Major transitions in the plant tree of life: insights from genes, genomes and traits

Alexander M. C. Bowles

A thesis submitted for the degree of Doctor of Philosophy

Department of Life Sciences

University of Essex

Date of submission December 2020

Abstract

The evolution of plants transformed the Earth's surface, atmosphere and climate and enabled the colonisation of new habitats, promoting the diversity of other lineages spanning the tree of life. The evolutionary history of plants has been marked by major transitions such as multicellularity, terrestrialisation and the origin of stomata, roots and seeds. These events have been accompanied by the gain, loss, expansion and contraction of gene families. Genome sequencing has increased the potential insights from evolutionary analyses which includes comparative genomics, gene family evolution and trait evolution. The overall aim of the research presented in this thesis is to improve our understanding of genes involved in the major transitions in plant evolution, by analysing plant genome data. First, I examine the broad scale evolution of genes across the plant tree of life, identifying two bursts of gene novelty that accompanied the origin of land plants. Second, I identify the modes of genome evolution underpinning the evolution of water relations in land plants, through the morphological innovations of stomata, vascular tissue and roots. Third, I report the spread and evolution of drought tolerance across the plant phylogeny, a key stressor accompanying plant terrestrialisation. This leads to the discovery that the first land plants and vascular plants were desiccation and drought tolerant respectively. Finally, I detail an evolutionary approach for identifying uncharacterised drought tolerance genes, through incorporating trait evolution into a comparative genomics framework. Preliminary experimental analysis aims to provide support for this novel technique. This work, on the common theme of plant evolution, advances research into gene innovation and diversification as well as detailing a novel method to identify uncharacterised drought tolerance genes. Ultimately, the research presented in this thesis contributes to our understanding of the major transitions of plants via insights gained from the study of genes, genomes and traits.

Acknowledgements

I would like to thank my supervisors Dr Jordi Paps and Dr Ulrike Bechtold for providing me with this incredible opportunity and their guidance and support throughout my PhD. I am incredibly grateful for their time, patience and hard work in helping me to develop as a research scientist. I would also like to thank past and present members of the Paps and Bechtold laboratories for their support, encouragement and advice during my time at the University of Essex. Members who merit particular mentions include Cristi, for her help with bioinformatics analysis and Osi, for all his help in the lab.

I am also appreciative of feedback and advice from past and present members of the Plant and Genomics Groups. A particular mention goes to Stuart Newman for his support and help in using the Genomics High Performance Computing server. I am extremely grateful to Dr Matt Jones and Professor Philip Mullineaux for acting as my board members during the course of my PhD, providing invaluable feedback.

Finally, I would like to thank members of the Bowles and Worsley families. To my Mum and brother, for all their encouragement, inspiration and for listening to me prattle on about plants. To my father, Dr Michael Bowles, who passed away in 2000, for inspiring this whole journey. To the Worsley family, for all their support and frequent supply of bakes. And last of all, I will be forever grateful to my fiancée Sarah for her unwavering support and being a constant source of motivation.

Table of Contents

List of Figures

List of Tables

Abbreviations

µg	Microgram
µl	Microlitre
ANOVA	Analysis of Variance
BLAST	Basic Local Alignment Search tool
bp	Base pairs
cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotides
EDTA	Ethylenediaminetetraacetic acid
HG	Homology Group
LCA	Last Common Ancestor
MCL	Markov Chain Clustering
MYA	Million years ago
NCBI	National Center for Biotechnology Information
PCR	Polymerase Chain Reaction
qPCR	Quantitative polymerase Chain Reaction
RNA	Ribonucleic acid

Publications arising from this work

Chapter 1

Bowles AMC (2020). Understanding plant evolution in the genomic era. *In preparation for New Phytologist.*

Chapter 2

Bowles AMC, Bechtold U, Paps J (2020). The evolution of land plants is rooted in two bursts of genomic novelty. *Current Biology* **30**(3): 530-536

Chapter 3

Bowles AMC, Paps J, Bechtold U (2020). Different genome evolution modes underline the evolution of water relations in land plants. *In Review in PNAS*.

Chapter 4

Bowles AMC, Paps J, Bechtold U (2020). Life out of water: the origin of drought and desiccation tolerance in plants. *In review in Frontiers in Plant Science*.

Obomighie I, Lapenas K, Murphy BE, Bowles AMC, Bechtold U, Prischi F (2020). Ribosomal Protein S6 Kinases roles in plant homeostasis. *Submitted to Frontiers in Molecular Biosciences*

Chapter 1 General introduction

1.1 Introduction

For the first four billion years of life on Earth, the terrestrial surface would have been an inhospitable environment to inhabit with only bacteria and a few fungi able to survive in these relatively harsh conditions (Horodyski *et al.*, 1994; Betts *et al.*, 2018). However, around 500 million years ago, the first plants moved from aquatic environments onto land (Morris *et al.*, 2018), and, following this, an immense diversity of plant life evolved. Indeed, studies have estimated that the total number of extant plants could easily exceed 450,000 species (Pimm *et al.*, 2014, 2015). This evolutionary journey has seen multiple adaptations arise in the first plants that colonised land, the independent evolution of trees and finally the evolution of seeds and flowers (Leebens-Mack *et al.*, 2019). The diversification of plants has changed the Earth's atmosphere, climate and biogeochemical cycles (Lenton *et al.*, 2016) and has also promoted the evolution of a huge diversity of fungal and animal species (Lutzoni *et al.*, 2018; Li *et al.*, 2019).

At the core of this thesis, linking all research chapters, is the evolution of plants and the plant phylogeny. The current framework of plant evolutionary history forms the foundation of all analyses described in this thesis (Figure 1.1). Therefore, it is important to understand the current depth of knowledge of the evolutionary history of plants, the limits of this knowledge and areas of ambiguity and contention surrounding plant evolution. The major plant groups as well as their evolutionary relationships and defining characteristics are discussed briefly below.

1.2 The evolutionary history of plants

1.2.1 Archaeplastida

Eukaryotes are divided into six supergroups with all plant species placed in the Archaeplastida (Burki *et al.*, 2020). Archaeplastida (kingdom Plantae *sensu lato*) is a group of plants that consist of red algae (Rhodophyta), Glaucophyta and green plants (Viridiplantae) (Figure 1.1). The oldest fossil evidence of a member of the Archaeplastida is the red algae, *Bangiomorpha pubescens*, dated at approximately one billion years old

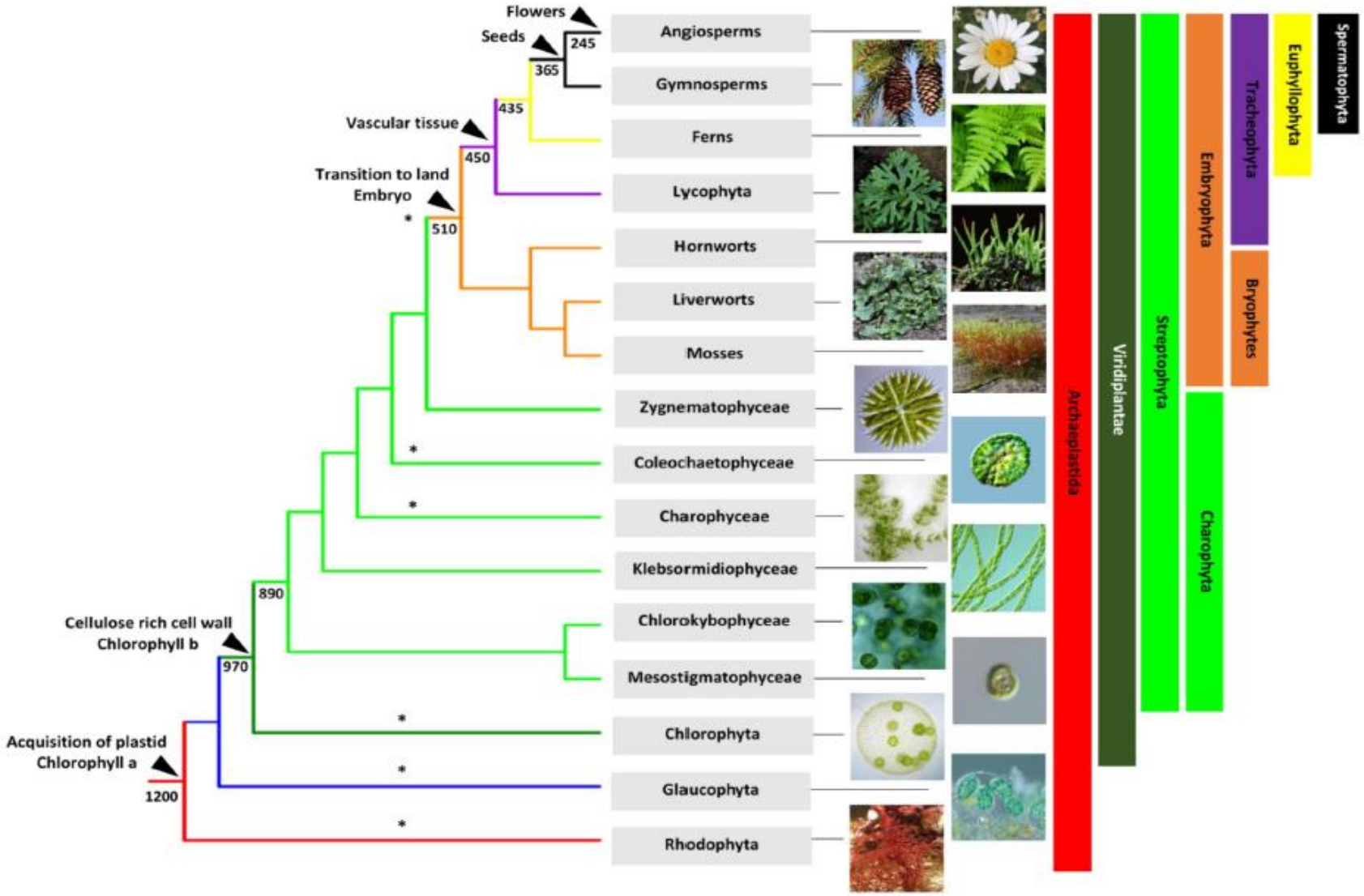


Figure 1.1. The evolutionary history of plants. Coloured branches in the tree corresponds to the group names on the right side of the figure. Arrowheads highlight important diversification events and the biological innovations associated with these events. Timing of diversification events is denoted by the numbers under arrowheads (in millions of years ago) (Morris *et al.*, 2018). Images beside group names illustrates species in these groups. Asterisks indicate groups with multicellular species.

(Butterfield, 2000; Gibson *et al.*, 2018). Over one billion years ago, the common ancestor of Archaeplastida acquired a plastid via the endosymbiosis of a cyanobacterium (Rodríguez-Ezpeleta *et al.*, 2005; Lee *et al.*, 2016b). This process involved the integration of a plastid into the cell of the ancestral Archaeplastida and would have allowed it to utilise this new photosynthetic organelle to convert light into chemical energy. This is marked as a pivotal event in plant evolutionary history as it gave rise to the first photosynthetic eukaryotes (Collén *et al.*, 2013).

There are an estimated 6,000 described Rhodophyta species (Guiry, 2012). Two species, recently described as the new phylum, Rhodelphidia, were identified as the sister group to red algae (Gawryluk *et al.*, 2019). Although red algae are mainly found in marine habitats, they have been identified in diverse environments that include hot acid springs (Matsuzaki *et al.*, 2004; Schönknecht *et al.*, 2013) and coastal caves (Azua-Bustos *et al.*, 2012). The multicellular red algae, *Porphyra umbilicalis*, or laver seaweed, inhabits intertidal zones, experiencing desiccation, osmotic stress and extremes of ultraviolet light (Brawley *et al.*, 2017). Despite their diversity, there are several characteristics such as pigmentation by phycobiliproteins (the red and blue pigments are phycoerythrin and phycoacyanin respectively, Sfriso *et al.*, 2018), the lack of cytoskeletal structures linked to motility and a reduced gene set that are shared between all red algae (Qiu *et al.*, 2015). Red algal morphology ranges from the unicellular class, Cyanidiophyceae (Bhattacharya *et al.*, 2013), to branched, multicellular species as large as 2m in size. This is an example of the convergent evolution of multicellularity in plants, with other transitions found in chlorophytes, charophytes and land plants (Parfrey *et al.*, 2013) which are discussed in later sections.

Glaucophyta are a small group of freshwater unicellular algae with 14 known species (Guiry, 2012). The branches of early plant evolution remain ambiguous, specifically with regards to the placement of glaucophytes and rhodophytes in relation to green plants (Palmer *et al.*, 2004). However, analysis of the first glaucophyte genome, *Cyanophora paradoxa*, provided strong evidence that glaucophytes and green plants are most closely related (Price *et al.*,

2012, 2019; Leebens-Mack *et al.*, 2019). Similar to red algae, glaucophytes contain phycobilins and a plastid. Additionally, glaucophytes have flagella which enables them to be motile (Price *et al.*, 2019).

1.2.2 Viridiplantae

Viridiplantae (kingdom Plantae sensu stricto) consists of two major groups, Chlorophyta and Streptophyta (Figure 1.1), that diverged approximately 970 million years ago (Ruhfel *et al.*, 2014; Morris *et al.*, 2018). There are now an estimated 450,000 – 500,000 species of Viridiplantae (Corlett, 2016; Lughadha *et al.*, 2016). The origin of Viridiplantae is marked by the loss of phycobiliproteins which, as mentioned, are found in rhodophytes and glaucophytes (Tomitani *et al.*, 1999). Also emerging in the ancestor of green plants is the photosynthetic pigment chlorophyll b which enabled the absorption of a greater spectrum of light than was possible for the ancestor of Archaeplastida, which only possessed chlorophyll a (Lewis *et al.*, 2004). Another defining feature of all Viridiplantae is the development of a more complex cell wall, which occurs through starch synthesis within the plastid (Popper *et al.*, 2011).

Although the relationships amongst Viridiplantae are fairly well understood, there remain areas of the plant tree of life that are contentious. Recent analysis of the genome of the marine green alga *Prasinoderma coloniale* identified a third phylum of Viridiplantae, Prasinodermatophyta, that emerged before the divergence of Chlorophyta and Streptophyta (Li *et al.*, 2020b).

Chlorophytes are a monophyletic group containing 8,000 described species, with a diversity of adaptations, morphologies and life histories (Guiry, 2012). The oldest fossil evidence for chlorophytes has been dated to 800-1000 million years old and these have been identified as multicellular organisms (Butterfield *et al.*, 1994; Tang *et al.*, 2020). The ancestor of Chlorophyta likely diversified in marine environments in the Neoproterozoic era (1000-541 mya) leading to the evolution of the core chlorophytes (Ulvophyceae, Trebouxiophyceae, Chlorophyceae) which subsequently radiated into marine, freshwater and terrestrial environments (Leliaert *et al.*, 2011, 2012; Fang *et al.*, 2017). Chlorophyte morphology ranges

from unicellular organisms (e.g. *Ostreococcus tauri,* Derelle *et al.*, 2006) to colonial (e.g. *Volvox carteri,* Prochnik *et al.*, 2010) and multicellular algae (e.g. *Ulva mutabilis,* or sea lettuce, De Clerck *et al.*, 2018). This represents another instance of the evolution of multicellularity in plants (Umen, 2014).

1.2.3 Streptophyta

Streptophyta diverged approximately 890 mya and consist of charophytes and embryophytes (land plants) (Figure 1.1). Unlike chlorophytes, charophyte algae are paraphyletic, which is defined as a group of organisms descended from a common ancestor but that does not include all descendants (in this case, Embryophyta, Civan *et al.*, 2014). Streptophyte algae are found in a range of brackish, freshwater and terrestrial habitats, which demonstrates the range of adaptations within this group to water availability (Fürst-Jansen *et al.*, 2020). Similar to chlorophyte algae, charophytes demonstrate a diverse range of morphologies further exemplifying the convergent evolution of multicellularity in plants (Umen, 2014). Six morphologically distinct groups have been identified (Figure 1.1): single celled Mesostigmales (e.g. *Mesostigma viride,* Liang *et al.*, 2019), sarcinoid (a cluster of cells) Chlorokybales (e.g. *Chlorokybus atmophyticus,* Wang *et al.*, 2019), filamentous Klebsormidiales (e.g. *Klebsormidium flaccidum,* Hori *et al.*, 2014), multicellular three-dimensional Charales (e.g. *Chara braunii,* Nishiyama *et al.*, 2018), multicellular two-dimensional Coleochaetales and filamentous Zygnematales (e.g. *Mesotaenium endlicherianum,* Cheng *et al.*, 2019). The interrelationships between these groups have been highly contested but the latest plant phylogenies, which are based on data from one thousand plant transcriptomes, place Zygnematophyceae as a sister group to land plants (Wickett *et al.*, 2014; Leebens-Mack *et al.*, 2019). Over 6,000 species of Charophyta have been described, with Zygnematales recognised as the most species rich group (Guiry, 2012).

Genome analysis has identified that the transition of plants from water onto land (terrestrialisation) was preceded by major innovations previously thought to be land plant specific (Xu *et al.*, 2011; Hori *et al.*, 2014; Nishiyama *et al.*, 2018; Wang *et al.*, 2019). These

include the symbiotic association of plants with beneficial fungi (Delaux *et al.*, 2015), a partial genetic toolkit for directing stress responses (Bowman *et al.*, 2017; de Vries *et al.*, 2018a, 2018b; Fürst-Jansen *et al.*, 2020) as well as cell wall modifications (Hori *et al.*, 2014; Mikkelsen *et al.*, 2014; Nishiyama *et al.*, 2018; Wang *et al.*, 2019; Jiao *et al.*, 2020).

1.2.4 Embryophyta

Over 500 million years ago, the first plants moved from aquatic environments onto land which is marked as one of the most important developments in plant evolution (Morris *et al.*, 2018). Terrestrialisation had major impacts on global biogeochemical cycles, leading to reductions in atmospheric CO_2 (Lenton *et al.*, 2012) and an increase in oxygen production (Lenton *et al.*, 2016). The conquest of land also resulted in the development of new habitats for animals (Labandeira, 2013) as well as changes to soil types and the formation of new river systems (Gibling *et al.*, 2012). Terrestrial colonisation has been attributed to a series of major innovations in plant anatomy and biochemistry. Common features required for plant life on land and therefore present in the first land plants are three dimensional growth, rhizoids (root-like structures), stomata (pores) and the alternation of generations (Harrison, 2017). The latter of these involves two distinct phases in the plant life cycle, alternating between sporophyte (non-sexual phase) and gametophyte (sexual phase) forms. Recent studies have also shown that the evolution of plants was coordinated by the evolutionary development of increasingly complex signalling molecules (Bowman *et al.*, 2017) and genetic networks (Catarino *et al.*, 2016).

Embryophyta (land plants), consisting of bryophytes (e.g. mosses) and tracheophytes (vascular plants), diverged approximately 450 mya (Figure 1.1). The phylogeny of early land plants is widely debated but the latest research classifies bryophytes as monophyletic (a group of organisms that share a most recent common ancestor) and a sister group to vascular plants (Puttick *et al.*, 2018; Harris *et al.*, 2020).

The bryophytes consist of liverworts (9000 species), mosses (12,700 species) and hornworts (225 species) which are all closely related to the first plants that colonised land (Figure 1.1) (Christenhusz *et al.*, 2016). Bryophytes lack vascular tissue and true roots but possess key innovations for life on land including the ability for 3-dimensional growth as well as specialised morphological and physiological adaptations, such as the ability to completely dehydrate and recover (Bowman *et al.*, 2017). Specific structures required for life in terrestrial environments are found in all land plants, such as rhizoids and root hairs, which are needed for water uptake and anchorage (Jones *et al.*, 2012). However, other structures show a marked phylogenetic distribution in bryophytes, for example there is evidence of reductive evolution or the loss of key traits in liverworts including rhizoid structures, sporangium development and spore wall structures (Puttick *et al.*, 2018). Stomata, the pores that regulate gas exchange in plants, were present in the ancestor of land plants and are present in every lineage apart from liverworts (Harris *et al.*, 2020). In liverworts, the air pore complex has instead independently evolved to enable gas exchange (Jones *et al.*, 2017). The evolutionary development of these analogous features likely required an individual genetic toolkit, facilitated by lineage specific gene group novelty and expansion.

1.2.5 Tracheophyta and Euphyllophyta

Tracheophytes can be divided into two major extant plant groups, the Lycophyta and Euphyllophyta (Figure 1.1). Distinguishing innovations of tracheophytes (also known as vascular plants) is a vascular system for the transport of water and minerals, bifurcation which is the division of shoots and roots into two branches, and a sporophyte dominated life cycle (Harrison, 2017). Additionally, there are several extinct lineages of land plants that diverged after the split of bryophytes and have been identified from fossil evidence. These include Horneophyton and Aglaophyton which form the protracheophyte group, as well as Rhyniopsids which are defined as early diverging tracheophytes (Kenrick *et al.*, 1997).

The Pteridophyte Phylogeny Group summarised that the 1290 lycophyte species can be placed into three orders, namely the Lycopodiales, Isoëtales, and Selaginellales (Christenhusz

et al., 2016; Schuettpelz *et al.*, 2016). There are several examples of independent evolution of important biological innovations in the lycophytes and other vascular plant groups. This includes the evolution of roots, which occurred once in the ancestor of Lycophyta and once in the ancestor of Euphyllophyta (Hetherington *et al.*, 2018). Leaves have also evolved independently at least three times, in the ancestor of Lycophyta, Monilophyta and Spermatophyta (Tomescu, 2009).

The group Euphyllophyta, which diverged approximately 435 mya, consists of Monilophyta (ferns and allies) and Spermatophyta (seed plants) (Figure 1.1). Monilophyta are a speciose plant group containing an estimated 10,560 species (Christenhusz *et al.*, 2016). As highlighted above, the innovation shared amongst all euphyllophytes are true roots (Doyle, 2017). Another example of convergent evolution in plants is the origin of lateral roots, which occurred in the ancestor of seed plants and on multiple occasions in ferns (Hetherington *et al.*, 2020).

1.2.6 Spermatophyta

Spermatophyta (seed plants) can be split into two major plant groups, the gymnosperms and angiosperms (flowering plants) (Figure 1.1). The extinct sister group of spermatophytes are early lignophytes such as Archaeopteris (Meyer-Berthaud *et al.*, 1999). Fossil evidence suggests that these plants reproduced in a similar way to lycophytes and ferns, via the dispersal of spores through the air (Meyer-Berthaud *et al.*, 1999). In the first seed plants, a different method of reproduction emerged through the development of seeds. Seeds are fertilised by pollen which can be transported by wind, water and animals (Linkies *et al.*, 2010). This revolutionary reproductive strategy minimised the influence of external environments and enabled plants to proliferate in terrestrial ecosystems as their dependence on water for reproduction was reduced. In addition to this, the ancestor of seed plants also possessed secondary xylem and phloem, collectively known as the vascular cambium. The emergence of secondary vasculature enabled the evolution of new plant forms, including large forest trees and woody vines such as lianas (Spicer *et al.*, 2010).

