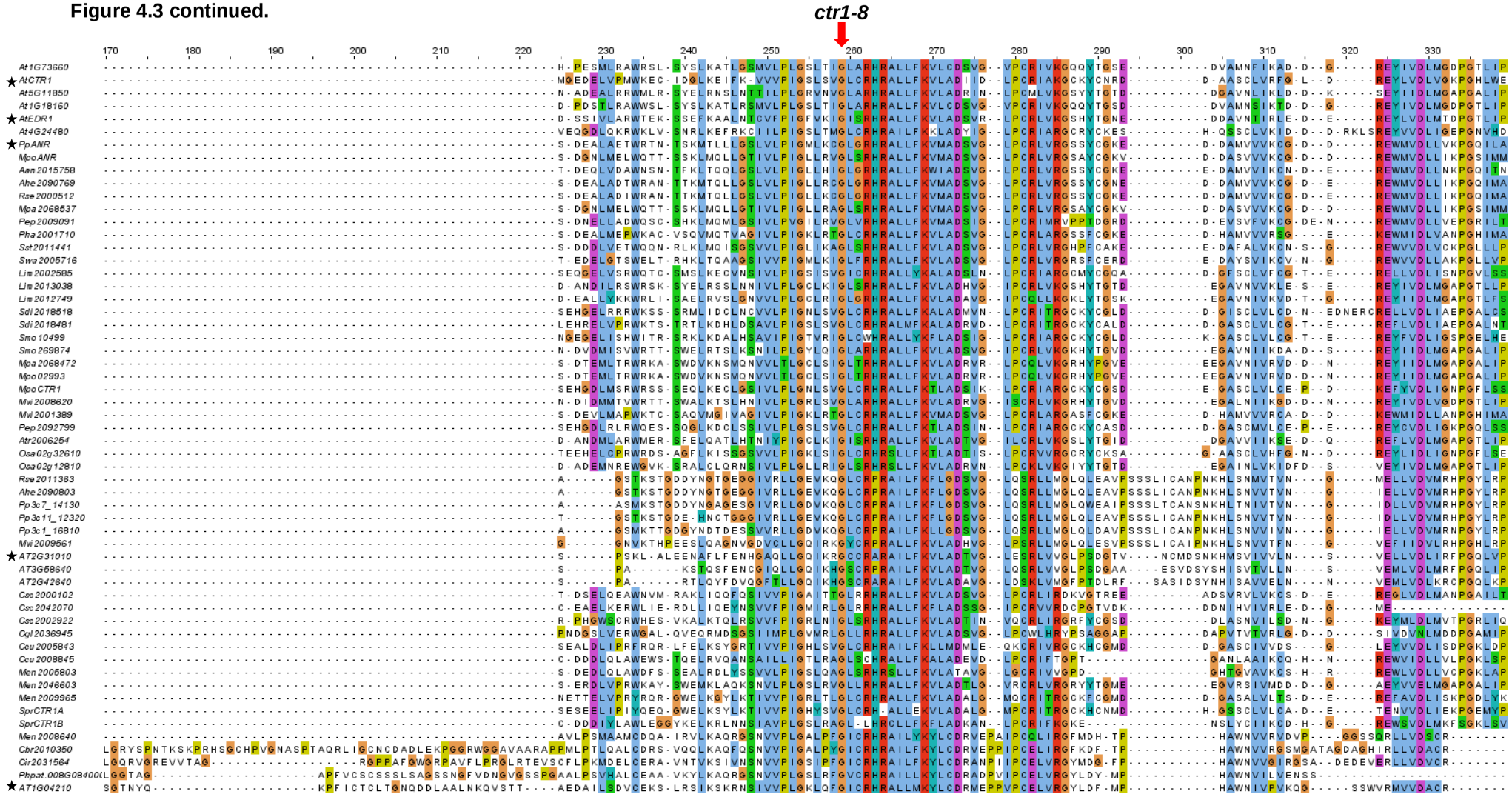


Figure 4.4A

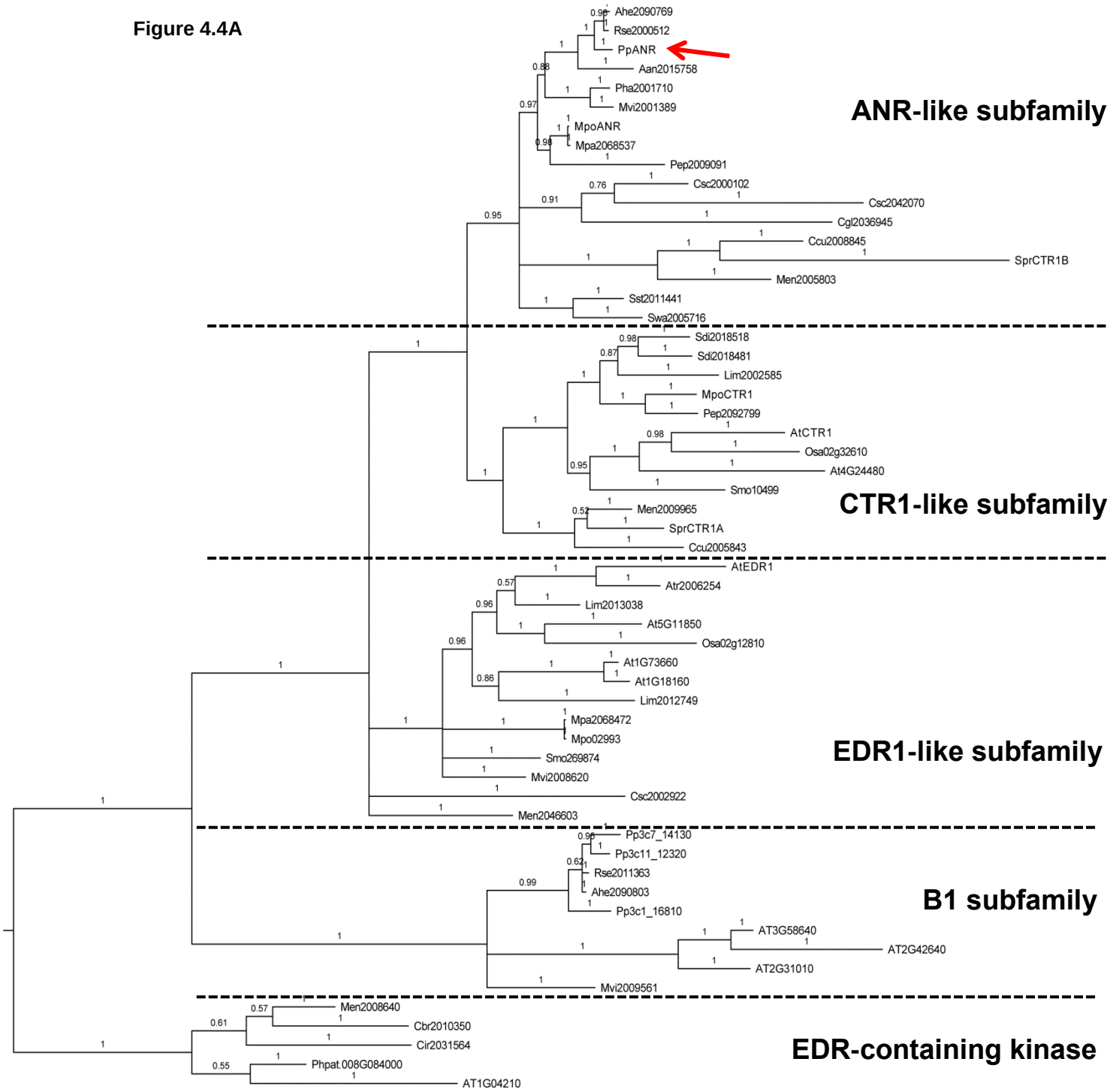

Ahe2090769
Rse2000512
PpANR
Aan2015758
Pha2001710
Mvi2001389
MpoANR
Mpa2068537
Pep2009091
Csc2000102
Csc2042070
Cgl2036945
Ccu2008845
SprCTR1B
Men2005803
Sst2011441
Swa2005716
ANR-like subfamily
Sdi2018518
Sdi2018481
Lim2002585
MpoCTR1
Pep2092799
AtCTR1
Osa02g32610
At4G24480
Smo10499
Men2009965
SprCTR1A
Ccu2005843
CTR1-like subfamily
AtEDR1
Atr2006254
Lim2013038
At5G11850
Osa02g12810
At1G73660
At1G18160
Lim2012749
Mpa2068472
Mpo02993
Smo269874
Mvi2008620
Csc2002922
Men2046603
EDR1-like subfamily
Pp3c7_14130
Pp3c11_12320
Rse2011363
Ahe2090803
Pp3c1_16810
AT3G58640
AT2G42640
AT2G31010
Mvi2009561
B1 subfamily
Men2008640
Cbr2010350
Cir2031564
Phpat.008G084000
AT1G04210
EDR-containing kinase

Figure 4.4B

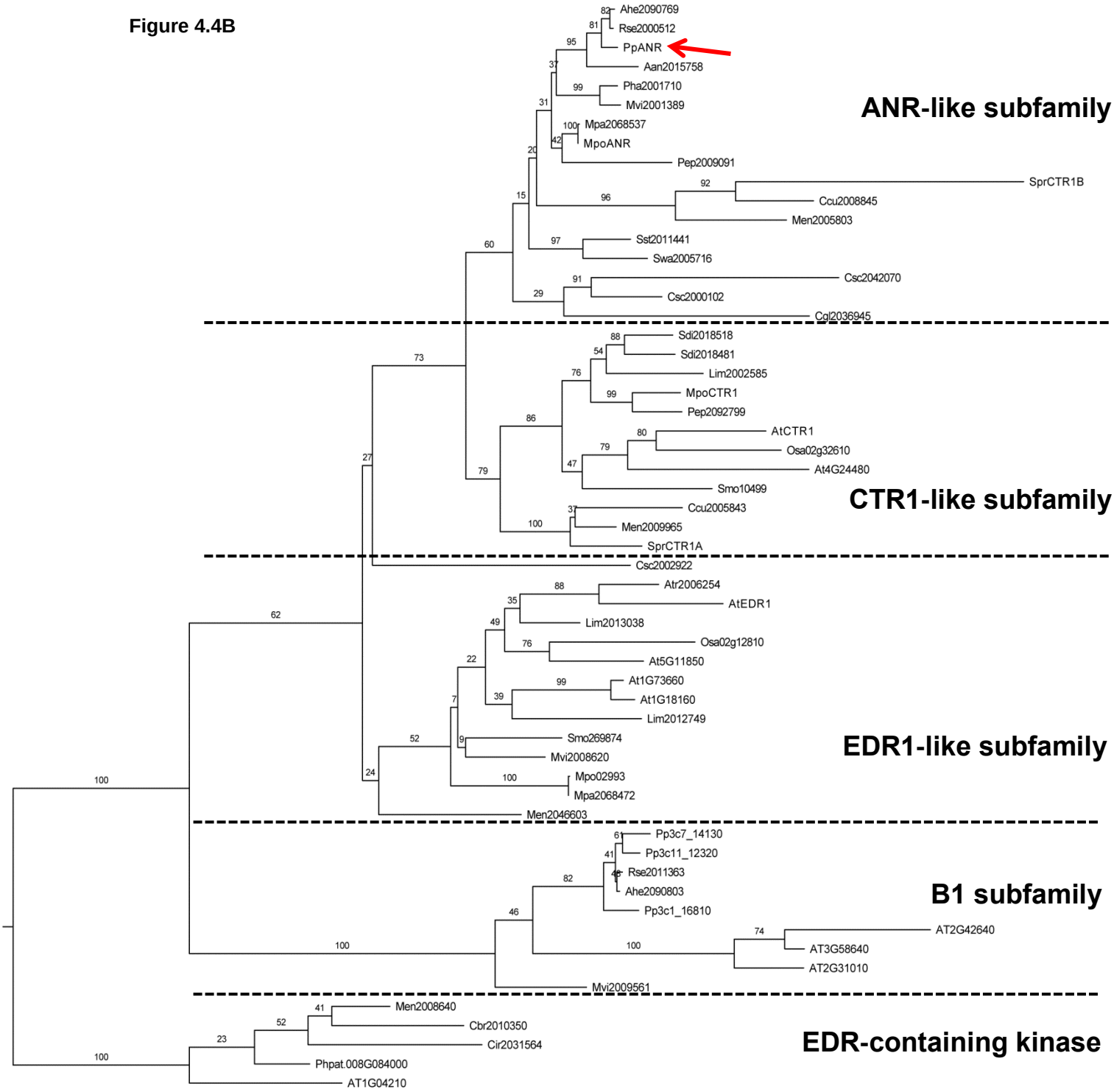


Figure 4.4. **Protein trees of the EDR domains from PpANR and other EDR-containing kinase including related MAP3 kinases.** These were produced using both a Bayesian **(A)** and Maximum Likelihood (ML) **(B)** approach rooted on a divergent EDR-containing kinase subfamily. Sequences from charophytes, bryophytes, lycophytes, ferns, gymnosperms and angiosperms were included. Both approaches produce almost identical topologies with the main difference being that the ML approach resolves the polytomies as short branches. The major clades are labeled: EDR-containing kinases (outgroup), B1, EDR1-like, ANR-like, CTR1-like and B2. Branches are labeled by posterior probability in the Bayesian tree **(A)** and with bootstrap values in the ML tree **(B)**. PpANR is found in a well supported moss specific sub-clade within the ANR-like clade. The EDR-like kinases assume a basal position. See table A1 for taxa details.

In general, the EDR domain trees largely support the findings from the kinase domain analyses (Figure 4.4). Again, the three “EK” proteins found in *Physcomitrella* are recovered in the very well supported B1 clade (100/1) alongside all three *Arabidopsis* members and other bryophyte sequences. The subdivisions of the B-group subfamilies are again recovered with good support as EDR1-like (52/1 – for land plants only), ANR-like (60/0.95) and CTR1-like (79/1) groups. What is apparent from both the kinase and EDR domain analyses is that many of the basal plant (especially the charophyte) EDR1-like sequences do not resolve into a well defined clade, unlike for all the other subfamilies, suggesting greater divergence in this group. Divergence of duplicated genes is known to greatly affect the rates of sequence evolution as a result of relaxed selection on one of the proteins (*e.g.* Conat and Wagner, 2002) and further analysis of the green algal sequences may reveal whether this has been a feature of the EDR1-like kinases. The recovery of the CTR1-like and ANR-like subfamilies as sister groups reiterates that the EDR1-like and CTR1-like subfamilies are distinct but also shows that without the inclusion of the B2 subfamily sequences, the recovery of the ANR-like subfamily as ancestral to both CTR1-like and B2 subfamilies cannot be identified. Again, with the exception of the EDR1-like subfamily, charophyte sequences are recovered within the subfamily clades in support of the kinase domain analysis.

There are, however, some important unique features of the EDR domain analyses. This stems from the inclusion of two putative orthologues from *Physcomitrella* and *Arabidopsis* which are the only ‘EDR-containing proteins’ outside the MAP3 kinase family; but are however annotated as Raf-like. Both their kinase and EDR domains show low homology with the respective domains in the B-group subfamilies analysed however they form a very well supported group with a number of algal kinases (100/1) which together have been used as the outgroup to root the trees. While the kinase domain analysis recovered algal sequences in the B1 subfamily, the EDR domain analysis has not, which necessitates further analysis. This may reflect the relatively fragmentary nature of algal sequence resources available. The B1 algal sequences came from scaffolds containing a kinase domain only and the improvement of assemblies and gene models will help clarify this. The presence of the EDR domain in a divergent group of proteins also necessitates further analysis to uncover the origin of this domain.

4.3.3 PAS domain

The PAS domain of PpANR is the most enigmatic of the domains given the impressive versatility of PAS functions (see Henry and Crosson, 2010 for review). This versatility is based on the highly conserved ‘PAS-fold’ structure, which is adopted by the PpANR PAS domain (see Chapter 5), but this same conformation can arise despite low

homology at the amino acid sequence level, making sequence-based phylogenetic analyses challenging across the evolutionary gaps. A thorough search of the *Physcomitrella* proteome to find PAS domain-containing proteins identifies 8 putative phytochromes (with one incomplete gene model) and 9 phototropins of which two (Phpat.015G037400 and Phpat.009G023100) lack the serine/threonine kinase domains in current models, which makes PpANR the only PAS-containing protein that is not a light receptor in *Physcomitrella*. Wider BLASTP searches of plant databases with the PpANR PAS sequence finds the Raf-like B2 subfamily members as the most similar but the PASb and PASc domains of phytochromes (with PASb to a slightly higher degree) show some homology. On the other hand, the PAS domains of phototropins show almost no homology (despite high structural similarity – see Chapter 5). Further sequence database searches reveal that there are a number of algal PAS-containing histidine kinases (HKs) which themselves show high homology with some bacterial two-component-like HKs. The phytochromes are believed to have bacterial and indeed two-component origins showing high sequence and structural homology with the bacterial light sensors, the (cyano)bacteriophytochromes; also PAS-containing HKs which lack the PASb and PASc domains. The inclusion of these classes of PAS sequences was carried out to resolve their relationships and in order to identify the origin of the PpANR PAS domain.

The Clustal omega alignment showed that the PAS has strong structural conservation in these sequences based on the similar lengths of secondary structure motifs and indels were mostly incorporated in loop regions (regions in between β-strands/α-helices that are usually less evolutionarily constrained) (Figure 4.5). The secondary structure motifs were derived from the PpANR PAS crystal structure (Chapter 5). The alignment was analysed using both a Bayesian and ML approach to assess the effect the different approaches would have on phylogenetic inferences. The phylogenetic analysis itself supported the findings of the kinase domain tree in that the B2 and ANR-like subfamilies are distinct (Figures 4.6). As in the EDR and kinase domain analyses, these findings across all three domains, combined with the identification of a number of potentially full length algal kinases, demonstrates that the subfamilies were established in their current structures at least in the charophytes; these include the “EK” (B1, EDR1-like and CTR1-like), the “PEK” (ANR-like) and the “PK” (B2) MAP3 kinases.

Figure 4.5. **Alignment of PAS domain protein sequences used in phylogenetic analyses.** Alignment was done using the clustal omega tool and coloured using the clustalx option in Jalview. The secondary structure of the PpAPD PpANR PAS domain is shown below the alignment, showing β-strands and α-helices which highlight the regions in which the automatic sequence alignment introduces gaps within secondary structures (red arrows) but these were not deemed problematic. PAS domains have a highly conserved PAS fold structure.

Figure 4.5.

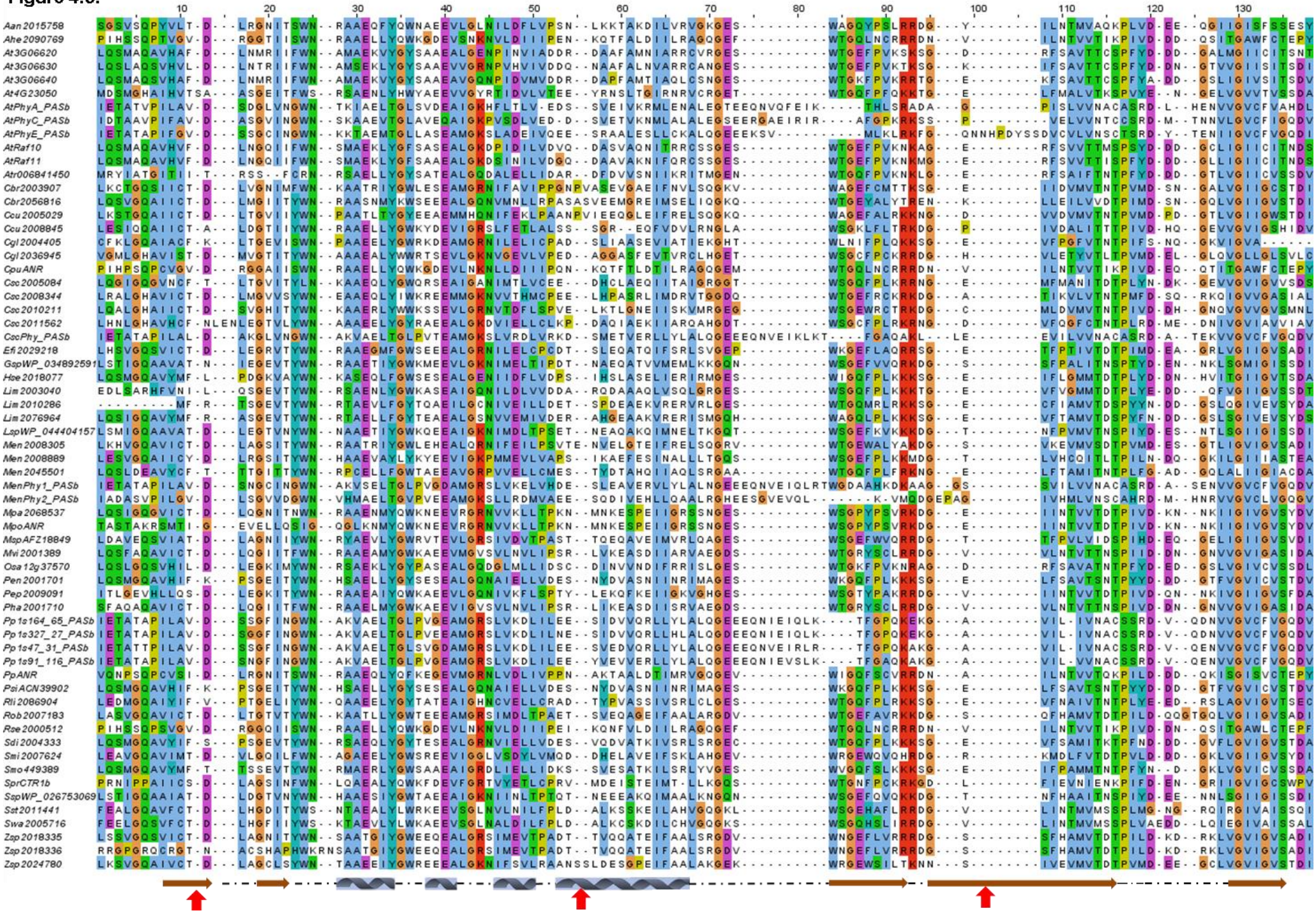
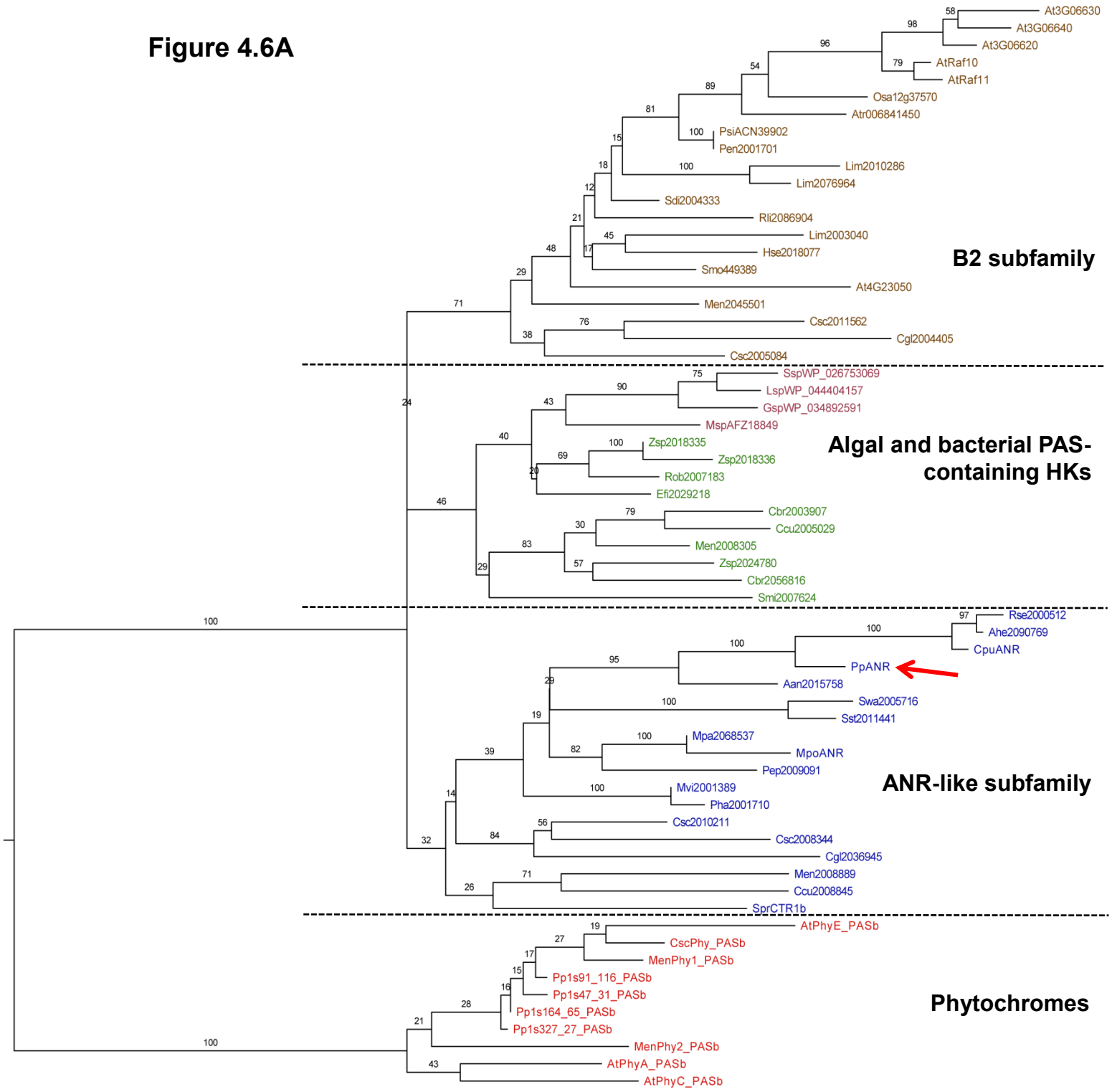

10 20 30 40 50 60 70 80 90 100 110 120 130
Aan2015758
Ahe2090769
At3G06620
At3G06630
At3G06640
At4G23050
AtPhyA_PASb
AtPhyC_PASb
AtPhyE_PASb
AtRaf10
AtRaf11
Atr006841450
Cbr2003907
Cbr2056816
Ccu2005029
Ccu2008845
Cgl2004405
Cgl2036945
CpuANR
Csc2005084
Csc2008344
Csc2010211
Csc2011562
CscPhy_PASb
Efi2029218
GspWP_034892591
Hse2018077
Lim2003040
Lim2010286
Lim2076964
LspWP_044404157
Men2008305
Men2008889
Men2045501
MenPhy1_PASb
MenPhy2_PASb
Mpa2068537
MpoANR
MspAFZ18849
Mvi2001389
Osa12g37570
Pen2001701
Pep2009091
Pha2001710
Pp1s164_65_PASb
Pp1s327_27_PASb
Pp1s47_31_PASb
Pp1s91_116_PASb
PpANR
PsiACN39902
Rli2086904
Rob2007183
Rse2000512
Sdi2004333
Smi2007624
Smo449389
SprCTR1b
SspWP_026753069
Sst2011441
Swa2005716
Zsp2018335
Zsp2018336
Zsp2024780

Figure 4.6A



ANR-like (blue), B2 (orange) and PAS-containing HKs (green=algal and burgundy=bacterial)

Figure 4.6B

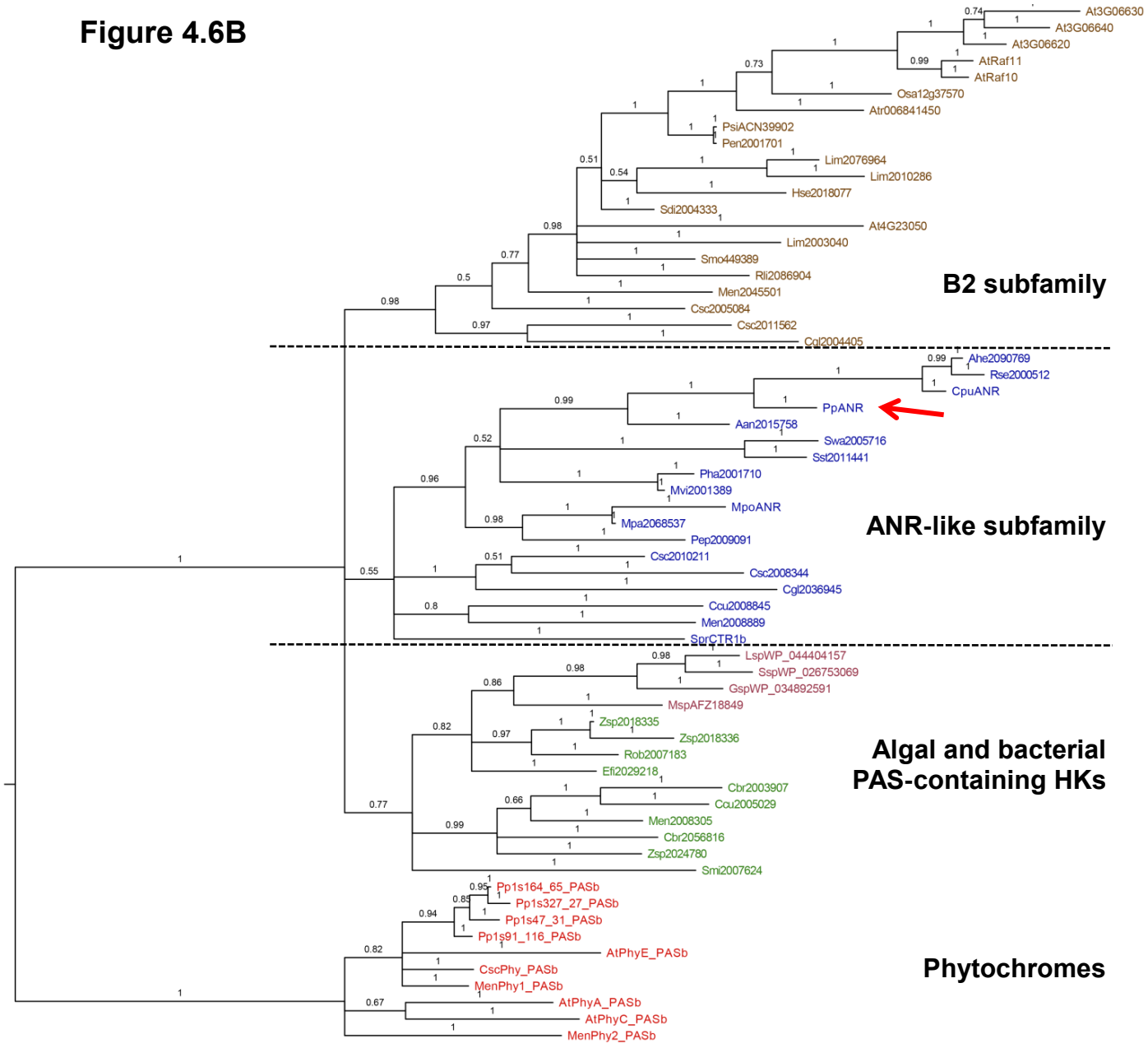


Figure 4.6. **Protein trees of the PAS domains from PpANR, related MAP3Ks and other PAS-containing histidine kinases.** Phylogenetic relationship of PAS domains from MAP3K and HK protein kinases carried out using a sequence alignment by Clustal omega. The relationships between ANR-like (blue), B2 (orange) and PAS-containing HKs (green=algal and burgundy=bacterial) clades are not fully resolved in either the Bayesian **(A)** or Maximum Likelihood **(B)** approaches but found to be polytomies. This may reflect the high similarity among all three groups resulting in insufficient data to fully resolve branch order. These are distinct from the phytochromes (red) which are found to be divergent following midpoint rooting.

These full length algal kinases include: SprCTR1a and Ccu2005843 (“EK”); Csc2005084 (“PK”) and Csc2011562 (“PK”); Ccu2008845 (“PEK”) and SprCTR1b (“PEK”) and so the full structural complement is also found before the origin of land plants. The inclusion of other PAS domain sequences that showed sequence homology with the MAP3K PAS domains demonstrated some interesting relationships. A number of algal and bacterial HK-PAS sequences were found to be closely related, not only to each other, but also to the MAP3K PAS domains with the phylogenetic analyses failing to resolve their relationship among the MAP3K PAS sequences. While all PAS domains arose in prokaryotes, this shows a clear phylogenetic between a subset of HK PAS domains from bacteria and algae and those found in the MAP3Ks. As previous analyses have indicated, the MAP3K subfamilies arose and evolved in the green algae with an “EK” kinase gaining a PAS domain to form the “PEK” kinases.

These PAS domains are found to be distinct from those in the phytochromes. Phytochromes are also structurally distinct having a billin (chromophore) binding light sensor domain complex made up of PASa-GAF-PHY domains with the PASb and PASc tandem repeat found in between this sensor domain and the C-terminus kinase domain. While phytochromes too have bacterial origins, their PASb domain was clearly distinct from the HK and MAP3K PAS domains analysed which supports the early origins of (cyano)bacteriophytochromes and the subsequent evolution of plant specific phytochromes. (Cyano)bacteriophytochromes lack the PASb and PASc domains which suggests these were plant specific additions with the canonical phytochromes originating in the early charophytes (Li *et al*., 2015). What is interesting is that homology searches show that all PAS-containing HKs (not including phytochromes) were lost in land plants, although 11 HKs are still found in *Arabidopsis* for example which include the ethylene and cytokinin receptors as well as osmosensors.

4.3.4 Comparison with original B1, B2 and B3 phylogenetic definitions

To check whether the inclusion of the ANR-like kinases was the cause of the rearrangement of the previously defined Raf-like B3 and B2 MAP3K subfamilies (Ichimura *et al*., 2002), which were also largely supported by phylogenetic analysis incorporating MAP4Ks (Champion *et al*., 2004b), a replicate subtree of that original study was produced employing both a Bayesian and ML approach. These original studies used a very limited number of angiosperm taxa and were carried out using the neighbour-joining program ClustalW which is computationally faster but less powerful than ML approaches. Both the Bayesian and ML approaches used here produced identical topologies (Figure 4.7) which included the same major nodes and relationships found in the novel kinase analyses presented here which included ANR-like sequences and many more taxa showing the findings were not an effect of the

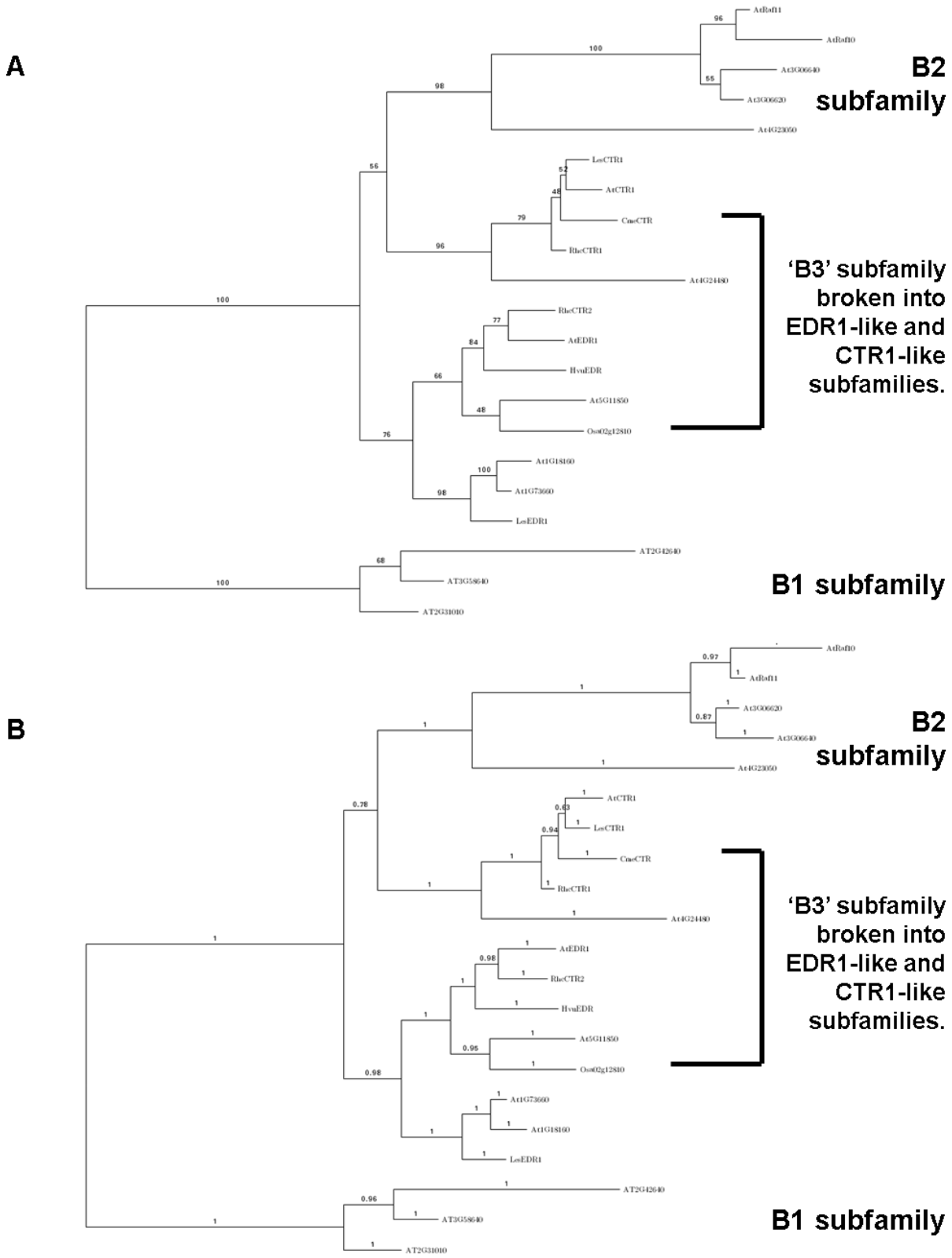


Figure 4.7. Re-analysis of original B group subfamily relationships confirms these were incorrect. Comparison of two phylogenetic methods support the reassignment of 'B3' subfamily into EDR1-like and CTR1-like subfamilies. Same proteins as used previously (Ichimura *et al*., 2002) were used taking only the kinase domain as input. Both Bayesian **(A)** and Maximum Likelihood (ML) **(B)** approaches find good support for branches with overall topology in support of large scale analyses for ANR-related MAP3 kinases. Importantly, even without the inclusion of ANR-like kinases the EDR1-like subfamily is found as sister to both the CTR1-like and B2 subfamilies showing it as a likely ancestor for these groups. Support values are shown as bootstrap values for ML tree (%) and as posterior probabilities for Bayesian tree (proportion of 1).

input data. Another large scale study primarily interested in *Gossypium raimondii* found a similar pattern to the MAPK group paper with CTR1-like and EDR1-like sub-clades being recovered as sister groups which were themselves sister to the B2 group (see Figure 1 from Yin *et al*., 2013) however this study also used a neighbor-joining method (MEGA v5) and used whole protein sequences which is not advisable given the bias that different modules can incorporate; especially with these modular kinases where presence/absence of PAS/EDR domains is common. Yin and colleagues obtained very low support for the B2 and B3 grouping (BS=36) and for the B3 clade itself (BS=21). Taken together, this suggests that the original definition of the B3 subfamily may not have been correct, highlighting the importance of using multiple statistical approaches to elucidate phylogenetic relationships and suggesting that the different and new relationships found in this study are more likely to be correct.

4.3.5 Wider analysis of MAPK-related kinases in *Physcomitrella*

In an effort to get a better understanding of the MAPK-related kinases in *Physcomitrella*, global phylogenetic analyses were conducted using a Bayesian approach. Two separate analyses were done with the primary aim of identifying all MAP kinases, MAP2 kinases, MAP3 kinases and MAP4 kinases as well as any closely related kinases in *Physcomitrella*. Sequences were mined from the V3.0 proteome using *Arabidopsis* queries on the Phytozome BLASTP server. Representative *Arabidopsis* sequences were used from each of the previously defined MAP kinase families (Ichimura *et al*., 2002; Champion *et al*., 2004b) to create a non-redundant library of *Physcomitrella* sequences. The general terminology for subfamily names comes from these papers. The initial collection contained a large number of kinases that fell into a receptor-like kinase (RLK) clade and most of these were selectively removed to leave only a ‘skeleton’ clade to indicate its relationship to the Raf-like MAP3 kinases. The tree was broadly split into two with one half containing the Raf-like MAP3Ks and allied clades while the other half contained the MEKK-like MAP3Ks which are found to be related to MAPKs, MAPKKs and MAP4Ks (Figure 4.8). The *Physcomitrella* MAPKs are not shown in this tree but some *Arabidopsis* members are shown as the MAPKs are analysed separately (Figure 4.11).

4.3.5.1 Raf-like MAP3 kinases

The RLKs were found to be closely related to the B group MAP3Ks and the B4 subfamily, with 2 members, is found to be sister to the RLKs (Figure 4.9). As observed in the more in-depth analysis of the MAP3 kinases most closely related to PpANR, *Physcomitrella* has a much reduced complement of B group MAP3K subfamilies when compared with *Arabidopsis* (6 vs 22). Again, no kinases are found in the B2 subfamily while only PpANR groups with AtEDR1 (AT1G08720) which represents the B3

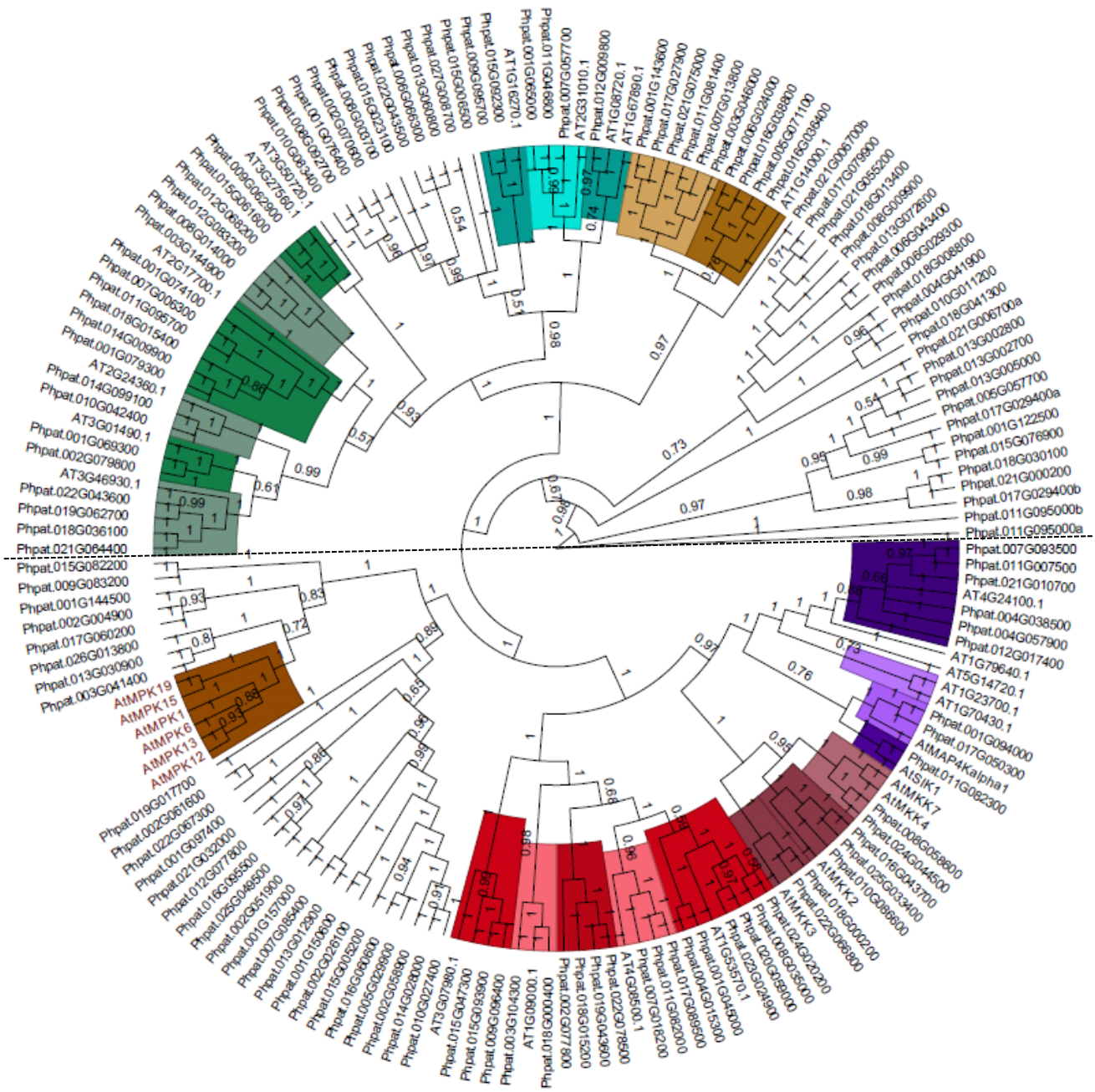


Figure 4.8. Large scale analysis of all *Physcomitrella* kinases showing significant similarity to *Arabidopsis* MAP3 kinases. Super-clade containing canonical MAP kinases (MAP4K (purple), MEKK-like (A-group MAP3Ks) (pink), MAP2K (burgundy), MAPK families is recovered (below black dotted line). Clade of *A. thaliana* only MAPKs are also shown (dark brown). B (blue) and C-group (green + brown) MAP3Ks are recovered in a Raf-like super-clade (above black dotted line) which includes the receptor-like kinases (RLKs). Alternating shades of colours represent the likely divisions of MAP kinase subfamilies as anchored by *A. thaliana* representatives. Clades without *A. thaliana* sequences indicate *P. patens* specific subfamilies likely the result of gene duplication events after the *P. patens*/*A. thaliana* divergence. Conversely, clades containing only *A. thaliana* sequences represent subfamilies that are *A. thaliana* specific. Gene IDs starting 'Pphat' are from the V3.0 *P. patens* assembly while those starting 'AT' are *A. thaliana* models found on TAIR. See sub-trees for more details.