Based on fossil evidence (from Elkinsia and Moresnettia) and molecular dating, Spermatophyta emerged around 365 million years ago (Serbet *et al.*, 1992; Morris *et al.*, 2018). There are approximately 1000 species of gymnosperm which can be grouped into five subclasses, which are the Pinaceae, Cycads, Ginkgos, Gnetophytes and Cupressophytes (Zhong *et al.*, 2010; Lu *et al.*, 2014; Wang *et al.*, 2014e; Christenhusz *et al.*, 2016).

1.2.7 Angiosperms

Flowering plants, or angiosperms, diversified only 209 mya (Figure 1.1) but represent the most successful group of land plants in terms of both distribution and number of species (Barba-Montoya *et al.*, 2018; Morris *et al.*, 2018; Li *et al.*, 2019). Charles Darwin described the diversity and speed of the evolution of angiosperms as the “abominable mystery” (Davies *et al.*, 2004). Considering angiosperms are the plant group that diversified most recently, their diversity is unparalleled with approximately 350,000 - 500,000 extant species (Christenhusz *et al.*, 2016).

Reports from the Angiosperm Phylogeny Group (APG) have improved our understanding of flowering plant evolution (Bremer, 1998; Bremer *et al.*, 2003, 2009). Current thinking assigns the ANA grade angiosperms (Amborellales, Nymphaleales, Austrobaileyales) as the sister group to the Mesangiosperms. Subsequently Magnoliids and Chloranthales are sister to a clade containing Monocots, Ceratophyllales and Eudicots. According to the latest APG report, Monocots are the sister group of Ceratophyllales and Eudicots (Chase *et al.*, 2016). However, there is contention over these relationships, particularly the placement of Ceratophyllales and Chloranthales, as highlighted by analysis from the one thousand plant transcriptomes project (Leebens-Mack *et al.*, 2019).

Diverse and species rich families within the flowering plants include the orchid (Orchidaceae), sunflower (Asteraceae), sedge (Cyperaceae) and mustard families (Brassicaceae) (Christenhusz *et al.*, 2016). Additionally, many flowering plants are of high economic, agricultural and cultural importance (Chen *et al.*, 2018) which is reflected in the number of sequenced representatives, with well sequenced plant families including the legumes

(Fabaceae) (Griesmann *et al.*, 2018) and the grasses (Poaceae) (Goff *et al.*, 2002; Yu *et al.*, 2002; Vogel *et al.*, 2010; Zhang *et al.*, 2014b; Stein *et al.*, 2018).

Despite their diversity, angiosperms have many shared innovations that facilitated their rise to ecological dominance. These mostly relate to the evolution of their reproductive biology (Jiao *et al.*, 2011). The carpel, a closed structure containing the ovules, is found only in flowering plants. Additionally, in most flowering plants, double fertilisation occurs with the first event producing the seed and the second event producing the endosperm, a nutritive tissue which feeds the growing seed (Endress, 2011; Soltis *et al.*, 2016). Double fertilisation and protected ovules are both common to flowering plants, although examples of double fertilisation are found elsewhere in the plant phylogeny (e.g. Gnetales) (Wan *et al.*, 2018a). This provided a competitive advantage for angiosperms by enabling plants to establish in previously hostile environments.

An additional trait to emerge in the last common ancestor (LCA) of flowering plants were leaves with reticulate veins, which are web or network-like patterns, enabling controlled movement of water and food (Boyce *et al.*, 2009). Stems with specialised xylem vessels also emerged in the ancestor of flowering plants. These structures are in contrast to the tracheids found in non-flowering plants which were simpler in comparison (Trueba *et al.*, 2019). These innovations enabled the first flowering plants to efficiently transport water throughout the plant.

Finally, the LCA of angiosperms were the first plants to evolve flowers. The earliest fossil evidence identifies unequivocal evidence of angiosperm flowers at around 125 mya, whilst fossilised pollen grains have been dated at ~135 mya during the early Cretaceous (Sun *et al.*, 1998, 2002). Recent analysis of floral traits from across the flowering plant tree of life suggested that the ancestral angiosperm flower was likely bisexual and radially symmetric (Sauquet *et al.*, 2017). The emergence of these innovations enabled the rapid diversification of angiosperms and had important implications for global biodiversity. Flowering plants have complex interactions with microbes (Rebolleda-Gómez *et al.*, 2019), fungi (Lutzoni *et al.*,

2018), pollinators (van der Kooi *et al.*, 2020) and seed dispersers (Eriksson, 2016) and are therefore fundamentally important for terrestrial ecosystems (Crane *et al.*, 1995).

1.3 Genes and genomes

Although there are several areas of contention in the plant phylogeny (for example bryophytes as a sister group to vascular plants), the major relationships are well resolved, particularly in comparison to the evolutionary history of animals (Jékely *et al.*, 2015; Pisani *et al.*, 2015; Feuda *et al.*, 2017; Simion *et al.*, 2017; Whelan *et al.*, 2017). Many of the major evolutionary steps in the plant tree of life are also well characterised, for example the transition from water onto land as well as the evolution of roots, seeds and flowers. Our understanding of plant evolution is now at its most advanced, partially due to the revolution of DNA sequencing technologies. This technological revolution has made the sequencing of problematic taxa, more feasible, in terms of cost, speed and accuracy (Koonin *et al.*, 2000) and has also improved the quality and quantity of plant transcriptome data which, in turn, has improved our understanding of the phylogenetic relationships between species (Wickett *et al.*, 2014; Puttick *et al.*, 2018; Leebens-Mack *et al.*, 2019). This revolution has also increased the availability and quality of plant genomes which is described in more detail below.

1.3.1 Revolution in genome sequencing

The first plant to have its genome sequenced was the model organism, *Arabidopsis thaliana*, in 2000 (The Arabidopsis Genome Initiative, 2000). Following on from this, the first crop plants to have their genomes sequenced were two subspecies of *Oryza sativa,* or cultivated rice, in 2002, costing approximately $100 million (Goff *et al.*, 2002; Yu *et al.*, 2002). In comparison, a genome of a similar size to rice (approximately 420 megabases, or Mb) can now be constructed de novo for around $10,000 (Li, 2018). This decline in cost can be attributed to the improvement of existing sequencing technologies (e.g. Sanger sequencing) and the development of Next Generation Sequencing approaches (e.g. Illumina, PacBio and Oxford Nanopore) (Goodwin *et al.*, 2016; van Dijk *et al.*, 2018).

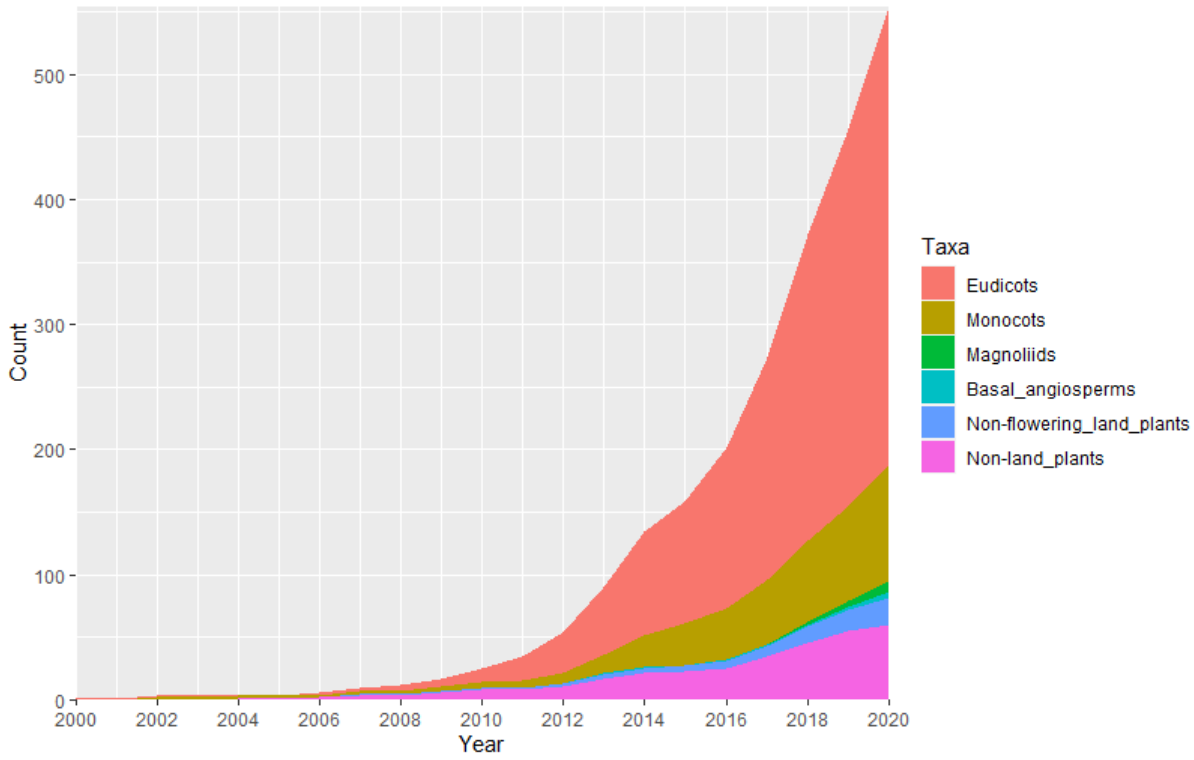


Figure 1.2. The increasing availability of plant genomes, coloured by taxonomic group. Information sourced from plabipd.de.

As a result of these declining costs, genome data from across the plant tree of life has been produced at an increasingly high rate (Figure 1.2). For example, at the beginning of my PhD research (October 2017), 178 well annotated plant genomes were available (Bowles *et al.*, 2020) whereas, at the time of writing this thesis, over 550 plant genomes have now been sequenced (Table 1.1). Considering approximately 550 plant genomes have been sequenced in the last twenty years, the availability of genomic data is predicted to increase exponentially in the next five to ten years. This combination of a well resolved phylogeny and unprecedented amounts of whole genome data has allowed us to begin to ask questions about the molecular evolution of plants.

Table 1.1. The number of publicly available, assembled genomes in 2017 and 2020.

Plant group	Number of genomes 2017	Number of genomes 2020
Rhodophyta	4	10
Glaucophyta	1	1
Chlorophyta	14	42
Charophyta	1	7
Bryophyta	2	8
Lycophyta	1	3
Monilophyta	0	3
Gymnosperms	3	8
Basal Angiosperms	1	4
Magnoliids	0	9
Monocots	43	92
Early diverging eudicots	5	9
Rosids	64	217
Asterids	39	141
Total	**178**	**554**

1.3.2 Genome diversity

Genomic innovation, variation and complexity is increasingly being recognised as a significant factor in the diversification of life on Earth. For example, whole genome duplications are considered fundamental to the expansion of many plant lineages (Clark *et al.*, 2018). There is large variation in both the size and structure of plant genomes, with these differences ranging from single nucleotide polymorphisms (SNPs) to genome wide duplications, deletions and rearrangements (Saxena *et al.*, 2014). Whilst land plant genomes have a 2,400 fold range in size (Pellicer *et al.*, 2018), there are only a few species that exhibit gigantism as maintaining a large genome is costly (Simonin *et al.*, 2018). *Paris japonica* has the largest genome (149 gigabases) of any plant, as well as any eukaryotic organism, recently surpassing the last record holder, the marbled lungfish, *Protopterus aethiopicus* (Pellicer *et al.*, 2010). Ferns also typically have larger than average genomes. For example, the genome of the whisk-fern

Tmesipteris obliqua is only marginally smaller than *Paris japonica* (Hidalgo *et al.*, 2017; Pellicer *et al.*, 2018). This has acted as a technological barrier to sequencing many fern genomes (Sessa *et al.*, 2014), meaning that to date, only three fern genomes have been sequenced (Li *et al.*, 2018a; Marchant *et al.*, 2019). At the other end of the scale, *Genlisea tuberosa*, a carnivorous bladderwort which inhabits nutrient poor environments, has the smallest genome of any plant at 61 Mb (Leushkin *et al.*, 2013). Finally *Utricularia gibba*, a member of the same family as *Genlisea tuberosa* has a very small genome of 82 Mb (Ibarra-Laclette *et al.*, 2013), which contains 28,500 protein coding genes; these represent 97% of the genome and therefore only 3% of its genome is non-coding DNA.

This large variation in the size of plant genomes can often be attributed to the frequency of Whole Genome Duplication (WGD) events also known as polyploidy (Wendel *et al.*, 2016). The extra genetic material in organisms with duplicated genomes provides the potential for evolutionary innovation of new traits that can provide a competitive advantage. A WGD event leads to an organism gaining an extra copy of every single gene in its genome, during the cell division phase of sexual reproduction (Jiao *et al.*, 2011). Genome analysis of *Amborella trichopoda*, an early branching angiosperm, suggested that the first flowering plants arose when an ancestral spermatophyte underwent WGD around 150 million years ago (DePamphilis *et al.*, 2013). The extra genetic material gained through paleopolyploidy, an ancestral WGD, allows for functional diversification of genes. In the case of the first angiosperms, this provided the ability to evolve new structures such as flowers. This finding has since been supported through the analysis of one thousand plant transcriptomes, which has identified an additional 244 ancient WGDs across all Viridiplantae (Leebens-Mack *et al.*, 2019).

In addition to experiencing WGD, genomes can undergo a series of restructuring events by families of lineage-specific transposable elements. It is important to note that these events can lead to both genome shrinkage as well as expansion (Wendel *et al.*, 2016). Genome stabilisation, the regulatory processes that govern genome structure and content through

mechanisms such as gene silencing and gene conversion, can also lead to the novel or partial expression of genes, which are respectively known as neo- or sub-functionalisation. The global diversification of plants can be largely attributed to the frequency of these events and their subsequent genomic plasticity (Soltis *et al.*, 2009).

Understanding the consequences of WGDs for gene evolution and the proliferation of gene families are still major unanswered questions in evolutionary biology. The increasing availability of genomic data means that our understanding of how plant genomes have evolved has improved greatly and the variation in both genome size and structure can be explored (Palmer *et al.*, 2004). Crucially, we can use this information to begin to identify the genomic changes that have accompanied the origin of different plant groups and therefore also unravel the molecular basis of biological innovations and adaptations.

1.3.3 Gene family evolution

Maynard Smith and Szathmary's seminal work highlighted the synthesis of DNA, the origin of the eukaryotic cell and the evolution of multicellularity as major transitions in evolution (Szathmáry *et al.*, 1995). The evolution of genes underpin these major transitions. For example, genome analysis of the multicellular chlorophyte algae, *Volvox carteri*, identified the expansion of gene families associated with multicellularity (Prochnik *et al.*, 2010). In particular, genes involved in the production of cyclin proteins, which are linked to the coordination of cell division, and pherophorins, involved in increasing the complexity of the cell wall, had diversified when compared to those in the unicellular *Chlamydomonas reinhardtii*. An additional example comes from an analysis of the genome of the most morphologically complex charophycean algae, *Chara braunii*, which found that multiple gene families, linked to organismal complexity, had emerged and expanded (Nishiyama *et al.*, 2018). For example, 730 transcription factors and regulators were identified, compared to 627 and 542 found in *Klebsormidium nitens* and *Chlamydomonas reinhardtii* respectively.

With a particular focus on plant evolution, other major evolutionary transitions have been identified, including the transition of plants from aquatic environments to land and the origin of stomata, vascular tissue, roots, seeds and flowers (as highlighted above) (Ligrone *et al.*, 2012; Harrison, 2017). Again, these transitions have also been accompanied by the evolution of new genes. For example, it has recently been identified that almost all transcription factor families, which are associated with multiple developmental processes, were present in the last common ancestor of land plants (Catarino *et al.*, 2016). Additionally, an analysis of Zygnematophyceae genomes, the closest extant relatives of land plants, found that many phytohormone signalling and biosynthesis genes emerged in the first land plants (Cheng *et al.*, 2019; Jiao *et al.*, 2020). Genome analysis of the first lycophyte genome, *Selaginella moellendorffii*, identified that the ancestor of Tracheophyta contained at least 7247 gene families; 27 of these emerged alongside the origin of Tracheophyta and were linked to developmental innovations of vascular plants (Banks *et al.*, 2011). Finally, analysis of an early diverging flowering plant, the blue-petal water lily (*Nymphaea colorata*), found that the ancestor of the genus experienced a whole genome duplication event and that the retained genes from this event included those involved in regulating flowering transition and development (Zhang *et al.*, 2019a). Such changes in the gene content of organisms form the foundation for biological innovation. These examples of gene family emergence or expansion can be seen across the plant tree of life and are linked to a plethora of major transitions.

Equally important for the evolution of plants are the genes that underpin key adaptations and traits. Many interesting examples of genes underlying adaptations have been elucidated by studying extremophile plants and crop wild relatives (Oh *et al.*, 2013; Bechtold, 2018; Boulc'h *et al.*, 2020). For example, there are only two flowering plant species found in the Antarctic, *Colobanthus quitensis* (Cho *et al.*, 2018) and *Deschampsia antarctica* (Lee *et al.*, 2013), which both exhibit high expression of core environmental stress response genes, for example Late Embryogenesis Abundant proteins and Ice Recrystallisation Inhibition proteins, that are vital for surviving low temperatures. Wild relatives of the model organism *Arabidopsis thaliana* have

also been studied extensively to understand the molecular basis of extreme adaptations. Genome analysis of a close relative of *A. thaliana*, *Crucihimalaya himalaica,* which has been proposed as a model for high altitude adaptation, revealed that gene families linked to intense radiation, DNA repair and low temperatures showed signs of positive selection (Zhang *et al.*, 2019c). The genetic diversity of wild relatives can, in turn, be utilised to understand and improve stress tolerance in model and crop species. A comparison of two drought tolerant Brassicaceae species, *Arabidopsis lyrata* and *Eutrema salsugineum*, with the drought sensitive *Arabidopsis thaliana*, found that increased water use efficiency and drought resistance could be attributed to high expression of key signalling genes. These genes were predominantly found in the abscisic acid (ABA) signalling pathways known to be intrinsically linked to drought tolerance (Marín-de la Rosa *et al.*, 2019).

As highlighted in the examples above the genetic basis of plant diversification and adaptation can be illuminated by investigating gene evolution and this can also improve our understanding of how genes have evolved in relation to traits of interest. As highlighted in the last example above, drought tolerance is a major constraint on crop productivity but the genetic basis of this trait is poorly understood (Fahad *et al.*, 2017). With the ever-growing need to develop stress tolerant crop varieties in the face of global climate change, it has been selected from the many traits that threaten global food security, to be examined in this thesis. Application of the evolutionary thinking described above could aid in the identification of novel genes responsible for drought adaptations.

1.4 Outline and aims of thesis

The overarching goal of this thesis is to explore the evolution of genes involved in the major transitions in plant evolution and the consequences of these events for some of the traits associated with these transitions, including drought tolerance. The approach begins by analysing plant genome data, which leads on to an experimental analysis that attempts to validate computational findings. The research detailed in this thesis will accomplish this goal by addressing the following aims:

- Examine gene gains and losses across the plant tree of life.
- Investigate gene group dynamics in relation to innovations involved in drought tolerance.
- Understand how drought tolerance as a trait has evolved.
- Discover unknown drought tolerance genes by incorporating trait evolution into a comparative genomic framework.

More specifically, the work described in the first research chapter of this thesis (Chapter 2) aims to examine the evolution of genes across the plant tree of life, thereby establishing the genetic innovations that appear during the major transitions in plant evolutionary history. Comparing the genomes of diverse plant species enables broad scale evolutionary patterns to be identified. In Chapter 3 I then move on to investigate the evolutionary history of genes that are important for the biological innovations intrinsically associated with drought tolerance, which include stomata, vascular tissue and roots. Thirdly, the research described in Chapter 4 aims to understand how drought adaptations have emerged and evolved across the plant tree of life. Finally, in Chapter 5, I aim to identify and characterise unknown drought tolerance genes, by combining the definition of drought tolerance outlined in Chapter 4 with the comparative genomic approach developed in Chapter 2. The ultimate objective of this work is to produce plants that have a greater tolerance of drought. As such, the chapter details preliminary experimental work which aims to identify the function of candidate drought tolerance genes through expression analysis and the generation of genetic mutants. Research described in Chapters 2 to 5, will be introduced and discussed in detail individually in each respective chapter, before a general discussion and conclusion of the thesis findings are presented in Chapter 6.

Chapter 2 The evolution of land plants is rooted in two bursts of genomic novelty

Bowles AMC, Bechtold U, Paps J. 2020. The evolution of land plants is rooted in two bursts of genomic novelty. *Current Biology* **30**(3): 530-536 (Appendix 1).

Authors

Alexander M. C. Bowles[1], Ulrike Bechtold[1] & Jordi Paps[123]

Affiliations

1. School of Biological Sciences, University of Essex, Colchester CO4 3SQ, UK

2. School of Biological Sciences, University of Bristol, Bristol, BS8 1TQ, UK

3. Department of Zoology, University of Oxford, Oxford, OX1 3SZ, UK

Author Contributions

Conceptualization, A.M.C.B., U.B., and J.P.; Formal Analysis, A.M.C.B.; Visualization, A.M.C.B.; Writing of Original Draft, A.M.C.B., U.B., and J.P.; Review & Editing, A.M.C.B., U.B., and J.P.; Supervision, U.B. and J.P.

2.1 Abstract

Over the last 470 million years, plant evolution has seen major evolutionary transitions such as the move from water to land and the origins of vascular tissues, seeds, and flowers (Morris *et al.*, 2018). These have resulted in the evolution of terrestrial flora that has shaped modern ecosystems, and the diversification of the Plant Kingdom, Viridiplantae, into over 374,000 described species (Christenhusz *et al.*, 2016). Each of these transitions was accompanied by the gain and loss of genes in plant genomes. For example, whole genome duplications are known to be fundamental to the origins of both seed and flowering plants (Vanneste *et al.*, 2014; Clark *et al.*, 2018). With the ever-increasing quality and quantity of whole genome data, evolutionary insight into origins of distinct plant groups using comparative genomic techniques is now feasible. Here, using an evolutionary genomics pipeline to compare 208 complete genomes, the gene content of the ancestral genomes of the last common ancestor of land plants and all other major groups of plant was analysed. This approach reveals an unprecedented level of fundamental genomic novelties in two nodes related to the origin of land plants, the first in the origin of streptophytes during the Ediacaran (629 million years ago) and another in the ancestor of land plants in the Ordovician (473 million years ago). The findings highlight the biological processes that evolved with the origin of land plants and emphasise the importance of conserved gene novelties in plant diversification. Comparisons to other eukaryotic studies suggest a separation of the genomic origins of multicellularity and terrestrialisation in plants.