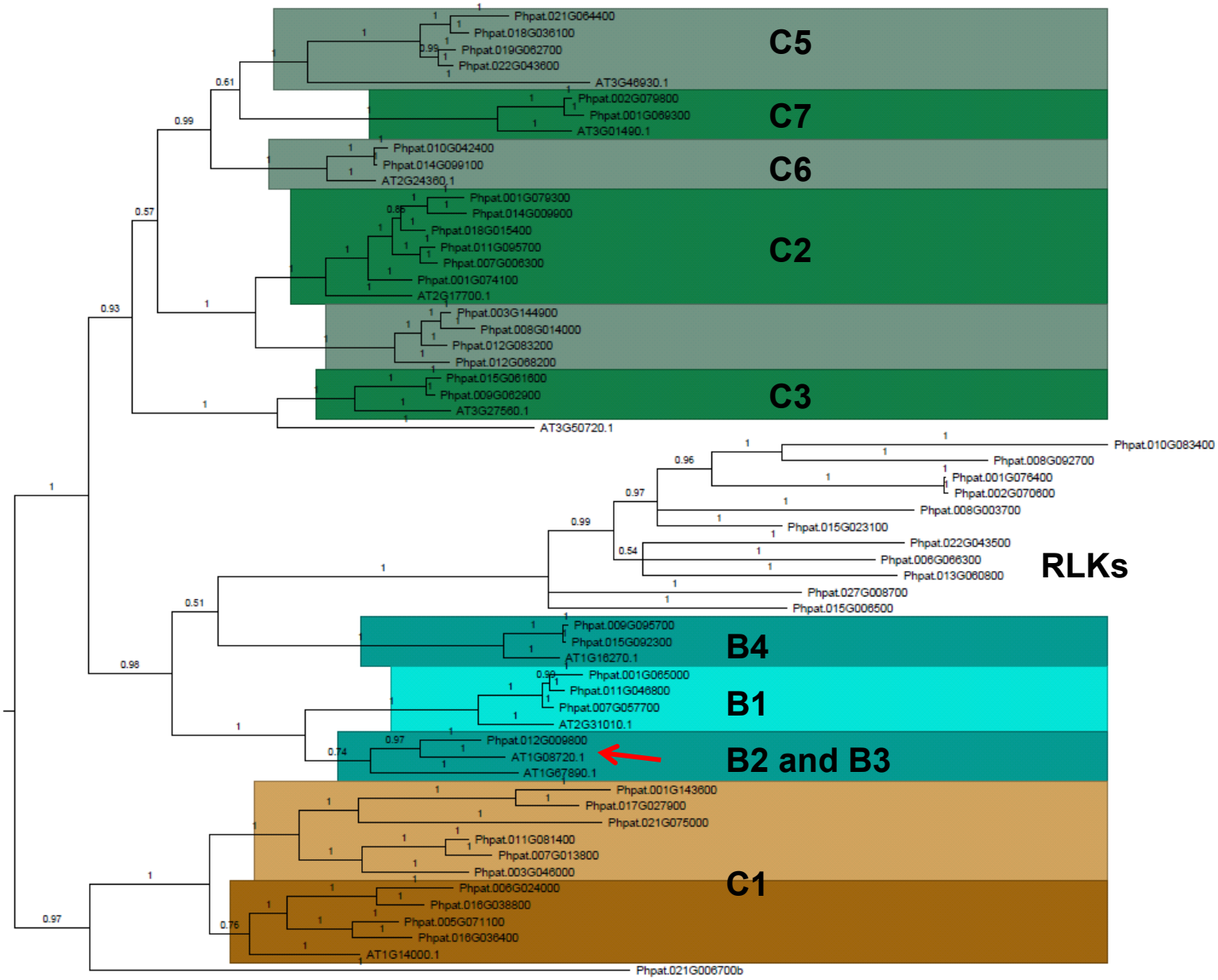


Figure 4.9. Subtree of the Raf-like MAP3Ks. Shown are the B (blue) and C-group (green + brown) MAP3 kinases from global analysis. The C1 subfamily (brown) is found as sister to all B and C group subfamilies. The other C group subfamilies all have representatives in *Physcomitrella* except C4 which appears to be an *Arabidopsis* specific subfamily. The B group subfamilies are greatly reduced in *Physcomitrella* as compared to *Arabidopsis* (6 vs 22) with no members of the EDR1-like or CTR1-like subfamilies (here referred to as the B3 subfamily as previously known) nor of the B2 subfamily. In confirmation with previous analyses, all three "EK" kinases in *Physcomitrella* fall into the B1 subfamily while the B4 subfamily has two *Physcomitrella* members. The receptor-like kinase (RLK) family is found to be related to the B group Raf-like MAP3Ks with the B4 subfamily being recovered as the sister group, although with low support. All branches are labelled with posterior probabilities as calculated by MrBayes.

subfamily (but now known to not be monophyletic but made up of the EDR1-like and CTR1-like subfamilies). Coupled with the previous identification of an ANR-like subfamily distinct from these subfamilies, this confirms that there are no canonical EDR1-like and CTR1-like homologues in *Physcomitrella*. Also in agreement with the previous study, is that the three “EK” kinases found in moss are members of the B1 subfamily.

In agreement with previous definitions, the C group Raf-like MAP3Ks are recovered as a sister group to the B group with most subfamilies being already established in the bryophytes. The exception is C4 which is a ‘higher’ plant specific subfamily likely arising from a duplication of a C3 member kinase. Interestingly, the C2 subfamily is enlarged in *Physcomitrella* with 10 members, all containing an N-terminal ACT domain which is a feature of this subfamily, compared with the 3 members found in *Arabidopsis*. The C1 subfamily is not recovered with the other C group subfamilies but this ankyrin domain-containing group is slightly expanded in *Physcomitrella* with 10 members (9 with the ankyrin domain and one kinase domain from a protein with a kinase domain repeat) compared with the 6 in *Arabidopsis*. This subfamily is found as sister to all other C and B group subfamilies. The final group in this Raf-like super-clade contains many kinases with leucine-rich repeats.

4.3.5.2 MEKK-like MAP3 kinases and related MAP kinases

On the other side of the major split between the Raf-like super-clade and the canonical MAP kinases lie the MEKK-like MAP3Ks (Figure 4.8) which is in support with previous analyses (Champion *et al*., 2004b). These have been previously been split into four subfamilies and all four contain 2 or more of the 20 *Physcomitrella* members. The relationships show much the same pattern as observed in previous studies (Ichimura *et al*., 2002 and Champion *et al*., 2004b) with the exception of an extension in the A3 subfamily which has a sister clade of 5 MEKK-like MAP3Ks. Contrary to previous studies, these MEKK-like MAP3Ks are not found as sister to the MAP4Ks, but rather are sister to both the MAP4K and MAP2K families. There are 7 identified *Physcomitrella* MAP2Ks; a number comparable with the 10 MAP2Ks found in *Arabidopsis*. This clade is found as a sister group to the MAP4K clade to which 9 *Physcomitrella* kinases belong, and this is again comparable with the 10 found in *Arabidopsis*. This large grouping of MAP2Ks, MAP4Ks and the MEKK-like MAP3Ks is found as sister to a clade within which the MAPKs lie (Figure 4.10). The clade containing the MAPKs also contains a number of non-MAP kinases annotated as being involved in a wide range of cellular and developmental regulatory processes. The relationship of the MAPKs and closely related kinases is explored in detail in the MAPK analysis.

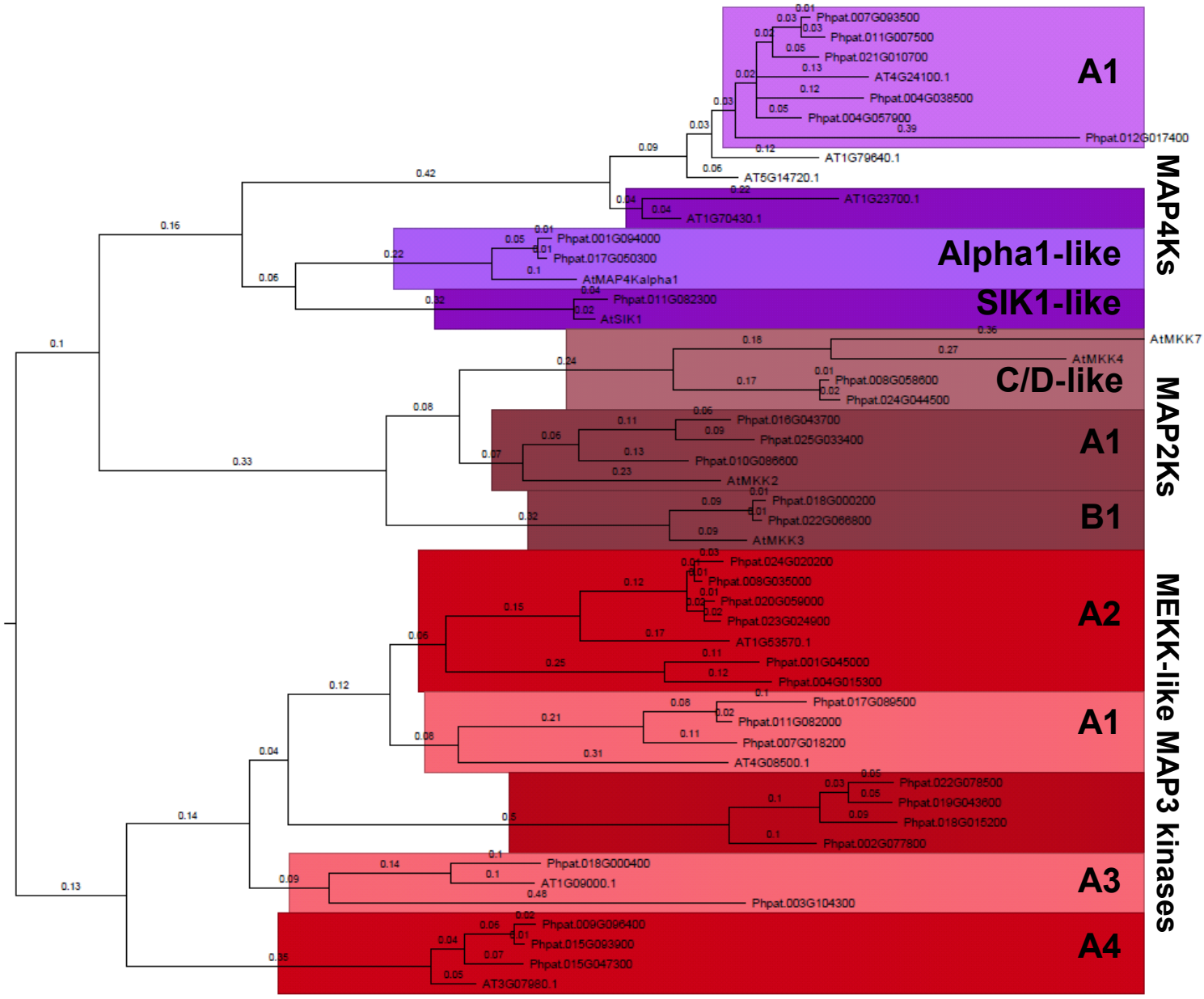


Figure 4.10. Subtree of the canonical MAP kinases. These include the MAP3 (MEKK-like), MAP2 and MAP4 kinases from the global analysis. The MEKK-like A group subfamilies are shown (A1-4) with a possible extra subfamily identified in *Physcomitrella* making up 18 in total. 7 *Physcomitrella* MAP2Ks are recovered with members from the A1 and B1 subfamilies identified. Two MAP2Ks are assigned as C/D-like given their relationship to both C and D group *Arabidopsis* MAP2Ks. 9 MAP4Ks are recovered and three new subfamilies are assigned: SIK1-like, alpha1-like and A1. All branches are labelled with posterior probabilities as calculated by MrBayes.

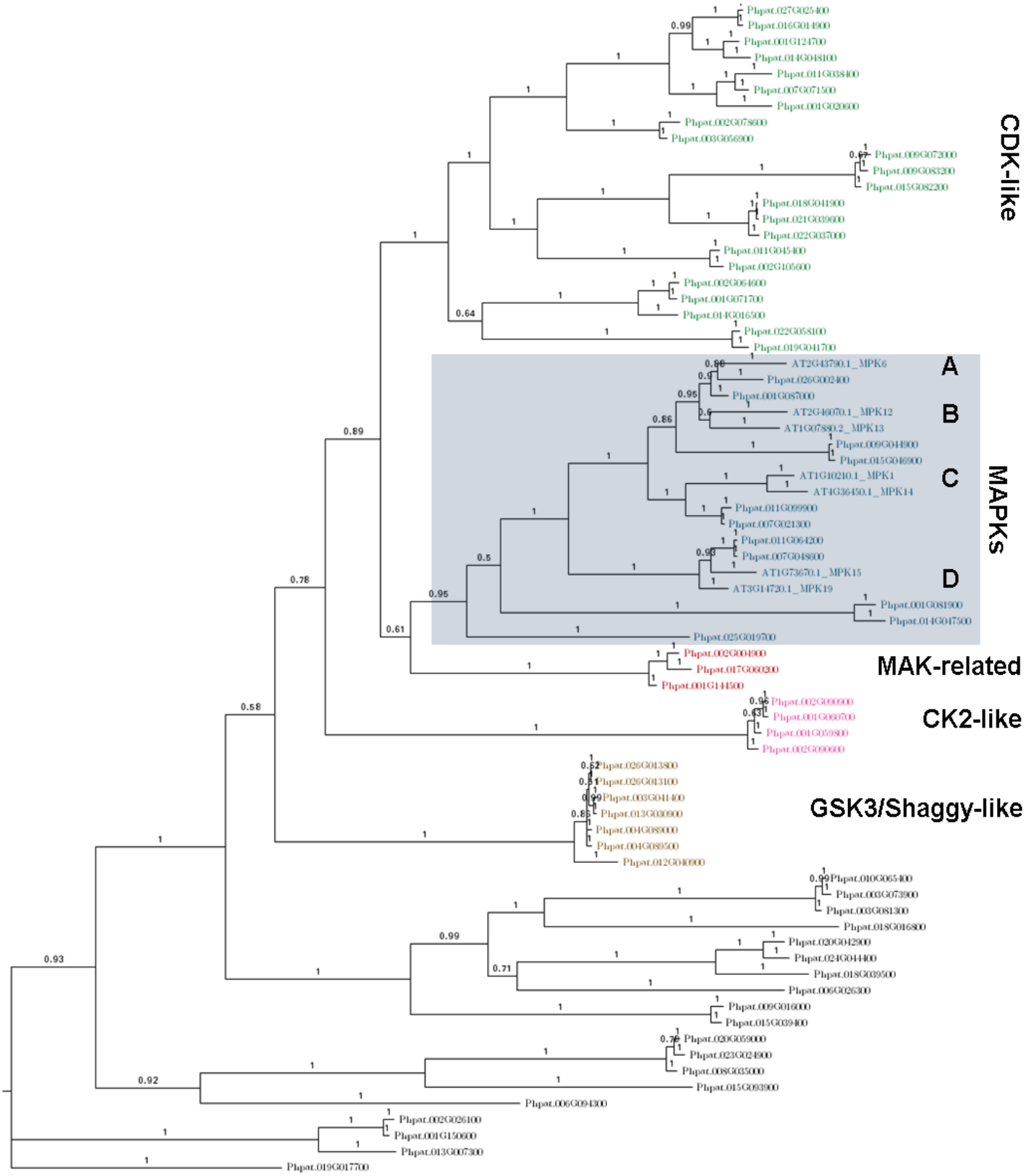


Figure 4.11. Tree showing the MAPKs and related kinases. *Physcomitrella* contains 11 MAPKs, about half that found in *Arabidopsis* (blue box). MAPKs from each of the four groups identified in *Arabidopsis* are shown (A, B, C, and D). A number of MAK-related kinases are most closely related to the MAPK family (red). A large clade of cyclin-dependent kinases (green) is found as sister to the MAK-related and MAPK kinases. Sister to these are a number of casein kinase2-like (CK2-like) kinases (pink). A clade of highly homologous GSK3/Shaggy-like kinases is also shown (brown).

4.3.5.3 MAP kinases

As seen in the MAP3K analysis, all MAP kinases can be broadly split into two lineages with the canonical MAP kinases on one side and the Raf-like kinases on the other. The MEKK-like MAP3Ks are known to be related to the MAPKs, the MAP2Ks and the MAP4Ks (Champion *et al*., 2004b) with all families sharing a common ancestor. While the likely full complement of the MAP2Ks and MAP4Ks in *Physcomitrella* was identified in the previous analysis, the MAPK complement was not and so a dedicated analysis was carried out. This analysis recovered a MAPK clade containing 11 *Physcomitrella* kinases which are significantly fewer than the 20 found in *Arabidopsis* (Figure 4.11). There are 4 defined groups (A-D) in *Arabidopsis* and this analysis suggests that in *Physcomitrella* there are 2 in group A, none in group B, 2 in group C and 2 in group D with the remaining 5 not clustering with *Arabidopsis* counterparts. The MAPK clade appears to share a common ancestor with a small subset of 3 MAK-related kinases which BLASTP searches reveal to have high homology to AtMHK which may be involved in meiosis regulation (Moran and Walker, 1993). A large clade of cyclin-dependent kinases (CDK-like) is found as a sister group to the MAPKs (and 3 MAK-related kinases). A further clade of 4 kinases is found, by BLASTP searches, to be related to casein kinase2 (CK2-like) which has roles in ABA and stress mediated tolerance in *Arabidopsis* (Wang *et al*., 2014b). Finally, a clade of 7 highly homologous paralogues is recovered which are most closely related to GSK3/Shaggy-like kinases putatively involved in osmotic stress responses in *Physcomitrella* (Richard *et al*., 2005). Many MAP kinases are known to have roles in stress signalling and the putative functional roles of these MAPK-related kinases indicate that this extends to the wider kinase superfamily.

4.4 Discussion and conclusions

4.4.1 The MAP kinase protein families in *Physcomitrella*

The wider phylogenetic analyses of related protein kinases across all four tiers of MAP kinases show that the subfamilies and groups were well established in the bryophytes. The numbers of members from each subfamily is comparable between *Physcomitrella* and *Arabidopsis* in all the MAP3K groups except for, interestingly, the B group subfamilies where the B2, EDR1-like and CTR1-like members are absent in moss; discussed in detail below. In general support of early phylogenies based on the MAP kinases, the MEKK-like MAP3Ks share a common ancestor with the MAP2Ks and MAP4Ks. *Arabidopsis* is found to show only slight increases in family sizes suggesting that the total numbers of MAP kinases may have been established early on in the land plant lineage although more in-depth analysis of other bryophyte taxa and charophytes would clarify this. There is a notable difference in the MAPKs however, with

Physcomitrella found to have almost half the number found in *Arabidopsis* which may indicate an increase in signalling complexity in higher plants with greater fine-tuning of outputs possible through greater substrate variation. In general, however, based on the number of MAP3Ks, the range of input signals that could be transduced is similar between the bryophyte *Physcomitrella* and the angiosperm *Arabidopsis*.

4.4.2 Phylogenetic robustness

The phylogenetic analysis represented here represents the most comprehensive analysis known to date on the B group Raf-like MAP3 kinases in plants. Previous studies have either only employed angiosperm samples (*e.g.* Ichimura *et al*., 2002; Champion *et al*., 2004b) or employed very few basal taxa (*e.g.* Yasumura *et al*., 2012). The inclusion of taxa from all major plant lineages including charophytes, bryophytes, lycophytes, ferns, gymnosperms and angiosperms has allowed robust conclusions to be drawn not only about the grouping of these B group subfamilies but the inference of evolutionary events within them. This thesis is focused on the moss ABA regulator PpANR and this study finds strong evidence for it being part of a novel MAP3K subfamily, unique to basal lineages. The use of two statistical approaches to infer phylogeny (Bayesian and Maximum Likelihood (ML)) has allowed more confidence in the findings. By their nature, they employ different statistical methodology and they were also found to employ different amino acid substitution matrices, however neither difference resulted in changes in topology and overall phylogenetic relationships. Various comparisons between the methods suggest that while there may be issues with the posterior probability (PP) values from Bayesian approaches, the overall performance and accuracy of the two is similar with the computationally faster Bayesian approach sometimes found to perform better (Beerli, 2006). As seen in the trees presented in this study, the PP values are almost always higher than the bootstrap (BS) values found by ML and occasionally greatly so; however they are generally correlated. This is a known issue that means the greater conservatism of ML approaches result in fewer false positive results (see Douady *et al*., 2003). On the other hand, PP values have been found to be more likely to support correct relationships than BS (*e.g.* Alfaro *et al*., 2003). This means that running both approaches in parallel on the same data input means we can be more confident of the ‘high confidence’ PP values when both approaches agree on a topology while at the same time remaining cautious of false positives from particularly low BS values.

4.4.3 PpANR’s place among the Raf-like MAP3 kinases

Homology searches with the “PEK” PpANR initially identified the Raf-like MAP3K “EK” ethylene response regulator AtCTR1 as the most similar protein in established plant databases. These searches also suggested the kinase domain itself was likely a

MAP3K-like serine/threonine kinase. Previous phylogenetic analyses found there to be 4 subfamilies of B group Raf-like MAP3 kinases with the “EK” kinases AtCTR1 and ATEDR1 belonging to the B3 subfamily while “PK” proteins made up the closely related B2 subfamily (Ichimura *et al*., 2002). The B1 subfamily, which also contains “EK” kinases, was included as an outgroup given previous findings that it was sister to the B3 and B2 subfamilies (Ichimura *et al*., 2002); a finding which was confirmed in this study. The B4 subfamily had previously been found to be sister to the B1, B2 and B3 subfamilies and was not included in the B group analyses, however the global analysis of *Physcomitrella* MAP kinases confirmed this relationship among the B-group MAP3Ks and also revealed the B-group to be sister to the RLKs. By extending the breadth of taxa and including the ANR-like kinases in the analyses of the B1, B2 and B3 kinases, these established relationships were found to be likely incorrect. Analysis of the kinase domain protein sequence revealed there to in fact be 6 B group subfamilies: the B1, B2, EDR1-like, CTR1-like and ANR-like (with the B4 subfamily presumed genuine and monophyletic). Using both Bayesian and ML approaches recovered well-supported clade nodes for 4 of the 5 defined subfamilies with only the EDR1-like subfamily not being recovered as monophyletic. The main finding, in reference to PpANR, was that it was part of the novel ANR-like subfamily unique to basal plant lineages including homologues form the charophytes, bryophytes and lycophytes (specifically *Selaginella*). The recovery of charophyte kinases in all subfamilies indicates these MAP3Ks were established in the green algae and that the negative ethylene regulator CTR1 clearly did not arise from ANR (through loss of the PAS domain) as has been suggested (Yasumura *et al*., 2015). While not all subfamilies had the PAS or EDR domains, the analyses on these domains revealed the subfamily divisions to be corroborated. The analysis also reveals that *Physcomitrella* lacks canonical members of the EDR1-like, CTR1-like and B2 subfamilies; with all three identified “EK” kinases found to be part of the B1 subfamily. This makes *Physcomitrella* (and possibly all mosses) a special case and the significance of this is discussed in more detail. One of the interesting features of this analysis was that the use of the B1 subfamily as an outgroup allowed the order of evolutionary events to be inferred. The presence of the EDR1-like kinases as ancestral to the ANR-like, CTR1-like and B2 subfamilies suggests that an “EK” ancestor gave rise to the B1 and EDR1-like subfamilies through gene duplication events. An EDR1-like “EK” kinase gained a PAS domain to form the newly defined ANR-like “PEK” kinases. The CTR1-like and B2 subfamilies were then likely formed from the loss of a PAS and EDR domain respectively from a “PEK” kinase. The finding of the “PEK” kinase PpANR as a dual ethylene/ABA regulator and the CTR1-like and B2 subfamilies as ethylene and ABA regulators respectively gives functional support to this hypothesis. In this light, it appears that the EDR domain is

likely the key regulatory module in mediating ethylene responses, given its known interaction with the ethylene receptors in both CTR1 and EDR1, while the PAS domain may be the crucial regulatory module in ABA/stress responses as the EDR domain is not present in the “PK” B2 ABA regulators Raf10/11. This makes elucidating the role and mode of action of the PAS domain vital for understanding this newly discovered means of ABA signalling.

4.4.4 A histidine kinase origin for the PpANR PAS domain?

Extensive database searches and phylogenetic analysis have shed light on the possible origins of the PAS domain. Plant PAS domains have been well characterized in the light sensors; the phototropins and phytochromes where they enable signal transduction of light through their interactions with FMN and billin respectively. These PAS domains appear quite distinct to those found in the MAP3Ks with only the PASb and PASc domains from phytochromes showing low, but significant, homology. The inclusion of a number of algal and (cyano)bacterial PAS domains from HKs that cluster with the MAP3K PAS domains, and importantly not the PASb phytochrome domains, suggests the MAP3K PAS domains may well have originated from HKs in the green algae. HKs are ancient signalling components that originated in prokaryotes and were incorporated into eukaryote genomes, likely through the endosymbiosis of cyanobacteria in the formation of plant chloroplasts, where they continue to play similar, but reduced, roles in signal transduction. It appears that the HKs have decreased in importance in more advanced eukaryotes as other signalling mechanisms have evolved. In bacteria, HKs are integral components of two-component regulatory systems mediating responses to a multitude of external stimuli and are found in large gene families; for example there are around 29 HKs in *E. coli* and 63 in *Pseudomonas aruginosa* (Kim and Forst, 2001). This is in stark contrast to the 11 found in *Arabidopsis* where they function as ethylene and cytokinin receptors as well as osmosensors. The finding that the MAP3K subfamilies were established in the green algae and that these basal plants have PAS-containing HKs suggests that a domain swap may have occurred between these kinase families. Of the 11 HKs in *Arabidopsis* and the 10 in *Physcomitrella*, none have PAS domains suggesting this group of HKs were lost prior to land plants. HKs represent the largest PAS-containing gene family in bacteria at over 40% (Henry and Crosson, 2011) and therefore the possibility of the MAP3Ks PAS domains having HK origins should not be that surprising. Further phylogenetic analysis of these algal HKs as genetic resources improve will help show the extent of this family and just how prevalent the PAS domain is.

The finding that PpANR and its homologues form a newly described subfamily related to the “EK” and “PK” Raf-like MAP3Ks means we can make some inferences about its

mode of action. The primary mode of action of AtCTR1 is to inhibit EIN2 so preventing activation of the transcription factor EIN3. It has also been suggested as being involved in a more canonical MAP kinase cascade by inhibiting the MKK9-MPK3/6 module which would otherwise activate EIN3. AtCTR1 is predicted to act on another unknown MAPK pathway which is thought to result in phosphorylation of EIN3 for degradation by the 26S proteasome (see review by Hahn and Harter, 2009). AtEDR1 has also been shown to be at the head of the EDR1-MKK4/MKK5-MPK3/MPK6 kinase cascade (Zhao *et al*., 2014) suggesting that this group of MAP3 kinases has the ability to act as canonical MAP3 kinases regulating a phosphorylation cascade. PpANR also shares the GTXXWMAPE conserved motif identified in plant MEKK-like and Raf-like MAP3 kinases (Rodriguez *et al*., 2010) further supporting the possibility of a MAP3K-like signalling mechanism for PpANR. This motif is highly conserved throughout the *Physcomitrella* MAP3Ks analysed from both the Raf-like and MEKK-like groups. The finding that PpANR is functionally homologous to CTR1 in its ethylene signalling role also suggests it may be involved in similar downstream signalling pathways (Yasumura *et al*., 2015). The functional characterisation of two B2 subfamily members, Raf10 and Raf11, as ABA regulators (Lee *et al*., 2015b) is an intriguing addition to this group's functions but these authors have not yet confirmed whether these kinases act as canonical MAP3Ks. These authors did, however, show they have typical kinase activity, being able to autophosphorylate and to phosphorylate a heterologous substrate (the myelin basic protein). The phylogenetic distribution of the ANR-like, CTR1-like and B2 subfamilies among plant lineages suggests that a certain degree of functional redundancy may occur with the dual role in ethylene and ABA signalling in ANR-like kinases split into the two subsequent subfamilies. The lack of ANR homologues in the spermatophytes raises the possibility that some B2 members, such as Raf10/11, can functionally cover the ABA signalling otherwise conferred by ANR while CTR1-like kinases cover the role in ethylene.

4.4.5 PpANR dual function a special case?

By contrast, in *Physcomitrella* there are neither canonical B2 members nor CTR1-like kinases which may well have necessitated the retention of dual function in PpANR. The finding of PpANR as an ethylene regulator in *Physcomitrella* that had a lower affinity for the PpETR7 ethylene receptor than its *Arabidopsis* functional counterpart AtCTR1 did for AtETR1 is interesting. While the assay used is not an accurate measurer of affinity it does hint at a very likely scenario of a trade off in dual signalling roles. The recent discovery that ethylene signalling in the charophyte *Spirogyra pratensis* is mediated by selective binding of the ETR receptor with a canonical CTR1-like homologue (SprCTR1a in these analyses) and not with a canonical ANR-like homologue

(SprCTR1b in these analyses) (Li *et al*., 2015) supports this, suggesting the specialized kinases are better able to carry out their respective functions than the dual function ANR-like kinases. Whether the ANR-like SprCTR1b can act as an ABA regulator remains an interesting question. ABA-signalling in algae does not appear to be important, although algae exhibit dehydration tolerance responses similar to those of land plants (Holzinger *et al*., 2014). However it will be interesting to determine whether an algal ANR kinase could complement the *Physcomitrella anr* mutant. The functions of these MAP3K subfamilies in other basal plants require elucidation to resolve their roles, especially in the liverworts which appear to have members from all the subfamilies analysed. Such a dual-function hypothesis would strongly depend on the mode of action of PpANR in ABA signalling. If the role in ethylene or ABA is mediated by the EDR and PAS domains respectively then there would be less functional constraints as each regulatory module can 'specialise'. Knowing what the interacting partner with the PAS domains is will also be crucial to understanding any potential conflicts such as localisation. The interaction of AtCTR1 and AtETR1 means that this kinase is largely anchored to the endoplasmic reticulum (ER) where it remains activated and inhibits downstream signalling. Only upon ethylene binding to AtETR1 does this interaction break down and downstream targets (EIN2) are released from inhibition. It may be possible that the upstream interaction of PpANR in its ABA signalling capacity requires it to be in another cellular compartment or simply to be free to interact with another ER associated partner. This could necessitate a weaker PpETR7-PpANR interaction to prevent sequestering of PpANR from potential ABA signalling. This remains highly speculative but is a mechanistic question that requires understanding and one which has certainly been completely overlooked in its initial characterisation (Yasumura *et al*., 2015). Understanding this process is also important for understanding how dual roles can be achieved in such important regulators which are so rare.

CHAPTER 5

Chapter 5 – Structural analysis of the N-terminal PAS domain in PpANR

5.1 Introduction

5.1.1 The PAS domain superfamily

The Per-ARNT-Sim (PAS) domain superfamily is a group of structurally similar protein sensor modules found in all kingdoms of life principally involved in signal transduction (Taylor and Zhulin, 1999) - enabling the detection of and response to a range of environmental stimuli. The three basic helix-loop-helix (bHLH)-PAS proteins which gave their name to the PAS domain, Per, ARNT and Sim, were the first three of this family to be discovered and the extended C-terminal region was found to have a 270bp region of homology with two short 50bp repeats within it called the PAS domain (Nambu *et al*., 1991); however this has since been redefined. Despite low sequence homology across the sub-families the PAS domain is now defined around a 100-120 amino acid core that forms a canonical 'PAS fold' including a 5 antiparallel strand β-sheet and (usually) 4 α-helices (A_β, B_β, C_α, D_α, E_α, F_α, G_β, H_β, I_β) (Taylor and Zhulin, 1999; Hefti *et al*., 2004). However, there is some variation in the number and size of the α-helices (Möglich *et al*., 2009a) (Figure 5.1). This PAS fold is often flanked by further α-helices which show much structural variation between the PAS subfamilies but in some cases are thought to be functionally important principally by interacting with the β-sheet in various ways (Möglich *et al*., 2009a), affecting both quaternary structure (Ayers and Moffat, 2008) and by propagating signals from the PAS fold as coiled-coil linkers. This relatively simple structure is highly versatile enabling the PAS domain to:

- Sense light by associating with flavin cofactors (flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD)) as well as redox potential (see Becker *et al*., 2011 for recent review).
- Sense oxygen using a haem cofactor (see Gilles-Gonzalez and Gonzalez, 2004 for review).
- Bind a range of small ligands (see Henry and Crosson, 2011 for review of bacterial structures).
- Interact with transcriptional co-effectors as well as other PAS domains to enable modulation of its signal output (see Partch and Gardner, 2010 and Bersten *et al*., 2013 for reviews in bHLH-PAS transcription factors).

Interestingly, when the crystal structure of the photoactive yellow protein (PYP), a bacterial light sensing PAS protein and the first PAS domain to be crystallised, domain was first resolved, its structural similarity to other phylogenetically un-related domains

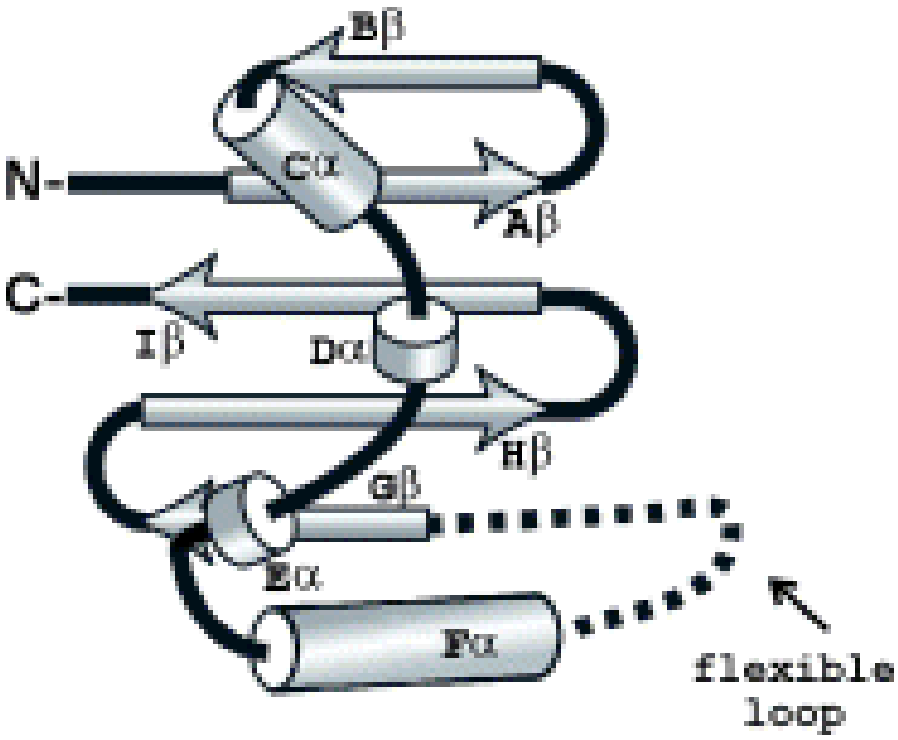


Figure 5.1. **Schematic representation of a PAS fold structure as defined by Hefti *et al*., 2004**. Some key features shown include: The alignment of the 5 β-strands to form an anti-parallel β-sheet, the dominant size and position of the highly variable F_α which acts as a 'gate' for ligand entry to the binding pocket, when present, enabling PAS subclass specificity and the flexible loop region of F_α-G_β which acts as a hinge for the F_α helix 'gate'. The PAS fold structure is found in all kingdoms in a varied array of proteins where they primarily act as sensor modules linked to effector modules enabling signal transduction. PAS domains are also often involved in dimerization with other PAS domains conferring another level of regulation. Figure taken from Hefti *et al*., 2004.

involved in protein-protein interactions was noted (Borgstahl *et al*., 1995) suggesting the basic structure of the PAS fold and similar structures is favoured in a range of contexts and roles. As more PAS structures have been solved, previously distinct domains have been found structurally similar enough to the PAS fold to be included as PAS subclasses such as the cache domains (divergent extracellular sensory domains) while the GAF domains, found adjacent to as PAS domain in phytochromes, are also suggested as distantly related (see Möglich *et al*., 2009a) further highlighting the conservation of a recognisable 3D structure despite almost no sequence homology. The F_α helix is sometimes referred to as the 'helical connector' and this relatively large helix, opposing the β-sheet as a thumb might the fingers of a partially closed hand, shows relatively low conservation among the PAS superfamily. Its position allows it to act as a 'lid' or 'gate' to the characteristic hydrophobic pocket within which ligands and cofactors can bind and its structural diversity (along with the E_α helix) plays a large part in conferring a PAS domain's specificity (Taylor and Zhulin, 1999).

The small PAS sensor module is rarely found alone in a protein but is found in a range of contexts befitting its range of functions including in: modular kinases (including the histidine kinase bacterial 'two-component' regulatory systems), transcription factors and circadian clocks. PAS domains are sometimes found as part of a signal sensor module complex with other sensor domains and are also frequently found with other PAS domains; occasionally ones with different origins and functions. PAS domains are almost always found intracellularly (see Chang *et al*., 2010 for extracytoplasmic examples), although they are occasionally membrane associated, despite their occasional sensing of external stimuli (such as light or toxins) meaning the cues must necessarily cross into the cell. More often than not however, proxies of environmental conditions such as changes to redox potential or in intracellular oxygen concentration are sensed. In summary, PAS domains usually act as input modules which modulate output domains.

5.1.2 bHLH-PAS family

The bHLH-PAS family, with 19 mammalian members, has been extensively studied in humans and mice due to the roles of family members in the hypoxia response and in cancer (see review Bersten *et al*., 2013). However, it is widespread throughout animals and has further roles in neuronal development and function and in circadian rhythms (reviewed in McIntosh *et al*., 2010). As seen, three members of this family in fact gave their names to the PAS acronym (Per-ARNT-Sim) highlighting their historical importance in this field. Given the extensive structural and functional data available on this family, they provide an excellent basis to understand PAS domains in more detail, especially considering that within this relatively small group the PAS domains can

regulate dimerisation, bind a range of exogenous ligands and even bind haem as a cofactor. bHLH-PAS proteins comprise a subfamily of the bHLH transcription factor (TF) superfamily and take the form of an N-terminal bHLH DNA binding domain followed by two PAS domains (PASa and PASb) and an extended C-terminus involved in transactivation. As in some other proteins in which multiple PAS domains are found, the two PAS domains have differing, but overlapping roles, with PASa primarily involved in specificity of both heterodimer formation and DNA binding to the enhancer-box motif (E-box), a common eukaryote *cis*-regulatory element, while PASb is more often a sensor module capable of binding ligands or co-effectors but still has a role in heterodimerisation. The heterodimers are formed between proteins from two distinct interacting classes, classes 1 and 2, in which only inter-class interactions can occur. Specificity is achieved through the various combinations as well as through spatial and temporal regulation. The most well studied members of these TFs are the class 1 aryl hydrocarbon receptor (AHR) and hypoxia inducible factor alpha's (HIF_{α}s) which interact with the class 2 aryl hydrocarbon receptor nuclear translocator (ARNT) to regulate xenobiotic metabolism and the hypoxia response respectively and both pathways are implicated in cancer (Bersten *et al*., 2013). The PAS domain has been the target for a number of drug treatments which have elucidated much about what molecules the PAS fold hydrophobic pocket can accommodate and functionally interact with as well as how interactions affect activity and how signal transduction is achieved and these particular members will be discussed in more detail.

5.1.2.1 The AHR-ARNT interaction

The PASb domain of the dioxin receptor AHR has been extensively studied as it is known to bind a range of both endogenous and exogenous ligands (Denison and Nagy, 2003), including the carcinogenic combustion-derived benzo(a)pyrene (BaP). While primarily studied for its capacity to regulate xenobiotic detoxification, AHR is also implicated in a number of normal developmental and physiological processes since a suite of pathologies are observed in AHR-null mice that gain dioxin resistance (see McIntosh *et al*., 2010 and Gu *et al*., 2000 reviews; Nguyen and Bradfield, 2008). AHR is associated with a multiprotein complex including HSP90, XAP2 and p23 with the PASb domain binding XAP2 directly which keeps AHR in the cytoplasm and spatially separate from its nuclear-located ARNT partner. Interestingly, it has also been suggested that the complex is necessary to keep the PASb domain in a ligand receptive conformation (Pongratz *et al*., 1992). This has important implications for ligand identification in novel PAS domains, such as in the PpANR PAS domain as unknown interactions may be required to allow binding of potential ligands. Upon ligand binding, PASb changes conformation, exposing a nuclear localisation site at the C-

terminus of the bHLH domain for phosphorylation and the complex moves into the nucleus and dissociates. In the nucleus, the active AHR forms a PASa-mediated heterodimer with ARNT which is able to induce gene expression as well as form an ubiquitin ligase complex for a range of substrates. This overview of a very well understood process not only demonstrates some of the mechanisms by which novel PAS domains may operate but also highlights just how hard these mechanisms are to predict. Besides the well documented affinity the AHR PASb has for dioxin and other toxins, a number of highly diverse endogenous ligands have also been identified including a range of indoles, tetrapyrroles and arachidonic acid, with the majority found to be agonists (Denison and Nagy, 2003) suggesting a flexible binding pocket. ARNT has not been shown to bind any ligands but rather is found to be an interacting partner with AHR, the two HIFα's (in the hypoxia response) as well as SIM2.

5.1.2.2 The HIFα-ARNT interaction

The HIFα proteins are class 1 bHLH-PAS members that stimulate the hypoxia (low oxygen) response by inducing gene expression and which are degraded and transcriptionally repressed under normoxic conditions (see reviews Semenza, 2012; Greer *et al*., 2012). HIFα factors, like all class 1 bHLH-PAS family members, require a class 2 interacting partner for heterodimerisation and, like AHR, this partner is ARNT (or HIFβ as it is sometimes known in this context). Given their activity during low oxygen conditions, the HIFα genes are found to be up-regulated in a number of cancers where they are implicated in tumorigenesis (see review by Kaelin, 2011 for details). While the mechanisms of HIFα regulation and associated cancer-related roles are outside our interest in PAS domains, this role has made it an attractive target for drugs that could inhibit its effects, preventing tumorigenesis. As we have seen, the PASa domain is vital for conferring specificity of dimerisation while the PASb domain is generally more involved in signal sensing and the existence of an endogenous ligand for the HIFα PASb hydrophobic pocket has been inferred making any antagonists a vital weapon in disrupting function. Ligand binding to HIFα PASb is thought to disrupt the dimerisation with ARNT, inhibiting any downstream signalling. Ligand screening identified a compound called THS-044 that specifically binds the internal cavity in HIF2α (Scheuermann *et al*., 2009). This specificity is interesting as no interaction is found in HIF1α which shares 74% amino acid identity with HIF2α. What this highlights is only a few amino acid changes, possibly in those that line the internal cavity, can determine whether or not a ligand will bind despite the overall structures remaining highly similar. THS-044 was found to disrupt wild-type HIF2α-ARNT PASb heterodimers most probably by inducing structural changes. This is thought of as a general mechanism of PAS ligand binding and one also documented in cofactor-bound

PAS domains such as the FMN-bound light receptors which change conformation following light induced structural changes in the flavin cofactor (Harper *et al*., 2003; Harper *et al*., 2004).

5.1.3 PAS role in PpANR and aims of study

This general introduction to PAS domains, including some more specific examples from bHLH-PAS TFs, shows that PAS domains clearly have a range of biological roles which can be broadly split between two modes of action: the first is the direct sensing of environmental cues enabling signal transduction which involves the binding of ligands or through associations with cofactors, while the second mode of action revolves around its β-sheet mediated dimerisation. These two properties are often combined. Dimerisation is found in many PAS domain crystal structures and has shown to be functionally significant in the bHLH-PAS TFs, in phototropins and in phytochromes. One feature of the PAS domain superfamily is the maintenance of very high conservation at the structural level while showing great variation at the sequence level. This means that predicting the function of the PpANR PAS domain (or, indeed, any PAS domain) from its sequence alone is difficult. Structural data were therefore required, in parallel with genetic data, to guide further studies into the role of the PpANR PAS domain. This chapter seeks to answer whether the PpANR PAS domain binds (or might bind) a ligand/ cofactor or whether its mode of action revolves around modulation of PpANR's dimerisation state – or a combination of both. The *Physcomitrella* genome does encode other PAS domain-containing proteins (see Chapter 4), however these are found in light receptors (phototropins and phytochromes) and are unlikely to functionally interact with PpANR. This means that PpANR is only likely to dimerise with itself if this mode of action has functional significance. That the PpANR PAS domain may be significantly involved in ABA signalling (possibly as a sensory module) is given greater credence by the finding of Raf10 and Raf11 as ABA regulators in *Arabidopsis* (Lee *et al*., 2015b). These are MAP3 kinases closely related to PpANR that also have an N-terminal PAS domain but which lack an EDR domain (see Chapter 4). Assuming that the conservation in function between these related MAP3Ks is mirrored by a conserved mode of action, elucidating the mode of action for the PpANR PAS domain will shed light on important and little understood ABA regulators which were likely important components in the conquest of land. The objective was to find possible interacting partners of the PpANR PAS domain and to determine its 3-dimensional structure by X-ray crystallography to elucidate how these interactions might be mediated and to guide the direction of future studies.

5.2 Methods and Materials

5.2.1 Bacterial strains and vector

The single expression construct used comprised the *PpANR* PAS domain sequence cloned into a modified pET-28a (Novagen) vector in which a SUMO fusion tag had been added in between the 6-His-tag and the multiple cloning site (MCS). Cloning was carried out in competent DH5α *Escherichia coli (E. coli)* cells (Invitrogen). The *E. coli* expression strain BL21 (DE3) was purchased from Stratagene.