2.2 Highlights

- Comparing 208 genomes gives insight into the role of gene novelty in plant evolution
- Two bursts of genomic novelty played a major role in the evolution of land plants (Figure 2.1)
- Functions linked to these novelties are multicellularity and terrestrialization
- The backbone of hormone signaling either predates or accompanies this transition

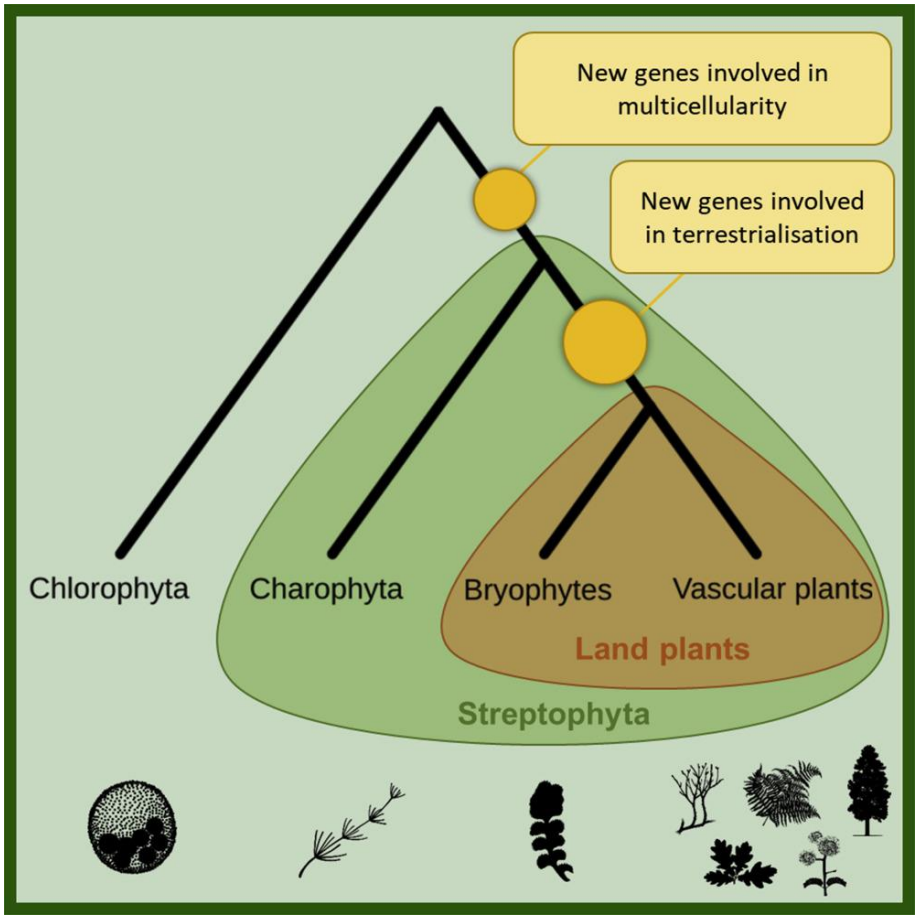


Figure 2.1 Graphical abstract summarising the key finding that the evolution of land plants was preceded by two bursts of genomic novelty.

2.3 Introduction

Understanding the diversification of plant life on Earth is still one of the major challenges in evolutionary biology. Although it is known that the diversification of plants has been driven by the evolution of key innovations, including roots, seeds and flowers, there remains much that is not known about the evolutionary history of plants. One lens through which to understand the diversification of plant life is the diversity of genes and genomes. Defining the genomic changes accompanying plant evolution is key to unravelling the molecular basis of biological innovations. Recent studies have used comprehensive taxonomic transcriptome data to understand angiosperm diversification rates and gene family expansion in the major plant groups (Landis *et al.*, 2018; Leebens-Mack *et al.*, 2019). Furthermore, reduced genomic datasets have been used to investigate whole genome duplications as well as gene family

gains and losses associated with plant diversification (Vanneste *et al.*, 2014; Wilhelmsson *et al.*, 2017; Li *et al.*, 2018a). However, the role of genomic novelty in the origins of distinct plant groups using an extensive sampling of complete genomes with a phylogenetically broad outgroup has not been fully evaluated.

2.3.1 Comparative genomics

Detailed below is a comparative genome approach to assign proteins into groups based on sequence similarity. These gene groups can be extracted based on taxonomic occupancy enabling patterns of gene gains and losses across the plant tree of life to be inferred. In this approach, similar, related proteins are clustered together into distinct groups. These proteins share a common ancestry, whether by gene duplication or speciation and, as such, are placed into groups of homology.

There are several approaches that aim to dissect homology relationships. Specifically, these relationships consist of orthologous genes which are related to one another by speciation and paralogous genes which are related by gene duplication (Koonin, 2005). These methods aim to assign proteins into orthologous groups and include packages such as OrthoDB (Kriventseva *et al.*, 2019), OrthoMCL (Li *et al.*, 2003) and OrthoFinder (Emms *et al.*, 2015, 2019). The most widely used and accurate approach for orthology assignment is OrthoFinder (Emms *et al.*, 2015). As such, it is compared in greater depth to the homology assignment approach described in this work, detailing the advantages and disadvantages of this method.

OrthoFinder begins with the reciprocal comparison of the sequence similarity between proteins. This process, similar to the homology assignment approach, uses BLAST to identify sequences of similarity within and between focal species. Next, a similarity matrix is produced based on the BLAST outputs and then Markov-chain clustering (MCL) is used to place proteins into groups. This step is also used in the homology assignment approach. However, in OrthoFinder, the BLAST bit-score output is additionally normalised by gene length. This extra step, as well as stricter BLAST and MCL parameters in OrthoFinder, enables the prediction of orthology groups, as opposed to homology groups (Emms *et al.*, 2015).

Although these approaches are highly popular, there are several limitations that are associated with most orthology assignment methods. For example these methods often struggle to detect gene fusion, fission and exon shuffling (Kuzniar *et al.*, 2008; Holland *et al.*, 2017). Additionally, there are issues in orthology detection for complex evolutionary scenarios. Lateral gene transfer (LGT), which is common amongst plants, particular between plant species (Yue *et al.*, 2012; Li *et al.*, 2014c; Cheng *et al.*, 2019; Dunning *et al.*, 2019), can be difficult to decipher (Glover *et al.*, 2019). Gene and genome duplication, a frequent feature of plant genome evolution (Flagel *et al.*, 2009; Clark *et al.*, 2018; Qiao *et al.*, 2019; Costello *et al.*, 2020), have also proved complex to disentangle, especially when this is followed by gene loss (Kapli *et al.*, 2020).

In light of the limitations surrounding orthology assignment detailed above, a homology group assignment approach was instead used in this study, incorporating both orthologous and paralogous genes in the same group. This approach is particularly suited to a plant genome dataset, due to the increased incidence of LGT and gene duplication. Additionally, this approach is less prone to the false positives and misassignments that are seen in orthology detection methods. Thus, homology group assignment was used to investigate the role of genomic novelty in plant evolution.

2.4 Results and Discussion

2.4.1 Analysing the ancestral plant gene content

Adapting a previously described (Dunwell *et al.*, 2017; Paps *et al.*, 2018) comparative genomics pipeline, 208 eukaryotic genomes, including a broad representation of animal (10), other unikont (11), and non-embryophyte bikont (29) genomes were compared (Methods, Supplementary Data 2.1, Appendix 2.1). Genome quality was assessed with BUSCO analysis, a quality control measure of genome sequencing and assembly, and genomes with more than 15% of BUSCO missing genes were discarded. Protein sequences were compared using BLAST and MCL to identify Homology Groups (HGs). To reduce the error produced by the complex evolutionary dynamics of genes involved in these transitions, further dissection of HG

was not conducted (Holland *et al.*, 2017; Paps *et al.*, 2018). Therefore, a single HG is defined as a set of proteins that have distinctly diverged from others. The 208 eukaryotic genomes contain ~9 million proteins which were clustered into ~650,000 Homology Groups. Using scripts incorporating a phylogenetic framework to inform comparative genomics, five evolutionarily distinct classifications of HG (Ancestral, Ancestral Core, Novel, Novel Core, Lost) were extracted (Supplementary Data 2.2, Appendix 2.2). Based on these outputs, patterns of large gene gains and losses were identified across the plant phylogeny (Figure 2.2).

The HG categorisation juxtaposes between the traditional gene classification (e.g. gene families, classes) and their evolutionary dynamics. Therefore a HG can either contain genes traditionally designated as subfamilies (e.g. GA3ox), gene families (e.g. Allene Oxide Cyclase) or gene superfamilies. This recovery of traditional gene classifications demonstrates the reliability of this clustering approach (Supplementary Data 2.3). There are limitations shared with other BLAST-based analyses, such as the impact of gene fusion, fission and lateral gene transfer. However, genes in broad HGs are less likely to be misassigned than orthologs and paralogs (e.g. OrthoFinder) (Pett *et al.*, 2019). The pipeline approach also tackles biases seen in tree reconciliation methods, which are prone to inaccurate assignments of gene gains and losses (Hahn, 2007).

2.4.2 The role of highly conserved gene groups in plant evolution

The evolution of Embryophyta (land plants) and Streptophyta (land plants and their closest algal relatives, Charophyta) are arguably the most dramatic transitions in the history of plants. These events have previously been linked with the expansion of many processes and developmental traits including embryogenesis (Nishiyama *et al.*, 2018), plant hormones (Wang *et al.*, 2015) and symbiotic interactions with arbuscular mycorrhizae and rhizobacteria (Field *et al.*, 2015). The analyses revealed that there was a substantial increase in the number of highly retained gene novelties in the Last Common Ancestor (LCA) of Streptophyta and the LCA of Embryophyta with 50 and 103 Novel Core HGs identified, respectively (Figure

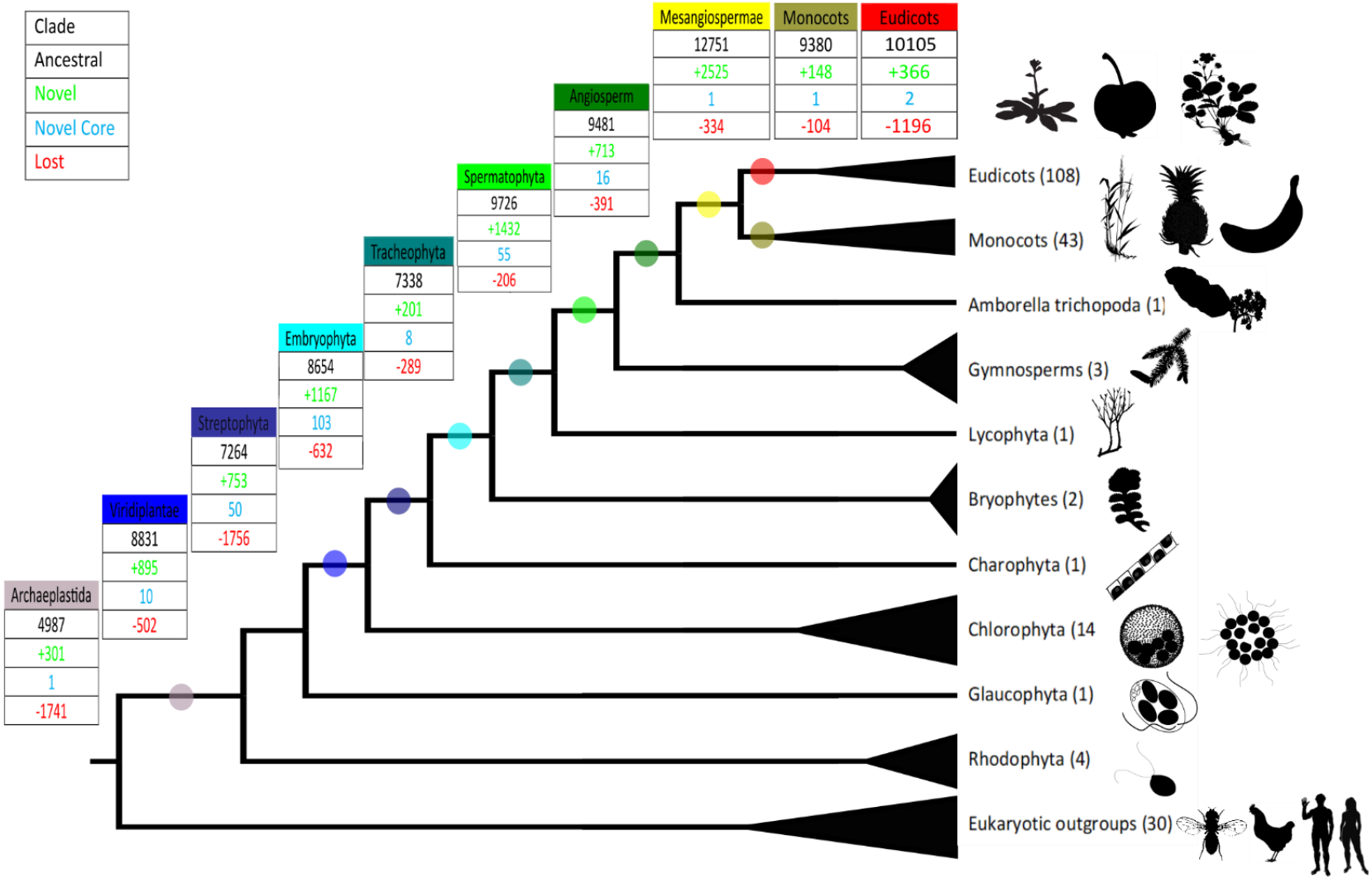


Figure 2.2 Analysis of the Gene Content of ancestral plant genomes. Evolutionary relationships of the major groups included in this study can be found in Supp. Data 2. Different categories of HG are indicated in each node, from top to bottom, Ancestral HG, Novel HG, Novel Core HG and Lost HG. Organism outlines are from phylopic.org.

2.2). Gene Ontology (GO) analyses using *Arabidopsis thaliana,* which has comprehensive GO annotations, were used to explore the modern functions of descendants of genes from Novel Core HGs (Supplementary Data 2.4, Figure 2.3). The Protein Class category was used as this classification is less prone to false assignments and biases (Paps *et al.*, 2018). All other GO categories including Molecular Function, Biological Process and Pathway were produced (Supplementary Data 2.4). HGs present in the LCA of embryophytes are abundant in classes involved in protein modification (e.g. transferase, oxidoreductase, ligase) and protein transport (e.g. transporter proteins, membrane traffic proteins) whilst HGs present in the LCA of streptophytes are abundant in gene regulation (e.g. transcription factor) and cell structure, movement and division (e.g. cytoskeletal proteins). The origins of Streptophyta were accompanied by the evolution of many plant-specific transcription factors (e.g. HD-ZIP) and an increasingly complex cell wall corresponding to the high number of the protein class hits seen in the Streptophyta Novel Core (NC) HGs (Hori *et al.*, 2014; Wilhelmsson *et al.*, 2017; Nishiyama *et al.*, 2018).

It is possible that the bursts of conserved genomic novelty could be explained by the presence of one or multiple whole genome duplications (WGDs). Inferring WGDs in these ancestral nodes is difficult with no events currently identified in the LCA of these groups (Van de Peer *et al.*, 2017; Zwaenepoel *et al.*, 2019). Analysis of over 1000 transcriptomes has identified 244 WGDs across the green plant phylogeny (Leebens-Mack *et al.*, 2019). These mostly occur after the origin of vascular plants and do not appear to coincide with the burst of novelties seen in this study. This supports the theory that there was a change in strategy from gene family birth and expansion to WGD along the backbone of the plant phylogeny. Another contributing factor that might explain the origins of some Novel Core HGs is the presence of horizontal gene transfer (HGT). BLAST searches against the Swissprot database confirmed the absence of all Novel Core HGs in outgroup taxa, validating the outputs of the pipeline approach (Supplementary Data 2.5). Queries using the pipeline approach revealed that 323 HGs were present in fungal and land plant genomes but absent in all other

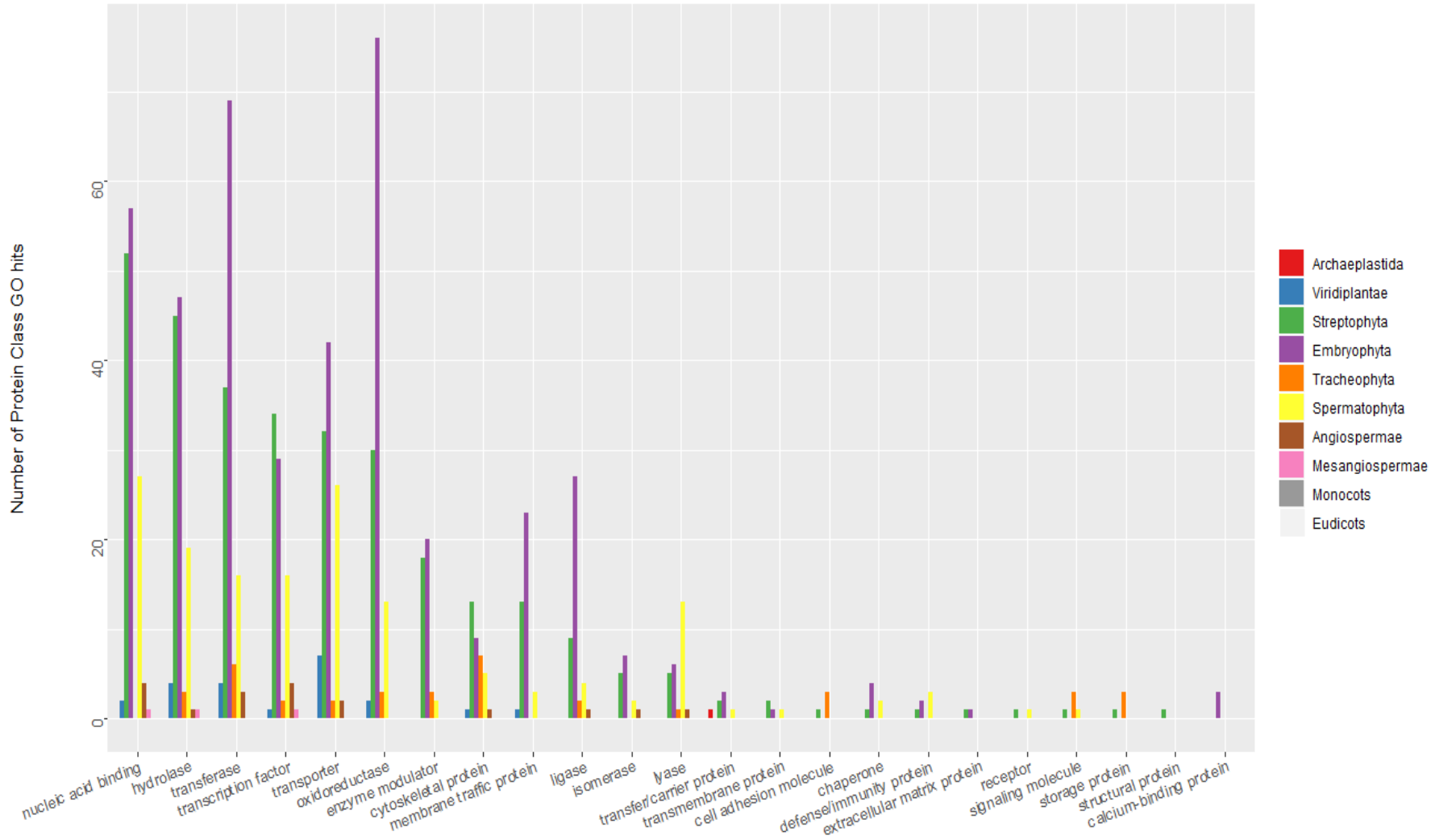


Figure 2.3. The number of Protein Class GO annotations for Arabidopsis thaliana as a representative for the Novel Core HGs at each phylogenetic node.

taxa in this study's dataset (Supplementary Data 2.1), suggesting widespread HGT in plants (Yue *et al.*, 2012; Lutzoni *et al.*, 2018). The last eukaryotic common ancestor (LECA) is the ancestor that connects all eukaryotes including plants and fungi. Either these HGs were in LECA and lost from all eukaryotic representatives aside from fungi and land plants or they are the product of HGT (Margulis *et al.*, 2006). GO analysis of 25 of the HGs that contained at least 100 embryophyte taxa revealed that they were associated with gene regulation and protein modification (Supplementary Data 2.6). Other possible HGT events that could explain the marked distribution of these Novel Core HGs include parasitism by other plants, symbiosis with other plants (e.g. transfer of a photoreceptor gene from bryophytes to ferns) and symbiosis with rhizobacteria (Yue *et al.*, 2012; Wickell *et al.*, 2019).

2.4.3 The functions of highly conserved gene groups

In streptophytes, Novel Core HGs were implicated in root, multicellular and lateral organ development (Supplementary Data 2.7; Figure 2.4). These terms were assigned based on the functions in extant *Arabidopsis thaliana* genes. In some cases, their evolutionary emergence predates the origin of the function with which they are often associated. For example, there is no evidence of roots outside the Tracheophyta, yet genes associated with root development are found in older nodes (Raven *et al.*, 2001; Hetherington *et al.*, 2018). Therefore these HGs are potential examples of co-option of old genes for new processes (Figure 2.5).

Other key functions include the increased complexity of the cell wall which is crucial for multidimensional cell growth (Becker *et al.*, 2009). Further indicators of multicellularity in the predecessor of land plants are HGs involved in the regulation of transcription, cell adhesion and division. The findings here also support an expansion of cellular signal transduction pathways associated with growth, development and stress responses in streptophytes.

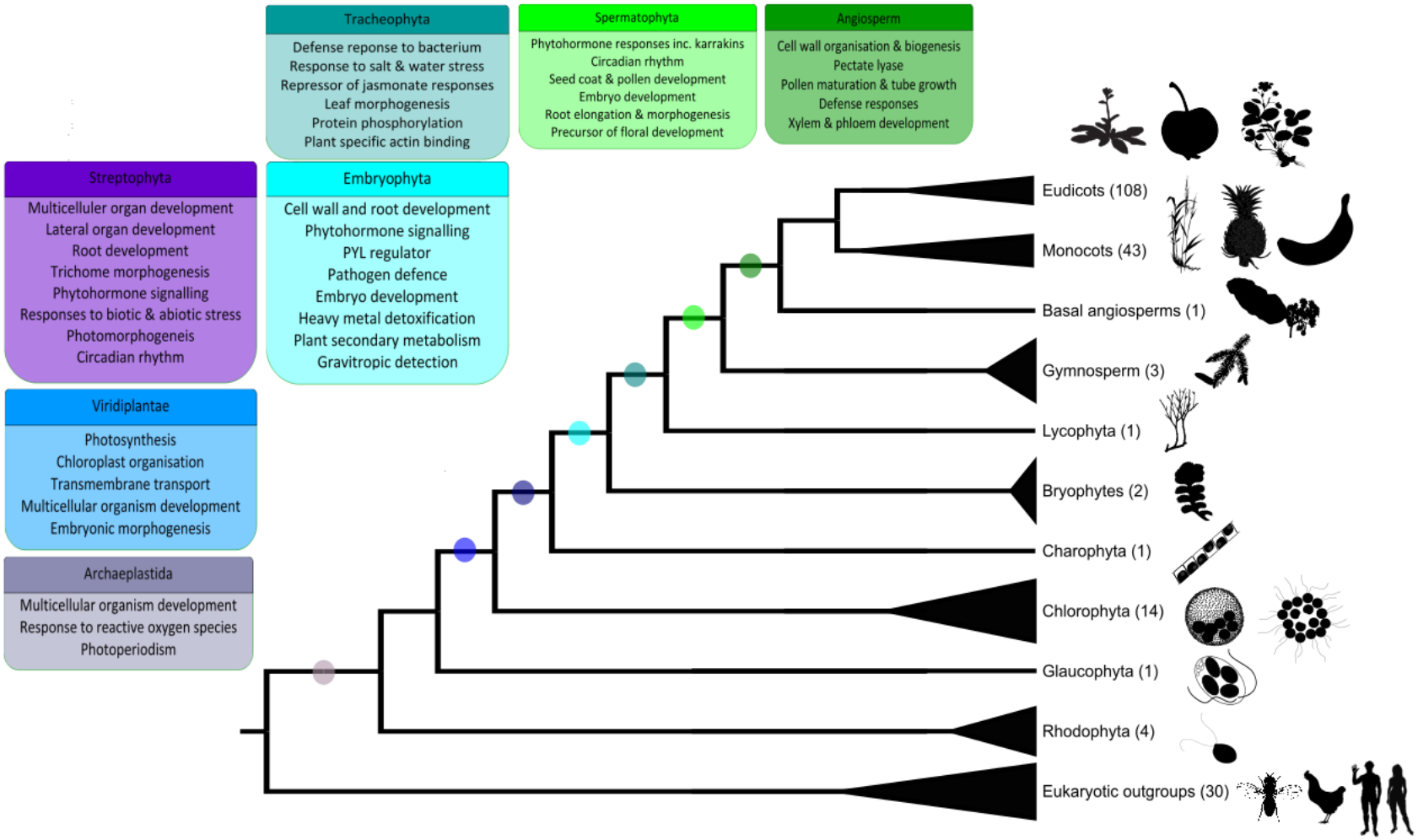


Figure 2.4 Terms identified from Novel Core Homology Groups for each phylogenetic group of plants. Evolutionary relationships of the major groups included in this study can be found in Supplementary Data 2.2. Organism silhouettes are from phylopic.org.

Many of the Novel Core HGs identified in the study have not previously been associated with the origin of land plants. These include proteins involved in plant organ development, cell wall construction and host microbe interactions (Raffaele *et al.*, 2007). Other HGs are related to terrestrialisation, with functions related to the synthesis of lignin, UV light protection and cell signalling. The latter comprise plant hormones (phytohormones) linked with growth such as auxin (body plan definition, Finet *et al.*, 2013), brassinosteroids (photomorphogenesis, Zhu *et al.*, 2013) and gibberellins, as well as those associated with environmental responses such as abscisic acid (ABA), salicylic acid and jasmonic acid (primordial root growth, Briggs *et al.*, 2006). Several Novel Core HGs including bHLH transcription factors, receptor like kinases (LRR-RLKs) and three families of heavy metal-associated isoprenylated plant proteins (HIPPs) have been previously linked to the origin of embryophytes, further validating the results (Supplementary Data 2.7) (Liu *et al.*, 2017a).

2.4.4 The evolution of phytohormone signalling

Some of these innovations have evolved in an incremental fashion. For example, phytohormone signalling genes identified as Novel Core to Streptophyta include Ethylene-overproduction protein 1, ETO1 and Ethylene Insensitive 3, EIN3 (Appendix 2.3). However, genes involved in ethylene signalling have been shown to originate before (1-aminocyclopropane-1-carboxylate synthase, ACS) and after (1-aminocyclopropane-1-carboxylate oxidase, ACO) this point in the evolutionary history of plants (Nishiyama *et al.*, 2018). Therefore these assigned functions do not demonstrate an establishment of these features but the additive developments contributing to their origin and evolution.

Using the same comparative genomics approach, the evolutionary origins and conservation of phytohormone pathways in plants were inferred (Appendix 2.3). The fundamental backbone of the biosynthesis and signalling pathways of all phytohormones either predates or accompanies the land plant transition (Ju *et al.*, 2015b; Bowman *et al.*,

2017; de Vries *et al.*, 2018b; Nishiyama *et al.*, 2018). Genes involved in gibberellic acid production and signalling originate with plant terrestrialisation (Figure 2.5). However, the role of hormones may have changed during land plant evolution, as recently highlighted for ABA signalling (McAdam *et al.*, 2016). Important innovations in land plants include tightly controlled responses to drought and salt stresses, which require the production and perception of ABA. The results show that ABA biosynthesis and perception evolved earlier than previously thought and is highly conserved across the plant phylogeny (Figure 2.5). The ABA receptor, PYL, has recently been identified in *Zygnema circumcarinatum* but is absent in other streptophyte algae (de Vries *et al.*, 2018b). In combination with the analysis presented here, this confirms that PYLs are conserved across Zygnematophyceae and Embryophyta. PP2Cs and SnRK2s, known to be present across Viridiplantae, are here supported as an Archaeplastida novelty (Bowman *et al.*, 2017). Identifying these HGs is a significant step in understanding the evolution of phytohormones and their implications for plant diversification.

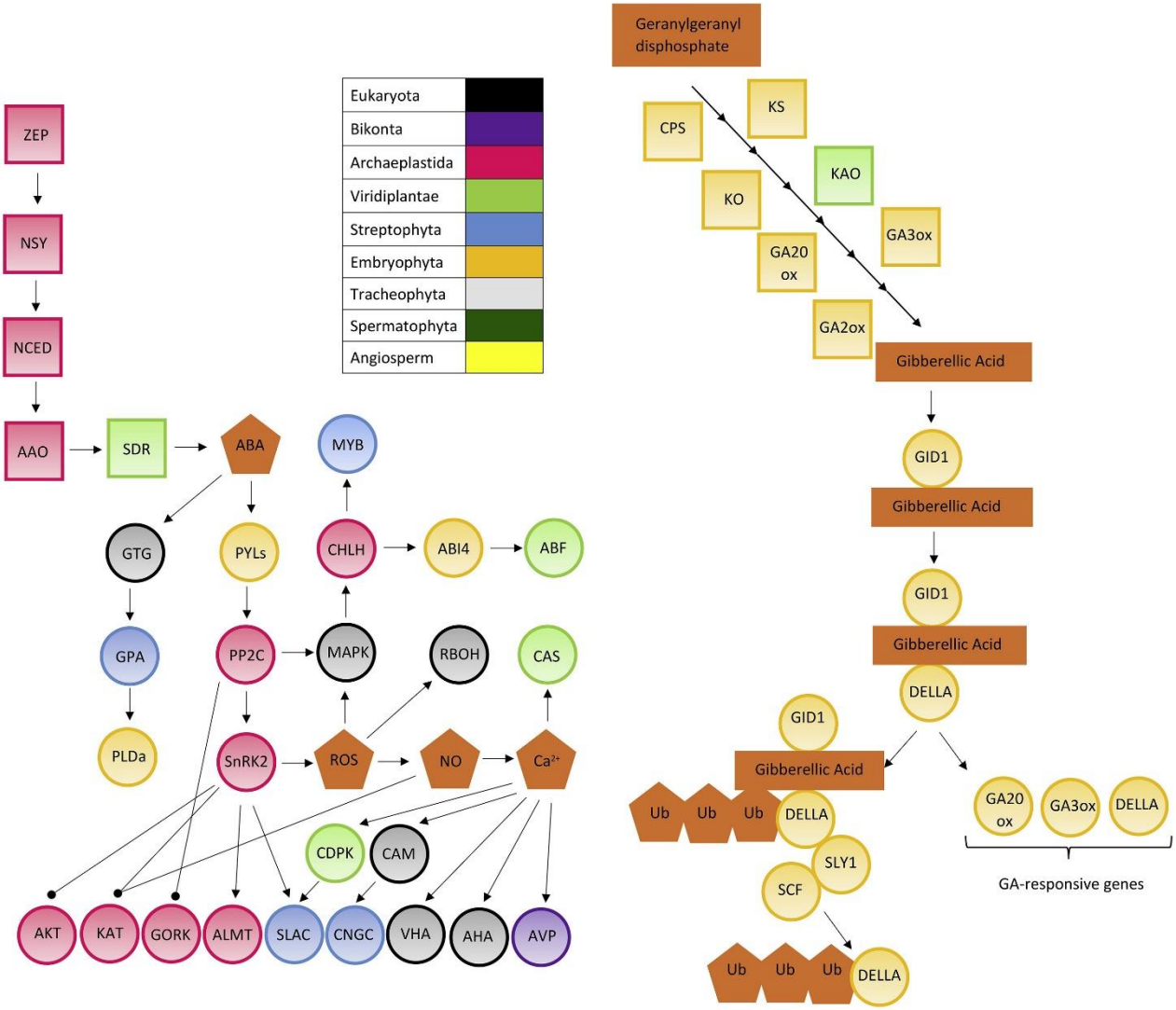


Figure 2.5 Evolution of abscisic acid (ABA) and gibberellic aicd (GA) biosynthesis and signalling. Squares indicate genes that are involved in biosynthesis whilst circles indicate genes involved in signalling. Dark orange shapes indicate non-genetic elements. Colour coding demonstrates that a gene was present in at least the last common ancestor of a clade. Arrows indicate positive regulation and circle ended lines indicate negative regulation. Acronyms for genes: ABA biosynthesis: *AAO, ABA-ALDEHYDE OXIDASE; NCED, 9-CIS-EPOXYCAROTENOID DIOXYGENASE; NSY, NEOXANTHIN SYNTHASE; SDR, SHORT-CHAIN ALCOHOL DEHYDROGENASE/REDUCTASE; ZEP, ZEAXANTHIN EPOXIDASE*. ABA signalling: *ABF, ABA RESPONSIVE ELEMENT-BINDING FACTOR; ABI4, ABA INSENSTIVE4; AHA, ARABIDOPSIS PLASMA MEMBRANE H+-ATPASE; AKT, SER/THR KINASE1; ALMT, ALUMINUM-ACTIVATED MALATE TRANSPORTER; AVP, ARABIDOPSIS VACUOLAR H+-PYROPHOSPHATASE; CAS, CALCIUM SENSING RECEPTOR; CHLH, PROTOPORPHYRIN IX MAGNESIUM CHELATASE, SUBUNIT H; CNGC, CYCLIC*

NUCLEOTIDE GATED CHANNEL; GORK, GATED OUTWARDLY RECTIFYING K+ CHANNEL; KAT, GUARD CELL INWARDLY RECTIFYING K+ CHANNEL; MAPK, MITOGEN ACTIVATED KINASE-LIKE PROTEIN; MYB, MYB DOMAIN PROTEIN; PLDa1, PHOSPHLIPASE Dα1; PP2C, PROTEIN PHOSPHATASE 2C; RBOH, RESPIRATORY BURST OXIDASE HOMOLOG PROTEIN; SLAC, SLOW ANION CHANNEL; VHA, VACUOLAR H+-ATPASE. GA biosynthesis: *CPS, ENT-COPALYL DIPHOSPHATE SYNTHASE; KS, KAURENE SYNTHASE; KO, ENT-KAURENE OXIDASE; KAO, ENT-KAURENOIC ACID OXIDASE; GA20ox, GIBBERELLIN 20 OXIDASE 1; GA3ox, GIBBERELLIN 3-BETA-DIOXYGENASE; GA2ox, GIBBERELLIN 2-BETA-DIOXYGENASE.* GA Signalling: *GID1, GIBBERELLIN-INSENSITIVE DWARD PROTEIN 1, DELLA; SLY1, SLEEPY1; SCF, SKP1-CULLIN-F-BOX.* This figure 2.has been adapted from previous publications for ABA (Yamauchi *et al.*, 2004; Cai *et al.*, 2017) and GA (Middleton *et al.*, 2012; Freschi, 2013). See also Appendix 2.4.

2.4.5 Other evolutionarily distinct gene groups of ancestral plant genomes

Genomic novelty is considered to have played an important role in the establishment of new features during the origins of land plants and other taxa. Genomic novelty in the LCA of distinct plant groups was substantial (Figure 2.2). In the LCAs of Streptophyta and Embryophyta, 753 and 1167 Novel HGs were identified respectively, similar to values found in other studies (Supplementary Data 2.4) (Li *et al.*, 2018a; Nishiyama *et al.*, 2018). In contrast to other plant nodes, these values are relatively low compared to the 2525 HGs identified in the origin of Mesangiospermae. As mentioned, WGD in plants is common and multiple events have been identified across the angiosperm phylogeny (Clark *et al.*, 2018). Two WGD events have been established in the ancestors of seed plants (Spermatophyta) and flowering plants (Angiospermae) which could explain the 1432 and 713 Novel HGs identified in these nodes (Jiao *et al.*, 2011; Ruprecht *et al.*, 2017).

The analyses also identify that the LCA of extant land plants (Embryophyta) contained at least 8654 Ancestral HGs (Supplementary Data 2.4). This number is likely lower than the total number of gene families present in the ancestral Embryophyta gene content because a HG can contain multiple genes, and HGs and genes can be lost from all extant

representatives. *Arabidopsis thaliana* and *Brachypodium distachyon* genomes contain 27,655 and 34,310 genes clustered into 13,345 and 14,235 HGs respectively, with 60-70% of their genes present in the LCA of land plants. 2254 of these ancestral HGs were retained (Ancestral Core) by at least 157 of the embryophyte genomes, demonstrating extensive gene loss has occurred across land plant evolution (Supplementary Data 2.4). GO analysis revealed genes derived from HGs present in the LCA of embryophytes are abundant in gene regulation (e.g. nucleic acid binding, transcription factors) and protein modification (e.g. hydrolase, transferase; Supplementary Data 2.4).

Furthermore, the analyses recognise HG losses (Supplementary Data 2.4). *Drosophila melanogaster* was used as a representative of a well-annotated non-plant genome in the GO analyses of HGs lost in plant evolution. A total of 1756 HGs were absent in the LCA of Streptophyta comprising protein classes involved in gene regulation (e.g. nucleic acid binding, transcription factor), cell signalling (e.g. enzyme modulator, signalling molecules) and catalytic activity (e.g. hydrolase, oxidoreductase). Lost HGs were also identified in Embryophyta suggesting that gene turnover was prolific during the evolution of the ancestors of streptophytes and land plants (Figure 2.2). Large losses were also identified in branches leading to the LCA of eudicots and Archaeplastida with 1196 and 1741 HGs respectively.

2.4.6 Comparisons with animal evolution

A previous study using the same comparative approach used in this study revealed an increase of genomic novelty during the origin of the Animal Kingdom, with an increase of conserved genomic novelty (Novel Core HGs) in a single node: the LCA of metazoans which comprises 25 Novel Core HGs associated with multicellular processes; this represents a 5-fold increase from previous ancestors (Paps *et al.*, 2018). The origin of land plants shows two nodes with an increase of conserved genomic novelty, one in the LCA of streptophytes (in the Ediacaran, 629 mya, Morris *et al.*, 2018) and another the LCA of land plants (Ordovician, 473 mya, Morris *et al.*, 2018). Moreover, plants show higher

numbers of conserved gene novelties than animals, representing a 10-fold increase compared to older ancestors (e.g. Novel Core HGs originating in the respective ancestors of Viridiplantae and Archaeplastida). In green plants, multicellularity has multiple independent evolutionary origins with chlorophycean and charophycean algae showing a patchy distribution, but is a trait that is conserved in all embryophytes (Umen, 2014; De Clerck *et al.*, 2018). Here, gene content data of the ancestral genomes of the Plant Kingdom (Viridiplantae) supports a decoupling between the emergence of multicellularity (streptophytes) and terrestralisation (embryophytes), which is in contrast to a single burst of novelty in the Animal Kingdom (Metazoa), whose origins did not involve a change of environment. In the future, the inclusion of new genomes may change the reconstruction of HGs at each node. Specifically recent sequencing of the first two fern genomes and a second charophyte genome would help to fill phylogenetic gaps (Li *et al.*, 2018a; Nishiyama *et al.*, 2018). Results from BLAST searches of Novel Core HGs against these phylogenetically important genomes supported the pipeline outputs, further validating the analyses (Supplementary Data 8, Appendix 2.5). In addition, this study solely focusses on protein-coding genes, however, non-coding genes, regulatory regions and epigenetic modifications most likely contributed to the diversification of plant life. The analysis presented here, which incorporates genomic data for 208 taxa from across the tree of life, provides new insight into the composition of ancestral plant genomes and emphasises the role of genome evolution in the emergence of terrestrial flora.

2.5 Methods

2.5.1 Materials Availability

Genome sources and software (e.g. BLAST) are listed (Supplementary Data 2.1) and referenced (Appendix 2.1) and all scripts used are available on Github listed below. This study did not generate any new, or unique reagents.

2.5.2 Compiling genomic dataset

A detailed description of the pipeline utilised here can be found elsewhere (Paps *et al.*, 2018). Briefly, the pipeline uses the protein coding genes of whole genome sequences to identify homologous groups of proteins within and between species (Appendix 2.1). Broad taxonomic sampling of genomic data was implemented to be able to accurately infer the phylogenetic origin of different HGs (Supplementary Data 2.1). 208 eukaryotic genomes were downloaded equating to 9,204,593 predicted proteins including 178 Archaeplastida genomes (including 158 land plant genomes) and 30 from a diverse representation of eukaryotic outgroups (Supplementary Data 2.1, Supplementary Figure 2.1). BUSCO analysis was used to assess the quality of the genome annotation, using a <15% of missing genes in the BUSCO Eukaryota dataset as a benchmark to accept a genome for further analysis (Supplementary Data 2.1, Supplementary Figure 2.2) (Simão *et al.*, 2015).

2.5.3 Homology assignment

Sequence similarity for all predicted proteins was identified with an all-versus-all BLAST (Altschul *et al.*, 1990) (version 2.7.1) using an e-value of 10^{e-5}, resulting in 84,724,532,295,649 comparisons with 3,680,714,880 significant BLAST hits. The BLAST search was launched on 7th February 2018 and therefore any genomes published after this date were not included in the analysis. Within the MCL protocols, it is recommended to assess the effects of changing of the granularity score which is the fineness of the clusters produced (Enright *et al.*, 2002). Outputs for granularity scores 1.2, 2, 4 and 6 were used to compare the phylogenetic appearance and clustering of plant gene families against published datasets of Banks et al (2011) and the transcription factor families from Catarino et al (2016) (Supplementary Data 2.3). After testing the impact of altering this inflation value, BLAST outputs were clustered using MCL with the default granularity score (I = 2.0, Supplementary Data 2.3) (Enright *et al.*, 2002). This approach identified 661,545 groups of homologous genes across all proteins.

2.5.4 Phylogenetically Aware Parsing Script

The MCL output was processed by modifying the Perl scripts described (Paps *et al.*, 2018) with Perl version 5. In the form of three Perl scripts, the pipeline can be used to identify the origin or loss of homologous groups of proteins (HGs) based upon their taxonomic occupancy (Supplementary Data 2.2). Different sets of HGs can be analysed (initially defined (Paps *et al.*, 2018));

- Ancestral (HGs present in the Last Common Ancestor of a clade),
- Ancestral Core (HGs present in every representative species within a clade or absent only in one genome),
- Novel (HGs present in the Last Common Ancestor of a clade and absent in all outgroup taxa),
- Novel Core (HGs present in every representative species within a clade or absent only once and absent in all outgroup taxa),
- Lost (HGs lost in the Last Common Ancestor of a clade).

A more detailed explanation of these query terms with examples is available (Appendix 2.2, Supplementary Figure 2.3). The main tree figures were made in FigTree (Rambaut A, 2012) and edited in Inkscape (The Inkscape Project, 2019).

2.5.5 Novel Core HG validation

To confirm accurate identification of conserved gene novelties, *Arabidopsis thaliana* (and *Brachypodium distachyon* for Liliopsida novelties) genes for each HG were tested, by performing BLASTP searches against the Swissprot database (Bairoch, 2000) (25th July 2018) excluding in-group sequences with the option negative_gilist (Altschul *et al.*, 1990) (Supplementary Data 2.5). This offers the maximum breadth of taxonomic sampling possible. Based on sequence similarity, e-value and taxonomic occupancy, BLAST searches further validated the identification of novel core Homology Groups.

Three evolutionarily significant genomes have recently been published, the first two fern genomes (Li *et al.*, 2018a) and the second charophyte genome (Nishiyama *et al.*, 2018). Novel Core HGs from all groups were BLASTP searched against the protein coding genes of these genomes (Supplementary Data 2.4). Based on sequence similarity, e-value and taxonomic occupancy, these BLAST searches refined the number of Novel Core HGs identified (Appendix 2.5).

2.5.6 Functional annotation

To obtain a functional description for all types of HG for every Archaeplastida node, their *Arabidopsis thaliana* genes were assessed using Panther GO (Mi *et al.*, 2017) (Version 11). The number of Gene Ontology hits for all GO classifications were collated: Protein Class, Molecular Function, Biological Process, Cellular Component, Pathways (Supplementary Data 2.4). A literature search further revealed the functions of the Novel Core Homology Groups (Supplementary Data 2.7). Graphics were produced in R (R Core Team, 2014) using packages tidyr (Henry, 2018) and GGplot2 (Wickham, 2016).

2.5.7 Inferring Horizontal Gene Transfer

Inferences about potential HGT were made. Based on the taxon sampling in the dataset, the pipeline was used to produce the query: Atleast1-fungi present, Atleast1-Embryophyta present and Outgroups absent. 323 HGs were identified which were subsequently whittled down to 25 HGs by stipulating that at least 100 land plant taxa must be present. Similar to the above, GO analysis was used to reveal the functions of these HGs (Supplementary Data 2.6).