The construct was made by inserting a 339bp fragment of cDNA containing the PAS domain with a reverse primer that incorporated a stop codon in a four base 3'-extension (TTAA) (PAS_F: AGCCAGCCATGTGTCTCCATCG; PAS_R: **TTAA**TAGTTCCGGGCCATAAGTTCAG). The fragment was blunt end ligated into the *Sac*I site, by using the isoschizomer *Ecl*136II, in the MCS. Clones were sequenced to ensure no errors were introduced.

5.2.2 Protein Expression

3 mL of overnight culture was used to inoculate 1L LB + kanamycin (50mM) in a 2L conical flask. Cultures were incubated at 37°C with aeration until an optical density (OD)600 0.6-0.8 was reached. Protein expression was induced by addition of sterile IPTG to the culture at a final concentration of 0.3 mM and the culture was then incubated at 18 or 25°C with aeration overnight. Cells were harvested by centrifugation at 3500 × g at 4 °C for 20 minutes. Cell pellets were resuspended in lysis buffer (20 mM Tris pH 7.9, 500 mM NaCl). 5 U DNase I and 2 mM $MgCl_2$ were added to the suspensions prior to lysing the cells. Cells were lysed by sonication. The lysate was centrifuged at 13000 × g for 25 minutes at 4°C. Both the pellet (insoluble fraction) and supernatant (soluble fraction) were analysed by SDS-PAGE for initial trials and then just the supernatant once suitable conditions were found.

5.2.3 OD600 measurement

The OD of cell cultures was read with a spectrophotometer (Pye Unicam 8600) set to 600nm by using 1mL of culture in a plastic cuvette after the spectrophotometer was zeroed against a blank sample (1mL of un-inoculated media).

5.2.4 Nickel affinity chromatography

For purification of protein with a His-tag from 2L cultures a 5mL Ni2+ sepharose HisTrap HP column (GE Healthcare) was used. Columns were attached to a peristaltic pump at room temperature. The maximum flow rate used was 5mL/min. Columns were washed with 5 column volumes of water, charged with 10ml 0.1M $NiSO_4$ and equilibrated with 5 column volumes of lysis buffer before loading the of filtered

supernatant. The filtered supernatant was passed through the column and the flow through was collected. The column was then washed with 5 column volumes of lysis buffer. Elution was carried out by an increasing concentration of imidazole in lysis buffer (50mM, 100mM and 300mM). All wash fractions were collected and analysed by SDS-PAGE. Columns were stored in 20% ethanol. Initial purified fusion protein was cleaved with SUMO protease and dialysed in lysis buffer over night at 5°C with magnetic stirrer giving constant agitation. Dialysis bags (Spectra/Por Dialysis Membrane MWCO: 3,500) contained around 50ml uncleaved protein were in 2L lysis/dialysis buffer.

5.2.5 Size exclusion chromatography

Size exclusion chromatography was performed on a Superdex 75 (26/60) column (GE healthcare) attached to an Akta prime system at 4°C. The absorbance of the sample was monitored at 280 nm throughout. The column, which is stored in 20% ethanol, was washed with water and equilibrated with appropriate buffer (20 mM Hepes pH 7.5, 100 mM sodium chloride, 1 mM DTT, 5 % glycerol). 5mL of protein sample (after affinity chromatography) was loaded onto the column using a 5mL injection loop. The basic process was as follows: All runs set to 2ml/min flow rate, sample injected, approximately 110ml run through before samples were collected as 3ml fractions until a single sample peak had run off. Upon completion of the method, the column was washed with water and stored in 20% ethanol. Each fraction from within the peak of the UV trace was analysed by SDS-PAGE.

5.2.6 Protein concentration

Protein samples were concentrated at 4°C in 15ml Centriprep centrifugal filters (EDM Millipore – 10kDa molecular weight cut-off) at 2770 × g until the desired concentration or volume was reached. The concentration of protein was routinely measured using a protein assay reagent dye (Bio-Rad) as follows: 200µL of dye was added to 800µL of water and then either 20µL of buffer (blank) or protein sample was added to this mix. The solutions were placed in a plastic cuvette and the absorbance measured at 595 nm in a spectrophotometer (Helious β) using the BioRad Protein Assay as recommended. Firstly the spectrophotometer was zeroed against the blank then the absorbance of the sample was measured. In the assay an absorbance of 1 signifies a concentration of 1mg/mL. The assay is accurate for readings between 0.2-0.8mg/mL, if outside this range the volume of protein sample added was diluted accordingly.

5.2.7 SDS PAGE gel electrophoresis

Sodium dodecyl sulphate polyacrylamide (SDS-PAGE) gels were made using a 12-15% gradient resolving gel with a 5.5% stacking gel summarised in Table 5.1. Samples

were mixed with loading buffer in a 1:1 ratio and boiled for 5 minutes. Gels were run in Tris-glycine-SDS running buffer (25 mM Tris pH 6.8, 0.1% (w/v) SDS, 190 mM glycine). A custom broad range molecular marker (14.2-66kDa) as well as a commercial marker (BioRad 7.1-209kDa) was included in gels. Proteins were visualized using Coomassie blue stain (45% (v/v) methanol, 10% (v/v) acetic acid, 0.25% (w/v) Coomassie Brilliant Blue R-250) and de-stained in 30% (v/v) methanol, 10% (v/v) acetic acid.

Table 5.1. Mixtures used for SDS-PAGE gels.

Solution	15% Resolving gel	10% Resolving gel	Stacking gel
H_2O	-	4 mL	10 mL
30% bis-acrylamide (37.5:1)	8.75 mL	5.85 mL	2.5 mL
1 M Tris (pH 9.4)	7.35 mL	7.35 mL	-
1 M Tris (pH 6.8)	-	-	1 mL
10% SDS	175 µL	175 µL	150 µL
10% APS	50 µL	50 µL	75 µL
TEMED	10 µL	10 µL	10 µL

5.2.8 ABA and FAD binding

Purified PpANR PAS protein was incubated with 10x molar excess ABA and FAD respectively before loading samples back through size exclusion chromatography to purify protein and any bound ligand. ABA binding was analysed by circular dichroism (CD) with the spectra of *apo*-PAS and co-purified PAS+ABA compared. FAD binding was analysed by Nanodrop spectrophotometry (ND1000) to assay the presence of FAD (which has characteristic absorbance maxima at 450nm and 535nm) in the recovered protein fractions.

5.2.9 Circular Dichroism (CD) Spectroscopy

Spectra were recorded on chirascan circular dichroism spectropolarimeter (Applied Photophysics), at 20°C, using 1mm cells and a scan speed of 5nm/min between 180-260nm. The spectra were averaged over 2 repeats with the buffer baseline subtracted. Protein concentrations of approximately 0.2mg/mL were used.

5.2.10 Crystallography

5.2.10.1 Factorials

Initial factorial screens were set up routinely using the commercially available screens: Hampton Research: Crystal Screen I and II, Index, Salt RX; Emerald Biosystems: Wizard I, II, III and IV; Molecular Dimensions: Morpheus. Screens were set up using a

Formulatrix NT8 robotics system. Typically drops of 0.1µL of protein (at 10mg/ml) in a 1:1 drop ratio with mother liquor were used in 96 well MRC 2-drop sitting drop plates (Molecular Dimensions). Wells were filled with 40µL mother liquor and sealed with Viewseal pressure adhesive clear seals and incubated at 18°C.

5.2.10.2 Optimisation

A single positive result from the factorial screens was optimised in 24 well hanging drop plates with 1µL protein (10mg/ml) mixed with 1µL mother liquor over 1mL mother liquor. This was condition 8 from the Wizard 4 screen: 0.1M Tris base, pH 8.5 (HCl), 2M $LiSO_4$ and 2% PEG400. Optimisations centred on the variation of precipitant concentration and pH. The conditions for this optimisation were across the pH range 6.5-9.5 and the Li_2SO_4 (instead of $LiSO_4$ due to availability) range 1.5-2.75M and using PEG500 rather than PEG400 due to availability.

5.2.10.3 Cryo cooling

Crystals were picking using a nylon loop (Hampton Research) and submerged into 1µL of mother liquor + glycerol in increasing concentrations of glycerol to a final concentration of 20%. Crystals were then immediately immersed in liquid nitrogen for storage.

5.2.10.4 PpANR PAS domain crystallography

Crystals diffracted to 1.7Å at diamond light source beamline i24 using a 0.97Å wavelength. The best dataset included 98,517 reflections of which 26,146 were unique. These were processed using the 3d program of xia2. The structure was solved by molecular replacement using the structure PDB:3LYX, using Phaser (McCoy *et al*.,2007) and built using arpwarp on the ccp4 suite (http://www.ccp4.ac.uk/). The structure was then refined using iterations of Refmac (Murshudov *et al*., 1997) and manual adjustments with coot (Emsley *et al*., 2010). Crystallography was carried out under the guidance of Dr C. Trinh.

5.2.11 Metabolite binding screen

Purified, but un-cleaved PpANR PAS protein was fixed on a 1ml cobalt column (HiTrap TALON) in loading buffer (20 mM TRIS pH 7.9, 500 mM NaCl). Moss tissue extract was prepared using an aqueous and a methanol extraction method. The aqueous extraction was carried out by re-suspending powdered tissue in loading buffer followed by centrifugation at 10000 × g for 5mins. The supernatant was boiled for 2mins to denature any soluble proteases that might interfere with the bound PpANR PAS. This was centrifuged at 10000 × g for 5mins and the supernatant loaded onto the column. The methanol extraction was based on previous moss metabolomic analyses (Erxleben *et al*., 2012). Sample loading, washing and elution was carried out as described before

(Jonker *et al*., 2008) in which a 10% methanol buffer was used to remove non-specific interactions and a low pH was used to strip off PpANR PAS protein and any bound compounds. The eluate was cleaned by Ziptip C_{18} pipette tips (Millipore) and analysed by MS-MS.

5.2.12 Structural Relatedness Analysis

This analysis was carried out based on a similar analysis by Henry and Crosson, 2011. PDB files were collected from the PDB database that showed initial similarity to the PpANR PAS domain. The analysis was carried out using the STAMP package (http://www.compbio.dundee.ac.uk/manuals/stamp.4.2/). First the pdbseq program was used to extract the appropriate regions from PAS structures that defined the core PAS fold and no additional sequence (Appendix 5.1). This prevented any biases being introduced from flanking regions which were not included in the PpANR PAS structure. The extracted sequences were initially aligned using Clustalx with default settings and the alignment was converted to amp format using the included aconvert program. Then the alignfit program was used which refines the alignment. This refinement was then run with the final stamp program which carries out the structural relatedness quantification. The stdout from this was piped to a text file and the root-mean-square-deviation (RMSD) values from each pairwise comparison were converted into a formatted distance matrix ready for input to phylip's (http://evolution.genetics.washington.edu/phylip.html) fitch program. This program clusters the distance matrix of RMSD values using the Fitch-Margoliash method (Fitch and Margoliash, 1967). The output tree was visualised and formatted using FigTree.

5.3 Results

5.3.1 PAS expression construct prediction and synthesis

In designing an expression construct for the PAS domain protein, a region that encompassed the minimal domain of interest was required. To choose what sequence was sufficient with minimal 'overhang' the structural prediction software I-TASSER was used to model the sequence against PAS domains in the PDB database. PAS domains have a highly conserved tertiary structure made up from 100-120a.a residues (Figure 5.1). Using the conserved domain tool (CD) on the NCBI server an initial estimate of the PAS domain sequence was found (encompassing 102a.a). A preliminary I-TASSER run was carried out which included a predicted overhang of 15 amino acids at either terminus revealing that like many PAS structures, the core PAS fold of 102a.a is likely flanked by helices (Figure 5.2). While the helices can be functionally relevant, such as in the LOV dimerisation in phototropins (see Christie *et al*., 2014 for recent review), the primary aim here was to define the core PAS fold structure. Based on the

A

```
                    20                  40                  60
                     |                   |                   |
Sequence      GRAGSTKQQSTVQNPSQPCVSIDLRGNITSWNRAAEQLYQFKEGEVMGRNLVDLIVPPNAAKTAALDTIM

Prediction    CHHHHHHHHHHHHHHCHHSSSSCCCCSSSSSCHHHHHHHCCCHHHHCCCCHHHHCCCHHHHHHHHHHHHH

Conf.Score    97899999999973140207999999476780899998892999994997899471923665689999999

                 80                  100                 120
                  |                   |                   |
Sequence      RVGQGEVWIGQFSCVRRDNAILNTVVTQKPILDDDQKISGISVCTEPYSMPELMARNYFDEGT

Prediction    HHHCCCCSSCSSSSSCCCCCSSSSSSSSSSSSCCCCCSSSSSSSSSCCHHHHHHHHHHHHHCC

Conf.Score    99869957886989979999999999999999999999999999998317999999999987069

              H:Helix; S:Strand; C:Coil
```

B

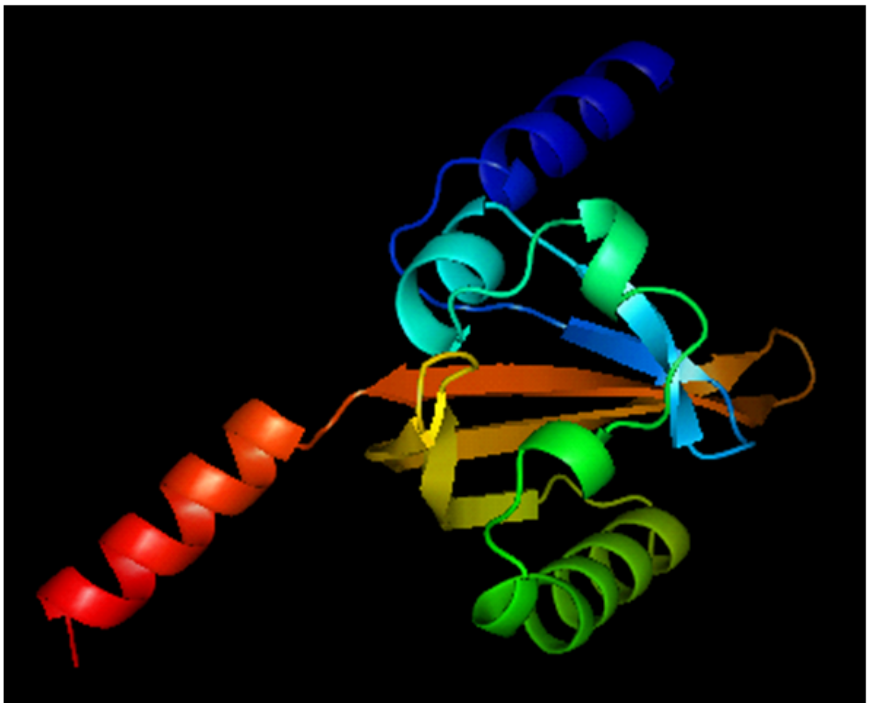


Figure 5.2. I-TASSER structure prediction of PAS fold and flanking 15a.a. A) Sequence and secondary structure prediction of conserved PAS domain and flanking 15a.a **B)** Best model for I-TASSER run with 15a.a overhangs. The overhangs are predicted to form helices which are not part of the 'PAS fold'. I-TASSER outputs: C-score of 0.74, a TM-score of 0.81+-0.09 and a RMSD value of 3.1+-2.2.

A

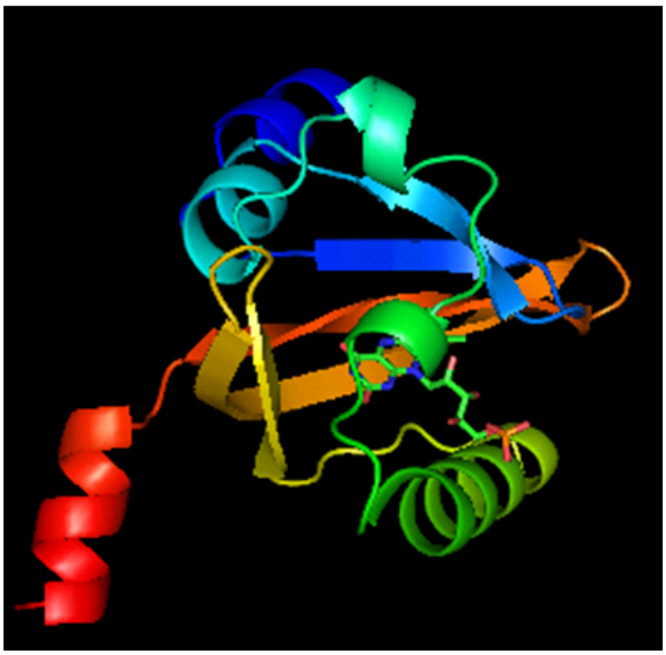
Predicted protein:
MGSSHHHHHHSSGLVPRGSHMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRR
LMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHRE|QIGGSEFESQPCVSIDLRGNI
TSWNRAAEQLYQFKEGEVMGRNLVDLIVPPNAKTAALDTIMRVGQGEVWIGQFSCVRRDNAILNTVV
TQKPILDDDQKISGISVCTEPYSMPELMARNY

B

```
                    20                  40
                     |                   |
Sequence    QIGGSEFESQPCVSIDLRGNITSWNRAAEQLYQFKEGEVMGRNLVDLIVPPNAKTAALD
Prediction  CCCSSSSCCCSSSSSCCCCSSSSSCHHHHHHHCCCHHHHCCCCHHHHCCCCHHHHHHHH
Conf.Score  901066217549999999979887288998889199999299879967391165699 99

            60                  80                  100                 120
            |                   |                    |                   |
Sequence    TIMRVGQGEVWIGQFSCVRRDNAILNTVVTQKPILDDDQKISGISVCTEPYSMPELMARNY
Prediction  HHHHHHCCCCSSCSSSSSCCCCCSSSSSSSSSSSCCCCCSSSSSSSSCCCHHHHHHHCC
Conf.Score  999998699488859999789999999999999999999999999999886888899860 69
            H:Helix; S:Strand; C:Coil
```

C



D

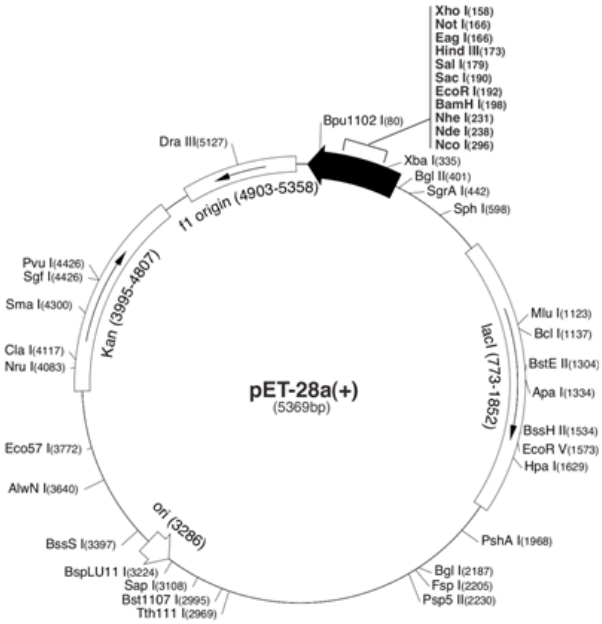


Figure 5.3. **I-TASSER structure prediction of PAS fold sequence used in expression construct. A)** Amino acid sequence of the fusion protein predicted to be expressed in the pET-28a-SUMO construct as confirmed by Sanger sequencing. 6-His tag is shown in red, the SUMO protein in green and PAS fold domain in blue. The cleavage site is shown at the 3' end of the SUMO sequence. **B)** Sequence of cleaved protein and its predicted secondary structure from I-TASSER. **C)** Best model from this cleaved protein used in crystal screens and binding assays with potential FMN ligand binding shown. C-score=0.85, TM-score=0.83+-0.08, RMSD=2.7+-2.0A. **D)** Vector map of the pET-28a(+) range from Novagen used in cloning.

assumption that these helices would decrease the probability of successful crystallisation of the core PAS fold, a final sequence was selected that encoded 112aa and contained minimal flanking helices apart from seven 3' residues immediately before the stop codon, introduced by the 3' primer. The 339bp insert was amplified from cDNA with the reverse primer containing the extension TTAA to introduce the stop codon. This 'excess' flanking 3' region was necessary to design a suitable primer. The 5' vector flanking region that would be included, in between the SUMO protease cleavage site and the start of the insert in the final cleaved protein, was also included in a final protein sequence of 120a.a and run through I-TASSER to ensure no unexpected features would be introduced (Figure 5.3B). This final protein sequence was again predicted to form a canonical PAS fold with the top PDB threading template being 3LYX which was also the top structural match of the best model (Figure 5.3C). I-TASSER also uses the program COACH which predicted FMN as a possible ligand/cofactor as found in the blue light sensing LOV subfamily of PAS domains with high probability (C-score=0.91). A comparison with the other PAS-containing MAP3Ks, Raf10 and Raf11 from *Arabidopsis*, show very similar structure modeling with FMN also predicted as a ligand with high confidence (C-scores for Raf11 and Raf10=0.97) as expected by their high sequence homology. Final clones were sequenced revealing no errors has been incorporated and a 234a.a protein was predicted to be translated including the 5' 6-His-tag, the SUMO tag and the PpANR PAS insert (Figure 3A).

5.3.2 Expression of recombinant PAS domain

Following transformation of the BL21_DE3 protein expression *E. coli* strain with the PpANR PAS expression construct, the expression conditions were checked. IPTG induction of cultures grown overnight at both 18 and 25°C produced a high yield of protein, with a slightly higher concentration obtained at 25°C which could be visualized by a large band matching the predicted 26.4kDa (Figure 5.4). The presence of salt (0.5M NaCl) also made little difference on protein expression but was included as it is known to have a stabilising effect on proteins. The insoluble fraction for each condition also showed that little of the expressed protein was insoluble. These preliminary tests showed the expressed protein to be readily expressed and almost exclusively in soluble form. Large scale expressions usually yielded 100-200mg/liter and the expressed protein was first purified on a 5ml nickel column with the 6-His-tagged and bound protein eluting in 5-30% 1M imidazole (Figure 5.5). This purified protein was then cleaved with a SUMO protease and re-run through the nickel column with the non-tagged PAS protein passing through while the 6-His-tag::SUMO fragments bound to the column (Figure 5.6A). This stage demonstrated that the cleaved PAS protein was well separated from the SUMO. Concentration of the cleaved PAS protein showed this

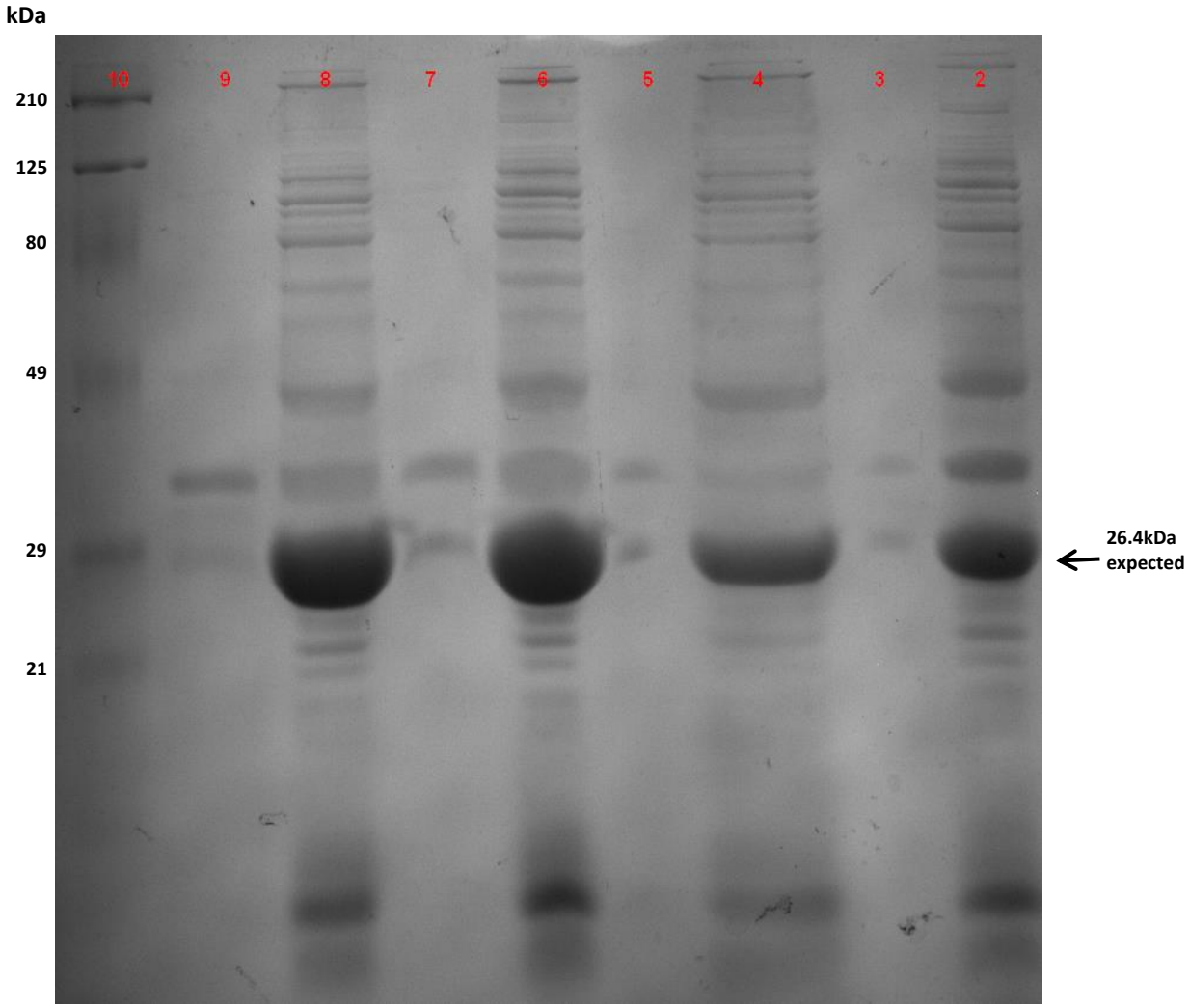


Figure 5.4. SDS-PAGE gel of initial PpANR PAS protein expression conditions. Initial conditions for protein expression were found to be favorable with large product band visible matching the predicted 26.4kDa. Lanes 3, 5, 7 and 9 represent the insoluble fractions and 2, 4, 6 and 8 the soluble fractions. As can be seen, the expressed protein was highly soluble. Overnight expression at 18°C (lanes 2-5) and 26°C (lanes 6-9) showed that 25°C resulted in greater protein concentrations while the presence (lanes 2, 3, 6, 7) or absence (4, 5, 8, 9) of 0.5M NaCl had little effect. Broad range BioRad molecular marker was used (lane 10). Samples were run on a 12-15% gradient SDS-PAGE gel and stained with coomassie blue.

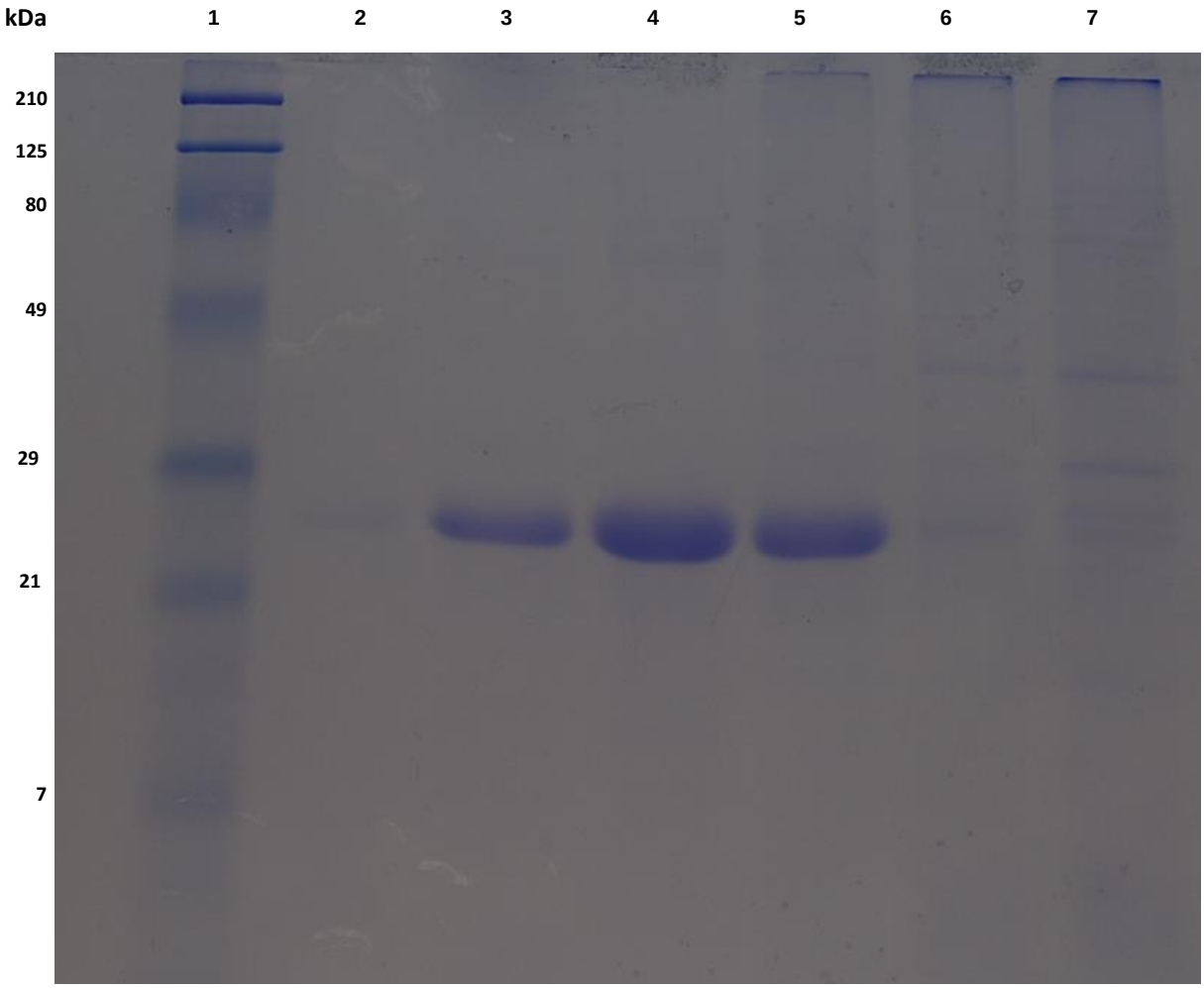
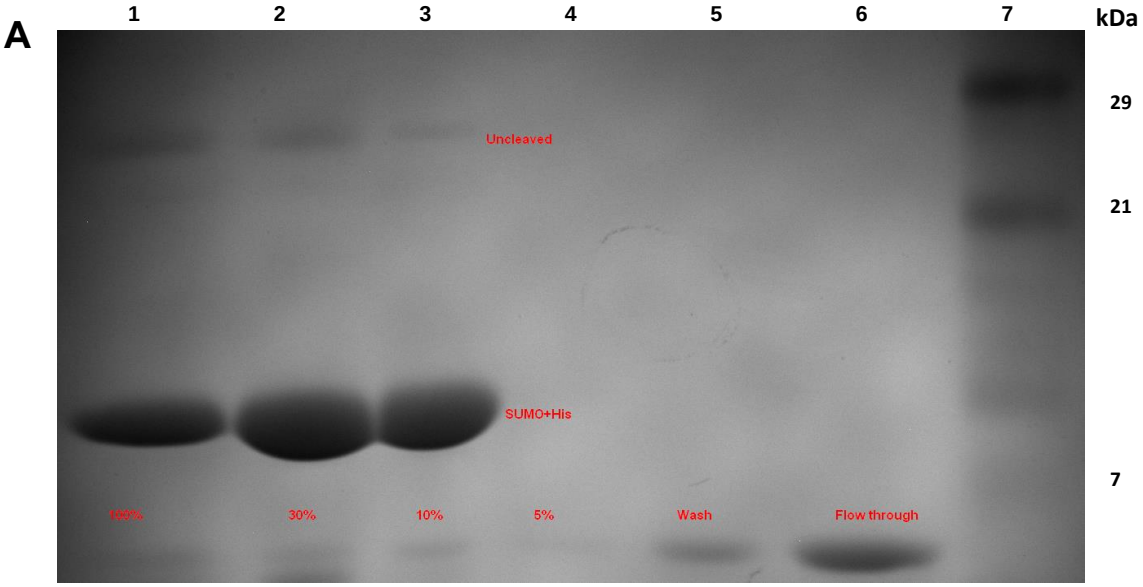


Figure 5.5. SDS-PAGE gel of fractions from first nickel column affinity purification of PpANR PAS protein. Flow through (lane 7) and wash step (lane 6) showed that little expressed protein was being lost. Elution at 5 (lane 5), 10 (lane 4) 30 (lane 3) and 100% (lane 2) 1M imidazole showed the expressed protein was successfully eluted with the vast majority eluting by the 30% fraction. Samples were run on a 12-15% gradient SDS-PAGE gel and strained with coomassie blue and run with a BioRad broad range molecular marker (lane 1).



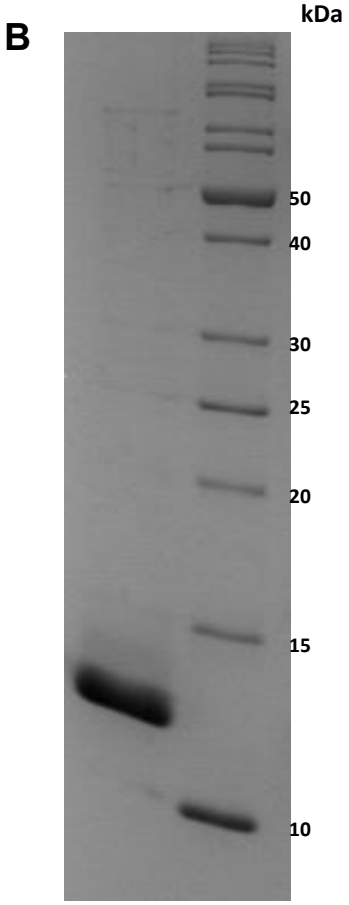


Figure 5.6. Fractions from second nickel column affinity purification yielded high purity of PpANR PAS protein. A) Fractions from second stage of affinity column purification after SUMO protease cleavage. The flow-through (lane 6) and wash (lane 5) contain the cleaved PAS protein while the bound 6-His-tagged SUMO was eluted by 5, 10, 30 and 100% 1M imidazole (lanes 1-4). Some un-cleaved protein is detectable in the elution fractions. Samples were run on a 12-15% gradient SDS-PAGE gel and strained with coomassie blue and run with a broad range BioRad molecular marker (lane 7). **B)** The purified PAS protein was concentrated and run alongside the PageRuler (Thermo scientific) marker with sizes shown. Purified protein is close to the expected 13.35kDa.

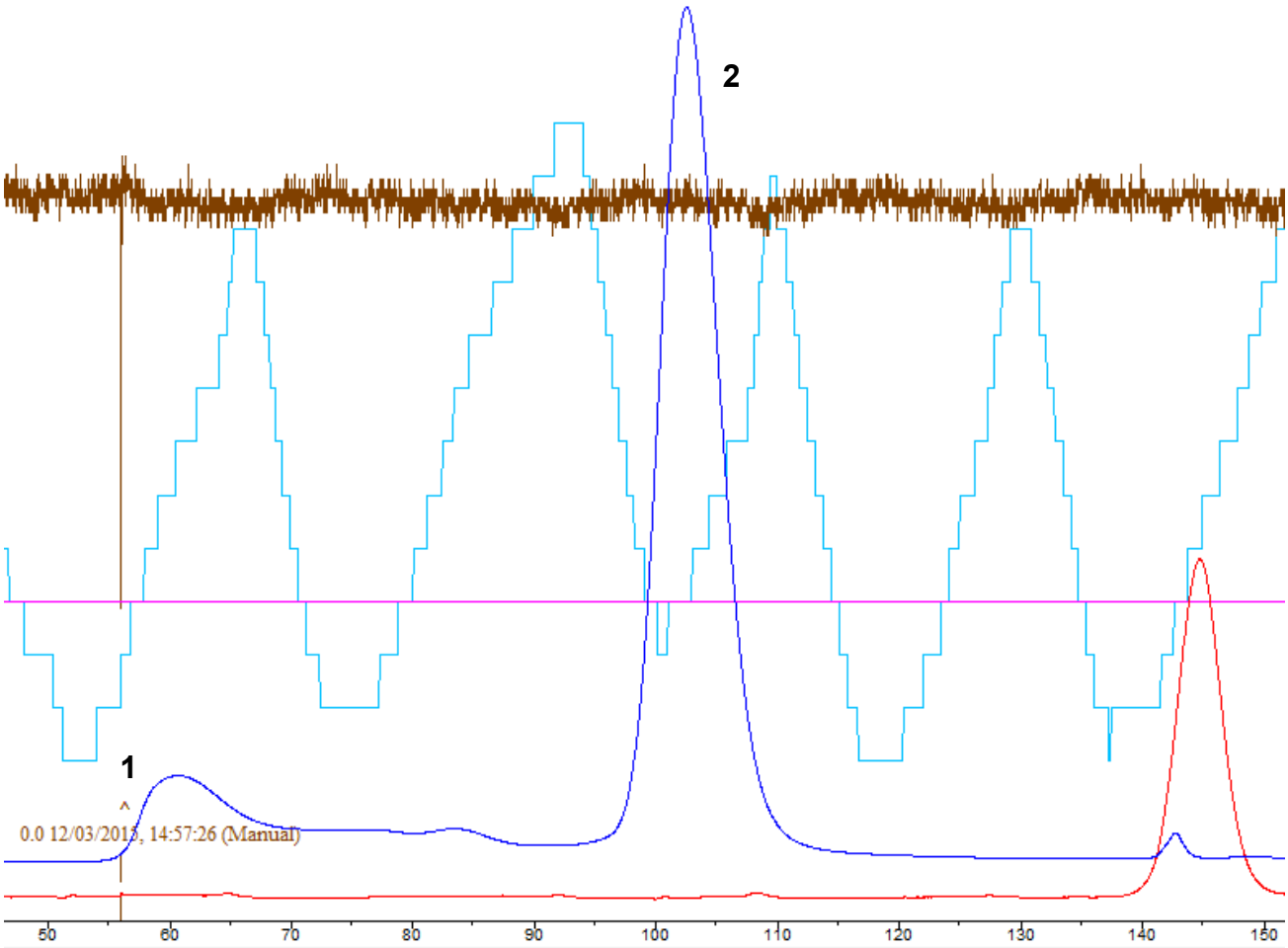


Figure 5.7. Size exclusion chromatography shows single peak of PpANR PAS protein. Akta prime output of size exclusion chromatography purification of PpANR PAS protein following purification by Ni column affinity with the UV trace shown by smooth dark blue curve. After approximately 100ml buffer has run through (50mins) following injection of protein sample **(1)**, the PpANR PAS protein is found coming off the column in a single peak and these fractions are collected and bulked **(2)**. The initial UV peak around injection represents the void from the column. The red curve is the conductivity with the peak representing the NaCl from the buffer coming off the column.

more clearly with the purified protein matching the expected 13.35kDa and being at high purity (Figure 5.6B). The concentrated protein was run through a gel filtration column to further purify the appropriate fractions from the single peak (Figure 5.7).

5.3.3 Metabolite interaction screen with PAS

The purified PpANR PAS protein was screened against moss tissue extract with the aim of identifying any interactions with metabolites, including with ABA itself. This approach aimed to prevent any bias in looking for specific interactions only which might miss less predictable interactions. This procedure used an affinity column ligand capture approach in which the 6-His-tagged PAS protein was bound to a cobalt column with potential ligands washed over to trap any interacting partners before eluting any complexes and analyzing these by mass-spectrometry. As ABA was still considered a possible interacting partner, standards of ABA were run on MS-ES both with and without a ‘ziptip’ cleanup and MS-MS to obtain a fragmentation pattern (Figure 5.8). While the cleanup step diminished the relative signal of ABA at ~263m/z, it was still present and the MS-MS fragments were found to be ~219m/z and ~153m/z in negative mode.

Two metabolite extraction methods (an aqueous and a methanol-based method) were tested to see how complex a molecular library was being screened. ~28 distinct high abundance peaks (defined as being over 10% of the maximum peak in negative mode) were found in the methanol extract compared to the 10 high abundance peaks in the aqueous extraction (Figure 5.9). Neither of the extractions therefore produced a particularly complex metabolome. Nevertheless, a pull-down was carried out on the aqueous extract and a number of peaks were obtained (Figure 5.10). Two of the four biggest peaks in the pull down (~237 and ~265 m/z in negative mode) were also the two biggest peaks in the whole extract. This suggested that nonspecific binding of high abundance components may have been an issue and that more stringent washes should be used in the future. Of these four, one peak (~148m/z) did, however, show clear enrichment having not been detected in the total extract profile. Another peak (~113m/z) showed moderate enrichment increasing from ~8% maximum to >50%. The ~148m/z peak had 9 matches on the METLIN (Scripps) metabolite database which were predominantly heterocyclic compounds with indole-related and benzene aromatic rings prominent (Figure 5.11). The ~113m/z peak was similarly queried and while 9 hits are presented (Figure 5.12) the tolerance of the search was greatly increased in order to find matches. These featured complex aromatic ring-containing compounds.

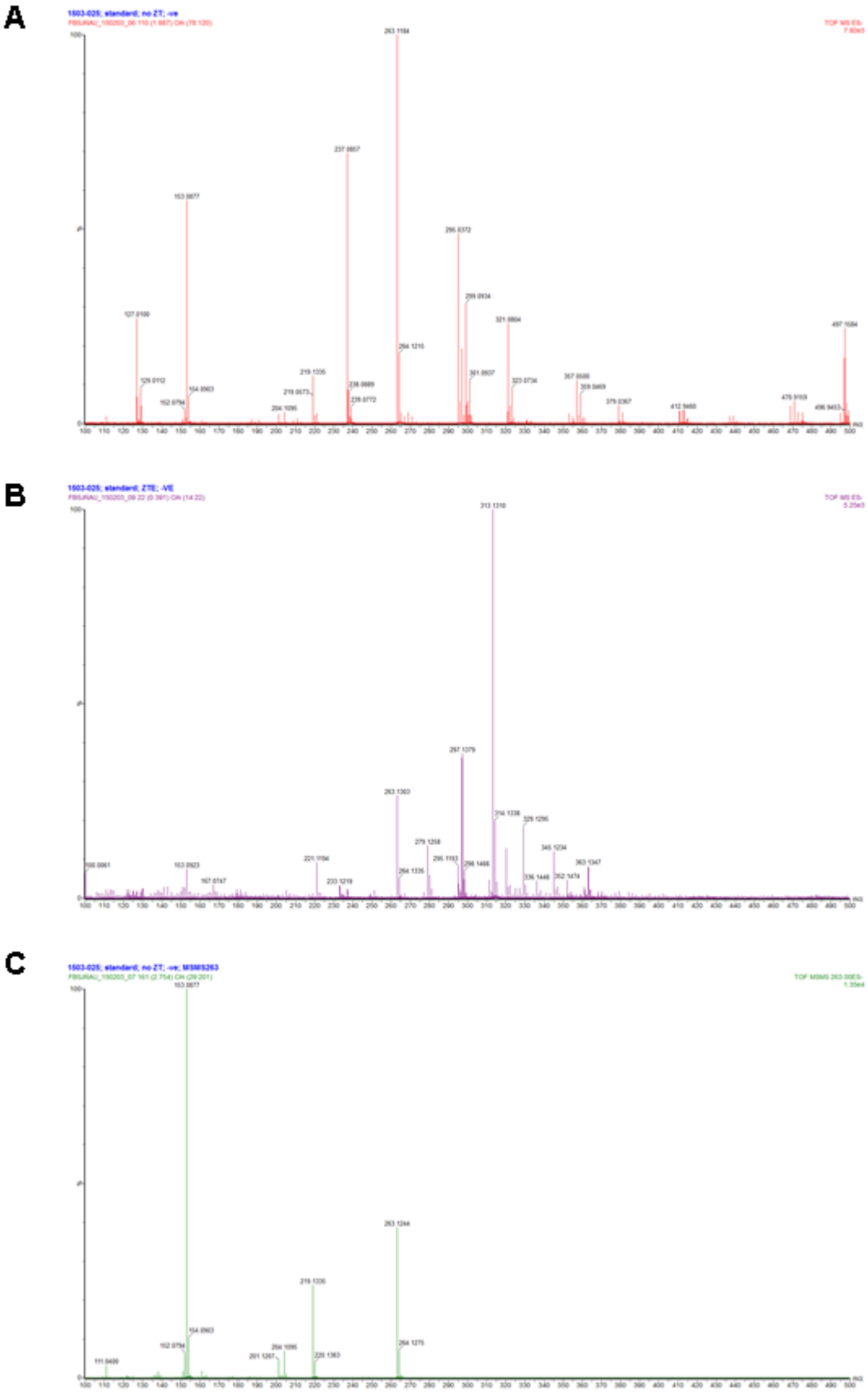
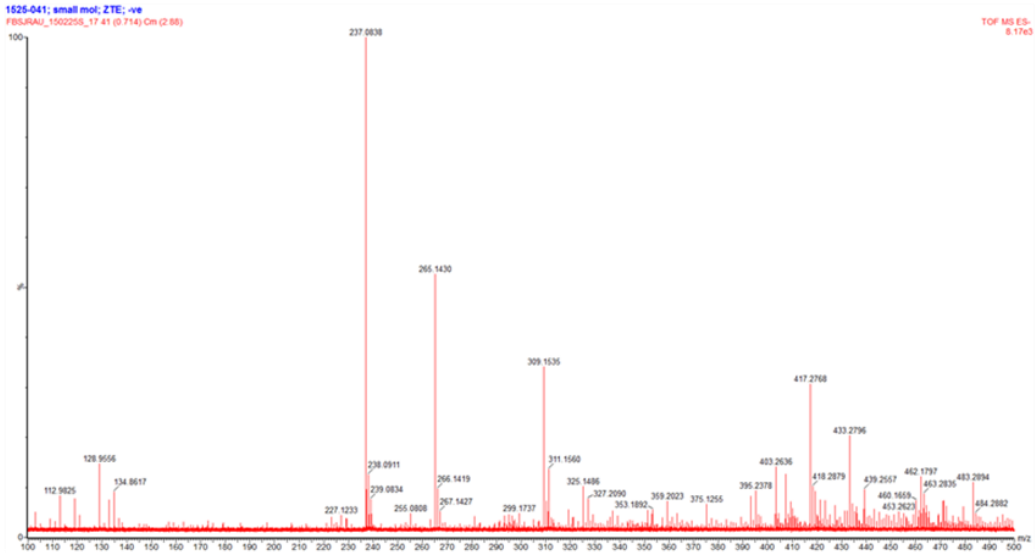


Figure 5.8. Mass-spectrometry of ABA standards. (+-)ABA standard run through MS-ES in negative mode with **(B)** and without **(A)** 'Zip-tip' C_{18} cleanup. ABA peak at ~263m/z is visible in both but cleanup step enriches a peak at ~313m/z. **(C)** ABA fragmentation pattern was also obtained by MS-MS showing fragments at ~219m/z and ~153m/z.