2.5.8 Data and Code availability

All genomic data used in the study is publically available with sources listed in Supplementary Data 2.1. The code used to process the outputs of MCL and extract the 5 evolutionarily distinct Homology Groups is available on Github at

https://github.com/AlexanderBowles/Plant-Evomics and in Supplementary Data 2.9 along with the outputs of MCL.

2.6 Acknowledgements

The author would like to thank Prof Phillip Mullineaux, Prof Leonard Schalkwyk, Prof Peter W.H. Holland, Prof Phillip Donoghue, Dr Nacho Maeso, Dr Ferdinand Marlétaz and Dr Sarah F. Worsley for their comments on the manuscript. I also would like to thank Stuart Newman for his support of the Genomics HPC server. A.M.C.B., U.B., and J.P. received funding from the School of Life Sciences (University of Essex).

Chapter 3 Different genome evolution modes underline the evolution of water relations in land plants

Bowles AMC, Bechtold U, Paps J. 2020. Different genome evolution modes underline the evolution of water relations in land plants. *In review in PNAS.*

Authors

Alexander M. C. Bowles[1], Jordi Paps[123], Ulrike Bechtold[1]

Affiliations

1. School of Biological Sciences, University of Essex, Colchester CO4 3SQ, UK

2. School of Biological Sciences, University of Bristol, Bristol, BS8 1TQ, UK

3. Department of Zoology, University of Oxford, Oxford, OX1 3SZ, UK

Author Contributions

A.M.C.B, J.P and U.B designed the study and analyses. A.M.C.B performed the analyses.

A.M.C.B, J.P and U.B wrote the manuscript.

3.1 Abstract

The relationships between plants and water have changed dramatically over the last billion years. The first land plants emerged from aquatic environments around 500 million years ago and, since then, their descendants have adapted to variable water availability through the evolution of key innovations. Although the origins of these adaptations have been characterised, the evolution of the genetic toolkit that underpins these adaptations, in terms of development and function, is less well understood. Here, by comparing 208 genomes, the evolutionary origin and diversification of genes involved in the development and regulation of stomata, vascular tissue and roots is investigated. This approach reveals that novel genes led to a single origin of stomata in the ancestor of land plants. However, stomatal control, which enables active regulation of water exchange, is the product of gene duplications in the ancestor of seed plants. Gene networks involved in vascular tissue development have emerged through a complex of evolutionary mechanisms. Root evolution has also been shaped by the emergence of key novel genes. The findings highlight the role of water as a driver of plant evolution and provides insights into the molecular mechanisms enabling plants to conquer land.

3.2 Introduction

Water is essential for life on Earth although, in the case of plants their relationship with water has changed dramatically over the last billion years. Within green plants, the divergence of chlorophyte algae, which are almost exclusively aquatic, and Streptophyta which include algae and land plants, occurred approximately 1 billion years ago (Morris *et al.*, 2018). Streptophyte algae are found in a range of brackish, freshwater and terrestrial habitats, which demonstrates their ability to adapt to a range of different niches that vary in terms of their water availability (Fürst-Jansen *et al.*, 2020). Emerging from aquatic environments 500 million years ago, the first land plants and their descendants have had to adapt to variable water availability in order to survive and conquer new terrestrial environments (Morris *et al.*, 2018). Genome analysis has identified that terrestrialisation

was preceded by major innovations previously thought to be land plant specific, for example the associations with substrate microbiota (Hori *et al.*, 2014; de Vries *et al.*, 2018a; Nishiyama *et al.*, 2018; Cheng *et al.*, 2019; Wang *et al.*, 2019). However, although many key genes evolved prior to the transition of plants onto land, specific adaptations and the genetic re-wiring of developmental and stress response pathways occurred later in land plant evolution, increasing their adaptive plasticity to water availability (Harrison, 2017; de Vries *et al.*, 2018a; Fürst-Jansen *et al.*, 2020).

Plant features that are important for water regulation and transport appear to have evolved in a stepwise manner. For example, three of the most important features for water regulation, and the focus of this study, are stomata, vascular tissue, and roots which evolved approximately 500 mya, 450 mya and 435 mya, respectively (Morris *et al.*, 2018). Stomata, vascular tissue, primary roots and lateral roots have emerged sequentially in the ancestors of Embryophyta, Tracheophyta, Euphyllophyta and Spermatophyta respectively (Harrison, 2017). Stomata, which are a key adaptation to dry environments and emerged in the ancestor of land plants, are pores in plant tissue that open to allow gas exchange and close to regulate water loss (Sussmilch *et al.*, 2019). Vascular tissue then enables efficient water transport throughout the plant, and is common to all tracheophytes (also known as vascular plants), although vascular like systems have also been identified outside of tracheophytes (Brodribb *et al.*, 2020). Roots, which emerged in the ancestor of euphyllophytes, provide multiple functions including anchorage, nutrient and water uptake (Kenrick *et al.*, 2014).

3.2.1 The genetic toolkit for root development

Briefly described below is our current understanding of the genetic pathways and mechanisms that are involved in the development of root hairs, primary roots, lateral roots and vascular tissue, as well as the development and functioning of stomata. They are described in the order in which water enters and leaves a plant, namely roots, followed by vascular tissues and then by stomata.

3.2.1.1 Root hairs

Root hair development begins with cell fate determination which determines whether an epidermal cell becomes a root hair cell or non-root hair cell. The determination of a root hair cell, triggers a transcription factor cascade, inhibiting GLABRA2 (GL2) expression. This inhibition leads to the expression of the ROOT HAIR DEFECTIVE 6/ ROOT HAIR DEFECTIVE 6-LIKE 1 (RHD6/RSL1) gene, which initiates the development of the root hair (Vissenberg *et al.*, 2020). This genetic toolkit is heavily influenced by signalling from phytohormones; these signals enable plants to control root hair development and function in relation to environmental factors, such as drought stress (Bobrownyzky, 2016).

3.2.1.2 Primary roots

The primary root is the foundation of dicot rooting systems and the initiator of monocot rooting systems. Primary root initiation involves the formation and maintenance of the stem cell predominantly through auxin (e.g. AUX1/LAX) and cytokinin gene signalling (Moubayidin *et al.*, 2009). The abundance of auxin controls the expression of indole-acetic acid (IAA) genes and auxin response factor genes (ARFs) (De Smet *et al.*, 2010). Under high levels of auxin, ARF proteins are released from IAAs and then are able to activate many root development genes (Goh *et al.*, 2012).

Additionally, important for the specification of the stem cell are the transcription factors encoded by the genes SHORTROOT (SHR) and SCARECROW (SCR). These genes are crucial for controlling primary root initiation, root patterning and regulation of cell division (Lucas *et al.*, 2011). The DELLA transcription factors, GIBBERELLIN INSENSITIVE (GAI), REPRESSOR OF GA1 (RGA) and RGA-LIKE (RGLs) are, in turn, important for cell division and root elongation (Ubeda-Tomás *et al.*, 2008; Fonouni-Farde *et al.*, 2019).

These steps involved in root initiation, development and elongation are essential for root system development as a whole. If these initial stages of root development are affected by environmental factors, then downstream development can be impacted, resulting in

irregularities in gravitropic growth, root elongation and lateral root emergence (Jung *et al.*, 2013).

3.2.1.3 Lateral roots

Lateral roots emerge from the primary root, enabling plants to access nutrients and water from a greater soil area (Nibau *et al.*, 2008). Similar to primary root development, auxin signalling and the IAA/ARF complex is also important for lateral root development (e.g. IAA8/ARF7) (Verstraeten *et al.*, 2014; Santos Teixeira *et al.*, 2019). Auxin concentrations define the cell fate of the lateral root founder cells, initiating the lateral root primordium (Van Norman *et al.*, 2013). Asymmetric division leads to the emergence of the lateral root primordium from the primary root (Benkova *et al.*, 2010). Auxin also tightly controls the patterning of the lateral root primordium and facilitates its emergence (Swarup *et al.*, 2008; Péret *et al.*, 2012).

Lateral root growth is regulated by a complex of molecular factors and regulators (Nibau *et al.*, 2008). For example, GATA23 is involved in the specification of the lateral root founder cell (Lavenus *et al.*, 2013) whilst Lateral Organ Boundaries-domain (LBD) proteins actively regulate lateral root formation (Okushima *et al.*, 2007). Similar to primary root development, the transcription factors SHORTROOT (SHR) and SCARECROW (SCR) are important for lateral root development, patterning and emergence (Lucas *et al.*, 2011).

Drought is known to inhibit the development and growth of lateral roots which is predominantly mediated through abscisic acid (ABA) (Shkolnik-Inbar *et al.*, 2010; Xu *et al.*, 2013b). Typically, drought stress can lead to shorter primary roots and a reduced number of lateral roots (Zolla *et al.*, 2010).

3.2.2 The genetic toolkit for vascular tissue development

The evolution of the vascular system enabled plants to efficiently transport water from root to shoot, approximately 450 million years ago (Morris *et al.*, 2018). The vascular system

consists of important tissue types including the xylem and phloem which are connected by the undifferentiated cambium. Procambial and cambial cells function as vascular stem cells which differentiate into specialised cells of xylem and phloem (Vaughan-Hirsch *et al.*, 2018). The development of vascular tissue involves a set of highly coordinated consecutive processes which are coordinated by the expression of particular genes and transcription factors which are detailed below.

Auxin signalling plays a crucial role in asymmetric cell division, leading to the development of provascular cells. The auxin response transcription factor MONOPTEROS (MP) is involved in the process of asymmetric cell division, which eventually leads to xylem cell development. Target genes of MP, the transcription factors TARGET OF MONOPTEROS 5 and LONESOME HIGHWAY (TMO5/LHW), are also involved in orientating the divisions of procambium cells which, in turn, establishes the different vascular cell types (Smet *et al.*, 2019). The maintenance of the vascular stem (cambial) cells is regulated by PHLOEM INTERCALATED WITH XYLEM (PXY). VASCULAR-RELATED NAC-DOMAIN 1-7 (VND1-7) genes specify the cell fate of xylem. Other NAC domain proteins, notably SECONDARY WALL - ASSOCIATED NAC DOMAIN PROTEIN1 (SND1) and NAC SECONDARY WALL THICKENING PROMOTING FACTOR 1 – 2 (NST1 - 2), regulate the differentiation of xylem (Ruonala *et al.*, 2017).

Vascular tissue emerges through this series of developmental stages which are regulated by a number of transcription factors. However, although much is known about the development of xylem, many of the genes involved in the development of the vascular system still remain uncharacterised.

3.2.3 The genetic toolkit for stomatal development and function

3.2.3.1 Stomatal development

The stomatal development process involves a series of cell-fate transitions leading to the formation of guard cells. This process begins with a meristemoid mother cell (MMCs).

MMCs divide asymmetrically to produce a smaller meristemoid and a larger stomatal lineage ground cell (SLGC). This process is regulated by the bHLH transcription factors SPEECHLESS (SPCH) and ICE/SCREAM (SCRM) (Chater *et al.*, 2017). The SLGC either forms a pavement cell or divides further to become a satellite meristemoid. Meristemoids next differentiate into a guard mother cell (GMC) which is regulated by the bHLH transcription factors MUTE and SCRM (Macalister *et al.*, 2011; Lau *et al.*, 2012). Finally, GMCs divide symmetrically to form the guard cells of stomata. This process is coordinated by activity of the bHLH transcription factors FAMA and SCRM (Le *et al.*, 2014).

3.2.3.2 Stomatal signalling

Stomatal closure, which enables plants to actively control water loss, is predominantly mediated by the phytohormone abscisic acid (ABA), particularly in flowering plants (Brodribb *et al.*, 2017; Cai *et al.*, 2017; Sussmilch *et al.*, 2017a, 2019). ABA is detected by the receptor PYR/PYL/RCAR (PYLs) (Gonzalez-Guzman *et al.*, 2012). When ABA is perceived, PYLs bind to ABA which, in turn, increases the binding affinity of PYLs to PROTEIN PHOSPHATASE 2Cs (PP2Cs). Together, they form a PYL-ABA-PP2C complex which inhibits the activity of PP2Cs (Park *et al.*, 2009; Komatsu *et al.*, 2013). Under normal conditions (in the absence of ABA), PP2Cs are bound to SNF1-RELATED PROTEIN KINASE 2s/OPEN STOMATA 1 (SnRK2s/OST1), a family of protein kinases which are enzymes that modify other proteins (Lind *et al.*, 2015). The binding of PYLs and PP2Cs releases SNRK2s/OST1 which then activates downstream genes including SLOW ANION CHANNEL-ASSOCIATED 1 (SLAC1) (Geiger *et al.*, 2009), GUARD CELL OUTWARD RECTIFYING K(+) (GORK) (Hosy *et al.*, 2003), POTASSIUM CHANNEL IN ARABIDOPSIS THALIANA 1 (KAT1) (Sato *et al.*, 2009) and QUICK-ACTIVATING ANION CHANNEL 1 (QUAC1) (Imes *et al.*, 2013) anion channels as well as membrane water channels (aquaporins) such as PLASMA MEMBRANE INTRINSIC PROTEIN 2-1 (PIP2-1) (Grondin *et al.*, 2015).

3.2.4 Insights into the evolution of plants on land

As highlighted above, roots, vascular tissue and stomata play a key role in regulating plant water uptake and loss and the genetic toolkits underpinning these innovations are fairly well characterised. However, the evolution of the gene networks underpinning these innovations remains poorly understood. To address questions surrounding the origins of these key biological innovations in the context of the evolution of water regulation, a comparative genomics approach was used to investigate the evolution of the genetic toolkit that regulates the development and function of roots, vascular tissue and stomata (Figure 3.1).

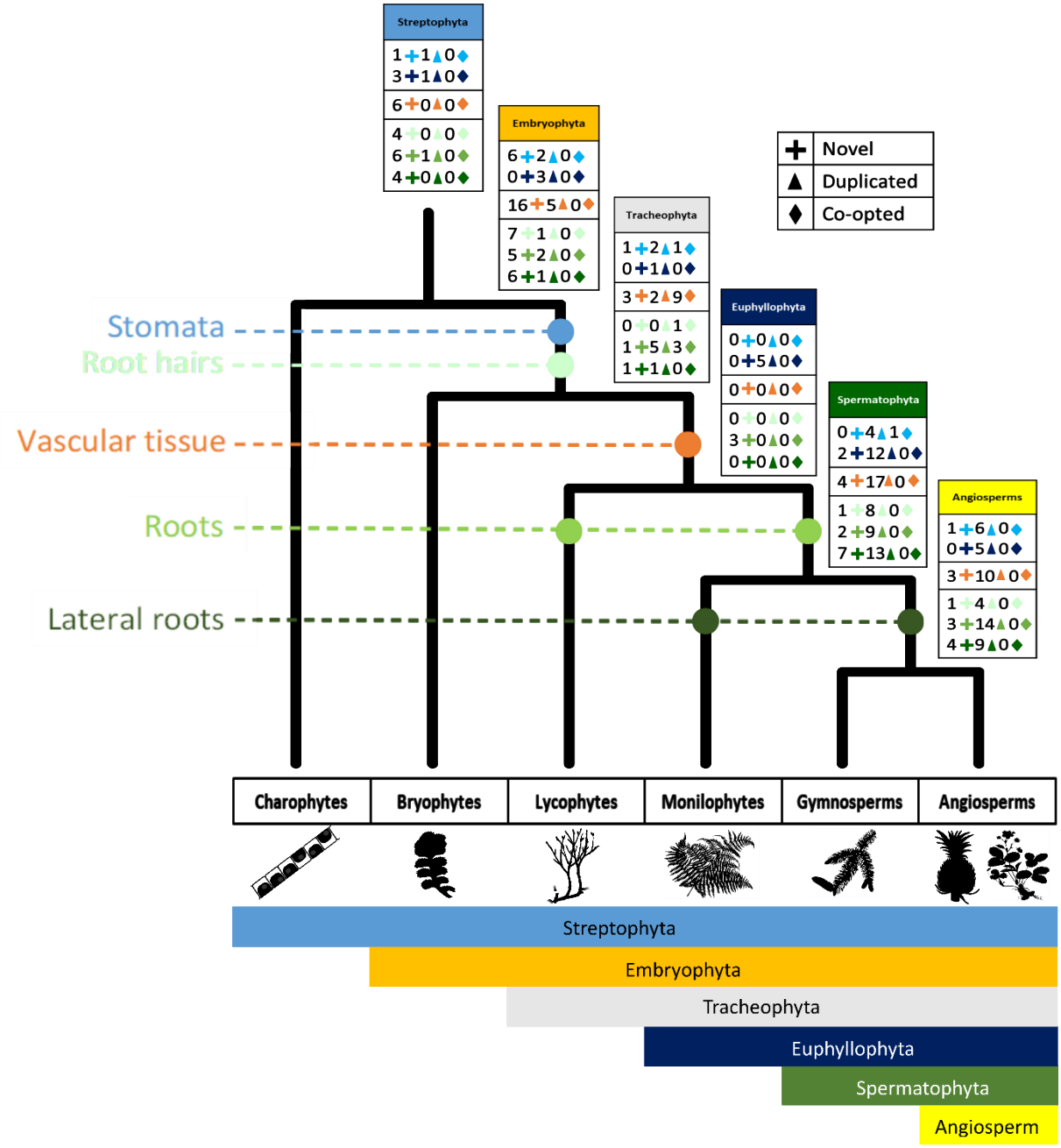


Figure 3.1. Plant-water relations have evolved in a stepwise manner. The tree demonstrates the evolutionary relationships of plants with silhouettes below (sourced from phylopic.org) illustrating species in each group. The dashed lines leading to each different water regulatory innovation denotes their origins. The colour is repeated within the boxes illustrating the genetic mechanisms associated with each innovation (Light blue: stomatal development, Dark blue: stomatal signalling, Orange: vascular tissue development, Light green: root hair, Green: primary root, Dark green: lateral root development). A key demonstrates the numbers of novel, duplicated and co-opted genes in the boxes.

3.3 Results and Discussion

3.3.1 Insights into root evolution

3.3.1.1 Novel genes in land plants enabled the development of root hairs

A bioinformatic approach was taken to identify homologous groups (HGs) of proteins of 208 genomes including 178 plant species and an outgroup of 30 eukaryotic species (Bowles *et al.*, 2020). HGs are identified as a set of proteins that have distinctly diverged from others. Genes important for root hair, primary root and lateral root development, vascular tissue development and stomatal development and signalling were identified in the literature (see Methods). The HGs containing these genes were extracted from the genomic dataset based on gene ID and taxonomic occupancy was analysed to identify novel and co-opted genes (Figure 3.1, Supplementary Data 3.1). A novel HG was defined as a set of genes present in the Last Common Ancestor (LCA) of a clade and absent in all outgroups. A co-opted HG was defined as a set of genes whose evolutionary emergence predates the function it is associated with (e.g. HG originating in the LCA of land plants linked to vascular tissue development). Co-opted HGs were identified as genes conserved across a clade of interest and present in sister group taxa. To investigate the diversification of genes, gene trees were inferred (Methods, Supplementary Data 3.2) and gene phylogenies were examined to identify the prevalence of gene duplications.

Roots have functions related to nutrient and water uptake as well as anchorage and symbiosis. Under water stress, plants adjust their root system architecture by stimulating deeper primary and lateral root growth (Uga *et al.*, 2013; Bao *et al.*, 2014; Orosa-Puente *et al.*, 2018; Ogura *et al.*, 2019; von Wangenheim *et al.*, 2020; Xiao *et al.*, 2020). Root hairs in vascular plants and rhizoids in bryophytes act as the interface between plants and soil, enabling water uptake and transport (Jones *et al.*, 2012; Kenrick *et al.*, 2014). The analyses show that most genes involved in root hair development emerge with or before the origin of land plants (Figure 3.2). CAPRICE (CPC), TRIPTYCHON (TRY) and

ENHANCER OF TRY AND CPC 1 (ETC1), are in a single HG and emerge in the ancestor of Spermatophyta, to promote root hair cell differentiation in flowering plants by repressing GLABRA 2 and 3 (Tominaga *et al.*, 2008). This suggests that non-spermatophyte land plants develop root hairs without these genes and that greater control of root hair development evolved in the last common ancestor (LCA) of seed plants.

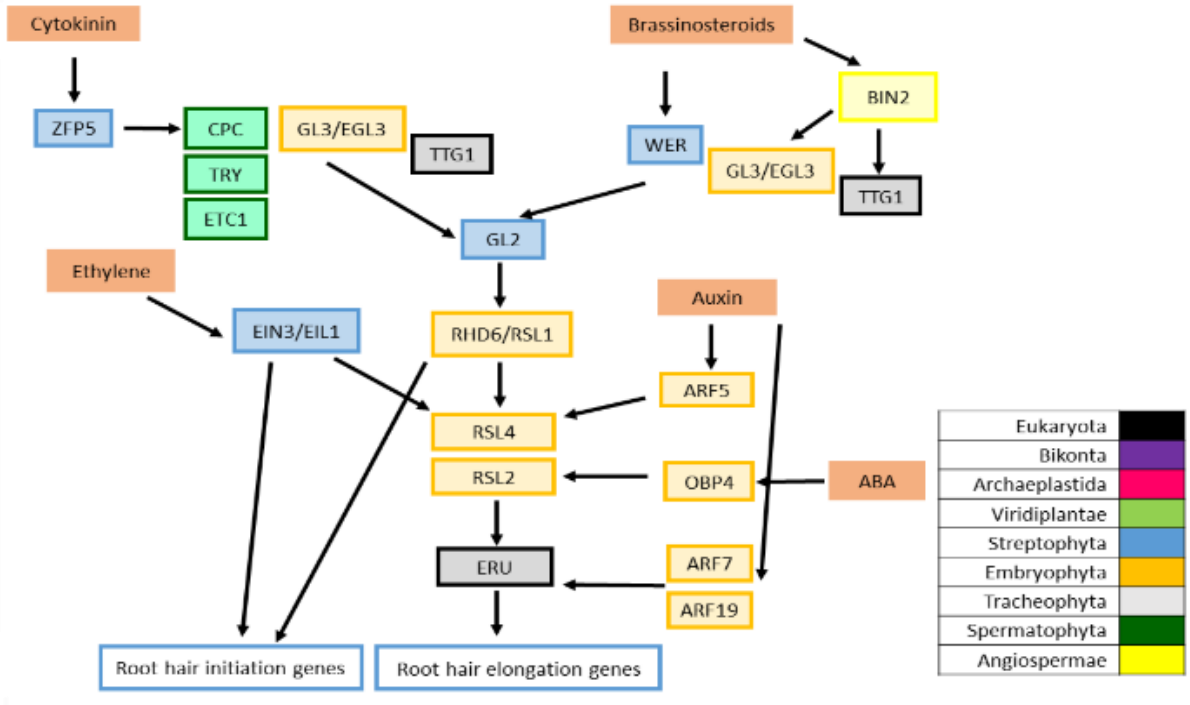


Figure 3.2. The genomic basis of the evolutionary development of root hairs. Genetic network leading to the development of root hairs. Each gene is coloured based on its phylogenetic appearance. Phytohormones are coloured in solid orange.