A



B

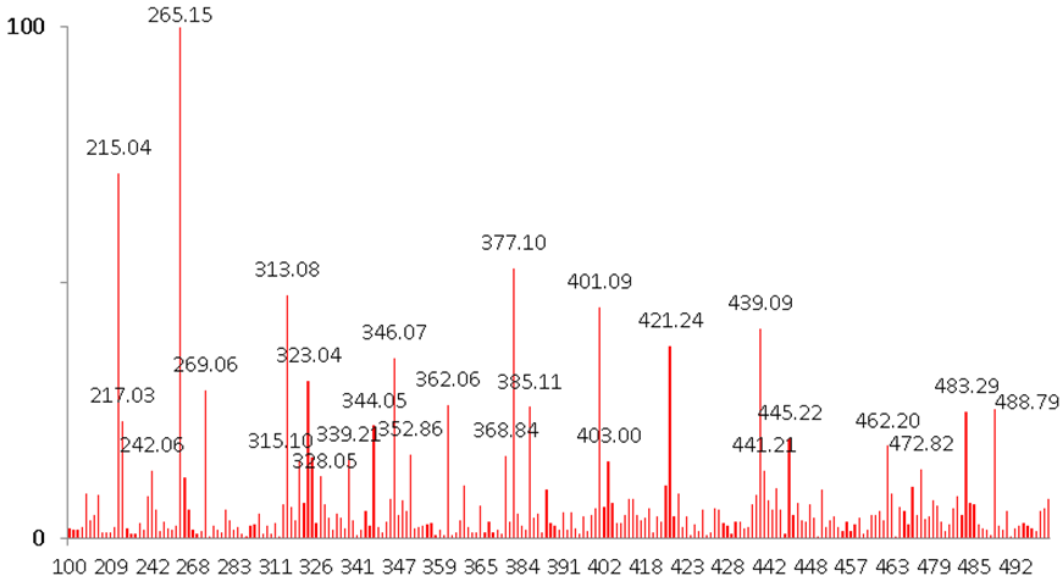


Figure 5.9. Metabolite profile of protonemal tissue from two extraction methods. Mass spectrometry of whole cell extract from an aqueous **(A)** and methanol based extraction **(B)**. The methanol extraction appears to have produced a more complex metabolite 'library' based on the number of peaks over 10% of the maximum peak. There were ~28 such peaks from the methanol extraction compared to the ~10 peaks from the aqueous extraction. Peaks relate to the m/z (mass to charge ratio) which is a proxy for mass.

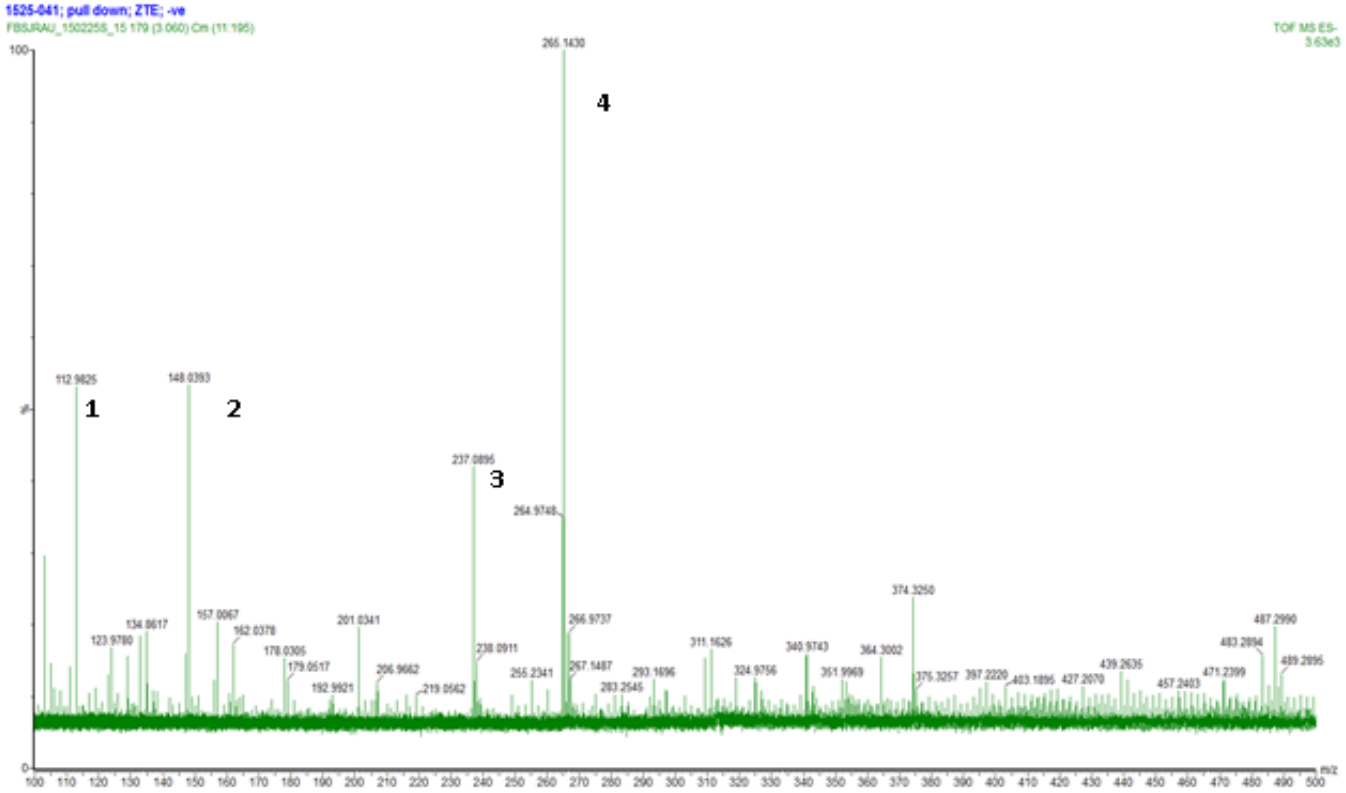


Figure 5.10. Enrichment of metabolites by protein affinity. Pull down of aqueous metabolite extraction with PpANR PAS-6-His-tagged protein bound to cobalt column. Mass spectrometry of putatively interacting metabolites shows four peaks. Peaks 3 and 4 are two of the most abundant compounds in total extract and so likely represent unspecific interactions. Peak 1 (~113m/z) is detectable in the total extract but shows relative enrichment. Peak 2 (~148m/z) is undetectable in the total extract and so is massively enriched and may be an interacting compound.

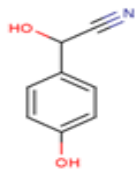
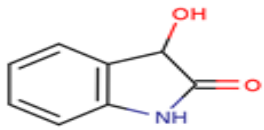
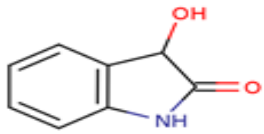
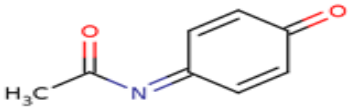
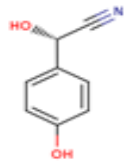
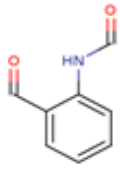
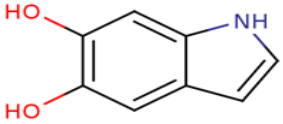
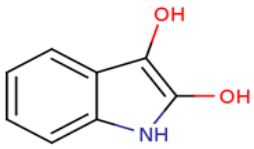
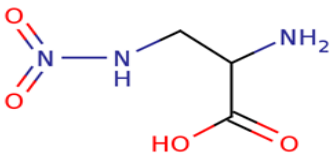
METLIN ID	MASS	Δppm	NAME	MS/MS	STRUCTURE
63628	[M-H]⁻ **m/z** **148.0404** M 149.0477	9	**4-Hydroxymandelonitrile** *Formula:* C8H7NO2 *CAS:*	NO	
64535	[M-H]⁻ **m/z** **148.0404** M 149.0477	9	**3-Hydroxyindolin-2-one** *Formula:* C8H7NO2 *CAS:* 61-71-2	NO	
68044	[M-H]⁻ **m/z** **148.0404** M 149.0477	9	**Gentianadine** *Formula:* C8H7NO2 *CAS:* 6790-32-5	NO	
578	[M-H]⁻ **m/z** **148.0404** M 149.0477	9	**N-Acetyl-p-benzoquinonimine** *Formula:* C8H7NO2 *CAS:* 50700-49-7		
63618	[M-H]⁻ **m/z** **148.0404** M 149.0477	9	**(S)-4-Hydroxymandelonitrile** *Formula:* C8H7NO2 *CAS:*	NO	
6090	[M-H]⁻ **m/z** **148.0404** M 149.0477	9	**2-Formylaminobenzaldehyde** *Formula:* C8H7NO2 *CAS:*	NO	
7009	[M-H]⁻ **m/z** **148.0404** M 149.0477	9	**5,6-Dihydroxyindole** *Formula:* C8H7NO2 *CAS:*	NO	
63545	[M-H]⁻ **m/z** **148.0404** M 149.0477	9	**2,3-Dihydroxyindole** *Formula:* C8H7NO2 *CAS:*	NO	
86243	[M-H]⁻ **m/z** **148.0364** M 149.0437	17	**(3-Nitroamino)alanine** *Formula:* C3H7N3O4 *CAS:* 58130-89-5	NO	

Figure 5.11. METLIN (Scripps) database search results with query of 148.04m/z and tolerance of +-30ppm. All but one compound is a variant of C8H7NO2. 3 contain an indole-related aromatic ring and 3 contain a benzene ring.

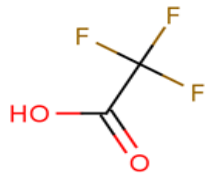
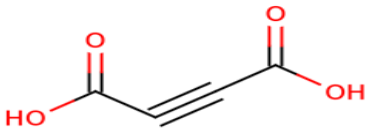
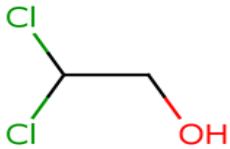
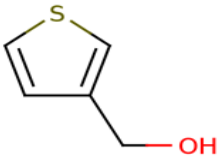
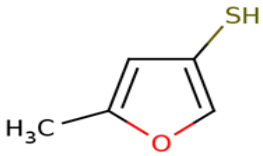
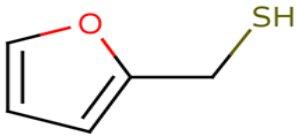
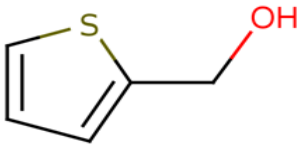
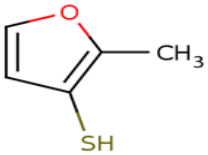
METLIN ID	MASS	Δppm	NAME	MS/MS	STRUCTURE
85337	$[M-H]^-$ **m/z** **112.9856** M 113.9929	49	**trifluoroacetic acid** *Formula:* C2HF3O2 *CAS:* 76-05-01	NO	
63238	$[M-H]^-$ **m/z** **112.9880** M 113.9953	71	**Acetylenedicarboxylate** *Formula:* C4H2O4 *CAS:* 142-45-0	NO	
2923	$[M-H]^-$ **m/z** **112.9566** M 113.9639	206	**Dichloroethanol** *Formula:* C2H4Cl2O *CAS:* 598-38-9	NO	
72897	$[M-H]^-$ **m/z** **112.9566** M 113.9639	206	**Bis(chloromethyl) ether** *Formula:* C2H4Cl2O *CAS:* 542-88-1	NO	
95889	$[M-H]^-$ **m/z** **113.0067** M 114.0139	235	**3-Thiophenemethanol** *Formula:* C5H6OS *CAS:* 71637-34-8	NO	
94347	$[M-H]^-$ **m/z** **113.0067** M 114.0139	235	**5-Methyl-3-furanthiol** *Formula:* C5H6OS *CAS:* 56078-97-8	NO	
88820	$[M-H]^-$ **m/z** **113.0067** M 114.0139	235	**2-Furanmethanethiol** *Formula:* C5H6OS *CAS:* 98-02-2	NO	
88889	$[M-H]^-$ **m/z** **113.0067** M 114.0139	235	**2-Thiophenemethanol** *Formula:* C5H6OS *CAS:* 636-72-6	NO	
91671	$[M-H]^-$ **m/z** **113.0067** M 114.0139	235	**2-Methyl-3-furanthiol** *Formula:* C5H6OS *CAS:* 28588-74-1	NO	

Figure 5.12. METLIN (Scripps) database search results with query of 112.98m/z and tolerance of +-300ppm. A large tolerance was required to match hits. Complex aromatic rings feature.

5.3.4 Circular Dichroism and ABA binding

Circular dichroism is an optical technique that can be used to probe molecular structural properties. For proteins, it produces well described spectral profiles that indicate the presence of α-helices and β-strands among other more complex structures and combinations. The purified PAS protein was analysed by CD both with and without (*apo*-PpANR PAS) ABA across 180-260nm. The PpANR PAS+ABA sample was the result of incubating protein and potential ligand together with excess ABA (10x molar excess (+-)ABA) before re-purifying the protein fractions from size exclusion chromatography so aiming to prevent excess and unbound ABA interfering with the results. The CD spectra for both samples were normalized for protein concentration and both showed very similar profiles except that a stronger signal was detected for the PpANR PAS+ABA sample which can indicate greater structural stability (Figure 5.13). The double minima at ~222nm and ~208nm are typical of a helix containing protein as expected in a typical PAS fold structure. The *apo*-PpANR PAS spectra showed that the purified protein was structured and had not suffered from degradation or any other structural issues while the stronger signal in the PpANR PAS+ABA sample suggested either an error in quantification of protein concentrations or an interaction between ABA and the PpANR PAS domain. To clarify whether an interaction had occurred, a thermal melt was carried out to test whether the PpANR PAS domain was binding ABA and therefore increasing its thermal stability. The thermal melt analysis allows the midpoint of the thermal denaturation curve (in linear phase) to be found. This is expected to be higher in a bound than an unbound protein due to greater structural stability being brought about by the interaction. This was not found to be case with PpANR PAS+ABA as this midpoint value did not increase with the addition of ABA as seen at both 208nm and 222nm (Figure 5.14). This suggested that the increase in signal seen in the +ABA sample in the CD spectra was unlikely due to any binding of the PpANR PAS domain and ABA but rather inaccuracies in protein concentration quantification.

5.3.5 FAD binding

Many PAS domains, such as the LOV subclass, bind flavins including flavin adenine dinucleotide (FAD) which can function as a redox sensor, such as in the bacterial FixL PAS domain. The onset of dehydration frequently alters redox potentials in cells (de Carvahlo, 2008) and so FAD was predicted as a plausible cofactor for the PpANR PAS domain; mediating ABA/stress signalling though detecting changes in redox potentials of dehydrating cells. Purified PpANR PAS was incubated with 10x molar excess FAD overnight before re-purifying the protein fractions from size exclusion chromatography to separate PpANR PAS and unbound FAD. The read-out from the AKTA UV sensor showed that the protein peak preceded the FAD peak (Figure 5.15). FAD has

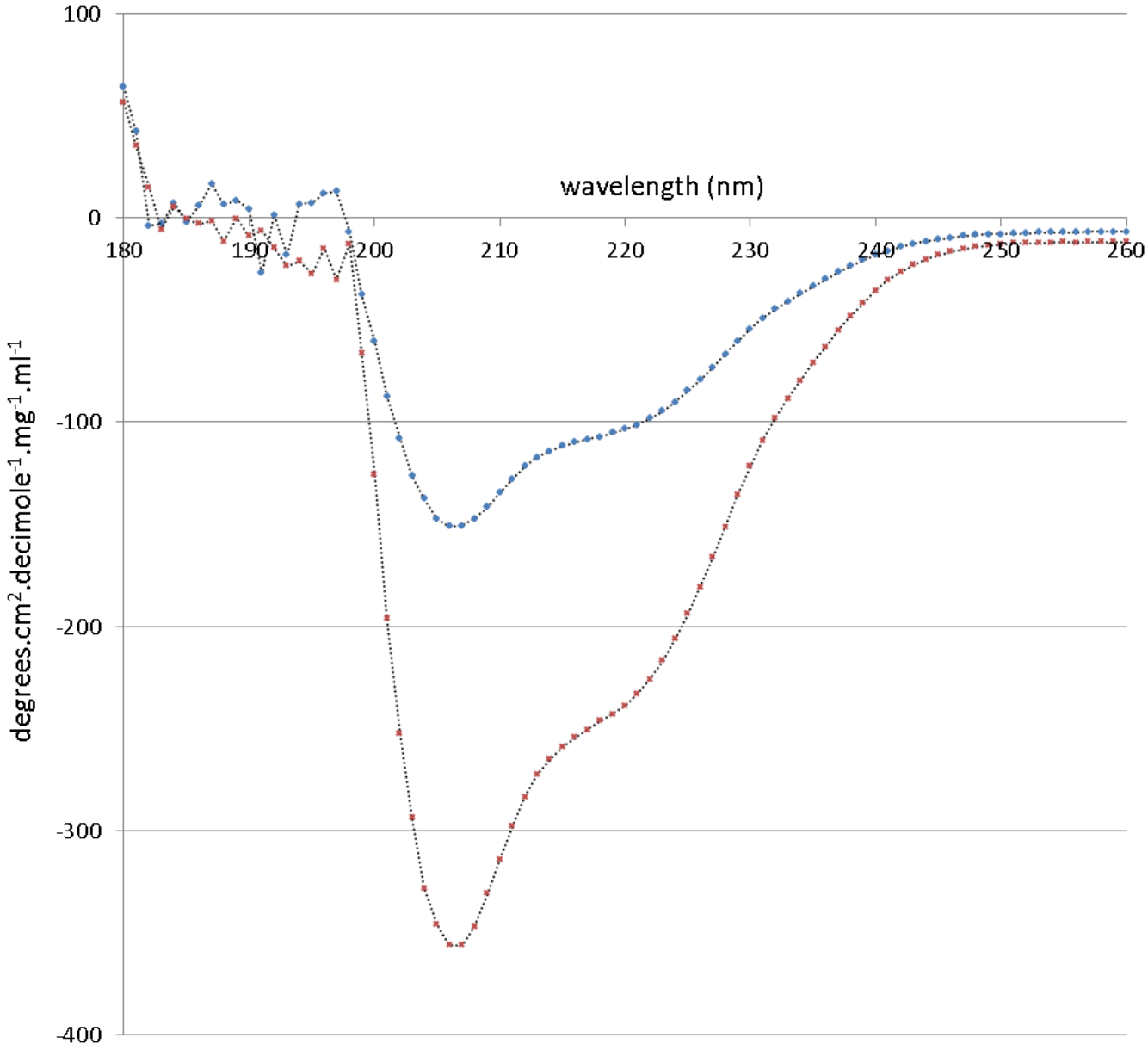


Figure 5.13. Purified PpANR PAS protein is structured. Circular dichroism (CD) spectra for *apo*-PpANR PAS (blue diamonds) and PpANR PAS+ABA (red crosses) across 180-260nm. CD values were normalized for protein concentrations however these quantifications may not have been accurate. The typical spectra minima of α-helices (208nm and 222nm) are clearly seen. Data are the mean of two technical repeats.

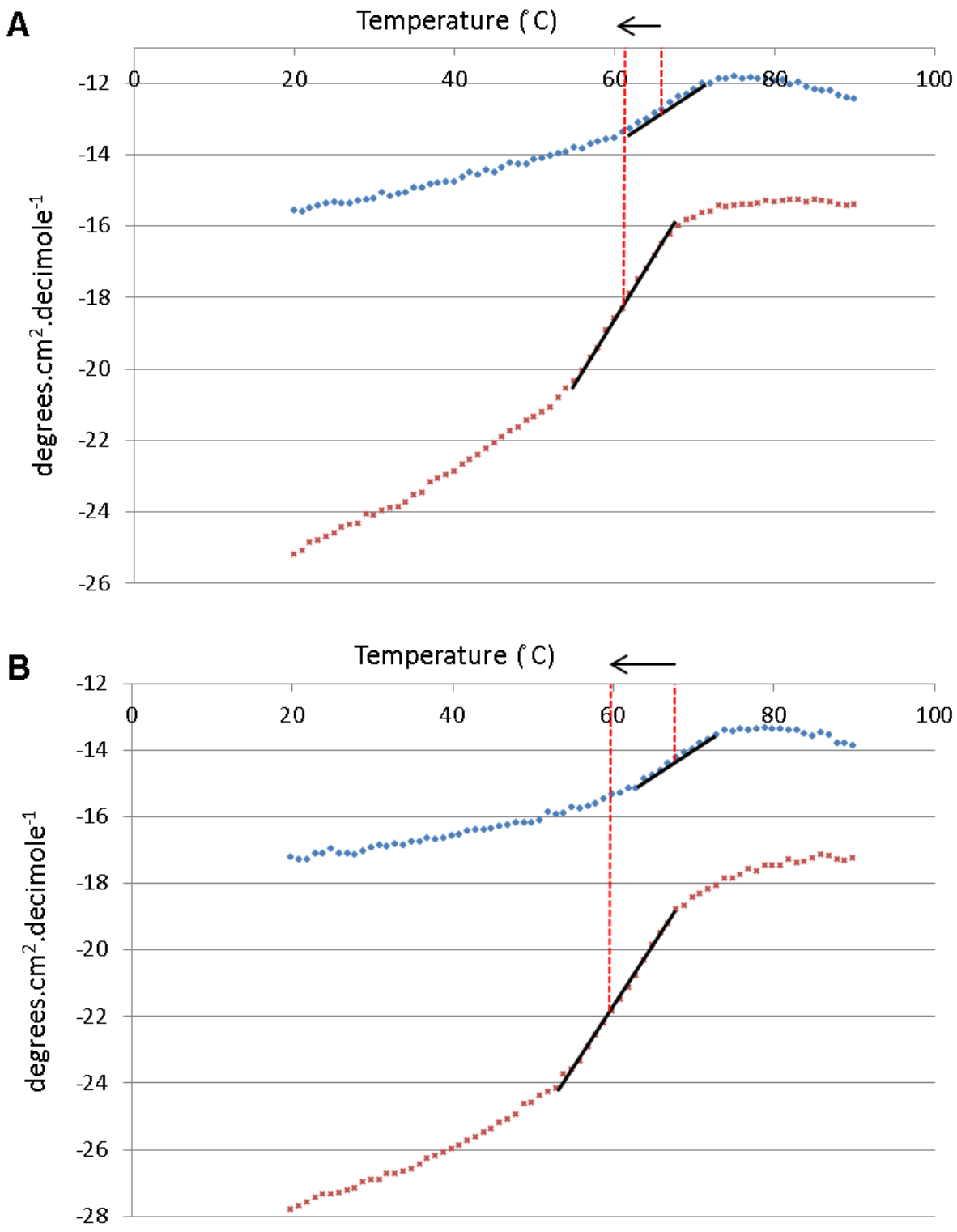


Figure 5.14. Thermal melt analysis reveals no PpANR PAS::ABA interaction. Circular dichroism (CD) thermal melt for *apo*-PpANR PAS (blue diamonds) and PpANR PAS+ABA (red crosses) across 20-90°C. No thermal stability is found in the PpANR PAS+ABA sample as seen at both 208nm **(A)** and 222nm **(B)**. Stability is implied by an increase in the temperature of the midpoint (red dashed lines) of the linear phase (black lines), however these decrease rather than increase with the inclusion of ABA.

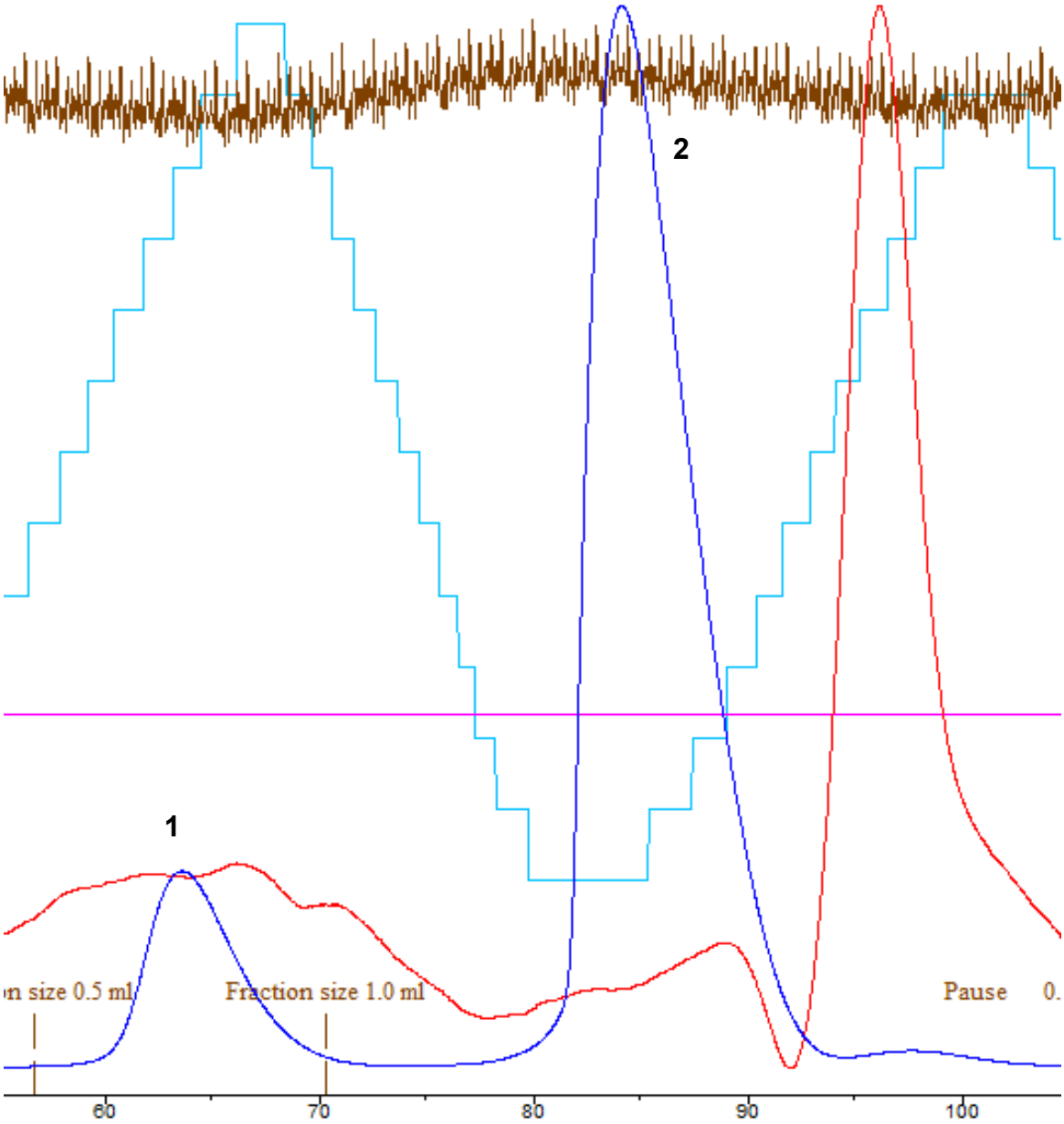


Figure 5.15. Size exclusion chromatography of PpANR PAS protein incubated with FAD. Akta prime output of size exclusion chromatography purification of PpANR PAS protein following incubation with excess FAD with the UV trace shown (dark blue curve). After approximately 100ml buffer has run through (60mins) following injection of protein sample the PpANR PAS protein is found coming off the column in a single peak and these fractions are collected and bulked **(1)**. A second UV peak is found as the non-bound FAD comes off **(2)** and these fractions were also collected. The red curve is the conductivity with the peak indicating the NaCl from the buffer coming off.

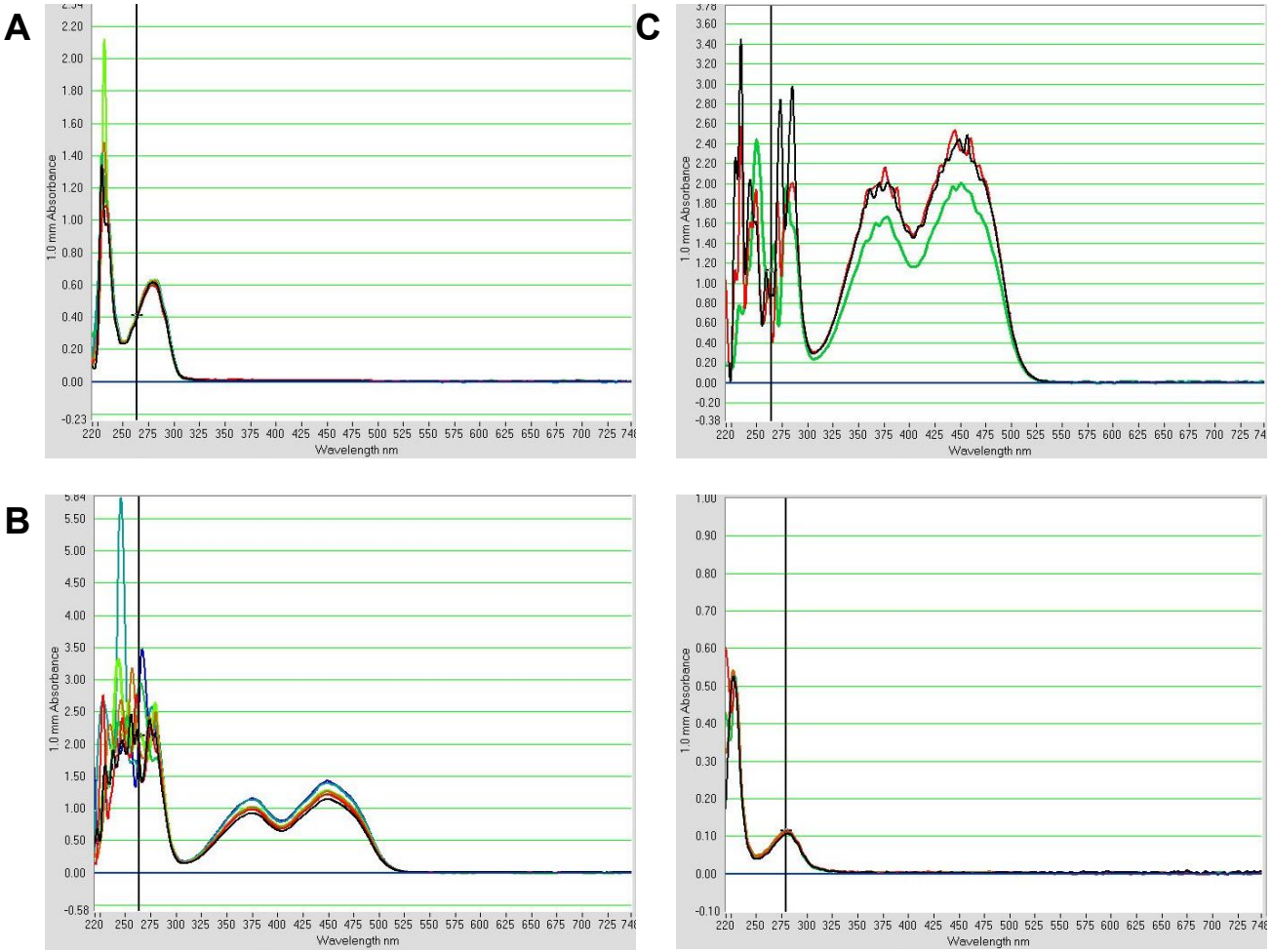


Figure 5.16. Spectrophotometry reveals no PpANR PAS::FAD interaction. Nanodrop spectrophotometer profiles for variable numbers of technical repeats for **A)** 8mg/ml PAS protein showing typical protein peak at 280nm. **B)** 2mM FAD showing typical absorption maxima at 375 and 450nm and 'noise'. **C)** Mix of 8mg/ml PAS and 2mM FAD prior to gel filtration shows clear presence of FAD. **D)** Protein fractions recovered from gel filtration clearly showing protein peak at 280nm but no FAD.

absorbance maxima at 450nm and 375nm and the protein fractions were tested for the presence of these indicating a PpANR PAS::FAD interaction alongside standards (Figure 5.16). The mixture prior to the gel filtration indicated the clear presence of the 375 and 450nm FAD absorption peaks however the 280nm protein peak was obscured by 'noise' produced from the FAD solution around this wavelength. Crucially, however, the protein fraction recovered by gel filtration contained no FAD signal suggesting no interaction occurred.

5.3.6 Crystal screens and optimisation

Since no interactions between the PpANR PAS domain and two potential ligands were found (ABA and FAD), structural data on the PpANR PAS domain became all the more valuable to help make predictions about both the protein's role and mode of action. The purified PpANR PAS protein was concentrated to ~10mg/ml and used in 6 crystal screens (96 conditions each) incubated at 20°C. Purified PpANR PAS was amenable to crystallisation with many conditions leading to the formation of small needle crystals of various shapes and sizes as seen by some of the more promising conditions (Figure 5.17A-C). One condition, however, resulted in the formation of a large plate crystal (Figure 5.17D) and this condition was further optimized to maximize the crystal size for X-ray crystallography. An optimisation array in a 24 well 1ml hanging dr op plate resulted in larger crystals forming after ~1 week. Four wells produced large plate clusters which were picked and cryoprotected in glycerol prior to liquid nitrogen freezing and X-ray crystallography. These conditions included those with the highest precipitant concentrations (2.6M Li_2SO_4) across all pHs. This indicated that the pH was not as important as using the highest concentration of precipitant although the presence of plate clusters at pH 6.5 and at 2M Li_2SO_4 indicated that a lower pH was also marginally favoured.

5.3.7 X-ray crystallography

5.3.7.1 General structure features

The crystal structure of the PpANR PAS domain was determined by molecular replacement using the PDB structure 3LYX, which I-TASSER found to be the best predictive structure (Figures 5.2 and 5.3). The structure was found to be a homodimer (the two parts referred to as chain A and chain B) in the C2 space group with a cell volume of 505,920 $Å^3$. The structure was solved to 1.7Å with 98.4% total completeness and 87.4% completeness at the lowest resolution representing a high quality, high resolution structure (Table 5.1). A single disulphide bond is formed between residues C^4 and C^{98} covalently binding the first and last β-strand of each chain. There are 182 water molecules, two SO_4 ions, a single PEG molecule and a glycerol molecule found

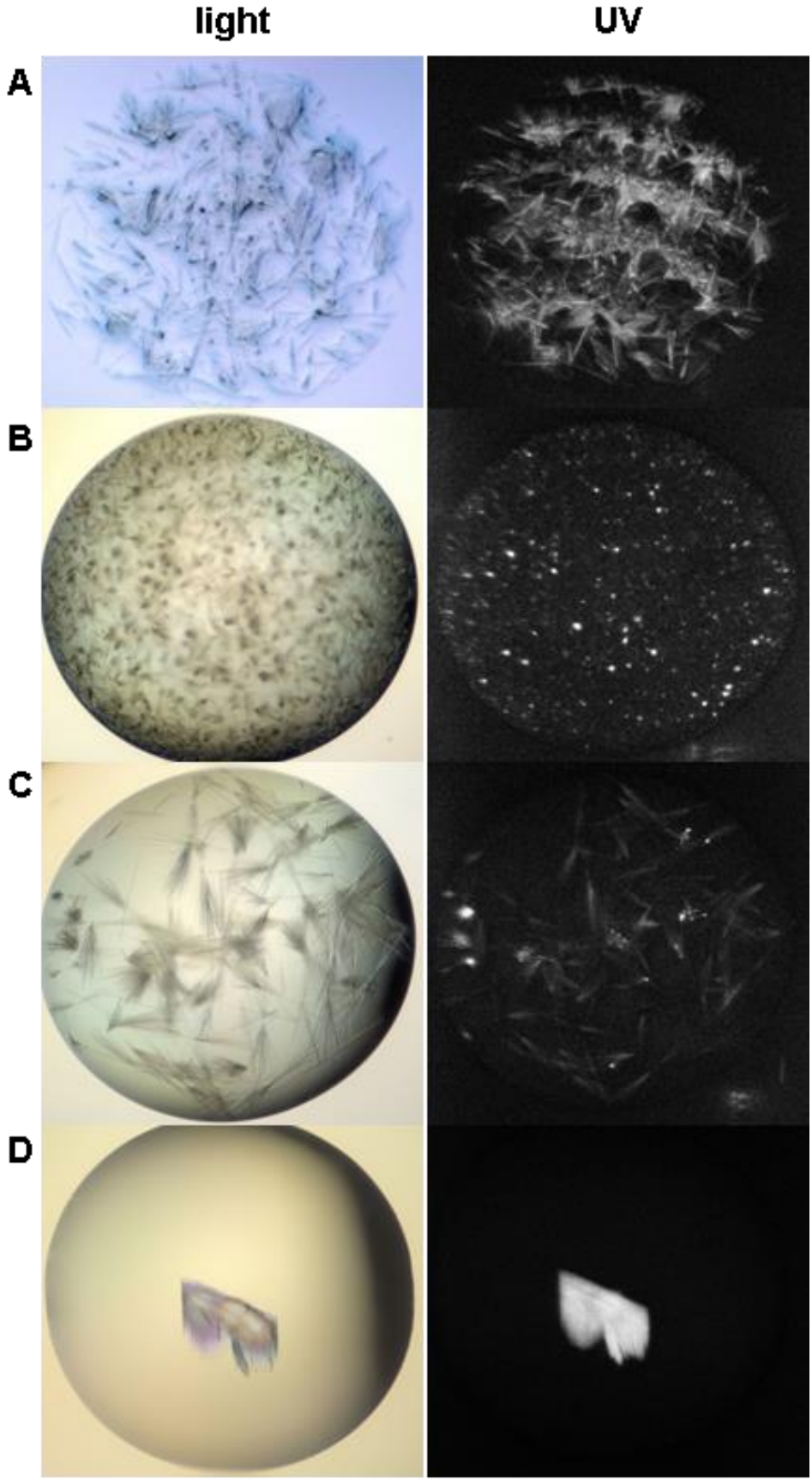


Figure 5.17. Crystal screening identifies favourable conditions. Light (left panels) and UV (right panels) camera images taken of promising PpANR PAS protein crystal screen conditions. **A)** Wizard 4 tube 39: 0.1M imidazole, pH 6.5 (HCl), 20% PEG8000, 3%MPDw/v. This condition produced some of the largest needle crystals which showed up strongly under UV. **B)** Wizard 2 tube 35: 0.1M NaAc, pH 4.5 (Acetic acid), 0.8M $NaH_2PO_4 \cdot H_2O$/1.2M K_2HPO_4. **C)** SaltRX tube 35: 0.1M BIS-TRIS propane, pH7, 1.4M sodium malonate, pH7. **D)** Wizard 3&4 E8: 0.1M Tris, pH8.5 (HCl), 2M LiSO4, 2% PEG400. This condition was further optimised.

as part of the structure. The domains in each homodimer are in a 180° rotation relative to each other such that the F_α helices are on the same side of the cell (Figure 5.18A). This dimerisation orientation, as well as stacking effects, introduces a number of asymmetries between the two domains in each homodimer. This can be visualised by

Table 5.2. Data collection, processing and refinement statistics for the PpANR PAS domain.

	PpANR PAS
Source	Diamond Light Source
Beamline	I24
Wavelength (Å)	0.97
Resolution range (Å) *	40.17–1.70 (1.74–1.70)
Space group	*C*2
Unit-cell parameters (Å)	*a*=85.2,*b*=64.9,*c*=51.9, α=90.0,β=121.4, γ=90.0
No. of observed reflections	98517
No. of unique reflections	26146
Redundancy	3.8 (2.8)
Completeness (%) *	98.4 (87.4)
$< I/\sigma(I) >$*	12.9 (2.4)
R_{merge} (%)§*	6.1 (45.2)
R_{pim} (%)¥*	3.5 (30.5)
Resolution range for refinement (Å)	40.17–1.70
R factor (%)	16.1
R_{free} (%)†	19.9
No. of protein non-H atoms	1960
No. of water molecules	182
No. of sulphate ions	2
No. of glycerol molecules	1
No. of PEG molecules	1
R.m.s.d bond lengths (Å) ξ	0.014
R.m.s.d bond angles (˚) ξ	1.613
Average overall *B* factor (Å^2)	
Protein (Chain A/Chain B)	18.6/25.4
Water	30.6
Sulfate ions	61
Ramachandran analysis, the percentage of residues in the regions of plot (%) ‡	
Favored region	98.7
Outliers	0

*Values given in parentheses correspond to those in the outermost shell of the resolution range.
§ $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl)-[I(hkl)]| / \sum_{hkl} \sum I_i(hkl)$
¥ $R_{pim} = \sum_{hkl} \{1/[N(hkl)-1]\}^{551/2} \sum_i |I_i(hkl) - [I(hkl)]| / \sum_{hkl} \sum_i I_i(hkl)$
† R_{free} was calculated with 5% of the reflections set aside randomly.
ξ Based on the ideal geometry values of Engh and Huber, 1991.
‡ Ramachandran analysis using the program MolProbity (Chen *et al*., 2010).

the b-factors scores of the residues whereby greater values (visualised as thicker regions on the b-factor ‘putty’ image) represent regions of greater flexibility. Chain B

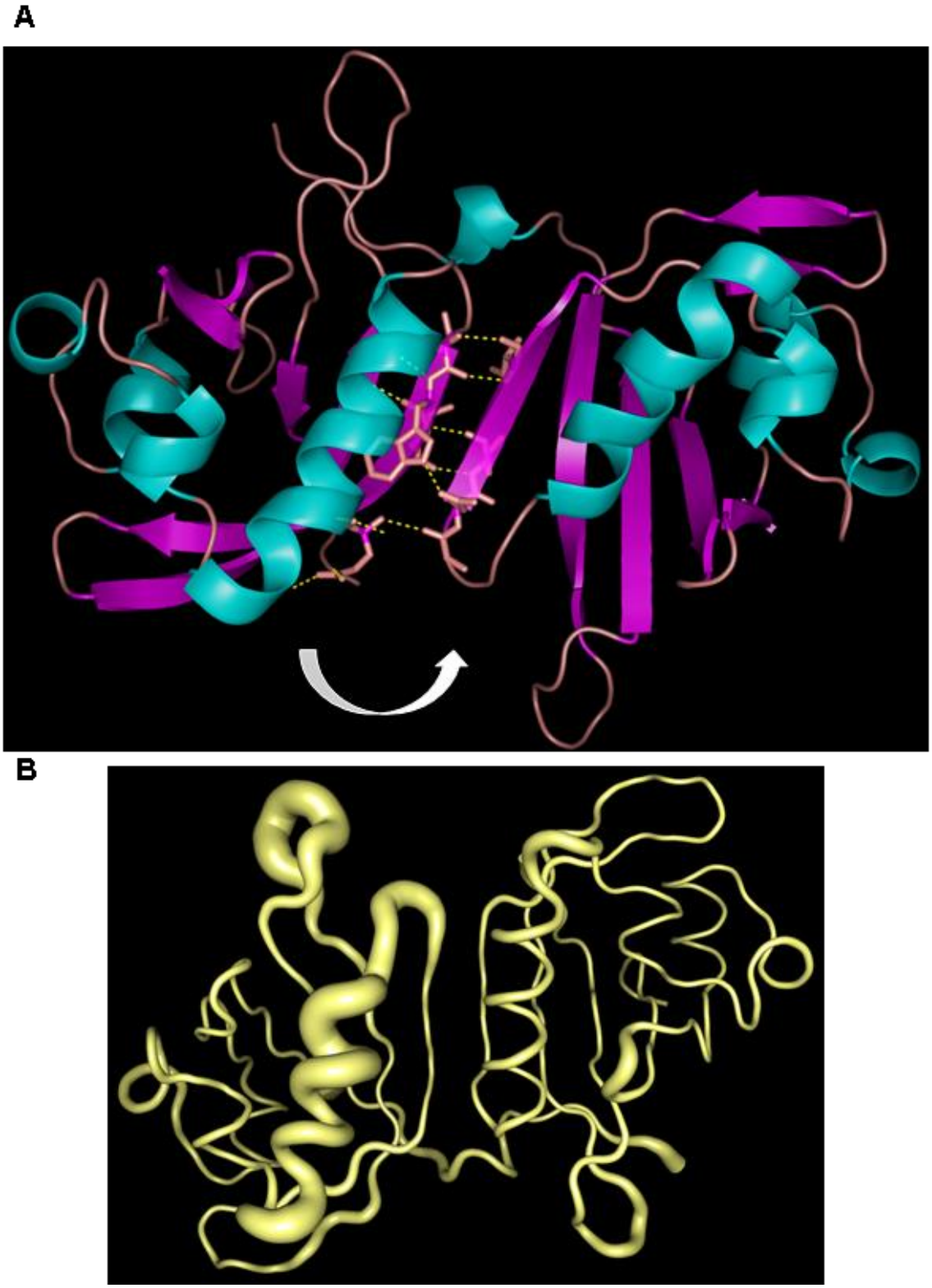


Figure 5.18. Crystallography reveals the PpANR PAS domain to be a homodimer with canonical PAS fold structure. A) Cartoon representation of the PpANR PAS homodimer. Chain B is on the left and chain A on the right with an arrow describing the rotation in symmetry between the two. The hydrogen bonds mediating the dimerisation interface between the two G_β strands are shown as dotted lines with the interacting residues shown. **B)** B-factor visualisation of the PpANR PAS homodimer. Chain B is on the left and chain A on the right. The thicker regions represent those with greater flexibility. The H_β-I_β loop and the F_α helix on chain B show the greatest flexibility.

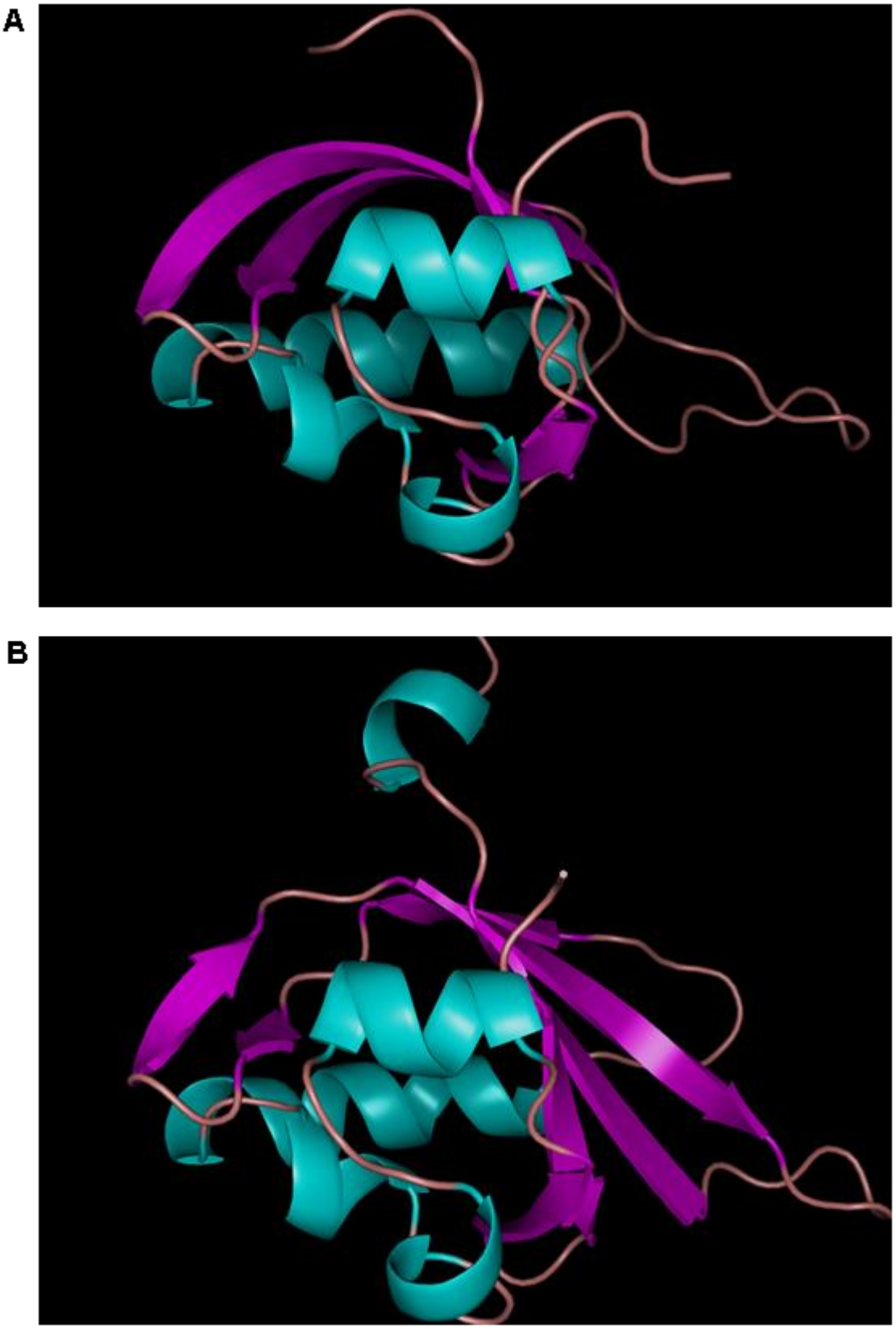


Figure 5.19. Asymmetries are found between the two chains. Comparison of the two chains from the same viewpoint highlights differences in the predicted secondary structures. **A)** Chain A has fewer β-strands than chain B forming the typical 5 antiparallel β-sheet. **B)** Chain B has two 'extra' b-strands in which 2 of the longer b-strands (G_β and H_β) are broken up into two smaller β-strands each. This doesn't significantly affect the overall tertiary structure. The main difference is the extension of the C-terminus in chain B.

has a higher b-factor average (25.4) than chain A (18.6). Specifically, this reveals two features on chain B, the H_β-I_β loop and the F_α helix, as being particularly flexible; a feature not mirrored on chain A (Figure 5.18B). The F_α helix, a dominant structural freature of PAS domains, is known to be variable and flexible in many PAS domains often regulating the internal cavity size and therefore specificity of potential ligands. This role in the PpANR PAS domain is discussed in more detail. Another interesting feature of this homodimer is that the asymmetries result in different secondary structure predictions despite the same amino acid sequences. Chain B has 2 'extra' β-strands as compared to chain A which changes the predicted composition of the β-sheet. However, the overall structures remain very similar and the secondary structure predictions are simply a guide for understanding the overall structure. The main structural difference is the extension of the C-terminus in chain B that interacts with chain A (Figure 5.19).

5.3.7.2 Dimerisation interfaces

The dimerisation interaction appears to be primarily mediated by multiple hydrogen bonds along the antiparallel G_β strands (Figure 5.18A). Analysis with the PISA program reveals that this dimerisation interface is made up of 6 hydrogen bonds between 4 residues from each domain (Figure 5.20A, B). As is common with other PAS dimers, this interaction involves a number of hydrophobic residues. This interface is also augmented by water mediated hydrogen bonds between G^{59} on chain A and Q^{65} on chain B. Interestingly, G^{59} is invariant among those PAS domains most related to the PpANR PAS suggesting it to be a crucial residue. While other interactions are mediated by G^{59}, it is involved in this dimerisation interaction in both chains suggesting this may be its primary role.

Another potential dimerisation interface is seen between the C-terminal flanking region of chain B and the H_β strand and the H_β-I_β loop of chain A with 5 hydrogen bonds identified (Figure 5.20 C, D). A potential water-mediated bond between N^{111} on chain B and S^{93} on chain A is also found. The C-terminal flanking region of chain A failed to be resolved crystallographically, suggesting it possessed greater flexibility. Analysis of the crystal packing reveals that the chain A flanking region appears to project into a solvent channel while that in chain B is 'sandwiched' between the chain B of a neighbouring homodimer. This suggests that this interface may either be an artefact on the one hand or disrupted on the other. At the very least, the interactions between the antiparallel G_β strands suggest a mechanism by which the PpANR PAS domains may form homodimers *in vivo*; potentially bringing the whole PpANR protein into a dimer. The asymmetry seen in the b-factor visualisation may be partially explained as the chain B

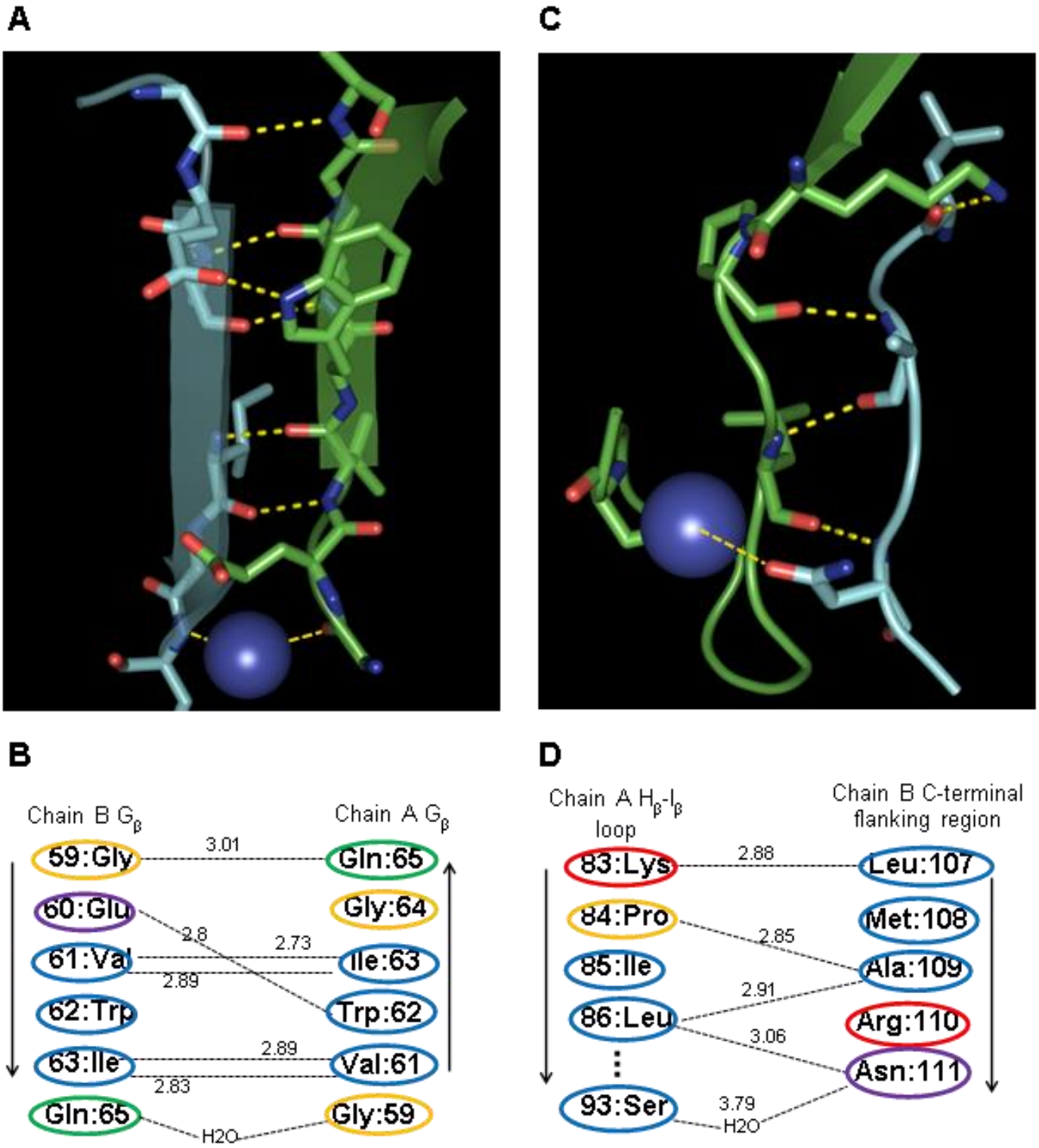


Figure 5.20. Analysis of dimerisation interfaces. A) Cartoon representation of the antiparallel G_β strand-mediated dimerisation interface with detailed schematic showing interacting residues and their relative distances **(B)**. Water (blue spheres) mediated hydrogen bonds between G^{59} on chain A and Q^{65} on chain B also contributes to interface but bond distance not calculated. **C)** Cartoon representation of interactions between chain A (H_β and H_β-I_β loop) and chain B (C-terminal flanking region) with detailed schematic showing interacting residues and their relative distances **(D)**. Another potential water-mediated hydrogen bond is shown between N^{111} on chain B and S^{93} on chain A. Cartoons and schematics are shown in relative orientations with residue numbers shown. Hydrogen bonds are indicated by dashed lines with the distance indicated in Ås as calculated by PISA. Residues are colour coded by side chain properties: green=polar uncharged, yellow=special cases, blue=hydrophobic, purple=negative charge, red=positive charge. Both interfaces feature hydrophobic residues.

H_β-I_β loop does not have the reciprocal dimerisation interaction with the C-terminal flanking sequence as seen in chain A, suggesting it lacks the same stabilisation.

5.3.7.3 Highly conserved residues

Some of the most highly conserved residues among those PAS domains most closely related to the PpANR PAS domain have been identified (see chapter 4). Two of these, a tryptophan at position 16 (W^{16}) and a glycine at position 33 (G^{33}), are found to mediate hydrogen bond bridges which enable a tertiary loop region to form (Figure 5.21A). This bends the secondary structure such that the B_β strand and the D_α helix run antiparallel with the turn occurring round the C_α helix. The interacting partners of W^{16} and G^{33} are also found to be highly conserved.

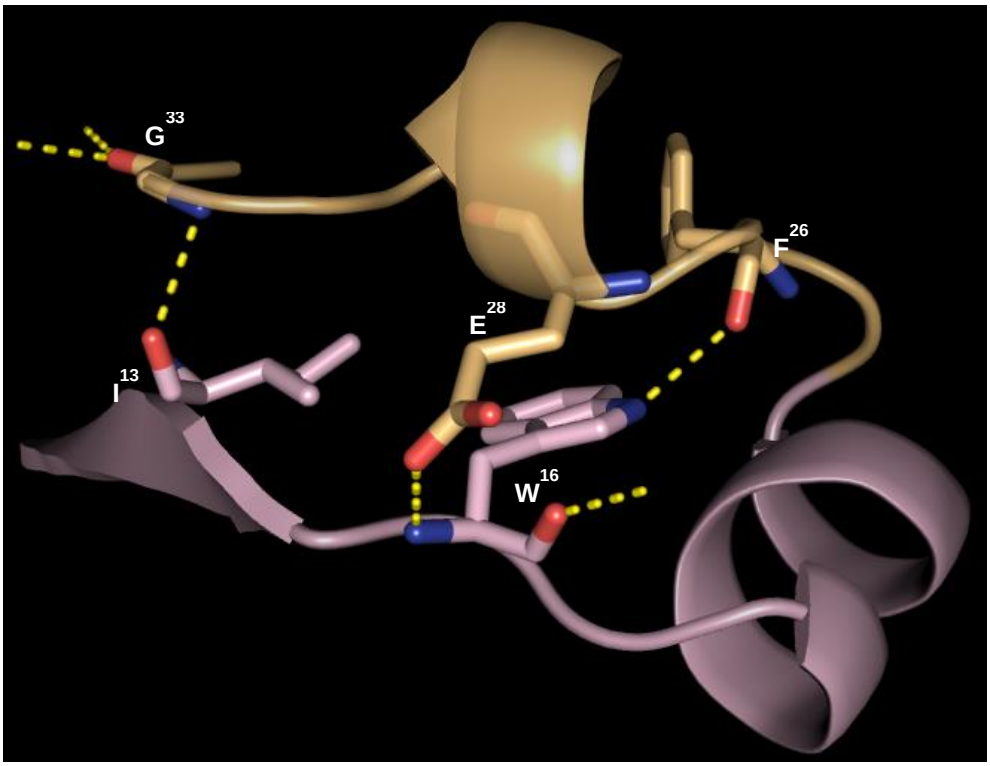
Another pair of highly conserved residues, a glutamic acid (E^{30}) and valine (V^{31}), mediates a network of interactions (Figure 5.21B). E^{30} interacts with the G_β-H_β loop through a two hydrogen bond connection with R^{71}. This interaction is possibly important for twisting the β-sheet to create the shell-like structure within which most PAS internal cavities are found. Its neighbour, V^{31}, interacts with R^{34} through two hydrogen bonds which bends the D_α-E_α loop region.

As already seen, G^{59} involved in the β-sheet mediated dimerisation is the only invariant residue among the PAS domains most closely related to the PpANR PAS domain (Chapter 4). In chain A it also appears to be important for interacting with R^{55} and V^{56} in the F_α helix through one hydrogen bond each and so enabling the bend in the F_α-G_β loop. This bend is potentially important for aligning the F_α in the correct orientation to regulate the internal cavity and the interactions are conserved with the structurally similar bacterial PAS domain (3LYX). It does not however seem to mediate this in chain B.

5.3.8 Structural Comparisons

Solving of the PpANR PAS crystal structure opened up the possibility of accurate structural comparisons to identify conservation of functionally important features between PAS domains. In order to find if the PpANR PAS domain has any clear structural relationships to other groups of PAS domains, a global analysis was carried out based on that by Henry and Crosson, 2011 but using a modified and updated set of PDB structures and including both domains in the PpANR PAS homodimer. The analysis included only those residues making up the core PAS fold (Appendix 5.1). The Henry and Crosson study, and others (Jaiswal *et al*., 2010), have previously uncovered clusters of PAS domains grouped largely by the cofactors they bound representing structural relatedness despite distant phylogenetic distance. After collecting a large array of PAS-related PDB structures, those that did not have sufficient similarity (as

A



B

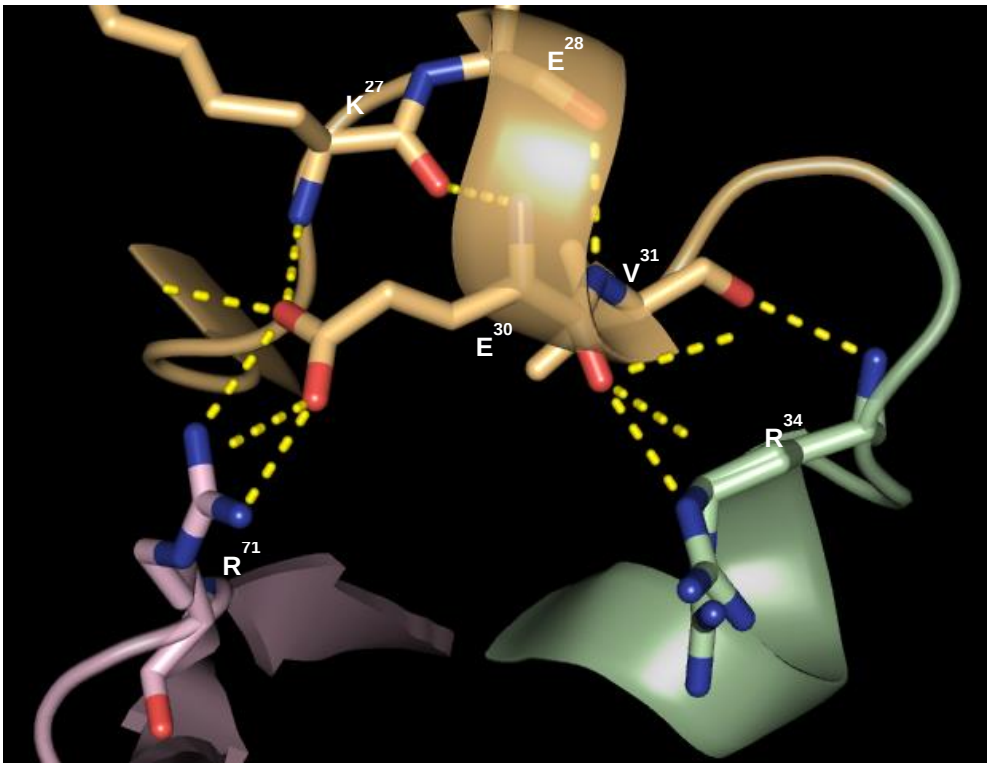


Figure 5.21. Analysis of highly conserved and structurally important residues. A number of highly conserved residues among PAS domains were analysed in more detail to understand their putative roles in conferring tertiary structure through intra-chain bonds. **A)** Interactions involving the W^{16} and G^{33} on chain A. The primary interaction is with F^{26} bridging the tertiary loop region shown but a secondary interaction found on chain A (but not chain B) with E^{28} likely strengthens this. G^{33} interacts with I^{13} which also reinforces this tertiary loop. **B)** Interactions involving the highly conserved E^{30} and V^{31} mediate a range interactions. E^{30} interacts with K^{27} through two hydrogen bonds but more importantly it mediates an interaction with the G_{β}-H_{β} loop through two hydrogen bonds with R^{71} possibly twisting the β-sheet enabling the internal cavity to form.

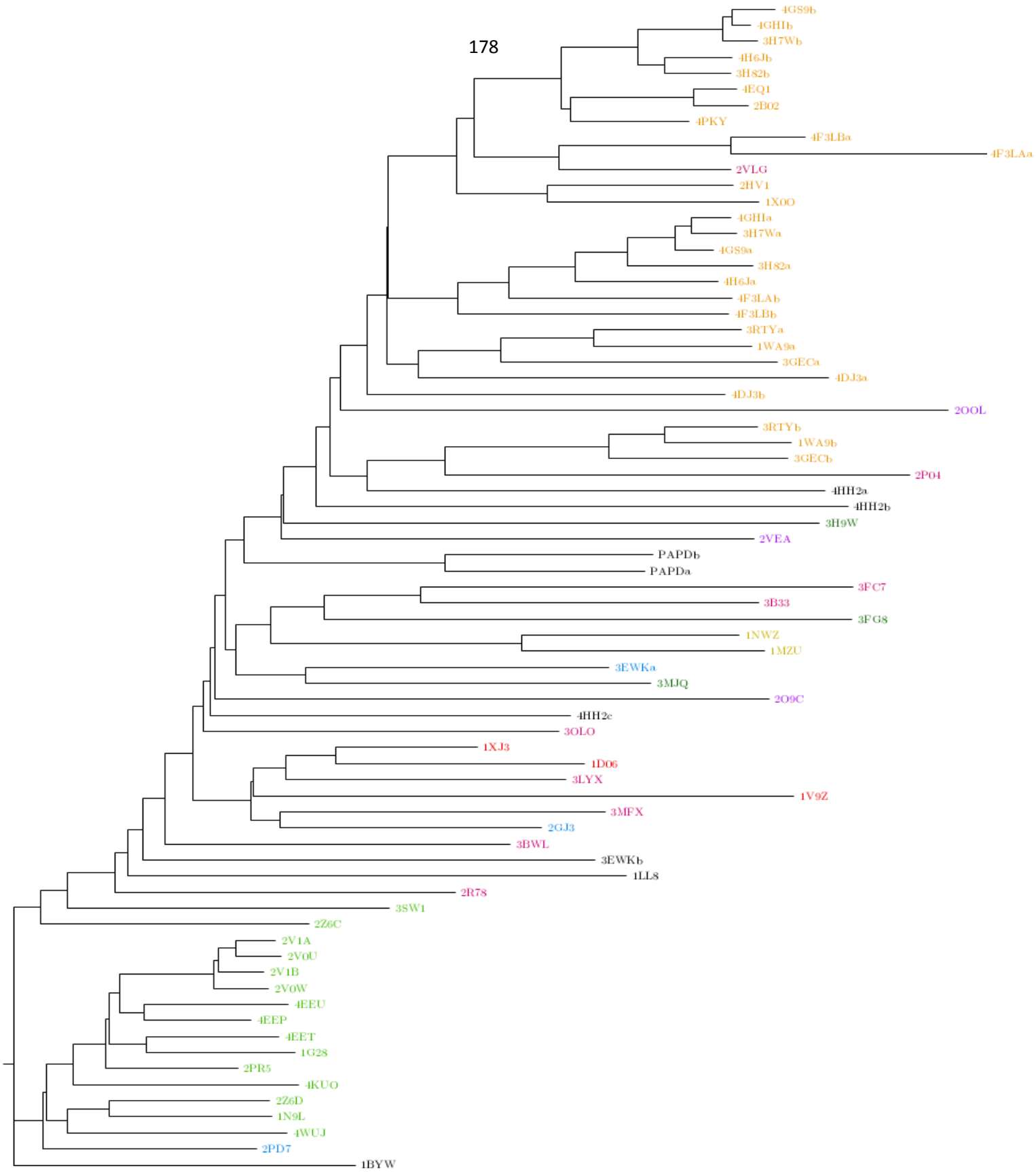


Figure 5.22. Structural similarity analysis of PAS structures reveals clusters of phylogenetically and functionally related structures. Structural similarity tree of PAS domains based on Henry and Crosson, 2011 including both chains of the PpANR PAS homodimer. Initial alignments using Clustal were carried out and the sequences trimmed to include the core PAS fold. The sequence alignment was then structurally aligned in STAMP from which a distance matrix of root-mean-square-deviations was clustered using the FITCH program (Phylip). A large cluster of FMN binding LOV structures are recovered (light green) and included is the fungal FMN/FAD dual binding VIVID structure (2PD7). Another large cluster of predominantly bHLH-PAS domains (orange) is recovered. Two bacterial clusters are recovered which contain a variety of PAS domains. One contains all three haem binding (red) PAS domains alongside a FAD binding redox sensor (blue) and uncharacterised histidine kinase PAS domains (pink). The other cluster contains the two PYP (yellow) PAS domains alongside two from nucleotide cyclases (dark green), 1 redox sensor (blue) and two histidine kinases (pink). The PpANR PAS domain structures cluster together but not with any other PAS domains. Also highlighted are the bacteriophytochromes (purple).

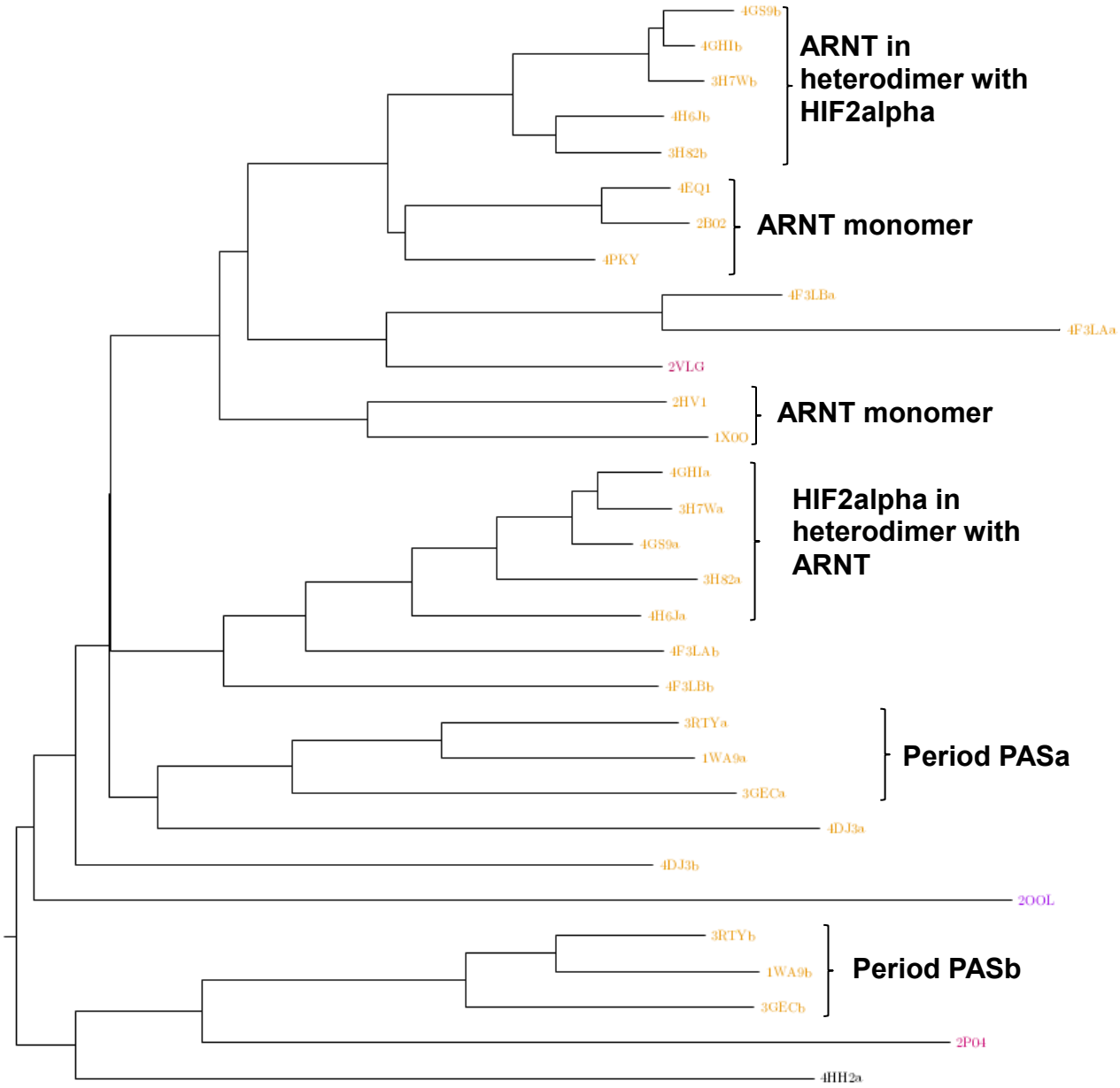


Figure 5.23. PAS structures as monomers or dimers have distinct structures. Subset of structural relatedness tree showing the bHLH-PAS PAS structures. The structures are recovered based on not only the protein they are found in, but whether they are PASa or PASb. A few non-bHLH-PAS structures were recovered in this grouping (non-orange – see figure 5.22). Despite identical sequences, this analysis recovers the ARNT PASb monomers and heterodimers separately. This suggests the interactions of dimerization affect tertiary structure in a measurable way.

flagged by the stamp program) with the PpANR PAS domain were discarded prior to the final tree (distance matrix of root-mean-square-deviation values clustered using the Fitch-Margoliash method) which predominantly included all extracytoplasmic structures used by Henry and Crosson (Figure 5.22).

This analysis was in general agreement with previous analyses (Henry and Crosson, 2001 and Jaiswal *et al*., 2010) although a few notable differences were found. Many more PAS structures from bHLH-PAS TFs were used which clustered not only by which gene they came from but, in the case of ARNT, whether they were in a heterodimer or not (Figure 5.23). This pattern has not been noted before and suggests that the dimerisation of PAS domains can induce measureable structural changes as these structures share 100% sequence identity. This also highlighted the sensitivity of this analysis. This analysis failed to group FAD binding structures as in previous analyses in which they were found to be related to FMN binding LOV domains. The only FAD binding structure found related to FMN binding domains was the fungal VIVID PAS domain (2PD7): a dual binder of FAD and FMN (Schwerdtfeger and Linden, 2003). A clear FMN binding cluster was, however, recovered which included LOV domains from phototropins as well as bacterial photoreceptors and VIVID. The analysis recovered two mixed clusters of bacterial PAS domains which included those structures with the highest pairwise similarities to the PpANR PAS domain. The first contained all three haem binding domains (1XJ3, 1D06, 1V9Z) alongside the FAD binding domain from NifL (2GJ3) and two histidine kinase (HK) PAS domains (3LYX and 3MFX); the former of which was used to solve the PpANR PAS structure by molecular replacement. The second cluster contains the two PYP domains (1NWZ and 1MZU) as well as domains from two nucleotide cyclases (3FG8 and 3MJQ), two HKs (3FC7 and 3B33) and the FAD binding MmoS (3EWKa). A number of other bacterial PAS domains fail to resolve into clusters and this is also the case for the two PpANR PAS chains which do cluster together and show the highest pairwise similarity as would be expected.

The failure of many bacterial PAS domains to cluster into clear groups, as with the PpANR PAS domain, possibly reflects a weakness of this approach. These structures come from a very diverse range of organisms including a number of bacteria which have themselves likely been separated for long evolutionary periods. The sampling rate across this diversity is far less dense than for sequence data and so the failure to recover clusters may reflect this. The recovery of a cluster of mammalian bHLH-PAS structures for example is in large due the relatively recent origins of the gene family and the gene duplications therein coupled with high sampling rate given the intense research into these proteins' PAS domain structures. While constituting a more

phylogenetically diverse cluster, this is likely also the case for the LOV FMN binding structures. Looking at the recovered relationships for these PAS structures we can see the strength of this approach to discern between family members and dimerisation states (Figure 5.23). While all ARNT structures derive from the same human protein sequence, the structural analysis is able to separate those in heterodimer complexes and those in homodimers. The LOV domains that make up the FMN binding structural cluster are clearly phylogenetically closely related, the majority being from plants, but the structural approach performed better at grouping the more distantly related structures as it clustered plant, fungal and bacterial structures together.

The finding of many bacterial structures having the highest pairwise similarity with the PpANR PAS domain however, is in support with the phylogenetic findings that the PAS domains in MAP3Ks, including the PpANR PAS domain, likely originated from algal PAS containing HKs; themselves originating from bacterial two-component regulatory systems. What is clear from this analysis – and in agreement with phylogenetic analyses - is that the PpANR PAS domain does not appear to be structurally related to light sensing domains in plants. The two PAS structures with highest pairwise similarity to both PpANR PAS chains are from an uncharacterised HK (3LYX) and the haem binding FixL (1XJ3); both from bacteria. Comparisons with a number of bacterial PAS structures and their modes of action are analysed in more detail.

5.3.8.1 Structures with the highest pairwise similarity to the PpANR PAS domain

In the structural analysis, the two chains of the PpANR PAS protein structure are found to be most similar to each other, however they clearly show differences as seen by some of the asymmetric features between them and they have a root-mean-square-deviation (RMSD) of 0.71. The RMSD is a measure of structural similarity which measures the differences in location of relative residues (based on the alignments) and as such a lower number represents less deviations between the two structures so a greater similarity. In contrast the RMSD between two human ARNT structures (3H7Wb and 4GHIb) in heterodimers with HIF2a is 0.1 representing a near perfect likeness.

The RMSD between the closely related bHLH-PAS PASb domains from human ARNT (4GHIb) and Drosophila Period (3GECb) is 1.63. The PpANR PAS domain is structurally most similar to a diverse set of PAS structures and these include: the haem binding FixL (1XJ3), an uncharacterised PAS from a bacterial HK (3LYX), a DNA binding protein (4HH2c), a PYP protein (1NWZ) and the FAD binding NifL (2GJ3) (Table 5.2). By computational structural comparisons alone, almost all known PAS ligands could be candidates based on these similar structures and so more detailed analysis is required.

Table 5.3. Key statistics and features of PAS domains showing the highest pairwise structural similarity to the PpANR PAS domain.

PDB	Gene name	PAS domain	PAS core length (a.a)	Fα helix length (a.a)	Cavity volume (Å3)	Interaction	RMSD to the PpANR PAS domain (chainA/chainB)
-	PpANR	-	96	16	193-231	?	
3LYX	CPS_1291	-	99	16	799-859	FAD?	1.21/1.27
2GJ3	NifL	PASa	99	14	529-548	FAD	1.33/1.47
1NWZ	PYP	-	96	12	185	4-HCA	1.31/1.34
1XJ3	FixL	PASb	100	17	965	Haem	1.2/1.22
4HH2	PpsR	PASc	97	16	56	DNA	1.32/1.32

5.3.9 PAS domain ligand binding mechanisms

5.3.9.1 The PpANR PAS domain binding cavity

One of the striking features of bacterial PAS structures, as seen, is their wide range of small molecule interacting partners; a trait clearly retained in the mammalian bHLH-PAS proteins (*e.g.* Scheuermann *et al*., 2009). To analyse the potential for ligand binding in the PpANR PAS domain, the features of its internal cavity were analysed. Both chains in the PpANR PAS homodimer form internal cavities, although as seen already, there are asymmetries between the two which is also seen in their respective cavities (Figure 5.24). While chain A has a large 'secondary' pocket not found in chain B, it has a canonical PAS cavity (defined as occurring adjacent to the F_α helix) as in chain B. In chain A this cavity is predicted to have a volume of 193Å^3 with another partially connected chamber around 44.5 Å^3. In chain B, the cavity is measured as 231 Å^3 while another internal cavity, like the small one in chain A but not connected, is around 35 Å^3. Comparisons with other PAS structure cavities show the overall cavity position, including its entrance, is similar to that seen in haem binding structures such as FixL (1XJ3) to which the PpANR PAS domain has the highest pairwise similarity and this is analysed in more detail (Figure 5.27). The presence of other small cavities in close proximity to the canonical cavity suggests that a larger ligand could be accommodated than the volumes measured in the *apo*- state as the F_α helix is known to be flexible, which may allow the cavities to combine in the bound state. There are three residues that have side chains in close proximity with the PpANR PAS domain cavity and that are also highly conserved among MAP3K PAS domains and the related PAS domains from algal and bacterial HKs: I^{53}, W^{62} and G^{64}. If any ligand binds in this

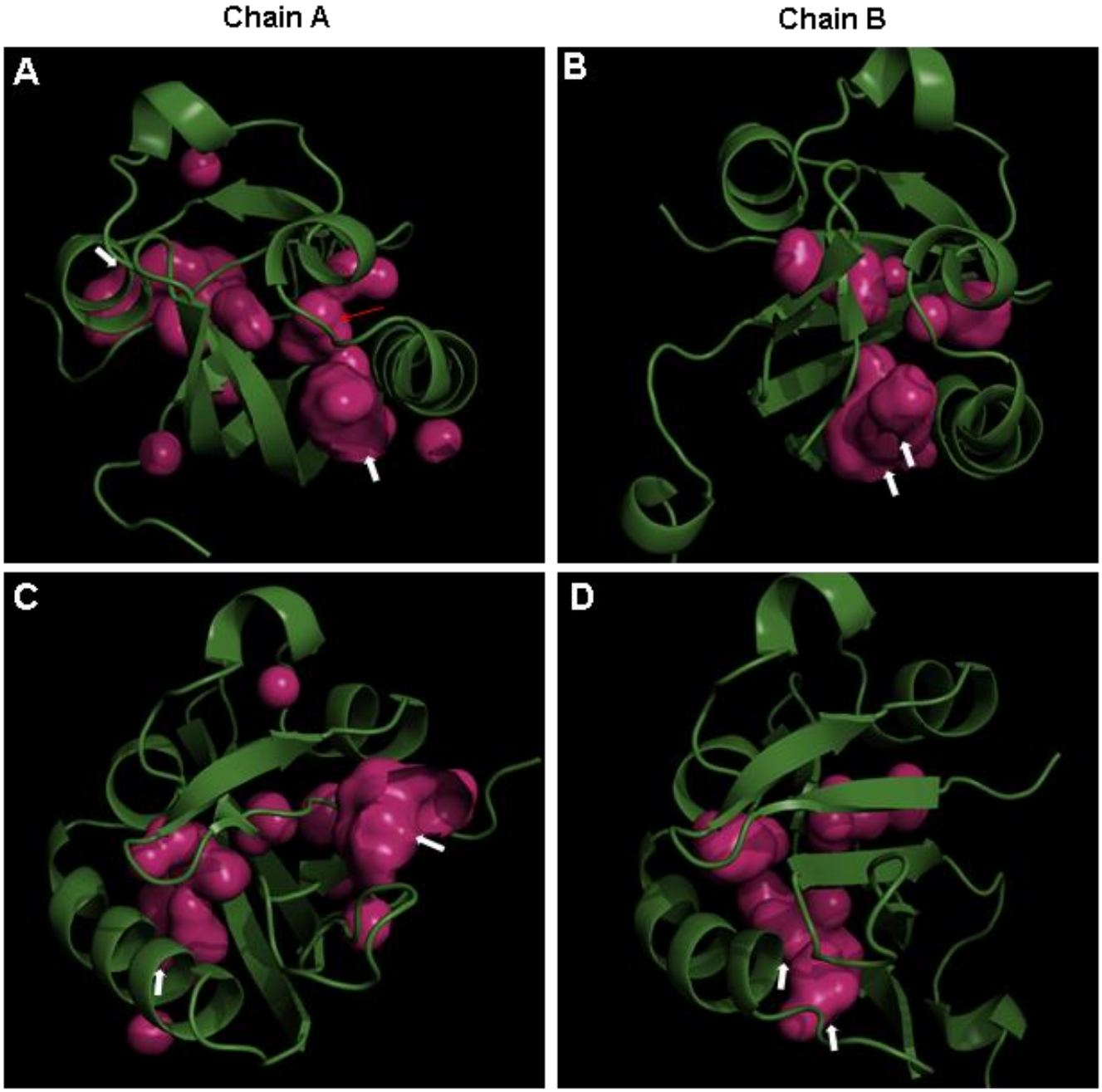


Figure 5.24. Both chains of the PpANR PAS domain have internal cavities. Internal cavities from both chain A and B shown in pink with white arrows indicating cavity entrances. Both chain A and B have similar canonical cavities lying adjacent to the F_{α} helix. The cavity in chain A **(A)** is slightly smaller than that in chain B **(B)**. The canonical cavity in chain A is partially connected to smaller cavities (red arrow). Chain A also has another pocket in a 'non-typical' position on the opposite side to the canonical PAS cavity **(C)** which is not found in chain B **(D)**. Cavities were visualised using the surface 'cavities and pockets only' option in PyMol with default settings.

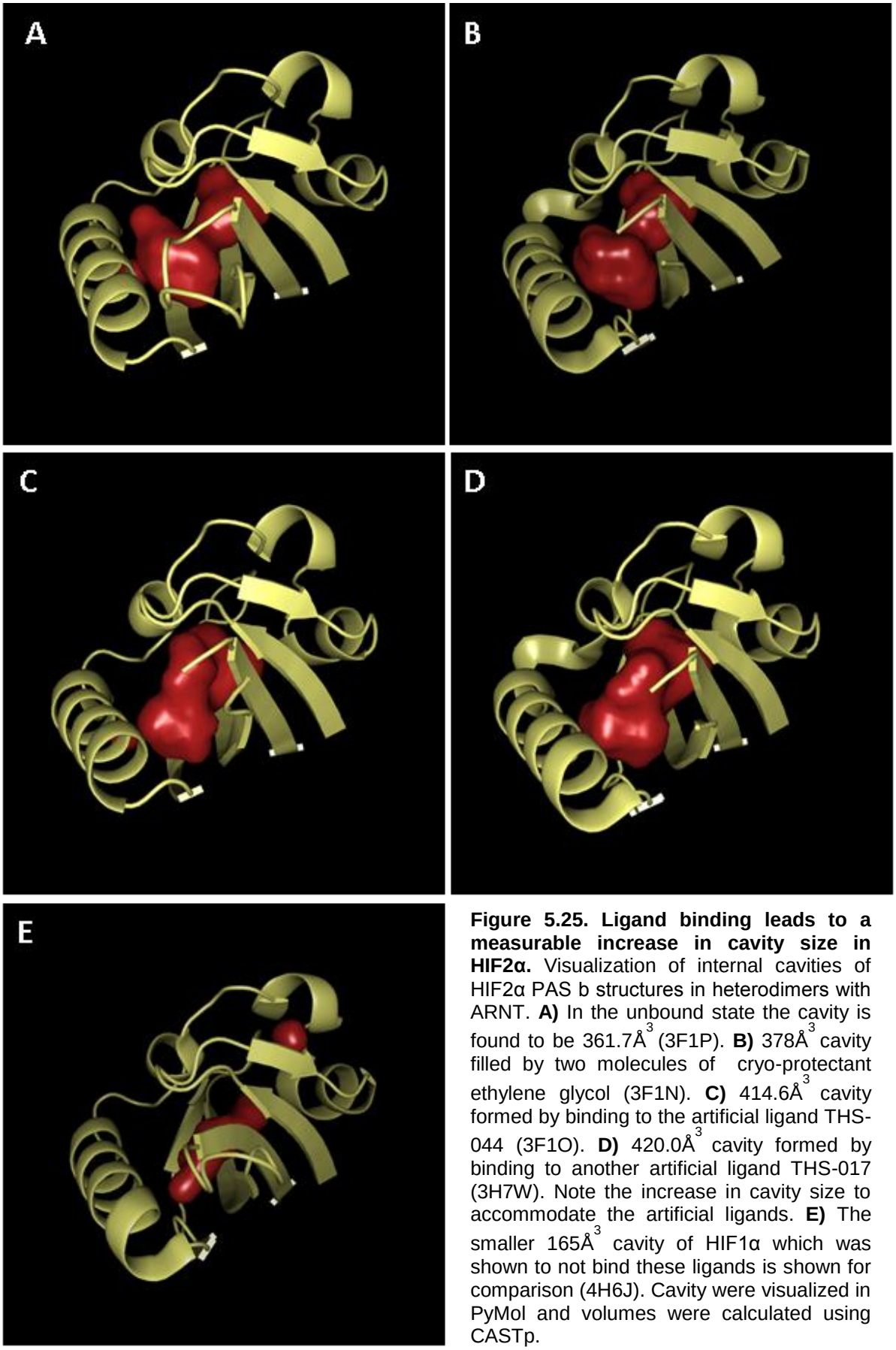


Figure 5.25. Ligand binding leads to a measurable increase in cavity size in HIF2α. Visualization of internal cavities of HIF2α PAS b structures in heterodimers with ARNT. **A)** In the unbound state the cavity is found to be 361.7Å^3 (3F1P). **B)** 378Å^3 cavity filled by two molecules of cryo-protectant ethylene glycol (3F1N). **C)** 414.6Å^3 cavity formed by binding to the artificial ligand THS-044 (3F1O). **D)** 420.0Å^3 cavity formed by binding to another artificial ligand THS-017 (3H7W). Note the increase in cavity size to accommodate the artificial ligands. **E)** The smaller 165Å^3 cavity of HIF1α which was shown to not bind these ligands is shown for comparison (4H6J). Cavity were visualized in PyMol and volumes were calculated using CASTp.

cavity then these residues may represent important functional residues and could be the targets for future mutagenesis work.