3.3.1.2 Novel genes in euphyllophytes enabled root growth towards water

Fossil evidence supports at least two origins of roots in the evolutionary history of plants, once in the ancestor of lycophytes and the other in the ancestor of euphyllophytes (Raven *et al.*, 2001; Hetherington *et al.*, 2018). Analysis of genes involved in primary root development revealed many HGs were conserved across land plants (Figure 3.3). The

search of HGs in two fern genomes (Li *et al.*, 2018a) shows that two HGs, ARABIDOPSIS RESPONSE REGULATOR 12 (ARR12) and LATERAL ROOT ORGAN DEFECTIVE (LATD), have emerged in the LCA of euphyllophytes (Supplementary Data 3.3). Both HGs have been shown to modulate primary and lateral root growth and development, with responses to ABA and water deprivation.

ARR12 regulates cell differentiation and meristem growth (Dello Ioio *et al.*, 2007; Yokoyama *et al.*, 2007; Moubayidin *et al.*, 2010). The dynamics between ARR12 and PLETHORA (PLT) control meristem expansion, particularly the rate of cell differentiation during early development (Xie *et al.*, 2018; Salvi *et al.*, 2020). *ARR12*, *ARR11* and *ARR1* triple mutants exhibit abscisic acid (ABA) hypersensitivity of primary root growth and increased drought tolerance (Huang *et al.*, 2018). Under drought, ARR12 is down-regulated as an adaptive mechanism to control root growth to cope with water deficit (Nguyen *et al.*, 2016). These genes therefore play crucial roles in plant growth and development but also in the response to changes in water availability. LATD is also required for root and nodule meristem development (Léran *et al.*, 2014). *LATD* mutants display defects in ABA responses and meristem organisation, arresting primary and lateral root growth (Liang *et al.*, 2007). LATD modulates reactive oxygen species (ROS) levels in the root and in concert with ABA, modulate primary root elongation (Zhang *et al.*, 2014a) thus providing a potential mechanism for controlling root growth under water stress conditions.

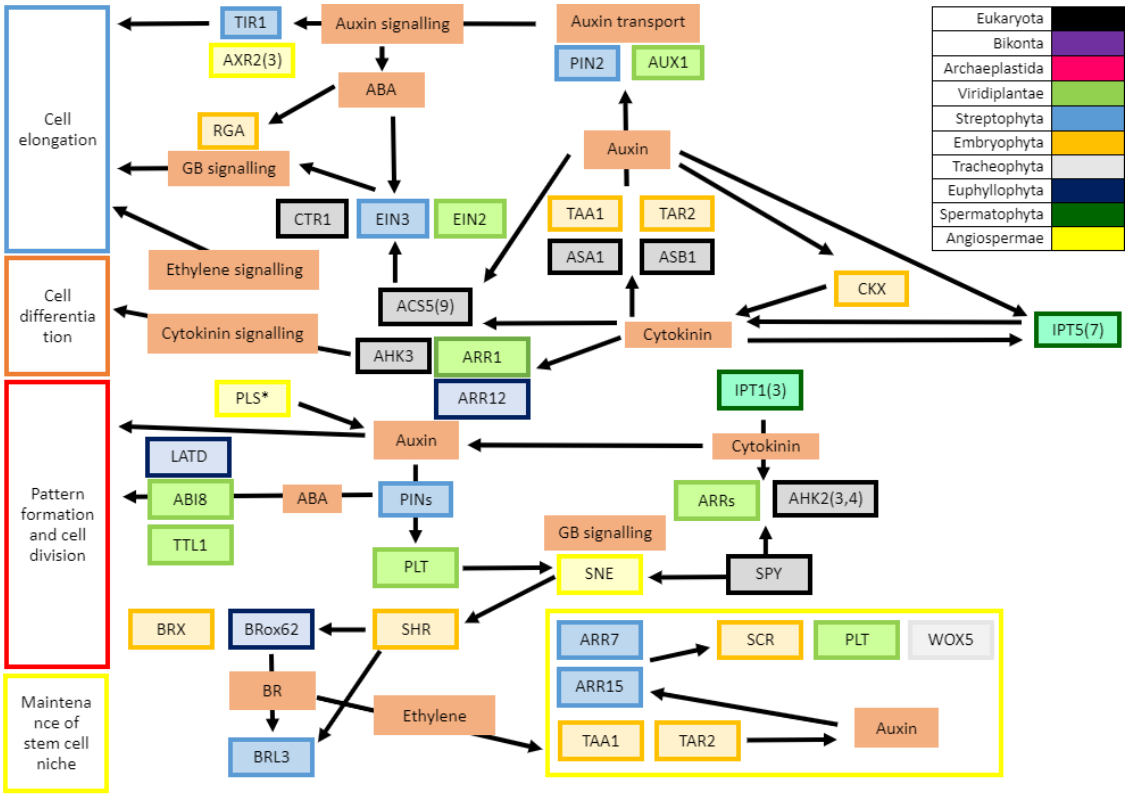


Figure 3.3. The genomic basis of the evolutionary development of roots. Genetic network leading to the development of roots. Each gene is coloured based on its phylogenetic appearance. Phytohormones are coloured in solid orange.

3.3.1.3 Novel and duplicated genes in seed plants enabled lateral root growth towards water

Lateral roots are found on most euphyllophytes (Liu *et al.*, 2018), but have evolved independently on multiple occasions (in ferns and in the LCA of seed plants) (Hetherington *et al.*, 2020) (Figure 3.1, Figure 3.4). The origin of lateral root branching mechanisms in seed plants would have enabled greater phenotypic plasticity to water availability, allowing the LCA of spermatophytes to adapt to more diverse environmental conditions (Motte *et al.*, 2019). The data presented here demonstrate that the majority of lateral root development genes predate the emergence of lateral roots, originating in the ancestors of Streptophyta (e.g. PIN 1, 2, 3) and Embryophyta (e.g. IAA 8, 14), which contribute to other

functions in these rootless plants (Mutte *et al.*, 2018; Vosolsobě *et al.*, 2020). Three key genes, which after confirmation of absence in the two fern genomes, appeared with the origin of lateral roots (INDOLEACETIC ACID-INDUCED PROTEIN 12 & 28 (IAA12, 28), INFLORESCENCE DEFICIENT IN ABSCISSION (IDA) and RAPID ALKALINIZATION FACTOR (RALF) (Supplementary Data 3.3)).

IAA12 and IAA28 are auxin-responsive proteins important in lateral root initiation and patterning (Stoeckle *et al.*, 2018). Specifically, IAA28, in conjunction with ARFs and GATA23, is involved in the spacing of the lateral root founder cell, optimising the distribution of new root organs (De Rybel *et al.*, 2010). IAA12, by suppression of ARF5, is involved in lateral root initiation, patterning and organogenesis, by activating the cell cycle to form lateral root primordia (Stoeckle *et al.*, 2018). IDA is required for cell wall dissolution, by facilitating the separation of epidermal tissues, enabling lateral root primordia emergence (Zhu *et al.*, 2019b). IDA is strongly induced by auxin, specifically through the module of IAA3 and ARF7 (Kumpf *et al.*, 2013). Lateral root branching to water is dependent on the regulation of this auxin module, which subsequently influences IDA (Orosa-Puente *et al.*, 2018). RALF, a signal peptide, regulates cell growth and expansion by interrupting brassinosteroid signalling (Bergonci *et al.*, 2014). RALF1 inhibits cell elongation for primary root growth, lateral root formation and lateral root density (Bergonci *et al.*, 2014; Murphy *et al.*, 2014). Under drought stress, RALF1 modulates root hair growth and cell size, and together with other RALFs is implicated in stress responses of lateral roots (Murphy *et al.*, 2014; Zhu *et al.*, 2020).

Hydrotropism is the directional growth of plant roots towards water and may help plants to efficiently obtain water under drought stress. However the precise signalling governing hydrotropism remains elusive (Dietrich *et al.*, 2017; Shkolnik *et al.*, 2018). Two genes, known to be involved in root development and hydrotropism, are MIZ1 and MIZ2/GNOM (Dietrich, 2018). Upon detection of a water gradient, the phytohormone, cytokinin, induces the expression of MIZ1 and negatively regulates auxin levels to control lateral root growth.

The data presented here infer that MIZ1 emerged in the ancestor of Embryophyta and diversified through a duplication event, also in the LCA of Embryophyta (Figure 3.4, Supplementary Data 3.2). Knockout mutants in *MIZ1* showed no major differences in root growth but display altered hydrotropism (Iwata *et al.*, 2013), suggesting that hydrotropism evolved in the ancestor of Embryophyta, and that root hydrotropism may have been crucial for the colonisation of land contributing to drought avoidance mechanisms (Supplementary Data 3.2) (Kobayashi *et al.*, 2007; Shkolnik *et al.*, 2018). The HG containing MIZ2/GNOM was present in the ancestor of Eukaryota with genes duplicating in the ancestor of seed plants, potentially responsible for hydrotropism of lateral roots (Figure 3.4, Supplementary Data 3.2).

WOX genes, a subgroup of eukaryotic homeobox transcription factors, have important roles in plant development, particularly stem-cell maintenance and organ formation (e.g. roots) (van der Graaff *et al.*, 2009). Due to their significance for plant development, all known *Arabidopsis* WOX genes were queried which were clustered into three HGs (Supplementary Data 3.1). Intermediate-clade WUSCHEL-RELATED HOMEOBOX (IC-WOX; WOX 8,9,11,12) and WUSCHEL-clade WOX (WC-WOX; WOX 1-7) genes are clustered into two distinct HGs, with IC-WOX originating in euphyllophytes (after confirmatory queries of the fern genomes, Supplementary Data 3.3) and WC-WOX as conserved across tracheophytes. WC-WOX genes also have roles in vascular tissue development suggesting this as the initial explanation for their emergence. The results presented here corroborate recent work to infer that the origin and divergence of IC-WOX genes in euphyllophytes promoted root initiation (Liu *et al.*, 2018; Yu *et al.*, 2020).

The development of increasingly morphologically complex rooting systems in the ancestors of Embryophyta, Euphyllophyta and Spermatophyta has enabled plants to access previously unavailable water sources and control growth in response to water availability. The data presented above demonstrates that a combination of novel and

duplicated genes is responsible for the emergence of plants with true roots with lateral branches.

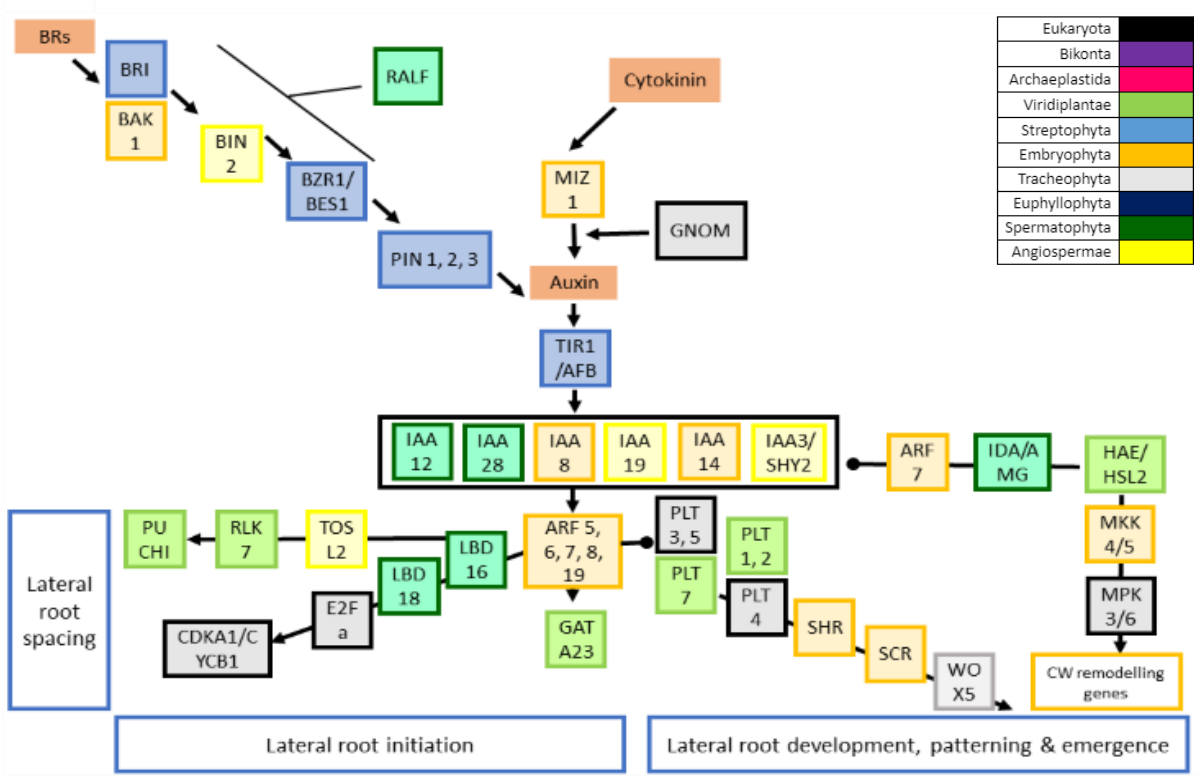


Figure 3.4. The genomic basis of the evolutionary development of lateral roots. Genetic network leading to the development of lateral roots. Each gene is coloured based on its phylogenetic appearance. Phytohormones are coloured in solid orange.

3.3.2 Insights into the evolution of vascular tissue

3.3.2.1 Vascular tissue evolved through a complex of genetic mechanisms

In tracheophytes, also known as vascular plants, transport tissues xylem and phloem enable water to be transported through the plant and evaporated through stomata (Lucas *et al.*, 2013). Lignified vascular tissue also provides mechanical support enabling plants to increase their body size and dominate terrestrial habitats. For example, the first trees in the fossil record are early vascular plants (Stein *et al.*, 2007).

Only 3 HGs involved in vascular system development originated in the ancestor of vascular plants (Figure 3.5, Supplementary Data 3.2). These included WC-WOX4, SUPPRESSOR OF ACAULIS (SACLs) and NAC45/NAC86-DEPENDENT EXONUCLEASE-DOMAIN PROTEIN (NEN) families which are crucial components in vascular development (Vera-Sirera *et al.*, 2015). Besides these novel genes, PHLOEM INTERCALATED WITH XYLEM (PXY) and TARGET OF MONOPTEROS 5 (TMO5) predate the origin of tracheophytes and are all involved in xylem differentiation and underwent duplications in the ancestor of tracheophytes (Figure 3.5, Supplementary Data 3.2). Finally, several genes involved in the vascular system emerged in the ancestors of land plants, showing a patchy distribution in non-tracheophytes but are present in all tracheophytes. This retention of genes in the LCA of Tracheophyta suggests a vital function in vascular plants. Nine HGs fitted this criterion of co-option with either losses in *Marchantia polymorpha* or *Physcomitrella patens* (Figure 3.1). These HGs were found in all elements of vascular development pathways. These findings support the importance of gene duplications as well as repurposing of old genes for novel functions in the evolutionary development of the vascular system. Together these genetic processes contributed to the origin and evolution of one of the most successful plant groups, approximately 450 million years ago (Morris *et al.*, 2018).

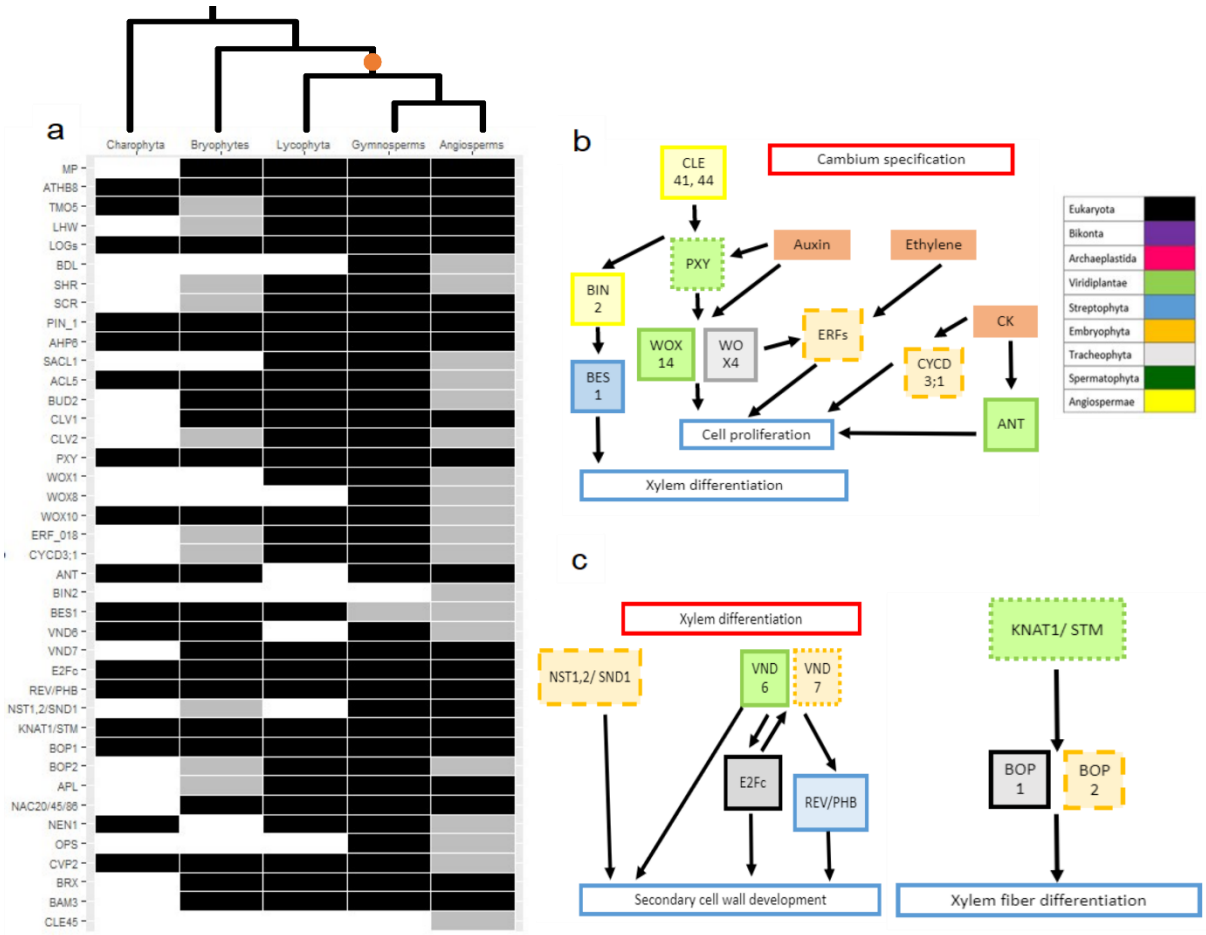


Figure 3.5. The genomic basis of the evolutionary development of vascular tissue. **A.** Heatmap displaying absence (white), partial presence (grey) and presence in all species (black) for genes involved in vascular tissue development. The tree at the top illustrates plant evolutionary relationships and the origin of vascular tissue (orange). **B.** Genetic network involved in cambium specification. Each gene, in figure 5B and 5C, is coloured based on its phylogenetic appearance. Phytohormones are coloured in solid orange. Duplicated genes are highlighted by boxes with dotted edges and co-opted genes are highlighted by boxes with dashed edges. **C.** Genetic network involved in xylem differentiation.

3.3.3 Insights into stomatal evolution

3.3.3.1 Stomata evolved once in the ancestor of land plants

Stomata in bryophytes demonstrate a patchy distribution with absences in all liverworts and some mosses, but are found across all tracheophytes (Duckett *et al.*, 2018). Functionally, stomata also differ between bryophytes and tracheophytes. In bryophytes, stomata are found only on sporangia and promote water loss for spore desiccation. In tracheophytes, stomata open to enable CO_2 uptake and close to prevent water loss (Sussmilch *et al.*, 2019; Harris *et al.*, 2020). This raised questions about the single origin of stomata in the LCA of land plants or convergent evolution in the ancestors of vascular plants, mosses and hornworts.

Of the 23 stomatal development genes, 21 predated or accompanied the origin of land plants (Figure 3.6, Supplementary Data 3.1). Using a genomic dataset for 178 plant genomes with comprehensive outgroup sampling (30 genomes), the stomatal development pathway was inferred to have originated in the LCA of land plants, reinforcing studies using transcriptome and other sequence data (Chater *et al.*, 2017; Harris *et al.*, 2020). Stomatal formation involves a sequence of cell-fate transitions, from a meristemoid mother cell, to a meristemoid, to a guard mother cell, and eventually to the guard cell. In *Arabidopsis thaliana*, the basic helix-loop-helix (bHLH) genes SPCH, MUTE and FAMA are required consecutively to determine stomatal development (Figure 3.6). Additionally, the bHLH transcription factor SCREAM interacts with SPCH, MUTE and FAMA (Lau *et al.*, 2012). In the analysis, SCREAM was identified in all land plants, even *Marchantia polymorpha*, which does not have stomata. MUTE, SPCH and FAMA were identified to have originated in the LCA of Embryophyta and subsequently lost in *Marchantia polymorpha*, potentially accompanying the loss of stomata in liverworts. This suggests that bryophyte stomata and by extension the stomata of the first land plants develop in a similar manner to vascular plant stomata. Finally STOMAGEN, known to positively regulate

stomatal density (Sugano *et al.*, 2010), appears in the origin of vascular plants, the first time stomata appear on leaf-like structures (Figure 3.6).

Of the three features investigated in this study, stomata were the only innovation with a strong association with the appearance of novel genes. In previous work, it was shown that the origin of the embryophytes was accompanied by a high number of gene novelties (Bowles *et al.*, 2020). These HGs were found in all, or all bar one, land plant species in the genomic dataset. The data show that SCREAM is characterised as a novel gene family present in all Embryophyta. However, FAMA, SPEECHLESS and MUTE are lost in species of Alismatales, an order of monocots which are mostly aquatic, corroborating findings of gene loss from the genomes of *Zostera marina* (Olsen *et al.*, 2016) and *Z. muelleri* (Lee *et al.*, 2016a). Equally STOMAGEN involved in the regulation of stomatal density was lost in Alismatales species with the transition into aquatic environments (Figure 3.6).

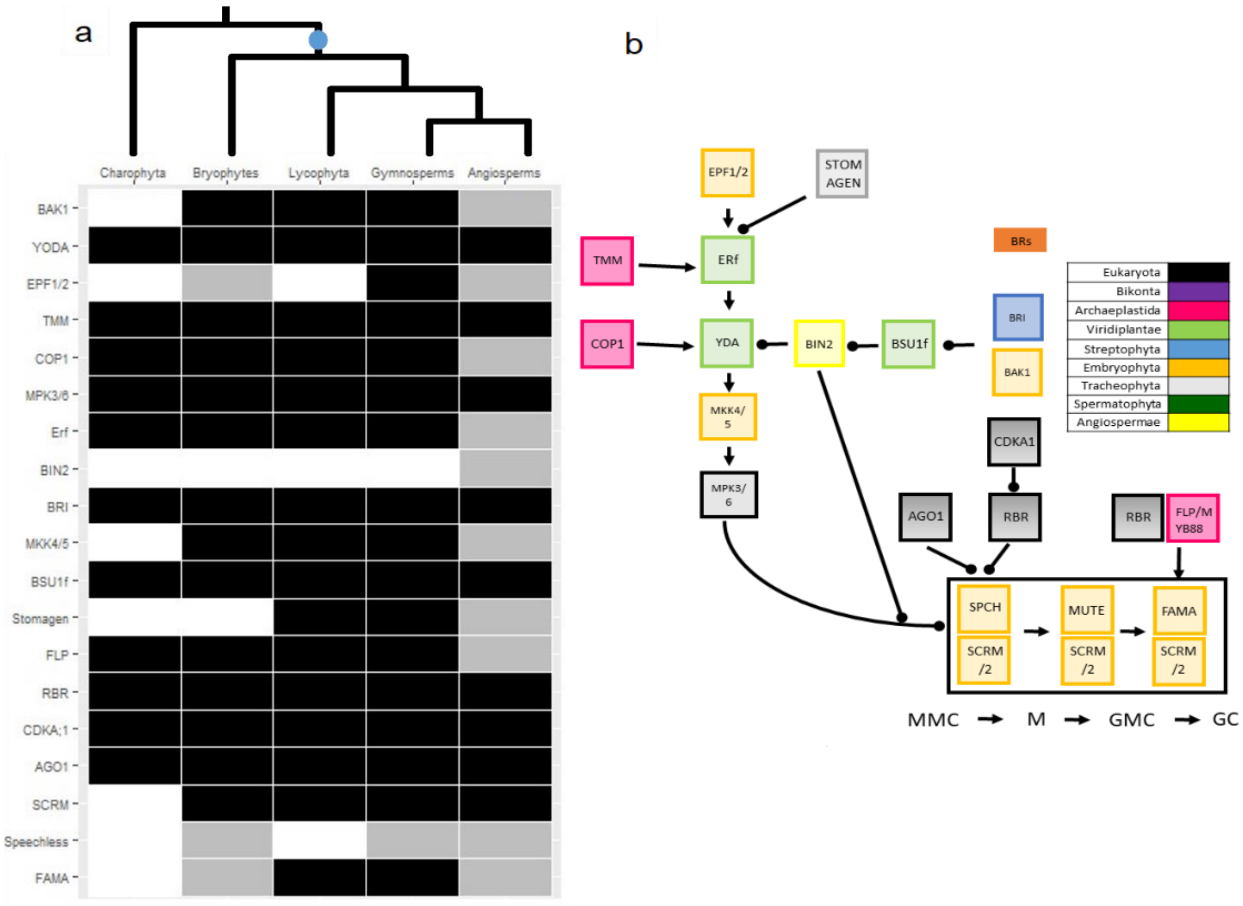


Figure 3.6. The genomic basis of the evolutionary development of stomata. A. Heatmap displaying absence (white), partial presence (grey) and presence in all species (black) for genes involved in stomatal development. The tree at the top illustrates plant evolutionary relationships and the origin of stomata (blue). **B.** Genetic network leading to the development of stomata. Each gene is coloured based on its phylogenetic appearance. Phytohormones are coloured in solid orange.

3.3.3.2 Gene duplication enabled active stomatal control in the ancestor of seed plants

Although stomata appear to have evolved in the ancestor of land plants, the regulation of stomatal function shows a clear distinction between different plant lineages. In bryophytes, fully differentiated stomata are unable to close, and are thought to have originally evolved to desiccate plants (Chater *et al.*, 2016; Renzaglia *et al.*, 2017). In vascular plants, they

are able to open and passively close offering some basic water saving capacity, although active stomatal closure has been identified in some fern species (Cai *et al.*, 2017). In seed plants, active regulation of opening and closing provides rapid responses to water availability triggered by the accumulation of abscisic acid (ABA) (McAdam *et al.*, 2013; Brodribb *et al.*, 2017). This draws into question the evolutionary origin of active stomatal closure.

Here it is demonstrated that genes involved in stomatal closure experience multiple duplications, predominantly in the ancestor of seed plants, suggesting potential neofunctionalisation of duplicate genes (Figure 3.7, Supplementary Data 3.2). Genes involved in the signalling of potassium and anion channels important for stomatal closure (QUAC1, KAT2, AKT1, GORK, CLC-C, CNGC) were present before the origin of land plants. Gene tree inferences revealed that these genes experienced duplications in the ancestor of seed plants (Figure 3.7, Supplementary Data 3.2). Importantly, the core SNRK2 (SNF1-RELATED PROTEIN KINASE 2) dependent ABA signalling pathway (PYL–PP2C–SNRK2) is found in all land plants, including in species without stomata. It has previously been shown the ancestor of land plants contained all genes necessary for ABA synthesis (Bowles *et al.*, 2020) and has been suggested that the downstream signalling pathway has subsequently been co-opted during the evolutionary history of plants to actively regulate gas exchange (Brodribb *et al.*, 2011; McAdam *et al.*, 2012; Ghosh *et al.*, 2016). The results presented here identify that SNRK2s are duplicated in the ancestor of Euphyllophyta. Furthermore, a HG containing PP2Cs (Protein Phosphatase 2Cs), that regulates ABA activation, duplicated twice, once in the ancestor of Euphyllophyta and again in the ancestor of Spermatophyta. Additionally, important in the ABA induced stomatal closure signalling pathway is a group of PYLs (1-3) (PYR1-LIKE 1), an abscisic acid receptor, identified as present in the ancestor of seed plants (Figure 3.7).

The data presented support the genetic re-wiring of ABA responses in the ancestor of Euphyllophyta, particularly the PP2C – SnRK2 complex. However further diversification of

genes in the ancestor of seed plants supports an evolutionary distinction of active stomatal control of monilophytes and seed plants. These duplications, in most notably anion and potassium channels, are associated with the specific mechanisms of stomata closure and therefore support the origin and evolution of active stomatal control in the ancestor of Spermatophyta. There is a possibility that there are unidentified genes in ferns that contribute to stomatal closure, that are absent in seed plants. This could mean that elements of the genetic toolkit for active stomatal closure evolved independently in the ancestor of ferns and seed plants, possibly with the convergent evolution of leaves (Tomescu, 2009). The significance of guard cell-specific expression of genes has been highlighted as important for stomatal closure and remains to be identified for many of these genes (Geiger *et al.*, 2009).

These findings suggest that gene duplications played an essential role in the transition from passive to active stomatal closure in the ancestor of Euphyllophyta and Spermatophyta, allowing plants to preserve water by actively restricting transpiration rates. These may be the product of whole genome duplication in the ancestor of seed plants (Jiao *et al.*, 2011; Ruprecht *et al.*, 2017; Leebens-Mack *et al.*, 2019) or frequent gene duplication.

3.4 Concluding paragraph

The ancestor of land plants likely lacked the inability to regulate water content via structural or functional methods (Proctor *et al.*, 2002). The evolution of stomata, vascular tissue and roots have increased the capacity of water transport and regulation of seed plants. The development of water regulatory features at every major step in the evolutionary history of plants highlights the role of water as a driver of plant evolution. Here the results demonstrate the role that gene novelty, gene duplication and gene co-option played in the evolution of water regulatory traits. Some of these genes evolved at the same time as the morphological innovation they are associated with (i.e., stomatal development genes), while others are older indicating that co-option was concomitant to the evolution of these

traits. Overall, the analyses shed new light on the evolution of the genetic basis of water regulation, highlighting the role of genome dynamics in the diversification of the Plant Kingdom.

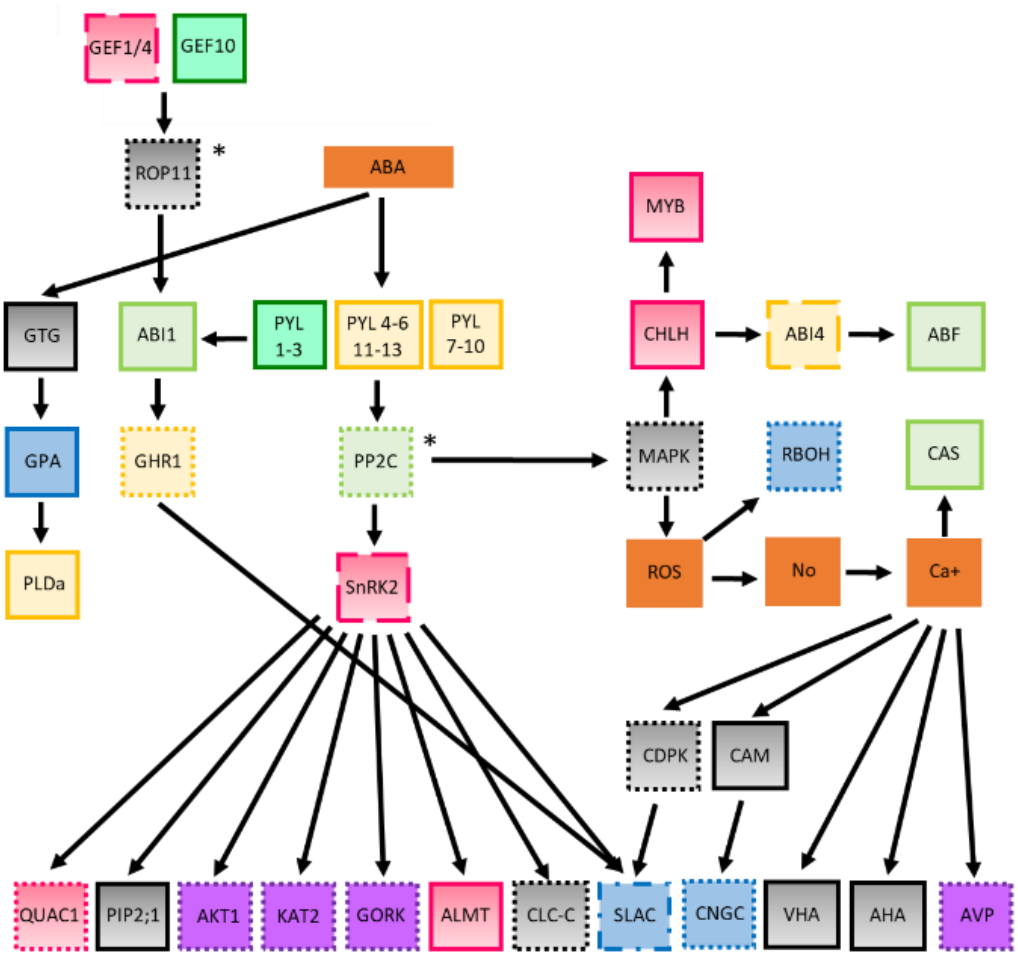


Figure 3.7. The genomic basis of the evolution of stomata signalling. Genetic network involved in stomatal signalling. Each gene is coloured based on its phylogenetic appearance. Phytohormones are coloured in solid orange. Duplicated genes in the ancestor of Spermatophyta are highlighted by boxes with dotted edges. Duplicated genes in the ancestor of Euphyllophyta are highlighted by boxes with dashed edges. Duplicated genes in the ancestor of Tracheophyta are highlighted by boxes with dot-dash-dot edges. Asterisks indicate a HG that duplicates twice once in the ancestor of Euphyllophyta and again in the ancestor of Spermatophyta.

3.5 Materials and Methods

3.5.1 Homology assignment

The pipeline approach has previously been described (Paps *et al.*, 2018; Bowles *et al.*, 2020). Briefly, proteins were extracted for 208 plant genomes and similarity between proteins was identified with an all-vs-all BLAST (Altschul *et al.*, 1990). Sequences were clustered into Homology Groups (HGs) using Markov Clustering (MCL) with a granularity score of 2 (Enright *et al.*, 2002). Additionally for this study, HGs were extracted based on Uniprot gene IDs using the MCL_search_by_gene_name_2.pl script (Supplementary Data 3.1) (UniProt Consortium, 2018).

3.5.2 Genetic toolkit of stomatal development and signalling, vascular tissue development and root development

Literature was searched to identify genes involved in developmental and signalling pathways of stomata, vascular tissue and roots. For stomatal development, a composite figure was made from Chater et al (Chater *et al.*, 2017), Lau et al (Lau *et al.*, 2012) and Le et al (Le *et al.*, 2014). Stomatal signalling genes were identified from Cai et al (Cai *et al.*, 2017), Albert et al (Albert *et al.*, 2017), Cotelle & Leonhardt (Cotelle *et al.*, 2016). For the development of vascular tissues, genes from Ruonala et al were used (Ruonala *et al.*, 2017). Root development genes were identified in Jung & McCouch (Jung *et al.*, 2013) whilst root hair development genes were identified in Vissenberg et al (Vissenberg *et al.*, 2020). For lateral root development, genes from Teixeira et al (Santos Teixeira *et al.*, 2019), Oh et al (Oh *et al.*, 2018) and Verstraeten et al (Verstraeten *et al.*, 2014) were used. The evolutionary development of stomata, vascular tissue and roots, is based on characterised genes in flowering plants, specifically based on *Arabidopsis thaliana.* Therefore, genes that are involved in stomatal, vascular tissue and root development in non-flowering plants are absent from the analysis.

3.5.3 Gene tree inference

A curated list of species was collated with representatives for each major plant group and was used to build gene trees. These were *Cyanidioschyzon merolae* (Rhodophyta), *Cyanophora paradoxa* (Glaucophyta), *Bathycoccus prasinos* (Chlorophyta), *Klebsormidium flaccidum* (Charophyta), *Marchantia polymorpha, Physcomitrella patens* (bryophytes), *Selaginella moellendorffii* (Lycophyta), *Picea abies* (Gymnosperms), *Amborella trichopoda* (ANA grade), *Oryza sativa indica, Brachypodium distachyon* (Monocots) and *Arabidopsis thaliana* (eudicots). Homology Groups were extracted based on the IDs for each gene and a file containing all gene IDs for all species was used to extract the protein sequences using the perl one liner 'perl -ne 'if(/^>(\S+)/){$c=$i{$1}}$c?print:chomp;$i{$_}=1 if @ARGV' ids.file all_fasta.file'.

Homologous sequences were aligned using MAFFT using –auto parameter which automatically selects an appropriate alignment strategy (Katoh *et al.*, 2002). Multiple sequence alignments were trimmed with Trimal using the heuristic method, automated1, to identify and remove poorly aligned positions. Gene trees (bootstrapped maximum likelihood phylogenies) were inferred using IQ-TREE, using the Bayesian Information Criterion (BIC) to select the best-fitting substitution model, specifying 100 bootstrap replicates (Nguyen *et al.*, 2015). Trees were rooted against the latest plant phylogeny to infer gene innovations, duplications and losses. Trees were visualised in iTOL (Letunic *et al.*, 2019).

Trees were analysed to understand the evolutionary relationships of HGs to the innovations they are associated. Novel, Duplicated and Co-opted HGs were identified using the search criteria from Figure 3.8 (Figure 3.1, Appendix 3.1, 3.2, 3.3, 3.4). The trees discussed in the text can be found in Appendix 3.5.

3.6 Acknowledgements

The authors would like to thank Prof. Phillip Mullineaux for his comments on the manuscript. We also would like to thank Stuart Newman for his support of the Genomics HPC server. A.M.C.B., J.P. and U.B. received funding from the School of Life Sciences (University of Essex).

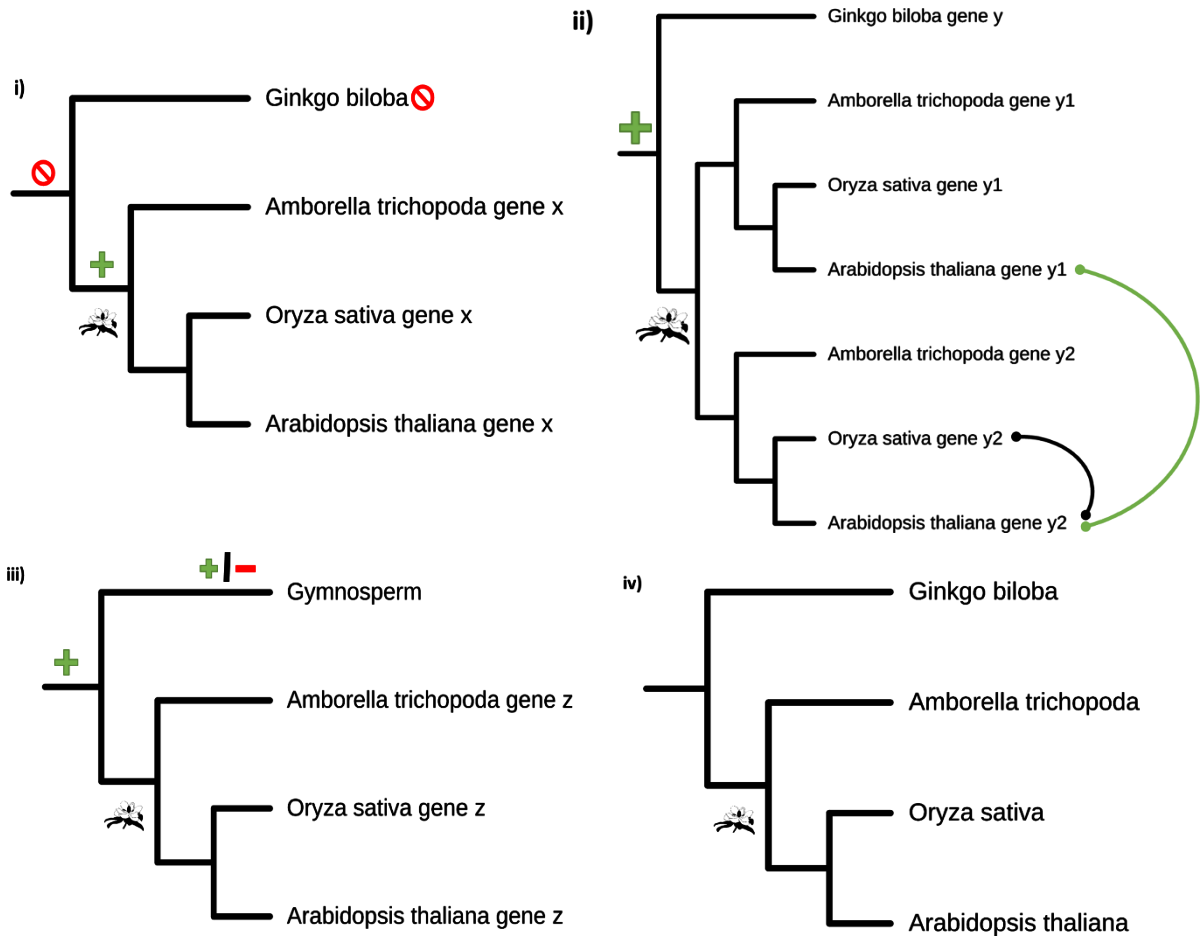


Figure 3.8. Evolutionary distinct classifications of HGs. For all examples of evolutionary distinct HGs, the genes relate to the origin of flowers, highlighted by the magnolia silhouette (sourced from https://phylopic.org). **i. Novel HGs:** as in Chapter 2, a Novel HG contains genes in the clade of interest, angiosperms, which here are represented by A. trichopoda, O. sativa and A. thaliana and absent outside the clade of interest (e.g. absent in G. biloba). **ii. Duplicated HGs:** These HGs contain genes that have duplicated in the ancestor of the clade of interest, in this case the LCA of angiosperms. These can be identified by comparing the gene tree (ii) to the species tree (iv). The black line connecting genes highlights two orthologous genes, related by

speciation. The green line connecting genes highlights two paralogous genes, related by duplication. **iii. Co-opted HGs:** Genes in these HGs predate the innovations they relate, e.g. the origin of flowers. Genes are present in the sister group (gymnosperms) to the clade of interest (angiosperms). However, in the sister group, they show a patchy distribution highlighted by plus (green) and minus (minus) symbol. Genes in these HGs are found in all species in the clade of interest suggesting that co-option of old genes for new functions occured. **iv. Species tree:** Species tree for comparison with example gene trees.

Chapter 4 Life out of water: the origin of drought and desiccation tolerance in plants

Bowles AMC, Paps J, Bechtold U. 2020. Life out of water: the origin of drought and desiccation tolerance in plants. *In review in Frontiers in Plant Science*.

Authors

Alexander M. C. Bowles[1], Jordi Paps[123] & Ulrike Bechtold[1]

Affiliations

1. School of Biological Sciences, University of Essex, Colchester CO4 3SQ, UK

2. School of Biological Sciences, University of Bristol, Bristol, BS8 1TQ, UK

3. Department of Zoology, University of Oxford, Oxford, OX1 3SZ, UK

4.1 Summary

It is commonly known that drought stress is one of the major constraints limiting crop production, decreasing productivity more than any other abiotic or biotic stress. Drought stress and associated drought resistance mechanisms are therefore under intense investigation with the view to future crop production with greater capacity for drought tolerance. Understanding the evolution of the highly complex and ambiguous trait, drought tolerance, may inform us about patterns of gene gain and loss in relation to diverse adaptations. By unravelling the evolutionary history of plants, the impacts of natural selection shaping genomes to adapt to stressful environmental conditions can be examined. Here, the evolution of drought and desiccation tolerance across the plant kingdom, Viridiplantae is investigated. First the transition of plants from water to land, and the role of desiccation tolerance in enabling this transition are considered, before discussing the first drought tolerant plant and common drought responses amongst vascular plants. Next, the distribution of a collective “drought adaptation” trait in ~180 extant plant species defined in broad terms to encompass different experimental systems and definitions used in the current literature is reviewed and classified. By completing ancestral state reconstruction incorporating these definitions, the evolutionary history of this drought tolerance trait is mapped onto the plant tree of life. With an ever growing population, novel approaches need to be considered to sustainably feed future generations. Finally comments are provided on how incorporating this information into an evolutionary genomics framework can provide insights into the molecular mechanisms underlying drought adaptations which could offer insight for producing crops with greater capacity for tolerating drought.