The extent to which this increase in cavity size can occur was explored in HIF2α which has both bound and unbound structures solved (Figure 5.25). In its unbound state, the cavity is calculated as being 362 $Å^3$ but this increases when artificial ligands are bound such as THS-044 (3F1O) and THS-017 (3H7W) to a maximum of 420$Å^3$ which is a 16% increase in volume. This increase appears to involve an expansion of the narrow connection between the two chambers in the *apo*- structure which may be the case for the PpANR PAS domain too as suggested. Interestingly, the internal cavity of HIF1α is found to be far smaller than HIF2α and this may partially explain the inability for HIF1α to bind ligands (Scheuermann *et al*., 2008). What is also evident from the ligand binding of HIF2α is that it does not appear to have any cavity entrance open to the outside as in the PpANR PAS domain. This suggests that the flexible F_α can temporarily 'open' the entrance to allow ligand entry before 'closing' again.

5.3.9.2 FAD binding

No PpANR PAS::FAD binding was found and the position of the internal cavities for the PpANR PAS and FAD binding structures are different, being on different sides of the F_α helix. Despite this, the PAS domains from MmoS (3EWK) and NifL (2GJ3) have high pairwise structural similarity to the PpANR PAS domain and so an analysis of FAD binding mechanisms was carried out. A number of key residues have been identified as being conserved in FAD binding (Henry and Crosson, 2011) and these include a tryptophan found on the F_α helix which mediates aromatic stacking with the FAD adenine (Figure 5.26). The PpANR PAS domain lacks conservation in these residues between either of the two binding mechanisms observed in MmoS and NifL. The fungal blue-light sensor VIVID (2PD7), which binds FAD and FMN, binds FAD through a different mechanism to the bacterial redox sensors. Instead it is found to have a modified D_α-E_α loop region involved in binding (Schwerdtfeger and Linden, 2003; Zoltowski *et al*., 2007) which shows that binding of certain ligands can re-evolve in PAS structures through new interaction mechanisms. From the superimpositions, it is evident that the F_α helix of the PpANR PAS domain has a distinct position meaning it effectively blocks the cavity entry for FAD binding in MmoS and NifL and conversely enables a cavity on the other side of the F_α helix. Taken together, the structural analysis of FAD binding supports the finding of no PpANR PAS::FAD binding as there does not appear to be any structural conservation in the binding mechanisms.

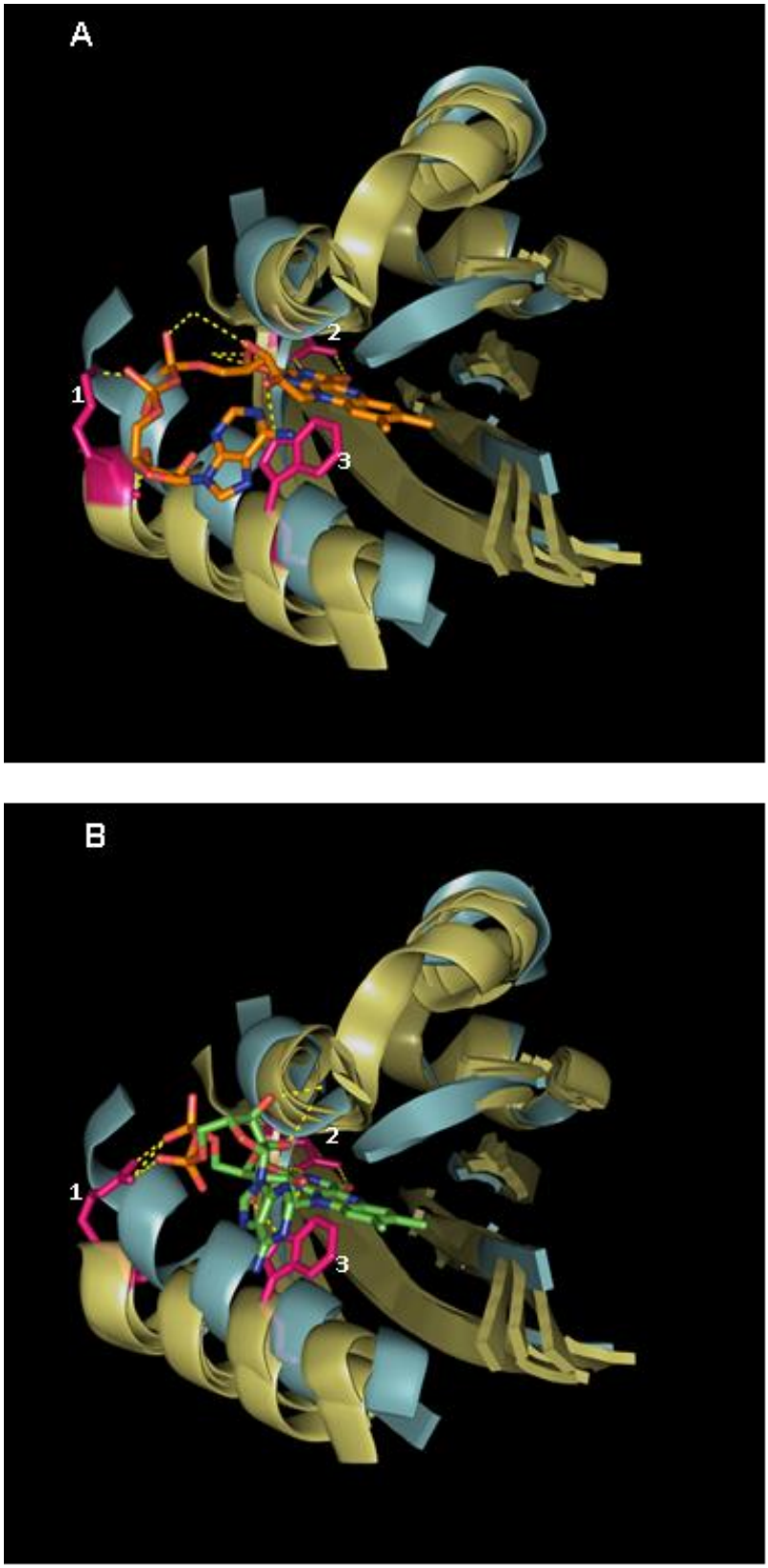


Figure 5.26. Conserved mechanisms of FAD binding not found in the PpANR PAS domain. Superimpositions of PpANR PAS-chain A (grey) with three FAD binding PAS domains (3EWK, 2PD7 and 2GJ3 all in pale yellow) showing two possible orientations of FAD binding. **A)** FAD binding as found in MmoS (3EWK) with key conserved residues highlighted in pink showing polar interactions between lysine (1) and asparagine (2) as well as aromatic stacking between a tryptophan and FAD adenine (3). **B)** FAD binding in NifL (2GJ3) is shown with key residues highlighted in pink showing polar interactions between arginine (1) and asparagine (2) as well as aromatic stacking between a tryptophan and FAD adenine (3). Loop regions have been omitted for clarity. Note the raised position of the F_{α} helix in the PpANR PAS domain which has no ligand/cofactor bound. The PpANR PAS domain lacks conservation of interacting residues, most notably the tryptophan on the F_{α} helix.

5.3.9.3 Haem binding

The PpANR PAS domain has the highest pairwise similarity with the haem binding FixL (1XJ3), although one of the lowest similarities with another haem binder; the DOS redox sensor (1V9Z), and the position and orientation of the internal cavity is shared between these structures. Haem binding is primarily mediated through a direct interaction of a highly conserved histidine residue on the F_α helix with the Fe^{2+} molecule (Figure 5.27). Many other haem binding proteins do so through interaction with a histidine residue such as in hemopexin (Morgan *et al*., 1993) suggesting this is a universal feature of haem binding. There are no histidine residues in the F_α helix of the PpANR PAS domain, or indeed in the entire PAS domain, so this mechanism is unlikely. As in the comparison of the PpANR PAS domain with FAD binding PAS structures, the F_α helix is distinctly orientated compared to the haem binding structures taking a more 'closed' conformation.

5.3.9.4 Indole binding

Column affinity ligand pull down assays suggested that a peak matching the mass of a number of indole-related compounds was being strongly enriched. While this potential interaction has not been tested, the mechanisms behind a single known indole-related (1H-indole-3-carbaldehyde(I3A)) binding PAS domain have been analysed. I3A binding in the bacterial HTR-like PAS domain (3BWL) occurs within a relatively small pocket (240$Å^3$), comparable to that in the PpANR PAS domain (Figure 5.28). Two hydrogen bonds are found between I3A and two negatively charged aspartic acid side chains at opposite sides of the cavity. There is no conservation of these residues in the PpANR PAS domain however suggesting, if binding were to occur it would do so by different mechanisms.

5.4 Discussion

The PpANR PAS protein was successfully expressed and purified with little need for optimisation of expression constructs, cell types or growth and expression conditions. The purification process produced sufficient amounts of protein on which binding and structural analyses could be carried out. This was not, however, unexpected for a small, soluble and highly structured protein. The expression and purification of the PpANR PAS protein aimed to complement genetic work to help understand whether it was integral to ABA signalling and to help understand its mode of action. While genetic work to mutate key residues (such as W^{16}) is ongoing, preliminary data suggests the PAS domain may indeed be involved in ABA signalling. The recent finding that the MAP3 kinases Raf10 and Raf11 act as ABA regulators in Arabidopsis (Lee *et al*., 2015b) makes this scenario all the more likely. These kinases, closely related to

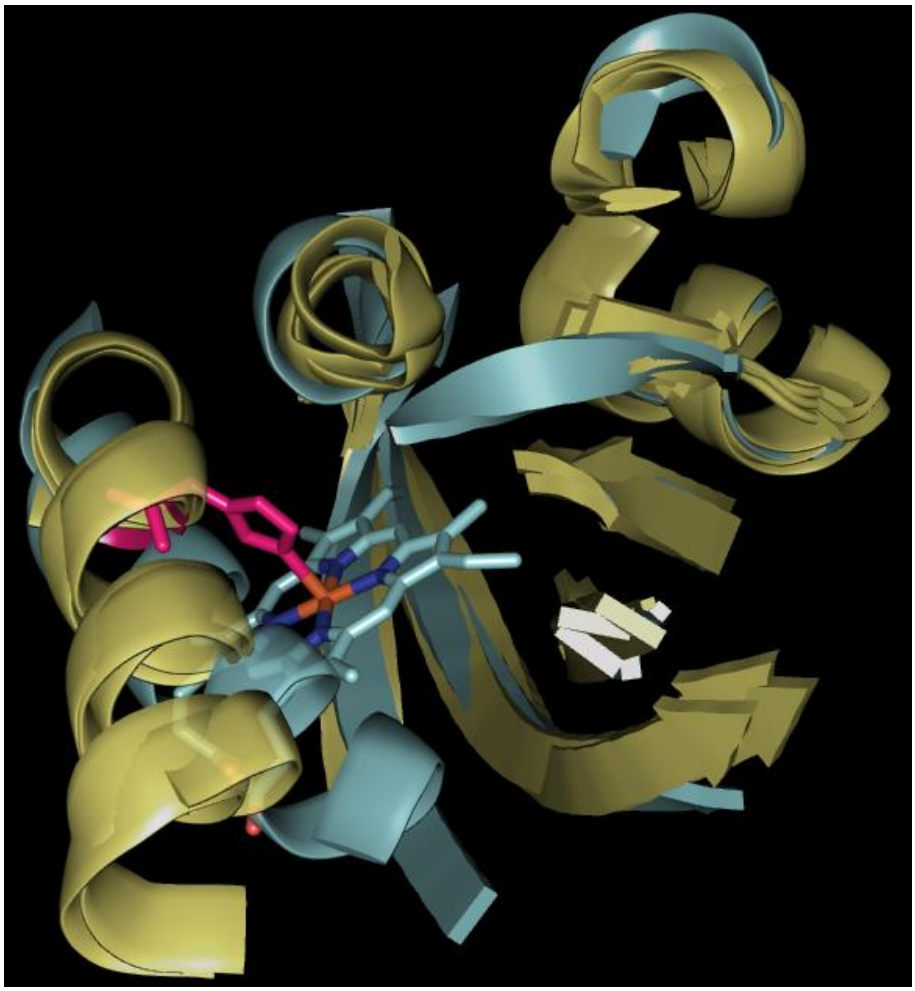


Figure 5.27. Conserved mechanisms of haem binding not found in the PpANR PAS domain. Superimposition of chain A (grey) with three haem binding PAS domains (1XJ3, 1D06 and 1V9Z all in pale yellow). The haem cofactor as found in the DOS redox sensor (1V9Z) is shown with the key histidine residue involved in haem binding highlighted in pink which interacts directly with the Fe molecule. Note the position of F_{α} in the PpANR PAS domain in comparison which is less 'open' but more importantly the PpANR PAS domain lacks the conserved histidine. Loop regions have been omitted for clarity

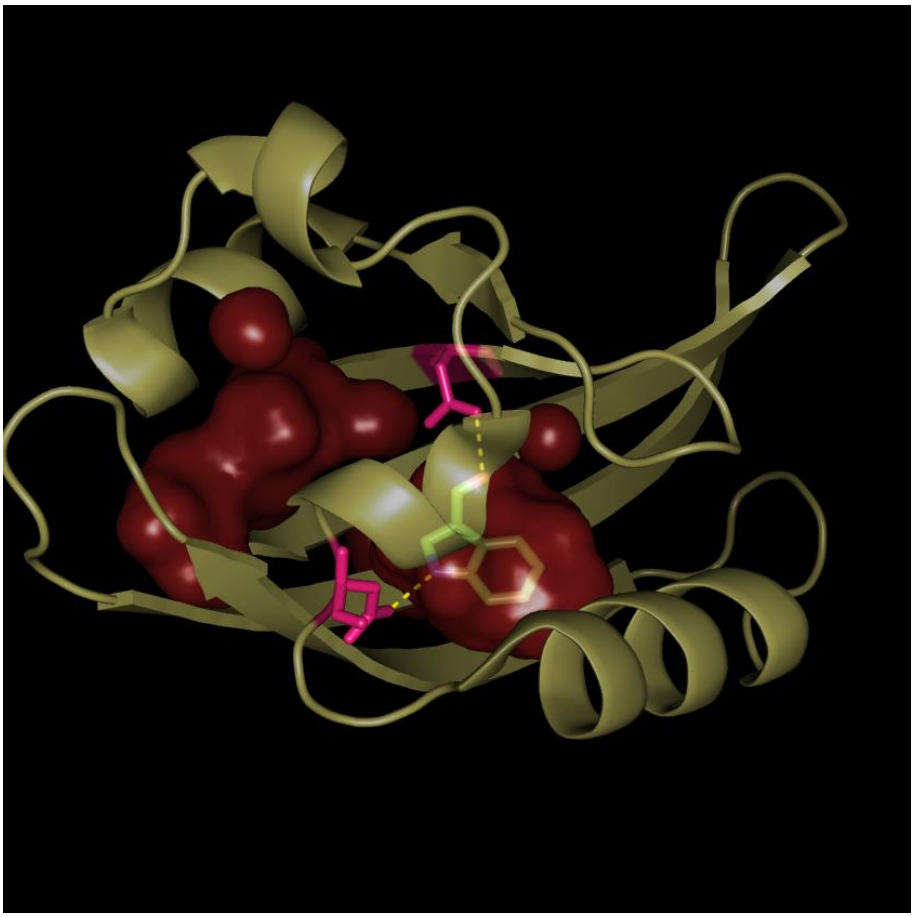


Figure 5.28. Structural analysis of indole binding in the bacterial HTR-like PAS domain (3BWL). The internal cavity within which the indole (I3A) molecule binds found to be 239.7Å^3. The interacting sites are two negatively charged aspartic acid residues shown in pink which bind the N (blue) and O (red) of I3A respectively.

PpANR-like ‘PEK’ kinases (see chapter 4), have a ‘PK’ structure suggesting the PAS domain may be the input module for ABA signalling while the EDR domain is involved in ethylene signalling; as suggested by the interaction between PpANR and PpETR7 (Yasumura *et al*., 2015) where the EDR domain is known to mediates this interaction between AtCTR1 and AtETR1. The solving of the crystal structure of the PpANR PAS domain provides vital data for predicting potential mechanisms and for explaining future findings in the context of its structure.

The structural similarity analysis suggested that the PpANR PAS domain shared greatest similarity with a range of bacterial structures predominantly from histidine kinases in two-component regulatory systems. High pairwise similarity is also found with some PASb structures from members of the bHLH-PAS family. Many of these structures have been intensively studied and many of their modes of action are known in detail. The analyses carried out have not demonstrated clear mechanistic conservation with any other PAS structures despite high overall structural similarity. The finding that the PpANR PAS domain does not appear to bind FAD can largely be supported by structural comparisons which found a lack of conservation in key binding interactions. What is clear is that small changes in key residues, such as the presence of a histidine on the F_{α} helix in haem binding, confer the impressive range of specific functions to PAS structures. As such it is prudent to incorporate as much information about the shared mechanisms of PAS structures to understand the possible roles of the PpANR PAS domain and the potential mode of actions in ABA signalling.

5.4.1 Possibility of PpANR PAS ligand binding

The limited binding assays so far have revealed no clear interaction between the PpANR PAS domain and any ligand. The negative results for ABA and FAD binding, while not completely ruling them out, throws open the question of what the PAS domain may bind and suggests a more complex or unexpected mechanism may be responsible. These two candidate ligands would have fit some of the more parsimonious mechanisms of PAS signalling in which a PpANR PAS::ABA interaction could potentially have mediated an ancestral ABA receptor signal transduction pathway. The combination of FAD being a common PAS cofactor and redox sensor and osmotic stresses being associated with changes in cellular redox potentials (see de Carvahlo, 2008 for review) meant a PpANR PAS::FAD interaction too was mechanistically possible. Screening of any other ligand interaction through high throughput screening techniques, such as metabolite PAS affinity purification, is required to identify what, if any, small molecule may be the input signal for the PpANR PAS domain and ABA signalling. The identification of a possible PpANR PAS domain interacting partner matching the mass of a number of indole-related compounds

requires further analysis; however no clear mechanistic explanation for this interaction presents itself. Another complementary technique not carried out to date would be to overexpress the 6-His-tagged PpANR PAS protein *in vivo* followed by column affinity purification of protein extracts to co-purify the PpANR PAS protein in complex with any interacting partners. Such a method may also allow the potential for protein-protein interactions to be identified through pull-downs, perhaps in conjunction with a yeast two-hybrid approach.

The structure of the PpANR PAS domain offers the opportunity to predict what ligands might bind based on structural similarity and by identifying key residues likely to be functionally shared with other ligand binding PAS domains. The paper by Henry and Crosson, 2011 identified conserved residues involved in ligand and cofactor binding among PAS domains with solved structures, albeit with a focus on bacteria. Using structural superimpositions with the PpANR PAS structure and other PAS domains known to bind ligands, equivalent residues can be assessed for likely functional conservation in the interaction. For example, two FAD binding bacterial PAS domain from NifL and MmoS have high resolution structures in the bound state solved revealing three conserved FAD interactions: aromatic stacking mediated by a tryptophan and two polar interactions mediated by an asparagine and arginine/lysine. As already stated, no conservation in these residues was found in the PpANR PAS domain and a comparison of the position and size of the internal cavities further showed FAD binding, at least by conserved mechanisms as the bacterial redox sensors, to be unlikely. Despite this, this kind of approach can be very useful for guiding the prediction and testing of specific interactions based on structural constraints implied by the crystal structure. The combination of the identification of possible indole-related compound binding and the solving of a crystal structure with an indole-related ligand offers a chance to predict and test a possible interaction between the PpANR PAS domain and these compounds. Another useful resource is the well-studied bHLH-PAS TFs which have demonstrated a number of key features of PAS functions and modes of action and that show a high degree of structural similarity with the PpANR PAS domain. While these TF functions and activities are principally regulated by PAS mediated heterodimerisation, a number of the PAS domains are known to bind ligands; both endogenous, such as the receptor AHR (see Denison *et al*., 2002 for review), and exogenous, such as the hypoxia response HIF2α TFs (Scheuermann *et al*., 2009; Scheuermann *et al*., 2013). The solving of structures bound to a range of ligands offers the chance to compare possibly conserved residues in the PpANR PAS domain in much the same way as described above that could uncover a potential binding mechanism; albeit to different ligands. The large scale screening of

ligands that can influence the function of HIF2α also offers the chance for similar approaches to identify chemicals that may interact with and alter the function of the PpANR PAS domain even if no native binding partners exist. This may open up the possibility of using a chemical genetics approach to probe the function and mode of action of the PpANR PAS domain, and indeed other MAP3K PAS domains, and for possible crop treatments to chemically confer tolerance.

5.4.2 Dimerisation

While the search for potential interacting partners goes on, the identification of dimerisation in the PpANR PAS domain suggests another possible mechanism for kinase activity regulation and ABA signalling. Many PAS-PAS interactions have been documented which can function as activity modulators and even regulate localisation. For example, plant photoptropins rely on the N-terminal PASa (LOV1) domain to mediate homodimerisation which regulates sensitivity to blue light (Salomon *et al*., 2004). Much like the PpANR PAS domain, phototropin LOV1 domains form homodimers which orient in a similar fashion with the β-sheets facing one another in antiparallel orientation. Solving of the crystal structures of *Arabidopsis* LOV1 domains (the LOV2 domains are primarily involved in blue light dependent activation of the serine/threonine kinase domain) from phot1 and phot2 identified a crucial cysteine mediated disulfide bond in phot1 (on the β-sheet) while phot2's dimerisation was mediated by hydrogen bonds across the interface (Nakasako *et al*., 2008). Disruption of these key dimerisation interactions breakdown signalling and so similar approaches can be taken for the PpANR PAS domain. The bHLH-PAS TFs, already discussed in the context of ligand binding and found to be structurally similar to the PpANR PAS domain, form heterodimers that define their functions and activity (see Bersten *et al*., 2013). Much like in the phototropins, the N-terminal PASa domain is predominantly involved in dimerisation while the PASb is more involved in ligand binding and signal activation. Some PASb domains, as in HIF2α and AHR, bind ligands that mediate effects by affecting heterodimerisation. The mechanisms by which this happens in these well studied PAS models, which principally involves ligand-induced structural changes, may be used to predict potentially similar events in the PpANR PAS domain. For example, the HIF2α-ARNT PASb heterodimer is disrupted by mutating residues on the solvent-exposed surface of the HIF2α β-sheet (Erbel *et al*., 2003) and similar approaches can be used to assess the role of dimerisation in the PpANR PAS domain on ABA signalling. It is worth also noting that the PAS crystal structure solved was only a fragment of PpANR encompassing the core PAS fold region. As other PAS structures have shown flanking regions to be important in dimerisation and function, it is possible that the dimerisation interface identified may not be the same as that in a native *in vivo*

(and indeed full protein) context, although the G_β-mediated dimerisation seems robust and unlikely to arise as an artefact of crystal stacking; unlike the other potential dimerisation interactions presented involving the chain B C-terminus.

5.4.3 General PAS domain mechanisms of signal transduction

While a few examples of PAS domains have a particular relevance to the PpANR PAS domain, many structural studies on a wide range of PAS domains have illuminated common themes and mechanisms primarily involving signal transduction through conformational changes in the β-sheet (reviewed by Möglich *et al*., 2009a).

Phototropin LOV2 domains have a conserved J_α helix that is thought to propagate the signal to the effector kinase domain when light induced conformational changes of the β-sheet disrupt the J_α-β-sheet interaction. This results in a dramatic change in angle between the LOV2 domain and the effector domains and prevents the LOV2 from inhibiting the kinase domain catalytic site (Harper *et al*., 2004). This effect has also been well described in bacterial histidine kinases in which a conserved DIT motif at the C-terminus of the I_β strand (near the loop region with J_α helix) is thought to be important (Möglich *et al*., 2009a). This motif is thought to mediate bonds between this region and the H_β strand so maintaining close and tight linkage with the C-terminal flanking J_α helix enabling changes in the PAS fold conformation to be propagated to the linker region and to induce conformational changes. A feature of these types of PAS domains, in which the J_α helix is important, is that the PAS fold is directly linked with the effector domains. As in the case of phototropins, it is the second PAS domain (LOV2) that is important for signal activation while the first or N-terminal PAS domain is more involved with dimerisation. By contrast, the PASb domain in AHR and HIF2a is separated from the bHLH DNA binding domain by the PASa but the important structural changes in this protein alter dimerisation states and complex formation with other cofactors through the PAS domains. This indicates that the PAS domain generally enables signal transduction through conformational changes mediated by close proximity of a PAS domain and the (usually) C-terminal effector domain – possibly propagated by rigid molecular rods known as coiled-coil linkers (Möglich *et al*., 2009a) or by altering other PAS-mediated interactions directly. Interesting, chimeric proteins have been constructed that also demonstrate that this conserved mechanism of signal transduction is likely to apply to many PAS domains. For example when a photo-sensory PAS domain (FMN binding) replaced the native redox sensing PAS domain in FixL (haem binding), its effector domain was now able to respond to blue-light signals rather than oxygen (Möglich *et al*., 2009b). Other synthetic proteins have been made that take advantage of this likely conserved signal transduction mechanism to put a number of effector domains under the control of light by fusing them with a photo-

sensing PAS domain and a coiled-coil linker (see Möglich *et al*., 2009b). While it appears that the PpANR PAS domain does have potentially helical flanking regions, the distance between the PAS domain and the kinase domain, separated by the EDR domain, suggests this mechanism is unlikely, assuming the kinase domain is the primary effector domain. This suggests that if structural changes occur in the PpANR PAS domain, these are likely to alter direct interactions possibly with other proteins unless such coiled-coil linkers propagate signals to the EDR domain.

While the signal transduction mechanisms may come from ligand binding or dimerisation the possibility of these upstream signals coming from a PpANR PAS domain interacting with another protein should be highlighted. Not many PAS structures are known to interact with other proteins outside of their own protein families through PAS dimerisation. Interactions between the PASb of certain bHLH-PAS proteins with coiled-coil coactivator proteins has been found that regulate transcriptional responses (*e.g.* Partch *et al*., 2009). These authors found the coactivators to interact with the opposite side of the PAS structure, an α-helical face, to that involved in dimerisation (usually the β-sheet). The failure to find a potential ligand for the PpANR PAS domain, despite structural evidence that an internal cavity is present, necessitates a wider evaluation of potential interacting partners that act upstream to signal in PpANR's ABA/stress responsive role. Nevertheless, the solving of a MAP3K PAS domain represents a novel contribution to the structural knowledge of PAS domains and sets a firm foundation against which to understand its role in ABA signalling.

CHAPTER 6

Chapter 6 – General discussion

6.1 Summary

This thesis has presented work on the novel ABA regulator PpANR that has revealed a number of interesting features about the evolution of ABA and stress signalling and their roles in the conquest of land. The discovery of PpANR was the result of the first successful implementation of forward genetics in *Physcomitrella patens*, and indeed in any bryophyte, and represents a potential watershed moment as now these basal land plants can be used for discovery and not just for comparison with *Arabidopsis*. The use of NGS approaches is highly recommended and will likely become an invaluable part of this approach in *Physcomitrella* for both rapid genotyping and gene expression analysis; able to simultaneously map and identify the causal mutation and obtain differential expression analysis for the molecular characterisation of the mutant. The molecular characterisation of *Ppanr* mutants has revealed an unexpected level of importance for PpANR in ABA-dependent stress responses as its absence leads to the almost complete abolition of both a growth and molecular response. The significance of this should not be overlooked and is discussed further. Phylogenetic analyses have revealed that PpANR, its putative orthologues and closely related MAP3 kinases were established in plants prior to the conquest of land. That many aspects of ABA signalling and indeed its role in the charophytes remains unknown, not least the apparent absence of putative PYR-like receptors, suggests that the 'core' ABA signalling pathway so well understood in angiosperms and increasingly in bryophytes was a late addition to the molecular toolbox and not part of the charophyte stress signalling networks; however it must have arisen from something. This makes the recently characterized MAP3 kinase ABA/stress regulators, PpANR and the B2 members such as Raf10/11, likely to be important components in the pre-adaptation of charophytes for life on land. The growing evidence of this pre-adaptation, such as the demonstration of conservation in the ethylene signalling mechanisms in *Spirogyra pratensis* (Yu *et al.,* 2015), is no doubt an important part of understanding the conquest of land. With growing genetic resources and molecular tools for not only the bryophytes but the charophytes, such as *Klebsormidium* and *S. pratensis*, we will be able to understand this process in more detail. It is clear that freshwater algae, with molecular mechanisms for coping with varying water conditions as water levels fluctuated following rain or drought, were one step away from life on dry land where all that changed was the duration and frequency of drought conditions.

This thesis has brought to light an interesting component of ABA and stress signalling in that many questions remain unanswered with implications for the origins and

evolution of ABA-mediated stress signalling. ANR-like genes are not found in vascular plants, with the exception of *Selaginella,* so an obvious first question is how this gene came to be lost and what were the implications of this for the evolution of ABA-mediated stress signalling in complex land plants? We know from comparative genomics that while functional conservation exists in ABA signalling and its components there are subtle but important differences between bryophytes and angiosperms reflecting the evolution of the pathways from aquatic origins. So what are the key changes that have occurred that first allowed osmotic stress signalling to be integrated with ABA itself and how did ABA signalling become so specialized in angiosperms? A question pertaining to PpANR is do ANR-like genes act in this osmotic stress-responsive pathway in the charophytes and if so how did this come to get recruited into the ABA-dependent pathway? Another key aspect relating to PpANR itself is that it is known to have roles in both ABA and ethylene signalling but how is this dual role carried out and is it a special case among ANR-like genes? Each of these questions is discussed in more detail to put the recent advances in ABA and ethylene signalling as well as the functions of a number of B group MAP3 kinases in context; in which PpANR has a central role.

6.2 A model for the evolution of ABA signalling in dehydration and desiccation tolerance.

As mentioned, a number of differences between bryophyte and angiosperm ABA signalling have been identified and these have led some authors to propose an evolutionary model upon which the model presented here builds (Komatsu *et al.,* 2013). The key difference noticed by these authors was in the role the two group A PP2C phosphatases, homologues of ABI1, in their regulation of the SnRK2 kinases, *e.g.* OST1. The classic model of ABA signalling involves these phosphatases inhibiting the phosphorylation of the OST1 kinases and when ABA is present this inhibition is lost due to sequestering of the phosphatases by PYR-like ABA receptors. The OST1 kinases in flowering plants were found to autophosphorylate and activate downstream targets. In *Physcomitrella*, however, this did not appear to occur as when both ABI1 homologues are deleted, the autophosphorylation does not occur but remains ABA-dependent; one of these subtle but important differences. This leads to the suggestion that the PP2C-mediated repression is acting on a pathway that responds to ABA but is likely activated by osmotic stress which confers dehydration/desiccation tolerance (D/DT). The model follows that this pathway existed prior to the incorporation of ABA and PP2C repression as an ancestral D/DT pathway leading to the expression of the typical molecular responses we see in *Physcomitrella* and indeed in the angiosperm seed. With the incorporation of the repression, or D/DT ‘braking mechanism’, the

pathway came under stronger ABA control and conversely weaker osmotic stress control which allowed the eventual specialisation of the D/DT pathway to regulating angiosperm seed maturation; a process itself dependent on the same ancestral molecular changes to enable dormancy and tolerance. This last bastion of true desiccation tolerance in (most) angiosperms is a key feature and one that makes them so successful, able to spread the next generation through time.

6.3 A role for PpANR and the ancestral DDT pathway

This model for the evolution of the 'core' ABA signalling pathway remains hypothetical and the evidence presented by Komatsu and colleagues (2013) is far from conclusive. The intellectual strength of the model comes from inferring what the likely D/DT mechanisms are for charophytes and so what was likely available for the first land plants. As mentioned, the mechanisms and functions of ABA signalling in charophytes remain unclear but there is a clear necessity for a signal transduction pathway linking osmotic stress, such as low water availability, in these transitory plants that eventually made the dry land their home. ANR kinases do not feature in the current ABA signalling model although the related MAP3Ks Raf10 and Raf11 have been shown to be ABA regulators (Lee *et al.,* 2015b). While *Physcomitrella* does not have Raf10/11 orthologues, PpANR is likely to be functionally equivalent. As already discussed, however, there are likely to be some differences as the vascular plants have placed more emphasis on ABA rather than osmotic stress regulation of the D/DT pathway. One consequence is that PpANR is found to be absolutely crucial in ABA-dependent D/DT in *Physcomitrella*. I propose that ANR-like and Raf10/11 orthologues (present in the other bryophyte lineages and charophytes) are part of the ancestral D/DT signal transduction pathway that connects osmotic stress with gene expression changes. These likely act upstream of the OST1 kinases based on the suggestion that the latter require upstream phosphorylation (Komatsu *et al.*, 2013) and that they are able to directly phosphorylate target TFs such as AREBs (Furihata *et al*., 2006).

6.3.1 A histidine kinase origin of D/DT signalling?

The RNA-seq analysis presented here revealed a putative homologue of the osmosensor histidine kinase AtHK1 as being very strongly ABA- and importantly, dehydration- and mannitol-up-regulated. This finding fits in with the emerging picture for a role of these histidine kinase osmosensors in transducing water stresses into responses. The role for the plasma membrane bound AtHK1 as a mediator of dehydration stress responses (Wohlbach *et al*., 2008) is supported by the finding that its putative rice orthologue OsHK3 is ABA- and stress-up-regulated and mediates oxidative stress responses (Wen *et al.,* 2015); a crucial component of D/DT.

Interestingly, Wohlbach and colleagues also found AtHK1 to function not only in seeds but also in the young seedlings when vegetative tissues are most at risk from dehydration stresses. They also found AtHK1 to be ABA-dependent, which is consistent with ABA being required to prevent the PP2C 'braking' of D/DT signalling in angiosperms. It has also been demonstrated that a number of key ABA/stress-responsive genes are up-regulated by AtHK1 such as *AREB1* and *DREB2* indicating its regulation of key stress response components (Tran *et al.,* 2007). The evidence for the role of this water stress osmosensor in mediating ABA-dependent stress responses strongly suggests it as the sensory component in an ancestral D/DT signalling pathway. Like the related HK ethylene sensors, the ETRs, AtHK1 is phosphorylated when no stimuli are present meaning that the phosphorelay activates downstream targets when top down phosphorylation stops upon signal detection. How might the signal therefore be transmitted to the putative downstream targets of PpANR and OST1-like kinases which require phosphorylation for activation? The typical pathway involves autophosphorylation of the HK receiver domain followed by phosphorelay intermediates (HPt) which allow the transfer of phosphoryl groups to response regulators (RRs). In the SLN1 osmotic stress pathway in yeast, which appears conserved with that headed by AtHK1, the detection of stress stops the phosphorelay, in the same way as following ethylene binding in the ETRs, and so a MAP kinase cascade is no longer inhibited (reviewed by Urao *et al.,* 2000). In this process, the direct protein-protein interaction of the now activated and unphosphorylated RR activates the MAP kinase cascade. Another feature of some of the osmotic stress responsive RRs (ARR1 and ARR2) identified in *Arabidopsis* (Urao *et al.,* 1998) is their potential intrinsic TF capabilities as they have domains with high similarity to MYB DNA binding motifs (Sakai *et al.,* 1998) suggesting a possible mechanism for ABA-independent D/DT signalling.

6.3.2 The PAS domain

I have proposed that the gain of a PAS domain possibly from a PAS-containing HK (since lost in land plants) to an "EK" EDR1-like kinase in the green algae created the ANR-like subfamily and conferred the ability to be involved in ABA-dependent stress signalling while the EDR domain conferred a role in ethylene signalling through an interaction with the ETR receptors (sensory HKs). PAS structures are involved in a wide range of functions, primarily in sensory capacities, but they also enable modulation of their effector domains through protein-protein interactions. In line with the failure to find any interaction between the PpANR PAS domain and ABA, or indeed the mechanistically plausible FAD, a possible role for the PpANR PAS domain in mediating ABA/stress signalling may therefore be through protein-protein interactions; possibly

with a downstream component of the osmosensing HK(s) and experiments to investigate this, such as yeast two-hybrid assays, should help clarify this. Another putative signalling component identified alongside the putative osmosensing HK1 homologue was a gene encoding a 14-3-3 general regulatory factor (GRF). These are general signalling components that function as phosphoserine-binding dimers and are known to have roles in a number of plant stress responses (reviewed by Roberts, 2003) and have even been shown to mediate the complex between ABA responsive bZIP and ABI3 TFs (Schultz *et al.,* 1998). The potential role of this 14-3-3 protein, and others, should be uncovered in *Physcomitrella* in an effort to understand this basal D/DT signalling pathway and whether it is a possible regulator of PpANR through the interaction with phosphoserines.

6.4 PpANR dual signalling role: an exception or the rule?

As already discussed, the trimodular PpANR seems likely to have its dual functions split between the two N-terminal input domains; the PAS and EDR domains regulating ABA and ethylene signalling respectively. This is based on the combined functional and phylogenetic evidence. ANR-like kinases have been lost in the vascular plants, with the exception of the lycophytes in the genus *Selaginella*. Clearly such a vital ABA-dependent stress regulator is not needed in the anatomically complex land plants and the presence of the CTR1 and Raf10/11 ethylene and ABA regulators respectively are able to functionally cover the missing component. The role of PpANR in ethylene signalling appears the same as for CTR1 in that it binds the ethylene receptors to function as a negative regulator through interaction between ETR and the EDR domain, and it would be interesting to determine whether the PAS domain is necessary to confer a positive ABA regulatory role in the Raf10/11 paralogues. Given the ability of the more anatomically complex land plants to avoid vegetative dehydration courtesy of vastly improved water retention, the selective pressure on components that confer D/DT is relaxed. Indeed, were these vascular plants to ‘allow’ desiccation to occur it would prove lethal given the development of air embolisms in xylem vessels. When this is coupled with molecular redundancy likely presented by Raf10/11, the loss of one such component becomes possible. Evidence suggests that the trimodular ANR-like kinases were the ancestors of the CTR1-like and B2 subfamilies through the loss of respective domains and these child kinases were likely more efficient at carrying out their respective functions, not being held back by having to compromise between the two roles. This is an assumption based on distinct modes of action and interacting partners for the two roles. While ANR-like kinases are lost from higher plants, quite the inverse situation is seen in *Physcomitrella* and possibly in all mosses (based on current database searches) where the EDR1-like, CTR1-like and B2 subfamilies are lost. So

without the CTR1-like and B2 kinases to function independently as efficient ethylene and ABA regulators respectively, PpANR has simply not been allowed to be lost and is forced into retaining both roles. It was recently demonstrated that the charophyte *Spirogyra pratensis* has conserved mechanisms of ethylene signalling mediated through a canonical CTR1 orthologue (*CTR1a*) interacting with an ETR receptor, whilst its “PEK” ANR-like kinase (*CTR1b*) did not show this interaction (Yu *et al.,* 2015). This indicates that ANR-like kinases do not necessarily have to retain a dual function and it would be interesting to see if the ANR-like orthologue functions as a positive dehydration stress regulator in the charophytes, as well as determining whether the *Spirogyra CTR1b* gene can complement the *Ppanr* mutation. While the sample size of functionally characterized B group MAP3 kinases is small, it appears that the role of PpANR as a dual ethylene and ABA regulator is indeed an exception; a likely consequence of the evolutionary compromises forced on the ABA- and ethylene signalling mechanisms. An understanding of the role of these B group MAP3Ks in liverworts such as *Marchantia polymorpha* would be very interesting as these bryophytes appear to have maintained members from all subfamilies. Has perhaps the ANR-like orthologue found itself another distinct role while allowing its better suited child CTR1-like and B2 kinases to carry out their roles as ethylene and ABA regulators respectively?

6.4.1 How might a dual role be accomplished?

One of the big questions that stems from the recent findings is just how such a dual role can be accomplished. While still hypothetical, it appears likely that PpANR acts in a linear fashion with the OST1 kinases to act as a positive ABA regulator which requires phosphorylation for its own activation and the subsequent activation of TFs and gene expression. Phosphorylation must also occur while PpANR functions as a negative ethylene regulator in which, assuming the mechanisms are conserved with CTR1, it is activated by the ETR receptor when no ethylene is bound; *i.e.* normal conditions. The detection of ethylene results in the dephosphorylation of the ETR and therefore PpANR, allowing downstream targets to become active (EIN2 translocates to the nucleus). There are two main ways such potentially different mechanisms can be resolved. The first is through spatial separation. The site of ethylene detection and signalling is the ER to which the ETR receptors are bound meaning that PpANR at the ER would associate with and become phosphorylated by the ETRs. This interaction is not just transitory and so these phosphorylated kinases will be largely fixed in place so mediating ethylene signalling through its phosphorylation state. If sufficient PpANR is produced, and the expression data show it to have constitutively high expression, then some may be active outside the ER in the nucleus, cytoplasm or even at the plasma

membrane where the HK osmosensors are found. When not interacting with the ETRs, PpANR would not become phosphorylated and so would be free for an upstream phosphorelay signal originating from osmotic stress detection to activate it resulting in OST1 and ABA-responsive TFs activation. There are a few variations of this spatial separation hypothesis. One possibility is that the upstream osmotic stress signal results in PpANR translocating from the ER to the nucleus where it is free to interact with OST1 and/or the target TFs. Another possibility is that the EDR domain interacts with the HK1-like osmosensors directly at the plasma membrane as it does with the ETR receptors. These are related HKs and have high homology and it could be that the EDR domain evolved as an HK interacting domain. This would leave the role of the PAS domain a little uncertain but it could confer translocation of PpANR to the nucleus following osmotic stimuli detection. In the Raf10/11 kinases, the lack of the EDR domain may fit in with a lack of direct osmotic stress regulation and no osmosensor interaction with the main regulation coming in the form of ABA-mediated phosphatase sequestering.

There is no evidence for an ER signal and/or retention signal so that PpANR does not appear destined to associate with ER interacting partners making the spatial separation hypothesis more plausible as some may just as easily bind to the ETRs at the ER or partners in the nucleus or at the plasma membrane. If PpANR was active in the ER where it functioned in both its roles it would require a sophisticated mechanism to translate the signal from one input into the specific response through specific interactions. The specificity could be due to the specificity of phosphoryl groups activating different active sites or modulating which partners can be bound. In this context, it is unlikely that PpANR can simultaneously be dephosphorylated and therefore activate ethylene signalling and phosphorylated and activate ABA-dependent stress signalling as this would require some extra mechanism to prevent one or the other being activated in normal conditions. If PpANR is required to be dephosphorylated for the activation of both pathways then some intermediate components are required to convert this into the phosphorylation of OST1 and/or the target TFs. There are many other possible ways of explaining the potential mechanisms involved but the most likely seems to be a spatial separation. However it is achieved, and achieved it must be, PpANR is likely compromised between the two roles and yet remains such a vital and global regulator of the ABA-dependent stress responses in moss.

6.5 Overview of the ABA-dependent signalling mediated by PpANR

As has been discussed, PpANR is essential for ABA-dependent stress responses and is therefore most likely to act in a linear fashion within the core ABA signal transduction pathway. The details of potential interactions have also been discussed including how a dual function might be resolved and what might act upstream and downstream of PpANR specifically. The suggested signalling framework is shown (Figure 6.1) however it will be summarized here with some key aspects important for testing specific hypotheses and guiding future work being discussed in more detail. This pathway has a special emphasis on *Physcomitrella* but is proposed to be the ancestral state of osmotic stress signalling likely present in the earliest land plants.

6.5.1 The activator module

A hypothetical signal transduction pathway that tightly integrates direct stress detection with ABA signalling is proposed. **(i)** The HKs homologous with the AtHK1 osmosensor are bound to the plasma membrane where extracellular domains may sense decreasing water availability and the onset of dehydration. **(ii)** These HKs regulate a canonical phosphorelay where dephosphorylation signals activation. **(iii)** The activated response regulator (RR) components (such as AtARR1 and AtARR2) activate PpANR (and possibly the Raf10/11 orthologues in other bryophytes) directly through protein-protein interactions with the PAS domain or possibly through intermediates although an interaction with the EDR domain remains a distinct possibility. (Some RRs are thought to have DNA binding activity similar to MYB binding TFs and so a possible ABA-independent and very direct pathway of osmotic stress signalling may exist. This might be linked to primary responses such as up-regulation of ABA biosynthesis and for example the strongly up-regulated *NCED1/2* genes in *Physcomitrella* (NCED1 and NCED2) have such MYB binding elements). **(iv)** Once PpANR and/or Raf10/11 is activated it starts the phosphorylation cascade that activates OST1 kinases although it is possible that other targets, such as those involved in canonical MAP kinase cascades (MAP3K-MAP2K-MAPK), might also exist. **(v)** The result is activation of key

Figure 6.1. Overview of ABA-dependent osmotic stress signalling model incorporating PpANR. PpANR is likely to act in a linear fashion with the ‘core’ pathway, possibly upstream of the OST1-like kinases. PpANR is represented as a trimodular structure in which the EDR domain mediates interactions with the ETR receptors and possibly the HK1-like osmosensors. The PAS domain’s function remains unknown but may function as regulator of ABA-dependent osmotic stress signalling given its presence in the *Arabidopsis* ABA regulators, Raf10/11. Cellular compartments are shown but many components are known to move in and out of the nucleus such as the ABI1/2 PP2Cs and the OST1-like kinases. The putative ancestral ‘activator’ module responsive to osmotic stress is highlighted in green while the more recently evolved ‘brake’ module is highlighted in red. Many interactions remain unknown as the interacting partners of PpANR remain unknown. The D/DT response involves up-regulation of genes encoding ‘protective’ proteins conferring tolerance while those encoding signalling components allow robust reinforcement of the signalling.

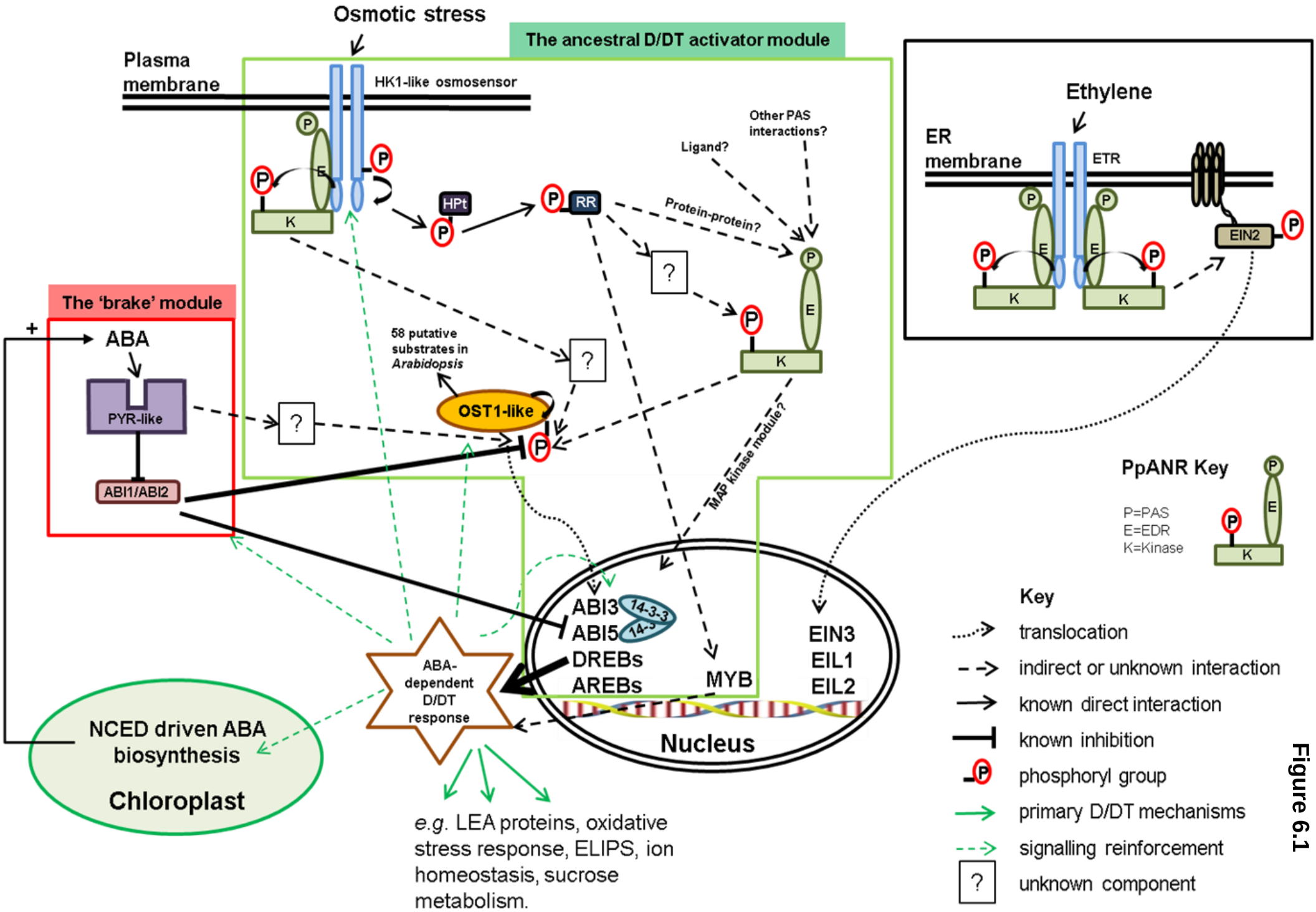

Osmotic stress
The ancestral D/DT activator module
Plasma membrane
HK1-like osmosensor
Ethylene
ER membrane
ETR
EIN2
Other PAS interactions?
Ligand?
Protein-protein?
HPt
RR
The 'brake' module
ABA
PYR-like
ABI1/ABI2
58 putative substrates in Arabidopsis
OST1-like
MAP kinase module?
PpANR Key
P=PAS
E=EDR
K=Kinase
ABI3
ABI5
14-3-3
DREBs
AREBs
MYB
EIN3
EIL1
EIL2
Nucleus
ABA-dependent D/DT response
NCED driven ABA biosynthesis
Chloroplast
e.g. LEA proteins, oxidative stress response, ELIPS, ion homeostasis, sucrose metabolism.
Key
translocation
indirect or unknown interaction
known direct interaction
known inhibition
phosphoryl group
primary D/DT mechanisms
signalling reinforcement
unknown component

Figure 6.1

ABA-dependent TFs which likely include the ABI3 homologues, ABI5-like bZIPs and possibly some DREB TFs. The resulting gene expression results in not only the expression of key ABA/stress-responsive genes involved in conferring D/DT but a number of key signalling components such as those involved in the 'core' pathway (including OST1-like, PYR-like, bZIPs), those inferred as important for ABA-dependent stress signalling (including the HK1-like osmosensor, a 14-3-3 GRF, DREB TFs) and importantly the key ABA biosynthetic NCED enzymes (as well as down-regulation of catabolism through CYOP707As).

6.5.2 The 'brake' module

This is the 'positive' pathway involved in the activation of D/DT but it must act against the repression by group A PP2Cs (ABI1 and ABI2). This acts as a brake and the balance of the brake against activation must be quickly broken for a rapid and robust ABA-dependent stress response conferring D/DT. In normal conditions, when there is no osmotic stress signalling, the phosphatase 'brake' must be sufficient to keep the whole pathway inactive. It is possible that this 'braking' mechanism is weaker in species with high constitutive levels of D/DT components such as *T. ruralis* either through the action of elevated ABA levels, gene dosage or through the strength of interactions. The breaking of the stalemate occurs primarily due to the up-regulation of ABA biosynthesis which allows these phosphatases to be sequestered from their inhibitory 'braking' role. The up-regulation of components all along this signal transduction pathway reinforces the response and the proteins involved in cellular stabilisation and protection and oxidative stress responses are synthesized minimizing the cellular damage experienced during dehydration and rehydration enabling the rapid resumption of normal photosynthesis and metabolism.

6.5.3 The problem: How does ABA alone exert regulation?

One issue the model does not explain is how exogenous ABA can strongly activate this pathway when no osmotic stress is experienced by the tissue in well hydrated, otherwise normal conditions. In this model, how can ABA itself activate PpANR and/or the OST1-like kinases and effectively mimic the 'ON' state of ABA-dependent signalling? Mechanisms within our current understanding, and specifically within the model presented here, do not fully explain this. One of the main difficulties stems from the apparently contradictory results observed in the *ppabi1/2* mutant. The first is that this mutant shows constitutive D/DT enabling it to survive complete desiccation – itself a trait only thought possible following ABA pretreatment – which is explained by the constitutive presence of key D/DT components such as LEA proteins and high levels of soluble sugars (Komatsu *et al*., 2013). This means that simply turning off the 'braking' module might be sufficient to induce D/DT responses if the 'activator' module has

‘intrinsic’ activity as suggested by Komatsu and colleagues. However, this did not appear to be due to the activity of the OST1-like kinases which were not constitutively phosphorylated but required exogenous ABA just like wild-type tissue (Komatsu *et al*., 2013). ABA mediates its regulation through the PYR-like receptors in the current model however no mutants in these receptors have been made in *Physcomitrella* or any bryophyte. The relatively high number of these components might suggest that more interactions might be important besides that with the ABI1/2 PP2Cs given there are 14 and 4 members in *Arabidopsis* and *Physcomitrella* respectively. A PYR-like sextuple mutant in *Arabidopsis* failed to show ABA-dependent OST1-like phosphorylation (Gonzalez-Guzman *et al*., 2012) suggesting that, if the same mechanisms are conserved with bryophytes, then it may be that there are other interactions between the PYR-like ABA receptors and possibly other PP2C members besides ABI1/2 (for example there are around 7 PP2Cs found to interact with one or more PRY-like receptor in *Arabidopsis*). Such interactions might mediate the phosphorylation and activation of OST1-like and possibly PpANR. For example, if there were multiple PP2Cs that repressed the autophosphorylation of OST1-like kinases and only two (ABI1/2) were removed then this might be sufficient for D/DT signalling through the intrinsic activation of the activator module but only when all PP2Cs are removed or sequestered will be see the full ABA-dependent phosphorylation of OST1-like kinases seen in the ppabi1/2 mutants (Komatsu *et al*., 2013). It is worth acknowledging too that other ABA receptors have been proposed which may be involved in the ABA-dependent phosphorylation of the OST1-like kinases however evidence for their roles and significance are lacking (reviewed by Cutler *et al*., 2010).

6.6 Evidence for the ancestral D/DT osmotic stress response

It could be suggested that a pathway involving direct osmotic stress signals through HK1-like osmosensors, the MAP3Ks PpANR and Raf10/11, OST1-like kinases, the TFs and the expression of the core stress-responsive genes conferring D/DT was likely present in the ancestral green algae as hypothesized by Komatsu and colleagues (2013) although a more extensive analysis of these components in charophytes is required. This is already starting to happen, however, as an OST1 homologue from *Klebsormidium nitens* has been shown to be functionally conserved in its ability to bind the SLAC1 channel protein: a key feature in conferring ABA dependent stomatal closure in bryophytes and vascular plants (Lind *et al*., 2015). Further, the transcriptomic response in *Klebsormidium crenluatum* has highlighted a putative AREB (bZIP TF), SnRK2 and PP2C components as being strongly up-regulated by desiccation (Holzinger and Becker, 2015). This pathway was likely functional before ABA became

so tightly integrated through the proposed land plant specific evolution of the PYR-like receptors and this possibility warrants further investigation.

6.7 Main mechanisms requiring answers

- Does the ANR-like orthologue in *S. pratensis* function as an ABA regulator?
- Do Raf10/11 orthologues share ABA-dependent stress signalling role in liverworts?
- Is the PAS domain crucial for ABA-dependent stress signalling?
- Are the HK1-like osmosensor homologues in bryophytes and charophytes the head of the ancestral osmotic stress D/DT pathway?

Of course, many more questions follow once any one of these questions are answered such as whether and how the PAS domain might function as an input domain? How is an active and dephosphorylated HK1-like pathway converted into an active and phosphorylated ANR-like/OST1/TF pathway? Does the EDR domain confer binding to the HK1-like osmosensors as well as the ETRs? Is phosphorylation of OST1 kinases and/or ABI3/AREBs lost in *PpanrKO*? One of the big unexplained issue's, as discussed, is how can ABA alone activate ABA-dependent phosphorylation of OST1-like kinases? Are there other PP2C phosphates that act to repress OST1-like kinase autophosphorylation? Do these OST1-like kinases have autophosphorylation activity in the bryophytes? The questions thrown up by the findings presented in this thesis and in recent papers are fascinating because they seek to understand not only fundamental questions about adaptive evolution but specifically how plant stress tolerance mechanisms came to be to allow first the conquest of land and then its domination.

References

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003) 'Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signalling', Plant Cell, 15(1), pp. 63-78.

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D. and Shinozaki, K. (1997) 'Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression', Plant Cell, 9(10), pp. 1859-1868.

Adamska, I. (1997) 'ELIPs - Light-induced stress proteins', Physiologia Plantarum, 100(4), pp. 794-805.

Aigner, S., Remias, D., Karsten, U. and Holzinger, A. (2013) 'Unusual phenolic compounds contribute to ecophysiological performance in the purple-colored green alga Zygogonium ericetorum (Zygnematphyceae, Streptophyta) from a high-alpine habitat', Journal of Phycology, 49(4), pp. 648-660.

Alfaro, M. E., Zoller, S. and Lutzoni, F. (2003) 'Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence', Molecular Biology and Evolution, 20(2), pp. 255-266.

Alpert, P. and Oliver, M. J. (2002) 'Drying without dying', Desiccation and survival in plants: Drying without dying, pp. 3-43.

Apel, K. and Hirt, H. (2004) 'Reactive oxygen species: Metabolism, oxidative stress, and signal transduction', Annual Review of Plant Biology, 55, pp. 373-399.

Ashton, N. W., Grimsley, N. H. and Cove, D. J. (1979) 'Analysis of gametophytic development in the moss, Physcomitrella patens, using auxin and cytokinin resistant mutants', Planta, 144(5), pp. 427-435.

Audran, C., Borel, C., Frey, A., Sotta, B., Meyer, C., Simonneau, T. and Marion-Poll, A. (1998) 'Expression studies of the zeaxanthin epoxidase gene in Nicotiana plumbaginifolia', Plant Physiology, 118(3), pp. 1021-1028.

Ayers, R. A. and Moffat, K. (2008) 'Changes in Quaternary Structure in the Signalling Mechanisms of PAS Domains', Biochemistry, 47(46), pp. 12078-12086.

Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F. and Covarrubias, A. A. (2008) 'The enigmatic LEA proteins and other hydrophilins', Plant Physiology, 148(1), pp. 6-24.

Becker, D. F., Zhu, W. and Moxley, M. A. (2011) 'Flavin Redox Switching of Protein Functions', Antioxidants & Redox Signalling, 14(6), pp. 1079-1091.

Beerli, P. (2006) 'Comparison of Bayesian and maximum-likelihood inference of population genetic parameters', Bioinformatics, 22(3), pp. 341-345.

Beike, A. K., Lang, D., Zimmer, A. D., Wuest, F., Trautmann, D., Wiedemann, G., Beyer, P., Decker, E. L. and Reski, R. (2015) 'Insights from the cold transcriptome of Physcomitrella patens: global specialization pattern of conserved transcriptional regulators and identification of orphan genes involved in cold acclimation', New Phytologist, 205(2), pp. 869-881.

Bensmihen, S., Rippa, S., Lambert, G., Jublot, D., Pautot, V., Granier, F., Giraudat, J. and Parcy, F. (2002) 'The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis', Plant Cell, 14(6), pp. 1391-1403.

Bersten, D. C., Sullivan, A. E., Peet, D. J. and Whitelaw, M. L. (2013) 'bHLH-PAS proteins in cancer', Nature Reviews Cancer, 13(12), pp. 827-841.

Bewley, J. D. (1979) 'Physiological-aspects of desiccation tolerance', Annual Review of Plant Physiology and Plant Molecular Biology, 30, pp. 195-238.

Bianchi, G., Gamba, A., Limiroli, R., Pozzi, N., Elster, R., Salamini, F. and Bartels, D. (1993) 'The unusual sugar composition in leaves of the resurrection plant Myrothamnus-flabellifolia', Physiologia Plantarum, 87(2), pp. 223-226.

Blackman, P. G. and Davies, W. J. (1985) 'Root to shoot communication in maize plants of the effects of soil drying', Journal of Experimental Botany, 36(162), pp. 39-48.

Borgstahl, G. E. O., Williams, D. R. and Getzoff, E. D. (1995) '1.4 angstrom structure of photoactive yellow protein, a cytosolic photoreceptor - unusual fold, active-site, and chromophore', Biochemistry, 34(19), pp. 6278-6287.

Champion, A., Jouannic, S., Guillon, S., Mockaitis, K., Krapp, A., Picaud, A., Simanis, V., Kreis, M. and Henry, Y. (2004a) 'AtSGP1, AtSGP2 and MAP4K alpha are nucleolar plant proteins that can complement fission yeast mutants lacking a functional SIN pathway', Journal of Cell Science, 117(18), pp. 4265-4275.

Champion, A., Picaud, A. and Henry, Y. (2004b) 'Reassessing the MAP3K and MAP4K relationships', Trends in Plant Science, 9(3), pp. 123-129.

Chang, C., Tesar, C., Gu, M., Babnigg, G., Joachimiak, A., Pokkuluri, P. R., Szurmant, H. and Schiffer, M. (2010) 'Extracytoplasmic PAS-Like Domains Are Common in Signal Transduction Proteins', Journal of Bacteriology, 192(4), pp. 1156-1159.

Chang, C.-Y., Lin, W.-D. and Tu, S.-L. (2014) 'Genome-Wide Analysis of Heat-Sensitive Alternative Splicing in Physcomitrella patens', Plant Physiology, 165(2), pp. 826-840.

Chater, C., Kamisugi, Y., Movahedi, M., Fleming, A., Cuming, A. C., Gray, J. E. and Beerling, D. J. (2011) 'Regulatory Mechanism Controlling Stomatal Behavior Conserved across 400 Million Years of Land Plant Evolution', Current Biology, 21(12), pp. 1025-1029.

Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) 'MolProbity: all-atom structure validation for macromolecular crystallography', Acta Crystallographica Section D-Biological Crystallography, 66, pp. 12-21.

Choi, H. I., Hong, J. H., Ha, J. O., Kang, J. Y. and Kim, S. Y. (2000) 'ABFs, a family of ABA-responsive element binding factors', Journal of Biological Chemistry, 275(3), pp. 1723-1730.

Christie, J. M. (2007) 'Phototropin blue-light receptors', Annual Review of Plant Biology, 58, pp. 21-45.

Civan, P., Foster, P. G., Embley, M. T., Seneca, A. and Cox, C. J. (2014) 'Analyses of Charophyte Chloroplast Genomes Help Characterize the Ancestral Chloroplast Genome of Land Plants', Genome Biology and Evolution, 6(4), pp. 897-911.

Clark, K. L., Larsen, P. B., Wang, X. X. and Chang, C. (1998) 'Association of the Arabidopsis CTR1 Raf-like kinase with the ETR1 and ERS ethylene receptors', Proceedings of the National Academy of Sciences of the United States of America, 95(9), pp. 5401-5406.

Colcombet, J. and Hirt, H. (2008) 'Arabidopsis MAPKs: a complex signalling network involved in multiple biological processes', Biochemical Journal, 413, pp. 217-226.

Collins, F. S., Brooks, L. D. and Chakravarti, A. (1998) 'A DNA polymorphism discovery resource for research on human genetic variation', Genome Research, 8(12), pp. 1229-1231.

Conant, G. C. and Wagner, A. (2003) 'Asymmetric sequence divergence of duplicate genes', Genome Research, 13(9), pp. 2052-2058.

Cooke, T. J., Poli, D., Sztein, A. E. and Cohen, J. D. (2002) 'Evolutionary patterns in auxin action', Plant Molecular Biology, 49(3-4), pp. 319-338.

Crowe, J. H., Crowe, L. M. and Chapman, D. (1984) 'Preservation of membranes in anhydrobiotic organisms - the role of trehalose', Science, 223(4637), pp. 701-703.

Cruz de Carvalho, M. H. (2008) 'Drought stress and reactive oxygen species: Production, scavenging and signalling', Plant signalling & behavior, 3(3), pp. 156-65.

Cuming, A.C. (1999) 'LEA proteins', In Shewry, P.R. Casey, R. eds, Seed Proteins. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 753–780.

Cuming, A. C., Cho, S. H., Kamisugi, Y., Graham, H. and Quatrano, R. S. (2007) 'Microarray analysis of transcriptional responses to abscisic acid and osmotic, salt, and drought stress in the moss, Physcomitrella patens', New Phytologist, 176(2), pp. 275-287.

Cutler, S. R., Rodriguez, P. L., Finkelstein, R. R. and Abrams, S. R. (2010) 'Abscisic Acid: Emergence of a Core Signalling Network', Annual Review of Plant Biology, Vol 61, 61, pp. 651-679.

Decker, E. L., Frank, W., Sarnighausen, E. and Reski, R. (2006) 'Moss systems biology en route: Phytohormones in Physcomitrella development', Plant Biology, 8(3), pp. 397-405.

Denison, M. S. and Nagy, S. R. (2003) 'Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals', Annual Review of Pharmacology and Toxicology, 43, pp. 309-334.

Douady, C. J., Delsuc, F., Boucher, Y., Doolittle, W. F. and Douzery, E. J. P. (2003) 'Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability', Molecular Biology and Evolution, 20(2), pp. 248-254.

Durrett, R. T., Chen, K. Y. and Tanksley, S. D. (2002) 'A simple formula useful for positional cloning', Genetics, 160(1), pp. 353-355.

Edwards, D., Duckett, J. G. and Richardson, J. B. (1995) 'Hepatic characters in the earliest land plants', Nature, 374(6523), pp. 635-636.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) 'Features and development of Coot', Acta Crystallographica Section D-Biological Crystallography, 66, pp. 486-501.

Engh, R. A. and Huber, R. (1991) 'Accurate bond and angle parameters for X-ray protein-structure refinement', Acta Crystallographica Section A, 47, pp. 392-400.

Erbel, P. J. A., Card, P. B., Karakuzu, O., Bruick, R. K. and Gardner, K. H. (2003) 'Structural basis for PAS domain heterodimerization in the basic helix-loop-helix-PAS transcription factor hypoxia-inducible factor', Proceedings of the National Academy of Sciences of the United States of America, 100(26), pp. 15504-15509.

Erxleben, A., Gessler, A., Vervliet-Scheebaum, M. and Reski, R. (2012) 'Metabolite profiling of the moss Physcomitrella patens reveals evolutionary conservation of osmoprotective substances', Plant Cell Reports, 31(2), pp. 427-436.

Feilner, T., Hultschig, C., Lee, J., Meyer, S., Immink, R. G. H., Koenig, A., Possling, A., Seitz, H., Beveridge, A., Scheel, D., Cahill, D. J., Lehrach, H., Kreutzberger, J. and Kersten, B. (2005) 'High throughput identification of potential Arabidopsis mitogen-activated protein kinases substrates', Molecular & Cellular Proteomics, 4(10), pp. 1558-1568.

Fitch, W. M. and Margolia.E (1967) 'Construction of phylogenetic trees', Science, 155(3760), pp. 279-&.

Flagel, L. E. and Wendel, J. F. (2009) 'Gene duplication and evolutionary novelty in plants', New Phytologist, 183(3), pp. 557-564.

Forde, B. G. and Lea, P. J. (2007) 'Glutamate in plants: metabolism, regulation, and signalling', Journal of Experimental Botany, 58(9), pp. 2339-2358.

Frye, C. A. and Innes, R. W. (1998) 'An Arabidopsis mutant with enhanced resistance to powdery mildew', Plant Cell, 10(6), pp. 947-956.

Frye, C. A., Tang, D. Z. and Innes, R. W. (2001) 'Negative regulation of defense responses in plants by a conserved MAPKK kinase', Proceedings of the National Academy of Sciences of the United States of America, 98(1), pp. 373-378.

Fujii, H., Chinnusamy, V., Rodrigues, A., Rubio, S., Antoni, R., Park, S. Y., Cutler, S. R., Sheen, J., Rodriguez, P. L. and Zhu, J. K. (2009) 'In vitro reconstitution of an abscisic acid signalling pathway', Nature, 462(7273), pp. 660-664.

Furihata, T., Maruyama, K., Fujita, Y., Umezawa, T., Yoshida, R., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2006) 'Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1', Proceedings of the National Academy of Sciences of the United States of America, 103(6), pp. 1988-1993.

Gao, B., Zhang, D., Li, X., Yang, H., Zhang, Y. and Wood, A. J. (2015) 'De novo transcriptome characterization and gene expression profiling of the desiccation tolerant moss Bryum argenteum following rehydration', Bmc Genomics, 16.

Gao, L. and Xiang, C.-B. (2008) 'The genetic locus At1g73660 encodes a putative MAPKKK and negatively regulates salt tolerance in Arabidopsis', Plant Molecular Biology, 67(1-2), pp. 125-134.

Gilles-Gonzalez, M. A. and Gonzalez, G. (2004) 'Signal transduction by heme-containing PAS-domain proteins', Journal of Applied Physiology, 96(2), pp. 774-783.

Gonzalez-Guzman, M., Pizzio, G. A., Antoni, R., Vera-Sirera, F., Merilo, E., Bassel, G. W., Fernandez, M. A., Holdsworth, M. J., Angel Perez-Amador, M., Kollist, H. and Rodriguez, P. L. (2012) 'Arabidopsis PYR/PYL/RCAR Receptors Play a Major Role in Quantitative Regulation of Stomatal Aperture and Transcriptional Response to Abscisic Acid', Plant Cell, 24(6), pp. 2483-2496.

Graham, L. E. (1996) 'Green algae to land plants: An evolutionary transition', Journal of Plant Research, 109(1095), pp. 241-251.

Graham, L. E., Cook, M. E. and Busse, J. S. (2000) 'The origin of plants: Body plan changes contributing to a major evolutionary radiation', Proceedings of the National Academy of Sciences of the United States of America, 97(9), pp. 4535-4540.

Gu, Y. Z., Hogenesch, J. B. and Bradfield, C. A. (2000) 'The PAS superfamily: Sensors of environmental and developmental signals', Annual Review of Pharmacology and Toxicology, 40, pp. 519-561.

Guenl, M. and Pauly, M. (2011) 'AXY3 encodes a alpha-xylosidase that impacts the structure and accessibility of the hemicellulose xyloglucan in Arabidopsis plant cell walls', Planta, 233(4), pp. 707-719.

Hahn, A. and Harter, K. (2009) 'Mitogen-Activated Protein Kinase Cascades and Ethylene: Signalling, Biosynthesis, or Both?', Plant Physiology, 149(3), pp. 1207-1210.

Hanada, K., Hase, T., Toyoda, T., Shinozaki, K. and Okamoto, M. (2011) 'Origin and evolution of genes related to ABA metabolism and its signalling pathways', Journal of Plant Research, 124(4), pp. 455-465.

Hanafusa, N. (1969) 'Denaturation of enzyme proteins by freeze drying part 1 freeze drying of myosin and catalase and the effect of some additives on the denaturation', Low Temperature Science Series B Biological Sciences, (27), pp. 11-22.

Harper, S. M., Christie, J. M. and Gardner, K. H. (2004) 'Disruption of the LOV-J alpha helix interaction activates phototropin kinase activity', Biochemistry, 43(51), pp. 16184-16192.

Harper, S. M., Neil, L. C. and Gardner, K. H. (2003) 'Structural basis of a phototropin light switch', Science, 301(5639), pp. 1541-1544.

Hartung, W. (2010) 'The evolution of abscisic acid (ABA) and ABA function in lower plants, fungi and lichen', Functional Plant Biology, 37(9), pp. 806-812.

Hashimoto, M., Negi, J., Young, J., Israelsson, M., Schroeder, J. I. and Iba, K. (2006) 'Arabidopsis HT1 kinase controls stomatal movements in response to CO2', Nature Cell Biology, 8(4), pp. 391-U52.

Hefti, M. H., Francoijs, K. J., de Vries, S. C., Dixon, R. and Vervoort, J. (2004) 'The PAS fold - A redefinition of the PAS domain based upon structural prediction', European Journal of Biochemistry, 271(6), pp. 1198-1208.

Henry, J. T. and Crosson, S. (2011) 'Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context', Annual Review of Microbiology, Vol 65, 65, pp. 261-286.

Hill, A., Nantel, A., Rock, C. D. and Quatrano, R. S. (1996) 'A conserved domain of the viviparous-1 gene product enhances the DNA binding activity of the bZIP protein EmBP-1 and other transcription factors', Journal of Biological Chemistry, 271(7), pp. 3366-3374.

Hirsch, R., Hartung, W. and Gimmler, H. (1989) 'Abscisic-acid content of algae under stress', Botanica Acta, 102(4), pp. 326-334.

Hiss, M., Laule, O., Meskauskiene, R. M., Arif, M. A., Decker, E. L., Erxleben, A., Frank, W., Hanke, S. T., Lang, D., Martin, A., Neu, C., Reski, R., Richardt, S., Schallenberg-Ruedinger, M., Szoevenyi, P., Tiko, T., Wiedemann, G., Wolf, L., Zimmermann, P. and Rensing, S. A. (2014) 'Large-scale gene expression profiling data for the model moss Physcomitrella patens aid understanding of developmental progression, culture and stress conditions', Plant Journal, 79(3), pp. 530-539.

Hoang, M. H. T., Nguyen, X. C., Lee, K., Kwon, Y. S., Pham, H. T. T., Park, H. C., Yun, D.-J., Lim, C. O. and Chung, W. S. (2012) 'Phosphorylation by AtMPK6 is required for the biological function of AtMYB41 in Arabidopsis', Biochemical and biophysical research communications, 422(1), pp. 181-6.

Hoekstra, F. A., Golovina, E. A., Van Aelst, A. C. and Hemminga, M. A. (1999) 'Imbibitional leakage from anhydrobiotes revisited', Plant Cell and Environment, 22(9), pp. 1121-1131.

Hohe, A. and Reski, R. (2003) 'A tool for understanding homologous recombination in plants', Plant Cell Reports, 21(12), pp. 1135-1142.

Holzinger, A. and Becker, B (2015) ‘Desiccation tolerance in the streptophyte green alga Klebsormidium: The role of phytohormones’, Communicative & Integrative Biology, 8:4, e1059978.

Holzinger, A., Kaplan, F., Blaas, K., Zechmann, B., Komsic-Buchmann, K. and Becker, B. (2014) 'Transcriptomics of desiccation tolerance in the streptophyte green alga Klebsormidium reveal a land plant-like defense reaction', Plos One, 9(10).

Huang, J., Yue, J. and Hu, X. (2014a) 'Origin of plant auxin biosynthesis in charophyte algae: a reply to Wang *et al*', Trends in Plant Science, 19(12), pp. 743-743.

Huang, Y., Li, C. Y., Qi, Y., Park, S. and Gibson, S. I. (2014b) 'SIS8, a putative mitogen-activated protein kinase kinase kinase, regulates sugar-resistant seedling development in Arabidopsis', Plant Journal, 77(4), pp. 577-588.

Huang, Y. F., Li, H., Hutchison, C. E., Laskey, J. and Kieber, J. J. (2003) 'Biochemical and functional analysis of CTR1, a protein kinase that negatively regulates ethylene signalling in Arabidopsis', Plant Journal, 33(2), pp. 221-233.

Ichimura, K., Shinozaki, K., Tena, G., Sheen, J., Henry, Y., Champion, A., Kreis, M., Zhang, S. Q., Hirt, H., Wilson, C., Heberle-Bors, E., Ellis, B. E., Morris, P. C., Innes, R. W., Ecker, J. R., Scheel, D., Klessig, D. F., Machida, Y., Mundy, J., Ohashi, Y., Walker, J. C. and Grp, M. (2002) 'Mitogen-activated protein kinase cascades in plants: a new nomenclature', Trends in Plant Science, 7(7), pp. 301-308.

Ingram, J. and Bartels, D. (1996) 'The molecular basis of dehydration tolerance in plants', Annual Review of Plant Physiology and Plant Molecular Biology, 47, pp. 377-403.

Iwasaki, T., Yamaguchishinozaki, K. and Shinozaki, K. (1995) 'Identification of a cis-regulatory region of a gene in Arabidopsis thaliana whose induction by dehydration is mediated by abscisic-acid and requires protein-synthesis', Molecular and General Genetics, 247(4), pp. 391-398.

Jackson, M. B. (2008) 'Ethylene-promoted elongation: An adaptation to submergence stress', Annals of Botany, 101(2), pp. 229-248.

Jaiswal, R. K., Manjeera, G. and Gopal, B. (2010) 'Role of a PAS sensor domain in the Mycobacterium tuberculosis transcription regulator Rv1364c', Biochemical and Biophysical Research Communications, 398(3), pp. 342-349.

Jonker, N., Kool, J., Krabbe, J. G., Retra, K., Lingeman, H. and Irth, H. (2008) 'Screening of protein-ligand interactions using dynamic protein-affinity chromatography solid-phase extraction-liquid chromatography-mass spectrometry', Journal of Chromatography A, 1205(1-2), pp. 71-77.

Ju, C. Van de Poel, B. Cooper, E.D. Thierer, J.H, Gibbons, T.R. Delwiche, C.F. Chang, C. (2015) ‘Conservation of ethylene as a plant hormone over 450 million years of evolution’, Nature Plants 1: 14004.

Kaelin, W. G., Jr. (2011) 'Cancer and altered metabolism: potential importance of hypoxia-inducible factor and 2-oxoglutarate-dependent dioxygenases', Cold Spring Harbor symposia on quantitative biology, 76, pp. 335-45.

Kamisugi, Y. and Cuming, A. C. (2005) 'The evolution of the abscisic acid-response in land plants: comparative analysis of group 1 LEA gene expression in moss and cereals', Plant Molecular Biology, 59(5), pp. 723-737.

Kamisugi, Y., Cuming, A. C. and Cove, D. J. (2005) 'Parameters determining the efficiency of gene targeting in the moss Physcomitrella patens', Nucleic Acids Research, 33(19), pp. 10.

Kamisugi, Y., Schaefer, D. G., Kozak, J., Charlot, F., Vrielynck, N., Hola, M., Angelis, K. J., Cuming, A. C. and Nogue, F. (2012) 'MRE11 and RAD50, but not NBS1, are essential for gene targeting in the moss Physcomitrella patens', Nucleic Acids Research, 40(8), pp. 3496-3510.

Kamisugi, Y., Schlink, K., Rensing, S. A., Schween, G., von Stackelberg, M., Cuming, A. C., Reski, R. and Cove, D. J. (2006) 'The mechanism of gene targeting in Physcomitrella patens: homologous recombination, concatenation and multiple integration', Nucleic Acids Research, 34(21), pp. 6205-6214.

Kamisugi, Y., von Stackelberg, M., Lang, D., Care, M., Reski, R., Rensing, S. A. and Cuming, A. C. (2008) 'A sequence-anchored genetic linkage map for the moss, Physcomitrella patens', Plant Journal, 56(5), pp. 855-866.

Kang, J., Hwang, J. U., Lee, M., Kim, Y. Y., Assmann, S. M., Martinoia, E. and Lee, Y. (2010) 'PDR-type ABC transporter mediates cellular uptake of the phytohormone abscisic acid', Proceedings of the National Academy of Sciences of the United States of America, 107(5), pp. 2355-2360.

Kanno, Y., Hanada, A., Chiba, Y., Ichikawa, T., Nakazawa, M., Matsui, M., Koshiba, T., Kamiya, Y. and Seo, M. (2012) 'Identification of an abscisic acid transporter by functional screening using the receptor complex as a sensor', Proceedings of the National Academy of Sciences of the United States of America, 109(24), pp. 9653-9658.

Kawai, H., Kanegae, T., Christensen, S., Kiyosue, T., Sato, Y., Imaizumi, T., Kadota, A. and Wada, M. (2003) 'Responses of ferns to red light are mediated by an unconventional photoreceptor', Nature, 421(6920), pp. 287-290.

Kenrick, P. and Crane, P. R. (1997) 'The origin and early evolution of plants on land', Nature, 389(6646), pp. 33-39.

Kenrick, P., Wellman, C. H., Schneider, H. and Edgecombe, G. D. (2012) 'A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic', Philosophical Transactions of the Royal Society B-Biological Sciences, 367(1588), pp. 519-536.

Khandelwal, A., Cho, S. H., Marella, H., Sakata, Y., Perroud, P. F., Pan, A. and Quatrano, R. S. (2010) 'Role of ABA and ABI3 in Desiccation Tolerance', Science, 327, pp. 546.

Kieber, J. J., Rothenberg, M., Roman, G., Feldmann, K. A. and Ecker, J. R. (1993) 'CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the Raf family of protein-kinases', Cell, 72(3), pp. 427-441.

Kim, D. and Forst, S. (2001) 'Genomic analysis of the histidine kinase family in bacteria and archaea', Microbiology-Sgm, 147, pp. 1197-1212.

Knight, C. D., Cove, D. J., Cuming, A. C. and Quatrano, R. S. (2002) 'Moss gene technology', Molecular plant biology: A practical approach. Volume Two, pp. 285–301.

Koiwai, H., Nakaminami, K., Seo, M., Mitsuhashi, W., Toyomasu, T. and Koshiba, T. (2004) 'Tissue-specific localization of an abscisic acid biosynthetic enzyme, AAO3, in Arabidopsis', Plant Physiology, 134(4), pp. 1697-1707.

Komatsu, K., Nishikawa, Y., Ohtsuka, T., Taji, T., Quatrano, R. S., Tanaka, S. and Sakata, Y. (2009) 'Functional analyses of the ABI1-related protein phosphatase type 2C reveal evolutionarily conserved regulation of abscisic acid signalling between Arabidopsis and the moss Physcomitrella patens', Plant Molecular Biology, 70(3), pp. 327-340.

Komatsu, K., Suzuki, N., Kuwamura, M., Nishikawa, Y., Nakatani, M., Ohtawa, H., Takezawa, D., Seki, M., Tanaka, M., Taji, T., Hayashi, T. and Sakata, Y. (2013) 'Group A PP2Cs evolved in land plants as key regulators of intrinsic desiccation tolerance', Nature Communications, 4.

Koster, K. L., Balsamo, R. A., Espinoza, C. and Oliver, M. J. (2010) 'Desiccation sensitivity and tolerance in the moss Physcomitrella patens: assessing limits and damage', Plant Growth Regulation, 62(3), pp. 293-302.

Krasensky, J. and Jonak, C. (2012) 'Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks', Journal of Experimental Botany, 63(4), pp. 1593-1608.

Kroj, T., Savino, G., Valon, C., Giraudat, J. and Parcy, F. (2003) 'Regulation of storage protein gene expression in Arabidopsis', Development, 130(24), pp. 6065-6073.

Kroken, S. B., Graham, L. E. and Cook, M. E. (1996) 'Occurrence and evolutionary significance of resistant cell walls in charophytes and bryophytes', American Journal of Botany, 83(10), pp. 1241-1254.

Kumar, K., Rao, K. P., Sharma, P. and Sinha, A. K. (2008) 'Differential regulation of rice mitogen activated protein kinase kinase (MKK) by abiotic stress', Plant Physiology and Biochemistry, 46(10), pp. 891-897.

Kuromori, T., Miyaji, T., Yabuuchi, H., Shimizu, H., Sugimoto, E., Kamiya, A., Moriyama, Y. and Shinozaki, K. (2010) 'ABC transporter AtABCG25 is involved in abscisic acid transport and responses', Proceedings of the National Academy of Sciences of the United States of America, 107(5), pp. 2361-2366.

Lang, D., Zimmer, A. D., Rensing, S. A. and Reski, R. (2008) 'Exploring plant biodiversity: the Physcomitrella genome and beyond', Trends in Plant Science, 13(10), pp. 542-549.

Lee, S.-b., Lee, S.-j. and Kim, S. Y. (2015a) 'AtERF15 is a positive regulator of ABA response', Plant Cell Reports, 34(1), pp. 71-81.

Lee, S.-J., Lee, M. H., Kim, J.-I. and Kim, S. Y. (2015b) 'Arabidopsis Putative MAP Kinase Kinase Kinases Raf10 and Raf11 are Positive Regulators of Seed Dormancy and ABA Response', Plant and Cell Physiology, 56(1), pp. 84-97.

Lee, S.-j., Park, J. H., Lee, M. H., Yu, J.-h. and Kim, S. Y. (2010) 'Isolation and functional characterization of CE1 binding proteins', Bmc Plant Biology, 10.

Lei, L., Li, Y., Wang, Q., Xu, J., Chen, Y., Yang, H. and Ren, D. (2014) 'Activation of MKK9-MPK3/MPK6 enhances phosphate acquisition in Arabidopsis thaliana', New Phytologist, 203(4), pp. 1146-1160.

Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F. and De Clerck, O. (2012) 'Phylogeny and Molecular Evolution of the Green Algae', Critical Reviews in Plant Sciences, 31(1), pp. 1-46.

Li, F.-W., Melkonian, M., Rothfels, C. J., Villarreal, J. C., Stevenson, D. W., Graham, S. W., Wong, G. K.-S., Pryer, K. M. and Mathews, S. (2015) 'Phytochrome diversity in green plants and the origin of canonical plant phytochromes', Nature Communications, 6.

Li, F.-W., Villarreal, J. C., Kelly, S., Rothfels, C. J., Melkonian, M., Frangedakis, E., Ruhsam, M., Sigel, E. M., Der, J. P., Pittermann, J., Burge, D. O., Pokornyk, L., Larsson, A., Chen, T., Weststrand, S., Thomas, P., Carpenter, E., Zhang, Y., Tian, Z., Chen, L., Yan, Z., Zhu, Y., Sun, X., Wang, J., Stevenson, D. W., Crandall-Stotler, B. J., Shaw, A. J., Deyholos, M. K., Soltis, D. E., Graham, S. W., Windham, M. D., Langdale, J. A., Wong, G. K.-S., Mathews, S. and Pryer, K. M. (2014) 'Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns', Proceedings of the National Academy of Sciences of the United States of America, 111(18), pp. 6672-6677.

Li, J., Mahajan, A. and Tsai, M.-D. (2006) 'Ankyrin repeat: A unique motif mediating protein-protein interactions', Biochemistry, 45(51), pp. 15168-15178.

Ligrone, R., Carafa, A., Duckett, J. G., Renzaglia, K. S. and Ruel, K. (2008) 'Immunocytochemical detection of lignin-related epitopes in cell walls in bryophytes and the charalean alga Nitella', Plant Systematics and Evolution, 270(3-4), pp. 257-272.

Lind, C., Dreyer, I., Lopez-Sanjurjo, E. J., von Meyer, K., Ishizaki, K., Kohchi, T., Lang, D., Zhao, Y., Kreuzer, I., Al-Rasheid, K. A. S., Ronne, H., Reski, R., Zhu, J.-K., Geiger, D. and Hedrich, R. (2015) 'Stomatal Guard Cells Co-opted an Ancient ABA-Dependent Desiccation Survival System to Regulate Stomatal Closure', Current Biology, 25(7), pp. 928-935.

Lindemose, S., O'Shea, C., Jensen, M. K. and Skriver, K. (2013) 'Structure, Function and Networks of Transcription Factors Involved in Abiotic Stress Responses', International Journal of Molecular Sciences, 14(3), pp. 5842-5878.

Liu, S., Wang, N., Zhang, P., Cong, B., Lin, X., Wang, S., Xia, G. and Huang, X. (2013) 'Next-generation sequencing-based transcriptome profiling analysis of Pohlia nutans reveals insight into the stress-relevant genes in Antarctic moss', Extremophiles, 17(3), pp. 391-403.

Lopez-Molina, L., Mongrand, B., McLachlin, D. T., Chait, B. T. and Chua, N. H. (2002) 'ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination', Plant Journal, 32(3), pp. 317-328.

Lu, Y. and Xu, J. (2015) 'Phytohormones in microalgae: a new opportunity for microalgal biotechnology?', Trends in Plant Science, 20(5), pp. 273-282.

Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A. and Grill, E. (2009) 'Regulators of PP2C Phosphatase Activity Function as Abscisic Acid Sensors', Science, 324, pp. 1064-1068.

Mahan, M. J. and Csonka, L. N. (1983) 'Genetic-analysis of the proBA genes of Salmonella typhimurium - physical and genetic analyses of the cloned proB+ A+ genes of Escherichia coli and of a mutant allele that confers proline overproduction and enhanced osmotolerance', Journal of Bacteriology, 156(3), pp. 1249-1262.

Mascher, T., Helmann, J. D. and Unden, G. (2006) 'Stimulus perception in bacterial signal-transducing histidine kinases', Microbiology and Molecular Biology Reviews, 70(4), pp. 910-+.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) 'Phaser crystallographic software', Journal of Applied Crystallography, 40, pp. 658-674.

McIntosh, B. E., Hogenesch, J. B. and Bradfield, C. A. (2010) 'Mammalian Per-Arnt-Sim Proteins in Environmental Adaptation', Annual Review of Physiology, 72, pp. 625-645.

Meinke, D. W., Franzmann, L. H., Nickle, T. C. and Yeung, E. C. (1994) 'Leafy cotyledon mutants of Arabidopsis', Plant Cell, 6(8), pp. 1049-1064.

Melcher, K., Ng, L.-M., Zhou, X. E., Soon, F.-F., Xu, Y., Suino-Powell, K. M., Park, S.-Y., Weiner, J. J., Fujii, H., Chinnusamy, V., Kovach, A., Li, J., Wang, Y., Li, J., Peterson, F. C., Jensen, D. R., Yong, E.-L., Volkman, B. F., Cutler, S. R., Zhu, J.-K. and Xu, H. E. (2009) 'A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors', Nature, 462, pp. 602-608.

Mikkelsen, M. D., Harholt, J., Ulvskov, P., Johansen, I. E., Fangel, J. U., Doblin, M. S., Bacic, A. and Willats, W. G. T. (2014) 'Evidence for land plant cell wall biosynthetic mechanisms in charophyte green algae', Annals of Botany, 114(6), pp. 1217-1236.

Minami, A., Nagao, M., Arakawa, K., Fujikawa, S. and Takezawa, D. (2003) 'Abscisic acid-induced freezing tolerance in the moss Physcomitrella patens is accompanied by increased expression of stress-related genes', Journal of Plant Physiology, 160(5), pp. 475-483.

Minami, A., Nagao, M., Ikegami, K., Koshiba, T., Arakawa, K., Fujikawa, S. and Takezawa, D. (2005) 'Cold acclimation in bryophytes: low-temperature-induced freezing tolerance in Physcomitrella patens is associated with increases in expression levels of stress-related genes but not with increase in level of endogenous abscisic acid', Planta, 220(3), pp. 414-423.

Minic, Z. (2008) 'Physiological roles of plant glycoside hydrolases', Planta, 227(4), pp. 723-740.

Miyazono, K.-i., Miyakawa, T., Sawano, Y., Kubota, K., Kang, H.-J., Asano, A., Miyauchi, Y., Takahashi, M., Zhi, Y., Fujita, Y., Yoshida, T., Kodaira, K.-S., Yamaguchi-Shinozaki, K. and Tanokura, M. (2009) 'Structural basis of abscisic acid signalling', Nature, 462(7273), pp. 609-U79.

Möglich, A., Ayers, R. A. and Moffat, K. (2009a) 'Structure and Signalling Mechanism of Per-ARNT-Sim Domains', Structure, 17(10), pp. 1282-1294.

Möglich, A., Ayers, R. A. and Moffat, K. (2009b) 'Design and Signalling Mechanism of Light-Regulated Histidine Kinases', Journal of Molecular Biology, 385(5), pp. 1433-1444.

Moran, T. V. and Walker, J. C. (1993) 'Molecular-cloning of 2 novel protein-kinase genes from Arabidopsis thaliana', Biochimica et Biophysica Acta, 1216(1), pp. 9-14.

Morgan, W. T., Muster, P., Tatum, F., Kao, S. M., Alam, J. and Smith, A. (1993) 'Identification of the histidine-residues of hemopexin that coordinate with heme-iron and of a receptor-binding region', Journal of Biological Chemistry, 268(9), pp. 6256-6262.

Moustafa, K., AbuQamar, S., Jarrar, M., Al-Rajab, A. J. and Tremouillaux-Guiller, J. (2014) 'MAPK cascades and major abiotic stresses', Plant Cell Reports, 33(8), pp. 1217-1225.

Moustafa, K., Lefebvre-De Vos, D., Leprince, A-S., Savourée, A., Laurière, C. (2008) ‘Analysis of the Arabidopsis mitogen-activated protein kinase families: organ specificity and transcriptional regulation upon water stresses’, Scholarly Research Exchange. p 12.

Musgrave, A., Jackson, M. B. and Ling, E. (1972) 'Callitriche stem elongation is controlled by ethylene and gibberellin', Nature-New Biology, 238(81), pp. 93-&.

Nakamura, S., Lynch, T. J. and Finkelstein, R. R. (2001) 'Physical interactions between ABA response loci of Arabidopsis', Plant Journal, 26(6), pp. 627-635.

Nakasako, M., Zikihara, K., Matsuoka, D., Katsura, H. and Tokutomi, S. (2008) 'Structural basis of the LOV1 dimerization of Arabidopsis phototropins 1 and 2', Journal of Molecular Biology, 381(3), pp. 718-733.

Nambu, J. R., Lewis, J. O., Wharton, K. A. and Crews, S. T. (1991) 'The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development', Cell, 67(6), pp. 1157-1167.

Nguyen, L. P. and Bradfield, C. A. (2008) 'The search for endogenous activators of the aryl hydrocarbon receptor', Chemical Research in Toxicology, 21(1), pp. 102-116.

Ning, J., Li, X., Hicks, L. M. and Xiong, L. (2010) 'A Raf-Like MAPKKK Gene DSM1 Mediates Drought Resistance through Reactive Oxygen Species Scavenging in Rice', Plant Physiology, 152(2), pp. 876-890.

Nishiyama, T., Wolf, P. G., Kugita, M., Sinclair, R. B., Sugita, M., Sugiura, C., Wakasugi, T., Yamada, K., Yoshinaga, K., Yamaguchi, K., Ueda, K. and Hasebe, M. (2004) 'Chloroplast phylogeny indicates that bryophytes are monophyletic', Molecular Biology and Evolution, 21(10), pp. 1813-1819.

Nishizawa, A., Yabuta, Y. and Shigeoka, S. (2008) 'Galactinol and raffinose constitute a novel function to protect plants from oxidative damage', Plant Physiology, 147(3), pp. 1251-1263.

Ohno, S. (1970) 'Enormous diversity in genome sizes of fish as a reflection of natures extensive experiments with gene duplication', Transactions of the American Fisheries Society, 99(1), pp. 120-&.

Oldenhof, H., Wolkers, W. F., Bowman, J. L., Tablin, F. and Crowe, J. H. (2006) 'Freezing and desiccation tolerance in the moss Physcomitrella patens: An in situ Fourier transform infrared spectroscopic study', Biochimica et Biophysica Acta-General Subjects, 1760(8), pp. 1226-1234.

Oliver, M. J., Dowd, S. E., Zaragoza, J., Mauget, S. A. and Payton, P. R. (2004) 'The rehydration transcriptome of the desiccation-tolerant bryophyte Tortula ruralis: transcript classification and analysis', Bmc Genomics, 5.

Oliver, M. J., Velten, J. and Mishler, B. D. (2005) 'Desiccation tolerance in bryophytes: A reflection of the primitive strategy for plant survival in dehydrating habitats?', Integrative and Comparative Biology, 45(5), pp. 788-799.

Osakabe, Y., Yamaguchi-Shinozaki, K., Shinozaki, K. and Lam-Son Phan, T. (2014) 'ABA control of plant macroelement membrane transport systems in response to water deficit and high salinity', New Phytologist, 202(1), pp. 35-49.

Ouaked, F., Rozhon, W., Lecourieux, D. and Hirt, H. (2003) 'A MAPK pathway mediates ethylene signalling in plants', Embo Journal, 22(6), pp. 1282-1288.

Park, S. Y., Fung, P., Nishimura, N., Jensen, D. R., Fujii, H., Zhao, Y., Lumba, S., Santiago, J., Rodrigues, A., Chow, T. F. F., Alfred, S. E., Bonetta, D., Finkelstein, R., Provart, N. J., Desveaux, D., Rodriguez, P. L., McCourt, P., Zhu, J. K., Schroeder, J. I., Volkman, B. F. and Cutler, S. R. (2009) 'Abscisic Acid Inhibits Type 2C Protein Phosphatases via the PYR/PYL Family of START Proteins', Science, 324, pp. 1068-1071.

Partch, C. L. and Gardner, K. H. (2010) 'Coactivator Recruitment: A New Role for PAS Domains in Transcriptional Regulation by the bHLH-PAS Family', Journal of Cellular Physiology, 223(3), pp. 553-557.

Paul, M. J., Primavesi, L. F., Jhurreea, D. and Zhang, Y. (2008) 'Trehalose metabolism and signalling', Annual Review of Plant Biology, 59, pp. 417-441.

Peterbauer, T. and Richter, A. (2001) 'Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds', Seed Science Research, 11(3), pp. 185-197.

Pongratz, I., Mason, G. G. F. and Poellinger, L. (1992) 'Dual roles of the 90-kDa heat-shock protein hsp90 in modulating functional activities of the dioxin receptor. Evidence that the dioxin

receptor functionally belongs to a subclass of nuclear receptors which require hsp90 both for ligand-binding activity and repression of intrinsic DNA-binding activity', Journal of Biological Chemistry, 267(19), pp. 13728-13734.

Qin, X. Q. and Zeevaart, J. A. D. (1999) 'The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory step of abscisic acid biosynthesis in water-stressed bean', Proceedings of the National Academy of Sciences of the United States of America, 96(26), pp. 15354-15361.

Qiu, Y. L., Cho, Y. R., Cox, J. C. and Palmer, J. D. (1998) 'The gain of three mitochondrial introns identifies liverworts as the earliest land plants', Nature, 394(6694), pp. 671-674.

Rensing, S. A., Ick, J., Fawcett, J. A., Lang, D., Zimmer, A., De Peer, Y. V. and Reski, R. (2007) 'An ancient genome duplication contributed to the abundance of metabolic genes in the moss Physcomitrella patens', Bmc Evolutionary Biology, 7.

Richard, O., Paquet, N., Haudecoeur, E. and Charrier, B. (2005) 'Organization and expression of the GSK3/Shaggy kinase gene family in the moss Physcomitrella patens suggest early gene multiplication in land plants and an ancestral response to osmotic stress', Journal of Molecular Evolution, 61(1), pp. 99-113.

Richardt, S., Timmerhaus, G., Lang, D., Qudeimat, E., Correa, L. G. G., Reski, R., Rensing, S. A. and Frank, W. (2010) 'Microarray analysis of the moss Physcomitrella patens reveals evolutionarily conserved transcriptional regulation of salt stress and abscisic acid signalling', Plant Molecular Biology, 72(1-2), pp. 27-45.

Roberts, M. R. (2003) '14-3-3 Proteins find new partners in plant cell signalling', Trends in Plant Science, 8(5), pp. 218-223.

Rodriguez, M. C. S., Petersen, M. and Mundy, J. (2010) 'Mitogen-Activated Protein Kinase Signalling in Plants', in Merchant, S., Briggs, W.R. & Ort, D. (eds.) Annual Review of Plant Biology, Vol 61 Annual Review of Plant Biology. Palo Alto: Annual Reviews, pp. 621-649.

Roychoudhury, A., Paul, S. and Basu, S. (2013) 'Cross-talk between abscisic acid-dependent and abscisic acid-independent pathways during abiotic stress', Plant Cell Reports, 32(7), pp. 985-1006.

Sakai, H., Aoyama, T., Bono, H. and Oka, A. (1998) 'Two-component response regulators from Arabidopsis thaliana contain a putative DNA-binding motif', Plant and Cell Physiology, 39(11), pp. 1232-1239.

Sakata, Y., Komatsu, K., Taji, T. and Tanaka, S. (2009) 'Role of PP2C-mediated ABA signalling in the moss Physcomitrella patens', Plant signalling & behavior, 4(9), pp. 887-9.

Sakata, Y., Komatsu, K. and Takezawa, D. (2014) 'ABA as a Universal Plant Hormone', Progress in Botany 75, 75, pp. 57-96.

Sallon, S., Solowey, E., Cohen, Y., Korchinsky, R., Egli, M., Woodhatch, I., Simchoni, O. and Kislev, M. (2008) 'Germination, genetics, and growth of an ancient date seed', Science, 320(5882), pp. 1464-1464.

Salomon, M., Lempert, U. and Rudiger, W. (2004) 'Dimerization of the plant photoreceptor phototropin is probably mediated by the LOV1 domain', Febs Letters, 572(1-3), pp. 8-10.

Santiago, J., Dupeux, F., Round, A., Antoni, R., Park, S.-Y., Jamin, M., Cutler, S. R., Luis Rodriguez, P. and Marquez, J. A. (2009) 'The abscisic acid receptor PYR1 in complex with abscisic acid', Nature, 462(7273), pp. 665-U143.

Sasayama, D., Matsuoka, D., Oka, M., Shitamichi, N., Furuya, T., Azuma, T., Itoh, K. and Nanmori, T. (2011) 'MAP3K delta 4, an Arabidopsis Raf-like MAP3K, regulates plant growth and shoot branching', Plant Biotechnology, 28(5), pp. 463-470.

Schaefer, D. G. and Zryd, J. P. (1997) 'Efficient gene targeting in the moss Physcomitrella patens', Plant Journal, 11(6), pp. 1195-1206.

Scheuermann, T. H., Li, Q. M., Ma, H. W., Key, J., Zhang, L., Chen, R., Garcia, J. A., Naidoo, J., Longgood, J., Frantz, D. E., Tambar, U. K., Gardner, K. H. and Bruick, R. K. (2013) 'Allosteric inhibition of hypoxia inducible factor-2 with small molecules', Nature Chemical Biology, 9(4), pp. 271-276.

Scheuermann, T. H., Tomchick, D. R., Machius, M., Guo, Y., Bruick, R. K. and Gardner, K. H. (2009) 'Artificial ligand binding within the HIF2 alpha PAS-B domain of the HIF2 transcription factor', Proceedings of the National Academy of Sciences of the United States of America, 106(2), pp. 450-455.

Schultz, T. F., Medina, J., Hill, A. and Quatrano, R. S. (1998) '14-3-3 proteins are part of an abscisic acid VIVIPAROUS1 (VP1) response complex in the Em promoter and interact with VP1 and EmBP1', Plant Cell, 10(5), pp. 837-847.

Schwerdtfeger, C. and Linden, H. (2003) 'VIVID is a flavoprotein and serves as a fungal blue light photoreceptor for photoadaptation', Embo Journal, 22(18), pp. 4846-4855.

Seo, M. and Koshiba, T. (2011) 'Transport of ABA from the site of biosynthesis to the site of action', Journal of Plant Research, 124(4), pp. 501-507.

Seo, P. J., Park, J., Park, M.-J., Kim, Y.-S., Kim, S.-G., Jung, J.-H. and Park, C.-M. (2012) 'A Golgi-localized MATE transporter mediates iron homoeostasis under osmotic stress in Arabidopsis', Biochemical Journal, 442, pp. 551-561.

Sethi, V., Raghuram, B., Sinha, A. K. and Chattopadhyay, S. (2014) 'A Mitogen-Activated Protein Kinase Cascade Module, MKK3-MPK6 and MYC2, Is Involved in Blue Light-Mediated Seedling Development in Arabidopsis', Plant Cell, 26(8), pp. 3343-3357.

Sharma, N., Jung, C.-H., Bhalla, P. L. and Singh, M. B. (2014) 'RNA Sequencing Analysis of the Gametophyte Transcriptome from the Liverwort, Marchantia polymorpha', Plos One, 9(5).

Sharrock, R. A. (2008) 'The phytochrome red/far-red photoreceptor superfamily', Genome Biology, 9(8).

Shen, Q. X. and Ho, T. H. D. (1995) 'Functional dissection of an abscisic-acid (ABA)-inducible gene reveals 2 independent ABA-responsive complexes each containing a G-box and a novel cis-acting element', Plant Cell, 7(3), pp. 295-307.

Shi, J., Zhang, L., An, H., Wu, C. and Guo, X. (2011) 'GhMPK16, a novel stress-responsive group D MAPK gene from cotton, is involved in disease resistance and drought sensitivity', Bmc Molecular Biology, 12.

Shitamichi, N., Matsuoka, D., Sasayama, D., Furuya, T. and Nanmori, T. (2013) 'Over-expression of MAP3K delta 4, an ABA-inducible Raf-like MAP3K that confers salt tolerance in Arabidopsis', Plant Biotechnology, 30(2), pp. 111-118.

Sinha, A. K., Jaggi, M., Raghuram, B. and Tuteja, N. (2011) 'Mitogen-activated protein kinase signalling in plants under abiotic stress', Plant signalling & behavior, 6(2), pp. 196-203.

Smirnoff, N. (1993) 'The role of active oxygen in the response of plants to water-deficit and desiccation', New Phytologist, 125(1), pp. 27-58.

Stone, S. L. (2014) 'The role of ubiquitin and the 26S proteasome in plant abiotic stress signalling', Frontiers in Plant Science, 5.

Strizhov, N., Abraham, E., Okresz, L., Blickling, S., Zilberstein, A., Schell, J., Koncz, C. and Szabados, L. (1997) 'Differential expression of two P5CS genes controlling proline accumulation during salt-stress requires ABA and is regulated by ABA1, ABI1 and AXR2 in Arabidopsis', Plant Journal, 12(3), pp. 557-569.

Suetsugu, N., Mittmann, F., Wagner, G., Hughes, J. and Wada, M. (2005) 'A chimeric photoreceptor gene, NEOCHROME, has arisen twice during plant evolution', Proceedings of the National Academy of Sciences of the United States of America, 102(38), pp. 13705-13709.

Suri, S. S. and Dhindsa, R. S. (2008) 'A heat-activated MAP kinase (HAMK) as a mediator of heat shock response in tobacco cells', Plant Cell and Environment, 31(2), pp. 218-226.

Szoevenyi, P., Perroud, P.-F., Symeonidi, A., Stevenson, S., Quatrano, R. S., Rensing, S. A., Cuming, A. C. and McDaniel, S. F. (2015) 'De novo assembly and comparative analysis of the Ceratodon purpureus transcriptome', Molecular Ecology Resources, 15(1), pp. 203-215.

Sztein, A. E., Cohen, J. D., Slovin, J. P. and Cooke, T. J. (1995) 'Auxin metabolism in representative land plants', American Journal of Botany, 82(12), pp. 1514-1521.

Szurmant, H., White, R. A. and Hoch, J. A. (2007) 'Sensor complexes regulating two-component signal transduction', Current Opinion in Structural Biology, 17(6), pp. 706-715.

Taji, T., Seki, M., Satou, M., Sakurai, T., Kobayashi, M., Ishiyama, K., Narusaka, Y., Narusaka, M., Zhu, J. K. and Shinozaki, K. (2004) 'Comparative genomics in salt tolerance between Arabidopsis and Arabidopsis-related halophyte salt cress using Arabidopsis microarray', Plant Physiology, 135(3), pp. 1697-1709.

Takezawa, D., Komatsu, K. and Sakata, Y. (2011) 'ABA in bryophytes: how a universal growth regulator in life became a plant hormone?', Journal of Plant Research, 124(4), pp. 437-453.

Takezawa, D., Watanabe, N., Ghosh, T. K., Saruhashi, M., Suzuki, A., Ishiyama, K., Somemiya, S., Kobayashi, M. and Sakata, Y. (2015) 'Epoxycarotenoid-mediated synthesis of abscisic acid in Physcomitrella patens implicating conserved mechanisms for acclimation to hyperosmosis in embryophytes', New Phytologist, 206(1), pp. 209-219.

Tan, B. C., Schwartz, S. H., Zeevaart, J. A. D. and McCarty, D. R. (1997) 'Genetic control of abscisic acid biosynthesis in maize', Proceedings of the National Academy of Sciences of the United States of America, 94(22), pp. 12235-12240.

Tang, D. Z., Christiansen, K. M. and Innes, R. W. (2005) 'Regulation of plant disease resistance, stress responses, cell death, and ethylene signalling in Arabidopsis by the EDR1 protein kinase', Plant Physiology, 138(2), pp. 1018-1026.

Taylor, B. L. and Zhulin, I. B. (1999) 'PAS domains: Internal sensors of oxygen, redox potential, and light', Microbiology and Molecular Biology Reviews, 63(2), pp. 479-+.

Timme, R. E., Bachvaroff, T. R. and Delwiche, C. F. (2012) 'Broad Phylogenomic Sampling and the Sister Lineage of Land Plants', Plos One, 7(1).

Tran, L.-S. P., Urao, T., Qin, F., Maruyama, K., Kakimoto, T., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2007) 'Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in Arabidopsis', Proceedings of the National Academy of Sciences of the United States of America, 104(51), pp. 20623-20628.

Tunnacliffe, A. and Wise, M. J. (2007) 'The continuing conundrum of the LEA proteins', Naturwissenschaften, 94(10), pp. 791-812.

Turmel, M., Otis, C. and Lemieux, C. (2006) 'The chloroplast genome sequence of Chara vulgaris sheds new light into the closest green algal relatives of land plants', Molecular Biology and Evolution, 23(6), pp. 1324-1338.

Urao, T., Yakubov, B., Satoh, R., Yamaguchi-Shinozaki, K., Seki, M., Hirayama, T. and Shinozaki, K. (1999) 'A transmembrane hybrid-type histidine kinase in arabidopsis functions as an osmosensor', Plant Cell, 11(9), pp. 1743-1754.

Urao, T., Yakubov, B., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1998) 'Stress-responsive expression of genes for two-component response regulator-like proteins in Arabidopsis thaliana', Febs Letters, 427(2), pp. 175-178.

Urao, T., Yamaguchi-Shinozaki, K. and Shinozaki, K. (2000) 'Two-component systems in plant signal transduction', Trends in Plant Science, 5(2), pp. 67-74.

Verbruggen, N. and Hermans, C. (2008) 'Proline accumulation in plants: a review', Amino Acids, 35(4), pp. 753-759.

Virk, N., Li, D. Y., Tian, L. M., Huang, L., Hong, Y. B., Li, X. H., Zhang, Y. F., Liu, B., Zhang, H. J. and Song, F. M. (2015) 'Arabidopsis Raf-Like Mitogen-Activated Protein Kinase Kinase Kinase Gene Raf43 Is Required for Tolerance to Multiple Abiotic Stresses', Plos One, 10(7).

Wang, C., Liu, Y., Li, S.-S. and Han, G.-Z. (2014a) 'Origin of plant auxin biosynthesis in charophyte algae', Trends in Plant Science, 19(12), pp. 741-743.

Wang, P., Xue, L., Batelli, G., Lee, S., Hou, Y.-J., Van Oosten, M. J., Zhang, H., Tao, W. A. and Zhu, J.-K. (2013) 'Quantitative phosphoproteomics identifies SnRK2 protein kinase substrates and reveals the effectors of abscisic acid action', Proceedings of the National Academy of Sciences of the United States of America, 110(27), pp. 11205-11210.

Wang, Y., Chang, H., Hu, S., Lu, X., Yuan, C., Zhang, C., Wang, P., Xiao, W., Xiao, L., Xue, G.-P. and Guo, X. (2014b) 'Plastid casein kinase 2 knockout reduces abscisic acid (ABA) sensitivity, thermotolerance, and expression of ABA- and heat-stress-responsive nuclear genes', Journal of Experimental Botany, 65(15), pp. 4159-4175.

Wen, F., Qin, T. T., Wang, Y., Dong, W., Zhang, A. Y., Tan, M. P. and Jiang, M. Y. (2015) 'OsHK3 is a crucial regulator of abscisic acid signalling involved in antioxidant defense in rice', Journal of Integrative Plant Biology, 57(2), pp. 213-228.

Wheeler, G. L. and Brownlee, C. (2008) 'Ca2+ signalling in plants and green algae - changing channels', Trends in Plant Science, 13(9), pp. 506-514.

Wickett, N. J., Mirarab, S., Nam, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M. S., Burleigh, J. G., Gitzendanner, M. A., Ruhfel, B. R., Wafula, E., Der, J. P., Graham, S. W., Mathews, S., Melkonian, M., Soltis, D. E., Soltis, P. S., Miles, N. W., Rothfels, C. J., Pokorny, L., Shaw, A. J., DeGironimo, L., Stevenson, D. W., Surek, B., Villarreal, J. C., Roure, B., Philippe, H., dePamphilis, C. W., Chen, T., Deyholos, M. K., Baucom, R. S., Kutchan, T. M., Augustin, M. M., Wang, J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong, G. K.-S. and Leebens-Mack, J. (2014) 'Phylotranscriptomic analysis of the origin and early diversification of land plants', Proceedings of the National Academy of Sciences of the United States of America, 111(45), pp. E4859-E4868.

Widmann, C., Gibson, S., Jarpe, M. B. and Johnson, G. L. (1999) 'Mitogen-activated protein kinase: Conservation of a three-kinase module from yeast to human', Physiological Reviews, 79(1), pp. 143-180.

Wingenter, K., Trentmann, O., Winschuh, I., Hoermiller, I. I., Heyer, A. G., Reinders, J., Schulz, A., Geiger, D., Hedrich, R. and Neuhaus, H. E. (2011) 'A member of the mitogen-activated protein 3-kinase family is involved in the regulation of plant vacuolar glucose uptake', Plant Journal, 68(5), pp. 890-900.

Wodniok, S., Brinkmann, H., Gloeckner, G., Heidel, A. J., Philippe, H., Melkonian, M. and Becker, B. (2011) 'Origin of land plants: Do conjugating green algae hold the key?', Bmc Evolutionary Biology, 11.

Wohlbach, D. J., Quirino, B. F. and Sussman, M. R. (2008) 'Analysis of the Arabidopsis histidine kinase ATHK1 reveals a connection between vegetative osmotic stress sensing and seed maturation', Plant Cell, 20(4), pp. 1101-1117.

Wood, A. J., Duff, R. J. and Oliver, M. J. (1999) 'Expressed sequence tags (ESTs) from desiccated Tortula ruralis identify a large number of novel plant genes', Plant and Cell Physiology, 40(4), pp. 361-368.

Wu, H.-P., Su, Y.-s., Chen, H.-C., Chen, Y.-R., Wu, C.-C., Lin, W.-D. and Tu, S.-L. (2014) 'Genome-wide analysis of light-regulated alternative splicing mediated by photoreceptors in Physcomitrella patens', Genome Biology, 15(1).

Xiong, L. M. and Zhu, J. K. (2003) 'Regulation of abscisic acid biosynthesis', Plant Physiology, 133(1), pp. 29-36.

Xiong, L. Z. and Yang, Y. N. (2003) 'Disease resistance and abiotic stress tolerance in rice are inversely modulated by an abscisic acid-inducible mitogen-activated protein kinase', Plant Cell, 15(3), pp. 745-759.

Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S. and Hattori, T. (2009) 'Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors', Plant Journal, 58(5), pp. 843-856.

Yasumura, Y., Pierik, R., Fricker, M. D., Voesenek, L. and Harberd, N. P. (2012) 'Studies of Physcomitrella patens reveal that ethylene-mediated submergence responses arose relatively early in land-plant evolution', Plant Journal, 72(6), pp. 947-959.

Yasumura, Y., Pierik, R., Kelly, S., Sakuta, M., Voesenek, L. A. C. J. and Harberd, N. P. (2015) 'An Ancestral Role for CONSTITUTIVE TRIPLE RESPONSE1 Proteins in Both Ethylene and Abscisic Acid Signalling', Plant physiology, 169(1), pp. 283-98.

Yin, P., Fan, H., Hao, Q., Yuan, X., Wu, D., Pang, Y., Yan, C., Li, W., Wang, J. and Yan, N. (2009) 'Structural insights into the mechanism of abscisic acid signalling by PYL proteins', Nature Structural & Molecular Biology, 16(12), pp. 1230-U42.

Yin, Z., Wang, J., Wang, D., Fan, W., Wang, S. and Ye, W. (2013) 'The MAPKKK Gene Family in Gossypium raimondii: Genome-Wide Identification, Classification and Expression Analysis', International Journal of Molecular Sciences, 14(9), pp. 18740-18757.

Yoshida, K., Igarashi, E., Mukai, M., Hirata, K. and Miyamoto, K. (2003) 'Induction of tolerance to oxidative stress in the green alga, Chlamydomonas reinhardtii, by abscisic acid', Plant Cell and Environment, 26(3), pp. 451-457.

Yoshida, T., Mogami, J. and Yamaguchi-Shinozaki, K. (2014) 'ABA-dependent and ABA-independent signalling in response to osmotic stress in plants', Current Opinion in Plant Biology, 21, pp. 133-139.

Yotsui, I., Saruhashi, M., Kawato, T., Taji, T., Hayashi, T., Quatrano, R. S. and Sakata, Y. (2013) 'ABSCISIC ACID INSENSITIVE3 regulates abscisic acid-responsive gene expression with the nuclear factor Y complex through the ACTT-core element in Physcomitrella patens', New Phytologist, 199(1), pp. 101-109.

Yu, L., Nie, J., Cao, C., Jin, Y., Yan, M., Wang, F., Liu, J., Xiao, Y., Liang, Y. and Zhang, W. (2010) 'Phosphatidic acid mediates salt stress response by regulation of MPK6 in Arabidopsis thaliana', New Phytologist, 188(3), pp. 762-773.

Yuan, F., Wang, M., Hao, H., Zhang, Y., Zhao, H., Guo, A., Xu, H., Zhou, X. and Xie, C. G. (2013) 'Negative regulation of abscisic acid signalling by the Brassica oleracea ABI1 ortholog', Biochemical and Biophysical Research Communications, 442(3-4), pp. 202-208.

Zeevaart, J. A. D. and Creelman, R. A. (1988) 'Metabolism and physiology of abscisic-acid', Annual Review of Plant Physiology and Plant Molecular Biology, 39, pp. 439-473.

Zhang, H., Cui, F., Wu, Y., Lou, L., Liu, L., Tian, M., Ning, Y., Shu, K., Tang, S. and Xie, Q. (2015) 'The RING Finger Ubiquitin E3 Ligase SDIR1 Targets SDIR1-INTERACTING PROTEIN1 for Degradation to Modulate the Salt Stress Response and ABA Signalling in Arabidopsis', Plant Cell, 27(1), pp. 214-227.

Zhang, J. Z. (2003) 'Evolution by gene duplication: an update', Trends in Ecology & Evolution, 18(6), pp. 292-298.

Zhao, C., Nie, H., Shen, Q., Zhang, S., Lukowitz, W. and Tang, D. (2014) 'EDR1 Physically Interacts with MKK4/MKK5 and Negatively Regulates a MAP Kinase Cascade to Modulate Plant Innate Immunity', Plos Genetics, 10(5).

Zimmer, A. D., Lang, D., Buchta, K., Rombauts, S., Nishiyama, T., Hasebe, M., Van de Peer, Y., Rensing, S. A. and Reski, R. (2013) 'Reannotation and extended community resources for the genome of the non-seed plant Physcomitrella patens provide insights into the evolution of plant gene structures and functions', Bmc Genomics, 14, pp. 20.

Zoltowski, B. D., Schwerdtfeger, C., Widom, J., Loros, J. J., Bilwes, A. M., Dunlap, J. C. and Crane, B. R. (2007) 'Conformational switching in the fungal light sensor vivid', Science, 316(5827), pp. 1054-1057.

Supplementary data

The RNA-seq data presented in this thesis has been submitted to the NCBI GEO database under the record GSE72583 to which temporary access can be found using the following link:
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=enajcgcelrknjyx&acc=GSE72583
See compact disk for spreadsheets of significantly regulated genes from wild-type and *PpanrKO* tissue. Supplementary data file 1 contains data from wild-type tissue and data file 2 contains data from *PpanrKO*. Each dataset contains the V3.0 genes ID, the log fold change for the treatment of interest, the P-value of this fold change (Fishers exact test), the false discovery rate P-value, the normalized count value for each gene in each library (*e.g.* WT_C is the count in the wild-type control library), the gene cluster to which the gene belongs and the annotation manually curated by A. Cuming.

Appendices

Table A1. List of sequences used in phylogenetic analyses. Where possible, sample IDs refer to their name in the databases they were retrieved from.

Name	Organism	Lineage	Source
At1G18160	Arabidopsis thaliana	Angiosperm	TAIR
At1G73660	Arabidopsis thaliana	Angiosperm	TAIR
AT2G31010	Arabidopsis thaliana	Angiosperm	TAIR
AT2G42640	Arabidopsis thaliana	Angiosperm	TAIR
At3G06620	Arabidopsis thaliana	Angiosperm	TAIR
At3G06640	Arabidopsis thaliana	Angiosperm	TAIR
AT3G58640	Arabidopsis thaliana	Angiosperm	TAIR
At4G23050	Arabidopsis thaliana	Angiosperm	TAIR
At4G24480	Arabidopsis thaliana	Angiosperm	TAIR
At5G11850	Arabidopsis thaliana	Angiosperm	TAIR
AtCTR1	Arabidopsis thaliana	Angiosperm	TAIR
AtEDR1	Arabidopsis thaliana	Angiosperm	TAIR
AtRaf10	Arabidopsis thaliana	Angiosperm	TAIR
AtRaf11	Arabidopsis thaliana	Angiosperm	TAIR
Atr006841450	Amborella trichopoda	Angiosperm	NCBI
Atr2006254	Amborella trichopoda	Angiosperm	NCBI
Osa02g12810	Oryza sativa	Angiosperm	Rice Genome Annotation Project

Osa02g32610	Oryza sativa	Angiosperm	Rice Genome Annotation Project
Osa12g37570	Oryza sativa	Angiosperm	Rice Genome Annotation Project
Cgl2001833	Chaetosphaeridium globosum	Coleochaetales	onekp
Cgl2005871	Chaetosphaeridium globosum	Coleochaetales	onekp
Cgl2008114	Chaetosphaeridium globosum	Coleochaetales	onekp
Cgl2036945	Chaetosphaeridium globosum	Coleochaetales	onekp
Cgl2004405	Chaetosphaeridium globosum	Coleochaetales	onekp
Csc2000102	Coleochaete scutata	Coleochaetales	onekp
Csc2005084	Coleochaete scutata	Coleochaetales	onekp
Csc2005085	Coleochaete scutata	Coleochaetales	onekp
Csc2009828	Coleochaete scutata	Coleochaetales	onekp
Csc2009975	Coleochaete scutata	Coleochaetales	onekp
Csc2011562	Coleochaete scutata	Coleochaetales	onekp
Csc2042070	Coleochaete scutata	Coleochaetales	onekp
Csc2002922	Coleochaete scutata	Coleochaetales	onekp
Csc2010211	Coleochaete scutata	Coleochaetales	onekp
Csc2008344	Coleochaete scutata	Coleochaetales	onekp
Cir2031564	Coleochaete irregularis	Coleochaetales	onekp
Lim2002585	Leucostegia immersa	Fern	onekp
Lim2003040	Leucostegia immersa	Fern	onekp
Lim2010286	Leucostegia immersa	Fern	onekp
Lim2012749	Leucostegia immersa	Fern	onekp
Lim2013038	Leucostegia immersa	Fern	onekp
Lim2076964	Leucostegia immersa	Fern	onekp
Sdi2004333	Sceptridium dissectum	Fern	onekp
Sdi2018481	Sceptridium dissectum	Fern	onekp
Sdi2018518	Sceptridium dissectum	Fern	onekp
Sdi2018518	Sceptridium dissectum	Fern	onekp
Sdi2018481	Sceptridium dissectum	Fern	onekp
Pen2001701	Picea engelmanii	Gymnosperm	onekp
Pen2011985	Picea engelmanii	Gymnosperm	onekp

PsiACN39902	Picea sitchensis	Gymnosperm	NCBI
Mvi2001389	Megaceros vincentianus	Hornwort	onekp
Mvi2008620	Megaceros vincentianus	Hornwort	onekp
Mvi2009561	Megaceros vincentianus	Hornwort	onekp
Pha2001710	Paraphymatoceros hallii	Hornwort	onekp
Mpa2008201	Marchantia paleacea	Liverwort	onekp
Mpa2068472	Marchantia paleacea	Liverwort	onekp
Mpa2068537	Marchantia paleacea	Liverwort	onekp
Mpo02993	Marchantia polymorpha	Liverwort	Phytozome
MpoANR	Marchantia polymorpha	Liverwort	Phytozome
MpoCTR1	Marchantia polymorpha	Liverwort	Phytozome
Pep2009091	Pellia epiphylla	Liverwort	onekp
Pep2092799	Pellia epiphylla	Liverwort	onekp
Rli2086904	Radula lindenbergia	Liverwort	onekp
Hse2018077	Huperzia selago	Lycophyte	onekp
Smo10499	Selaginella moellendorffii	Lycophyte	Phytozome
Smo269874	Selaginella moellendorffii	Lycophyte	Phytozome
Smo449389	Selaginella moellendorffii	Lycophyte	Phytozome
Sst2011441	Selaginella stauntoniana	Lycophyte	onekp
Swa2005716	Selaginella wallacei	Lycophyte	onekp
Aan2015758	Atrichum angustatum	Moss	onekp
Ahe2090769	Aulacomnium heterostichum	Moss	onekp
Ahe2090803	Aulacomnium heterostichum	Moss	Onekp
CpuANR	Ceratodon purpureus	Moss	Szövényi *et al.*, 2015*; C. purpureus* transcriptome
Pp3c1_16810	Physcomitrella patens	Moss	Cosmoss
Pp3c11_12320	Physcomitrella patens	Moss	Cosmoss
Pp3c7_14130	Physcomitrella patens	Moss	Cosmoss
PpCTR1L_ANR	Physcomitrella patens	Moss	Cosmoss
Rse2000512	Rhynchostegium serrulatum	Moss	onekp
Rse2011363	Rhynchostegium serrulatum	Moss	onekp
Men2004016	Mesotaenium endlicherianum	Zygnematales	onekp
Men2004539	Mesotaenium endlicherianum	Zygnematales	onekp

Men2005803	Mesotaenium endlicherianum	Zygnematales	onekp
Men2046294	Mesotaenium endlicherianum	Zygnematales	onekp
Men2009965	Mesotaenium endlicherianum	Zygnematales	onekp
Men2046603	Mesotaenium endlicherianum	Zygnematales	onekp
Men2008640	Mesotaenium endlicherianum	Zygnematales	onekp
Men2045501	Mesotaenium endlicherianum	Zygnematales	onekp
Men2008889	Mesotaenium endlicherianum	Zygnematales	onekp
Ccu2005843	Cylindrocystis cushleckae	Zygnematales	onekp
Ccu2008845	Cylindrocystis cushleckae	Zygnematales	onekp
SprCTR1A	Spirogyra pratensis	Zygnematales	Ju *et al.,* 2015; from alignment
SprCTR1B	Spirogyra pratensis	Zygnematales	Ju *et al.,* 2015; from alignment
Cbr2010350	Cylindrocystis brebissonii	Zygnematales	onekp
EDR tree			
AT1G04210	Ababidopsis thaliana	Angiosperm	TAIR
Phpat.008G084000	Physcomitrella patens	Moss	Cosmoss
PAS tree			
Zsp2018335	Zygnema sp.	Zygnematales	onekp
Cbr2056816	Cylindrocystis brebissonii	Zygnematales	onekp
Cbr2003907	Cylindrocystis brebissonii	Zygnematales	onekp
Efi2029218	Entransia fimbriata	Klebsormidiales	onekp
Smi2007624	Spirotaenia minuta	Zygnematales	onekp
Rob2007183	Roya obtusa	Zygnematales	onekp
Men2008305	Mesotaenium endlicherianum	Zygnematales	onekp
Zsp2018336	Zygnema sp.	Zygnematales	onekp
Zsp2024780	Zygnema sp.	Zygnematales	onekp
Ccu2005029	Cylindrocystis cushleckae	Zygnematales	onekp
MspAFZ18849	Microcoleus sp.	Cyanobacteria	NCBI
GspWP_034892591	Gillisia sp.	Flavobacteriaceae	NCBI
LspWP_044404157	Lacinutrix sp.	Flavobacteriaceae	NCBI
SspWP_026753069	Sediminibacter sp.	Flavobacteriaceae	NCBI
Pp1s91_116_PASb	Physcomitrella patens	Moss	Cosmoss
Pp1s47_31_PASb	Physcomitrella patens	Moss	Cosmoss
Pp1s164_65_PASb	Physcomitrella patens	Moss	Cosmoss
Pp1s327_27_PASb	Physcomitrella patens	Moss	Cosmoss
AtPhyA_PASb	Ababidopsis thaliana	Angiosperm	TAIR

AtPhyC_PASb	Ababidopsis thaliana	Angiosperm	TAIR
AtPhyE_PASb	Ababidopsis thaliana	Angiosperm	TAIR
CscPhy_PASb	Coleochaete scutata	Coleochaetales	onekp
MenPhy1_PASb	Mesotaenium endlicherianum	Zygnematales	onekp
MenPhy2_PASb	Mesotaenium endlicherianum	Zygnematales	onekp

Appendix 2.1 Python script that calculated the genotype ratios for each SNP marker enabling mapping of the co-segregating loci for *anr3* and *anr4* mutants.

```
from __future__ import print_function
import sys

mark = sys.argv[1]      # List of SNP marker names
anr = sys.argv[2]       # Matrix of genotyping data for each SNP marker separated for
anr and wild-type segregants

SNP_list=[]
order_list=[]
for SNP in open(mark):
        SNP = SNP.rstrip()
        order = SNP[3:7]
        SNP_list.append(SNP)
        order_list.append(order)

out = open('SNP_ratios','w')

snp=0
for line in open(anr):
        if 'anr' in line:
                continue
        if 'V_' in line:
                continue
        A=0.00
        B=0.00
        marker = line.split()
        for x in marker:
                if x == 'A':
                        A+=1   # Add one count to genotype A
                if x == 'B':
                        B+=1   # Add one count to genotype B
        if A == 0:
                A=1
        if B == 0:
                B=1
        #print(A,B)
        SUM = A+B    # Total progeny genotyped for SNP marker
        #print(SUM)
        #sys.exit()
        C = float((A/SUM)*100)          # Proportion of genotype A for SNP marker
        D = float((B/SUM)*100)          # Proportion of genotype B for SNP marker
        print(order_list[snp],SNP_list[snp],A,B,C,D,file=out)
```

```
        snp+=1
```

Appendix 4.1 Python script for the removal of redundant protein sequences from the large scale kinase analyses in *Physcomitrella.*

```
from __future__ import print_function
import sys

#removes duplicates from fasta file

total = sys.argv[1]

seqlist={}

out=open('stripped.fasta','w')

seqid=''

for line in open(total):
        line=line.strip()
        if '>' in line and line[35:]!=1:
                skip=1
        if '>' in line and line[35:]=='1':
                seqid=line[:17]
                skip=0
        if '>' not in line and skip==0:
                seq=line
                seqlist[seqid]=seq

for x in seqlist:
        print(x,seqlist[x],sep='\n',file=out)
```

Appendix 4.2 Python script to extract kinase domain coordinates from batch CD output.

```
from __future__ import print_function
import sys

CD_output = sys.argv[1]                 # Batch Conserved Domain output in tabular form
fasta = sys.argv[2]             # Fasta file with all full protein sequences

out=open('coordinates','w')
number=[]
x=0

# This part extracts the coordinates for the kinase domain.

for line in open(CD_output):
        if 'PKc_like' in line:
                line=line.split()
                x+=1
                print(line[2],line[5],line[6],line[0],line[10],x,sep='\t',file=out)
                number.append(line[0])
                if number.count(line[0])>1:
                        print(line)
```

```
# This part extracts the sequence IDs in the same order as the coordinates.

out2=open('seq_ids','w')

for line in open(fasta):
	if '>' in line:
		print(line.strip()[1:],file=out2)
```

Appendix 5.1. Input for STAMP defining PDB file sequences used for structural similarity analysis.

```
./pdbs/.pdb 1BYW { ALL }
./pdbs/.pdb 1D06 { A 149 _ to A 251 _ }
./pdbs/.pdb 1G28 { CHAIN A}
./pdbs/.pdb 1LL8 { A  10 _ to A 114 _ }
./pdbs/.pdb 1MZU { B  29 _ to B 123 _ }
./pdbs/.pdb 1N9L { A  21 _ to A 122 _ }
./pdbs/.pdb 1NWZ { A  29 _ to A 125 _ }
./pdbs/.pdb 1V9Z { A  32 _ to A 132 _ }
./pdbs/.pdb 1WA9a { A 240 _ to A 372 _ }
./pdbs/.pdb 1WA9b { A 390 _ to A 497 _ }
./pdbs/.pdb 1X0O { A  11 _ to A 112 _ }
./pdbs/.pdb 1XJ3 { A 154 _ to A 255 _ }
./pdbs/.pdb 2B02 { A 362 _ to A 463 _ }
./pdbs/.pdb 2GJ3 { A  37  _ to A 135 _ }
./pdbs/.pdb 2HV1 { A  11 _ to A 111 _ }
./pdbs/.pdb 2O9C { A  40 _ to A 129 _ }
./pdbs/.pdb 2OOL { A  44 _ to A 134 _ }
./pdbs/.pdb 2P04 { A  21 _ to A 105 _ }
./pdbs/.pdb 2PD7 { A  73 _ to A 183 _ }
./pdbs/.pdb 2PR5 { A  26 _ to A 125 _ }
./pdbs/.pdb 2R78 { A  26 _ to A 117 _ }
./pdbs/.pdb 2V0U { A 415 _ to A 515 _ }
./pdbs/.pdb 2V0W { A 415 _ to A 515 _ }
./pdbs/.pdb 2V1A { A 415 _ to A 515 _ }
./pdbs/.pdb 2V1B { A 415 _ to A 515 _ }
./pdbs/.pdb 2VEA { A  31 _ to A 124 _ }
./pdbs/.pdb 2VLG { A  17 _ to A 113 _ }
./pdbs/.pdb 2Z6C { A 197 _ to A 300 _ }
./pdbs/.pdb 2Z6D { A 134 _ to A 235 _ }
./pdbs/.pdb 3B33 { A  16 _ to A 110 _ }
./pdbs/.pdb 3BWL { A 413 _ to A 507 _ }
./pdbs/.pdb 3EWKa { A  99 _ to A 197 _ }
./pdbs/.pdb 3EWKb { A 221 _ to A 324 _ }
./pdbs/.pdb 3FC7 { A 147 _ to A 239 _ }
./pdbs/.pdb 3FG8 { A 130 _ to A 220 _ }
./pdbs/.pdb 3GECa { A 241 _ to A 369 _ }
./pdbs/.pdb 3GECb { A 393 _ to A 492 _ }
./pdbs/.pdb 3H7Wa { A 243 _ to A 343 _ }
./pdbs/.pdb 3H7Wb { B 362 _ to B 462 _ }
./pdbs/.pdb 3H9W { A  25 _ to A 125 _ }
./pdbs/.pdb 3H82a { A 243 _ to A 343 _ }
./pdbs/.pdb 3H82b { B 362 _ to B 463 _ }
./pdbs/.pdb 3LYX { A  19 _ to A 117 _ }
./pdbs/.pdb 3MFX { A  35 _ to A 138 _ }
./pdbs/.pdb 3MJQ { A  12 _ to A 106 _ }
```

./pdbs/.pdb 3OLO { A 22 _ to A 113 _ }
./pdbs/.pdb 3RTYa { A 240 _ to A 345 _ }
./pdbs/.pdb 3RTYb { A 390 _ to A 497 _ }
./pdbs/.pdb 3SW1 { A 17 _ to A 117 _ }
./pdbs/.pdb 4DJ3a { A 131 _ to A 253 _ }
./pdbs/.pdb 4DJ3b { A 269 _ to A 356 _ }
./pdbs/.pdb 4EEP { A 391 _ to A 491 _ }
./pdbs/.pdb 4EET { B 391 _ to B 491 _ }
./pdbs/.pdb 4EEU { A 391 _ to A 492 _ }
./pdbs/.pdb 4EQ1 { A 362 _ to A 463 _ }
./pdbs/.pdb 4F3LAa { A 120 _ to A 257 _ }
./pdbs/.pdb 4F3LAb { A 275 _ to A 375 _ }
./pdbs/.pdb 4F3LBa { B 158 _ to B 320 _ }
./pdbs/.pdb 4F3LBb { B 369 _ to B 440 _ }
./pdbs/.pdb 4GHIa { A 243 _ to A 343 _ }
./pdbs/.pdb 4GHIb { B 362 _ to B 463 _ }
./pdbs/.pdb 4GS9a { A 243 _ to A 343 _ }
./pdbs/.pdb 4GS9b { B 362 _ to B 463 _ }
./pdbs/.pdb 4H6Ja { A 241 _ to A 341 _ }
./pdbs/.pdb 4H6Jb { B 362 _ to B 463 _ }
./pdbs/.pdb 4HH2a { A 25 _ to A 119 _ }
./pdbs/.pdb 4HH2b { A 162 _ to A 255 _ }
./pdbs/.pdb 4HH2c { A 280 _ to A 376 _ }
./pdbs/.pdb 4KUO { A 36 _ to A 137 _ }
./pdbs/.pdb 4PKY { A 362 _ to A 463 _ }
./pdbs/.pdb 4WUJ { A 99 _ to A 204 _ }
./pdbs/.pdb PAPDa { A 4 _ to A 99 _ }
./pdbs/.pdb PAPDb { B 4 _ to B 99 _ }