4.2 The pathway to the first land plants

4.2.1 Water relations

Water is essential for life on Earth. However, the relationships of the plant kingdom (Viridiplantae) with water has changed dramatically. The common ancestor of extant

Viridiplantae were photosynthetic eukaryotes adapted for life in aquatic environments. The divergence of Chlorophyta and Streptophyta occurred approximately 1 billion years ago (Morris *et al.*, 2018). The evolutionary innovations associated with this development enabled ancestral streptophytes to respond to novel environmental challenges which included extremes of UV, temperature and light (de Vries *et al.*, 2018a, 2018b). Streptophyte algae are found in a range of brackish, freshwater and terrestrial habitats, which demonstrates their range of adaptations to water availability (Delwiche *et al.*, 2015; Fürst-Jansen *et al.*, 2020). Adaptations to these conditions involve many similar features of terrestrial stressors such as desiccation, salinity, pH and nutrient variation (de Vries *et al.*, 2016). As such, it has been remarked that the adaptations needed for plants on land and shallow or transient water are highly similar (Donoghue *et al.*, 2020).

Studies of charophyte algae are revealing that features once thought to be unique to land plants, in fact first appeared in the ancestor of close algal relatives (e.g. associations with substrate microbiota) (de Vries *et al.*, 2018a, 2018b; Nishiyama *et al.*, 2018; Cheng *et al.*, 2019; Liang *et al.*, 2019; Wang *et al.*, 2019). To a certain extent, the ancestors of Streptophyta (and subgroups e.g. Phragmoplastophyta) were pre-adapted to life on land (Delaux *et al.*, 2015). Sequencing of the *Klebsormidium flaccidum* genome revealed that ancestral charophytes acquired the fundamental machinery for land plant adaptation including hormone signalling, high intensity light and desiccation tolerance (Hori *et al.*, 2014). Analyses of further streptophyte algae have been shown they are able to tolerate periods of desiccation (Hori *et al.*, 2014; Holzinger *et al.*, 2015b). Whole genome sequencing of species either side of the transition to land is revealing much about the genetic innovations accompanying the development of land plants (Bowman *et al.*, 2017; De Clerck *et al.*, 2018; de Vries *et al.*, 2018a, 2018b; Nishiyama *et al.*, 2018).

By using comparative analysis, the genetic toolkit aiding the conquest of land is being elicited. It is becoming clear that the backbone of phytohormone signalling, required for stress responses, either predates or accompanies the transition to land (Wang *et al.*, 2015, 2019; Bowman *et al.*, 2017; Bowles *et al.*, 2020; Cannell *et al.*, 2020). However, there are

differences in responses to water availability between charophyte algae (extreme) and land plants (specialised). This means that although many key genes evolved prior to the transition to land, specific responses and genetic re-wiring of stress response pathways occurred later in land plant evolution, allowing for greater adaptive plasticity to water availability.

4.2.2 The first land plants

Based on the latest fossil evidence and molecular dating, the first plants transitioned from aquatic to terrestrial environments approximately 500 million years ago in the Ordovician – Cambrian period (Rubinstein *et al.*, 2010; Morris *et al.*, 2018). All extant land plants descend from a single common ancestor (Wickett *et al.*, 2014; de Vries *et al.*, 2018a) and have since diversified into almost 400,000 species that have shaped modern ecosystems (Kenrick *et al.*, 1997; Willis, 2017). Their rise to ecological dominance has enabled plants to colonize every continent on Earth. This involves adaptations to many extreme environments including arid deserts (Xiao *et al.*, 2015; Copetti *et al.*, 2017), marine environments (Olsen *et al.*, 2016) and even the Antarctic (Lee *et al.*, 2014). The origin of the first embryophytes was accompanied by the production of novel developmental and morphological mechanisms for adaptation to life on land (Bowman *et al.*, 2017). Analysis of fossils from the Rhynie Chert, a well maintained fossil deposit in Scotland, suggests that in the Early Devonian, ~400 million years ago, plants were tolerant to high salt levels and osmotic stress, a key component of drought stress (Channing *et al.*, 2009).

Land plants have many adaptations for surviving in water limited environments. A common feature of embryophyte life cycles are reproductive structures with the capacity to survive desiccation (e.g. spores, seeds). Based on phylogenetic evidence, the desiccation tolerance of early land plants derived from a mechanism first developed in spores. Furthermore it has been shown that the responses of extant bryophytes has changed very little to those of early land plants (Oliver *et al.*, 2005). It has been hypothesised that desiccation tolerance would have been an ancestral trait in green plants (Oliver *et al.*,

2000; Wood, 2007a). Therefore desiccation tolerance would have been a key component for the adaptations for life on land but importantly these plants would have lacked the ability to regulate water content, termed poikilohydry (Stevenson *et al.*, 2016; Becker *et al.*, 2020). Desiccation tolerance in bryophytes is common with over 200 of 2100 species verified as capable of this phenotype (Proctor *et al.*, 2007; Wood, 2007b; Gao *et al.*, 2017). In tracheophytes, also known as vascular plants, desiccation tolerance is less common. In the lycophytes, the majority of species are susceptible to desiccation, although a few tolerant species have been identified including *Selaginella lepidophylla* (Yobi *et al.*, 2013) and *Selaginella tamariscina* (Wang *et al.*, 2010). In the angiosperms, only 160 of 369,000 flowering plant species have been confirmed as desiccation tolerant including all species in the genera *Vellozia* and *Xerophyta* (Wood, 2007b). Based on cladistic thinking, this implies that desiccation tolerance was lost in the ancestor of tracheophytes. The major occurrence of desiccation tolerance of seed plants is in the reproductive structures of pollen and seed embryos which is thought to have derived from the desiccation tolerance of spores (Gaff *et al.*, 2013)*.*

4.3 Drought tolerance

In vascular plant evolution, desiccation tolerance in vegetative tissue has been lost, in place of desiccation tolerance in spores and seeds (Xu *et al.*, 2018). The responses to limited water in early tracheophytes diversified by increasing regulatory and morphological complexity (Lu *et al.*, 2020). The origin of tracheophytes was accompanied by the appearance of a sporophyte dominant life cycle and vascular tissue (Harrison, 2017). These two innovations enabled plants to tolerate dry conditions and to control the internal movement of water and nutrients. This suggests that during the evolution of tracheophytes, early forms of drought tolerance originated.

Throughout plant evolutionary history, controlled responses have evolved that allow plants to respond to the temporary lack of water (Park *et al.*, 2009; Wang *et al.*, 2015; Bowman *et al.*, 2017, 2019; Nishiyama *et al.*, 2018; Bowles *et al.*, 2020). The developmental and

morphological innovations linked to drought tolerance include stomata and roots. However, as these have been discussed in the previous chapter in depth, then they will be discussed no further. The evolution of other important features includes vascular tissue, specialised reproduction, euphylls and seeds (Harrison, 2017). The development and responses of these features are regulated by a network of genes and signalling molecules. Next, in this work, these key signalling pathways and genes for drought tolerance are discussed, investigating how these gene families have evolved across the plant phylogeny.

4.3.1 Phytohormone signalling

Studies have recently provided new insights into the evolution of phytohormone signalling pathways (Wang *et al.*, 2015; Bowman *et al.*, 2017, 2019; Nishiyama *et al.*, 2018). It is becoming apparent that many components of phytohormone signalling pathways were present in the LCA of streptophytes. In the context of drought tolerance, abscisic acid (ABA) signalling is of particular importance (Franks *et al.*, 2007). Comparison of phytohormone signalling in the streptophyte algae, *Chara braunii,* and the liverwort, *Marchantia polymorpha,* demonstrated that the majority of cytokinin, ethylene and ABA signalling pathways were present in the LCA of Streptophyta (Bowman *et al.*, 2017; Nishiyama *et al.*, 2018). PYLs, an important receptor in the ABA signalling process, have previously been identified to be highly conserved across all Embryophyta (Bowman *et al.*, 2017; De Clerck *et al.*, 2018; Nishiyama *et al.*, 2018). A recent study of six major streptophyte algal lineages identified the presence of PYLs in the charophycean algae, *Zygnema circumcarinatum* (de Vries *et al.*, 2018b). This indicates that drought stress responses developed earlier than previously thought and are a key factor that enabled streptophytes to colonise terrestrial habitats (Ruszala *et al.*, 2011).

Bryophyte ancestors recruited PROTEIN PHOSPHATASE 2Cs (PP2Cs) to be able to regulate the pre-existing ABA mediated desiccation tolerance signalling pathways (Komatsu *et al.*, 2013). Later in the evolution of land plants, response specific pathways

evolved (e.g. the SLAC1 anion channel modulated by SnRK2s for stomatal closure) (Brodribb *et al.*, 2011). This ability to provide tightly regulated responses to drought through distinct ABA signalling channels would have provided ancestral land plants and descendants with a competitive advantage, contributing toward their proliferation.

In extant plants, phytohormones function in drought responses of leaf shedding, cell division in plant roots and seed dormancy. It has been demonstrated that there are many land plant specific phytohormones allowing for tighter regulation of responses to abiotic stresses, for example, jasmonic and gibberellic acid were demonstrated to have emerged and evolved within embryophytes (De Clerck *et al.*, 2018). However, studies have also shown these pathways may have served different functions in ancestral plants to their functions in extant plants (McAdam *et al.*, 2016).

4.3.2 Drought and Desiccation Tolerance Gene families

Although, drought tolerance is a complex trait with many genes involved in the expression of responses, several gene families play a fundamental role. Cellular protection is known to be coordinated by LATE EMBRYOGENESIS ABUNDANT (LEA) proteins (Hundertmark *et al.*, 2008). Originally involved in desiccation tolerance, LEA proteins have since been shown to enhance drought and heat stress tolerance in the vegetative plant tissue (Delahaie *et al.*, 2013; Magwanga *et al.*, 2018). Evolutionary analysis of LEA proteins found that gene families underwent rapid diversification in the ancestor of land plants, enabling adaptations to water limited environments. Furthermore, the study also identified subsequent diversification of the Group 4 LEA protein family in angiosperms, which accumulate under water stress (Artur *et al.*, 2019). Detailed analysis of the expansions of LEA proteins in non-flowering plants and implications for plant adaptations to water limited environments remains to be determined.

DEHYDRATION RESPONSIVE ELEMENT BINDING (DREB) transcription factors are another important gene family for drought responses (Agarwal *et al.*, 2017) and are a subfamily of the AP2/ERF gene family (Xie *et al.*, 2019). It has been demonstrated that

AP2/ERF transcription factors originated in the ancestor of Viridiplantae and have subsequently diversified in the LCA of Embryophyta (Catarino *et al.*, 2016). It is yet unclear the specific timing of the diversification of DREB transcription factors and the biological implications of their diversification.

4.3.3 Distribution of drought adaptation in the plant phylogeny

As highlighted above, many important physiological, structural and regulatory adaptations have arisen in response to drought stress, over the course of plant evolutionary history. Analysing the distribution of drought and desiccation tolerance in embryophytes could give important insights into the evolution of land plants, as well as the origins of drought tolerance. A similar approach has previously been used to map the distribution of halophytes, salt-tolerant plants, onto the land plant phylogeny. This demonstrated that salt tolerance evolved independently across different plant lineages (Flowers *et al.*, 2010). In order to conduct such evolutionary analyses, it is important that the trait of interest (for example salt or drought tolerance) is clearly defined (Delaux *et al.*, 2019). As such, in this chapter, a collective “drought adaptation” trait for the plant species in the genomic dataset described in Chapter 2 is defined, by conducting a thorough literature search. This definition is then applied in Chapter 5 to investigate the distribution of drought tolerance genes across the plant phylogeny.

4.4 Methods

4.4.1 Defining a drought adapted plant

The great diversity of plant species means that it is easy to distinguish between plants requiring water to survive (defined as mesophytes) and plants adapted to live without it (xerophytes) (Xi *et al.*, 2018). For example, the desert species *Anabasis syriaca* is a xerophyte since it can live without water for long periods of time, whereas *Spinacia oleracea* is described as a mesophyte since it requires a constant supply of water (Baydoun *et al.*, 1985). However, plants with less extreme disparity in adaptation are much harder to classify. The diversity of plant species in a variety of climates indicates that there

are a plethora of drought responses that have evolved including morphological, physiological and regulatory adaptations.

The responses of individual plant species to drought are not always well-documented and, as such, the definition of drought adaptation is sparse across the plant phylogeny. Well studied species, such as *Arabidopsis thaliana* and *Oryza sativa,* have clearly defined and tested responses to drought stress. The drought responses of dehydration adapted plants are also evident, for example in the case of the resurrection plant, *Boea hygroemetrica* (Xiao *et al.*, 2015) and the desert tree, *Populus prunoisa* (Yang *et al.*, 2017b). However, there still remain many enigmatic plant species whose ability to tolerate drought has not been clearly evaluated or defined.

Therefore, in this chapter, a literature search is conducted to create a collective definition of a drought adapted (and a drought sensitive) plant. A drought adapted plant would be a plant adapted to periods of variable precipitation and that is able to maintain biomass under such conditions. The literature search was conducted using relevant search terms and in relation to a species name (Table 4.1; Appendix 4.1). Terms were inputted into PubMed, Web of Science and Google Scholar. Based on the resulting papers, species were then categorised into drought adapted or drought sensitive classes. Species were said to be undefined if a drought term could not be clearly associated with a particular species in a paper in the literature. Additionally species were said to be undefined if there had been no attempts to characterise their response to drought. Short of characterising the drought adaptation of species individually, as has been completed for species in the genus *Vigna* (Iseki *et al.*, 2016, 2018), this is an approach that can provide definitions for a range of well characterised and poorly studied plants.

Table 4.1. Terms used in the literature search to categorise plants according to their drought response.

Drought adapted	Drought sensitive
Drought tolerance	Drought sensitive
Drought avoidance	Drought susceptible
Drought escape	Drought prone
Desiccation tolerance	

4.4.2 Ancestral State Reconstruction

Following the classification of plant species as either drought tolerant or drought sensitive, both Maximum Likelihood (ML) and Bayesian approaches were used to reconstruct the ancestral states of drought characters. Both methods are commonly used in phylogenetics but they differ in terms of input data and their methodology (Svennblad *et al.*, 2006). Thus, outputs were compared between the two approaches. The ML approach aims to identify the character states at ancestral nodes that maximise the probability of the observed character states in extant plants (Felsenstein, 1981). Bayesian approaches aim to sample character histories from the posterior probability of the given data. Posterior probability is the probability of character states at ancestral nodes, given the distribution of character states in extant plants (Rannala *et al.*, 1996). Similar support from both approaches for the evolutionary history of drought tolerance would provide robustness to the findings.

4.4.2.1 Likelihood approach for ancestral state reconstruction

Character states for each species (Appendix 4.1) and species relationships were entered into Mesquite. The species relationships were based on the NCBI taxonomy used in Chapter 2 (Federhen, 2012) and were entered in the form of a species tree without branch lengths. The Trace Character History option was used to analyse the evolutionary history of the drought tolerant and sensitive characters using maximum likelihood approaches.

Additionally, a domestication status (cultivated or wild species) was incorporated to investigate the influence of domestication on the loss of drought adaptations. The definitions of a cultivated and wild species were sourced from the genome papers of these plant genomes detailed in Appendix 2.1.

4.4.2.2 Bayesian approach for ancestral state reconstruction

4.4.2.2.1 Concatenation approach to build a species tree

To complete the Bayesian approach to infer the evolution of drought tolerance, a species tree with branch lengths was required. To produce this tree, genes from 315 homology groups (HGs) present in all Archaeplastida were extracted from the computational pipeline described in Chapter 2. Specifying presence of all genes in all Archaeplastida species in the pipeline query ensures that data for all species is present. Due to the broad clustering of homology groups (compared to orthogroups), each HG contained more than one protein sequences per species. However, for gene tree inference, one gene per species per HG is required. As such, the first gene for each species for each HG was selected.

There are several methods with which to build species trees from multi-gene datasets. These include a single step coalescent, two step coalescent approach and a gene concatenation approach. The later of these, the concatenation approach, has several strengths and has been used to infer the origin and early diversification of land plants (Wickett *et al.*, 2014). This approach, incorporating phylogenetic inference from multiple gene alignments, provides a strong phylogenetic signal with which to build a robust species tree.

As in previous chapters, the genes from each HG were individually aligned using MAFFT with the --auto parameter (Katoh *et al.*, 2002) and trimmed using trimal with the automated1 option (Capella-Gutiérrez *et al.*, 2009). PhyUtility was used to concatenate the trimmed gene alignments into a supermatrix, with the command ‘phyutility -concat -in *aligned.trimmed.files -out concatenated_supermatrix.nexus’ (Smith *et al.*, 2008). Once this concatenated supermatrix was produced, a partition file was created which is used to

identify each trimmed gene alignment. This was completed with the bash command ‘head -n 3 concatenated_supermatrix.nexus | tail -n 1 | sed 's/^.*\[//g' | sed 's/\s\].*$//g' | sed 's/.fa.mafft.trimal_gene.\s/=/g' | tr '' '' "\n" | sed 's/^/AUTO,\s/g' >> partitions.txt’.

Different sequences have different rates of evolution (Lopez *et al.*, 2002). To account for this, parameters were altered in IQTree to determine the different rates of sequence evolution for each individual gene alignment (Nguyen *et al.*, 2015). To improve the speed of analysis, the command ‘iqtree –s concatenated_supermatrix.nexus -spp partitions.txt -m TESTMERGE –rcluster 10 -bb 1000‘ was used. The option –m TESTMERGE specifies that a subset of models of evolution, which are invariable site and Gamma rate heterogeneity, are used to save computational time (Lanfear *et al.*, 2017). The –rcluster 10 option specifies that only the top 10% of partition merging schemes are considered. The –bb 1000 option specifies 1000 ultrafast bootstraps. The species tree produced was used in subsequent analysis (Supplementary Data 4.2).

4.4.2.2.2 Ancestral State Reconstruction using the Bayesian approach

Phytools can be used to estimate ancestral character states for discretely valued traits (Revell, 2012), in this case drought adaptations. Phytools was run using a continuous-time Markov chain (MCMC) model. Initially, data were mapped onto the species tree built from the concatenation approach detailed above (Fig. 4.1). The MCMC approach is used to sample character histories from their posterior probability distribution, termed stochastic character mapping (Huelsenbeck *et al.*, 2003). To sample a greater portion of the distribution of the character history, 100 stochastic maps were produced and plotted (Fig. 4.2). The results of these sets of stochastic maps were then summarised and have been plotted in the results section (Fig. 4.5).

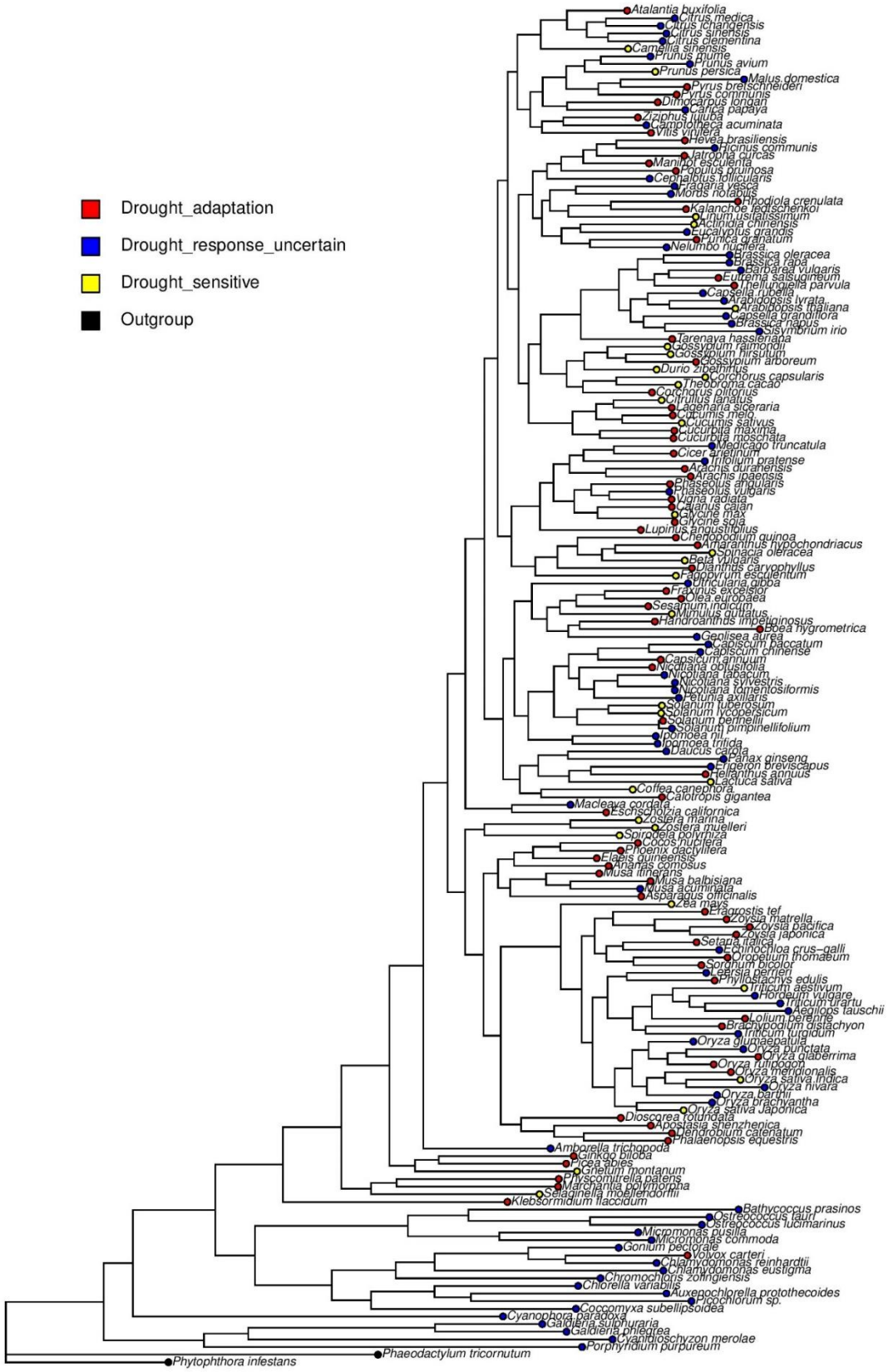


Figure 4.1. The distribution of drought response categories on a species tree of 178 plant species built from concatenation analysis of 315 universal genes. The colours of terminal nodes represent the drought response status of each species as defined by a literature search (See Appendices 4.1).

Figure 4.2. 100 stochastic character trees with the mapped drought response categories (Drought adapted: red, Drought sensitive: yellow, Drought response uncharacterised: blue, Outgroup taxa: black).

4.5 Results

4.5.1 Definition and distribution of drought tolerant plants

Of the 178 plant species included in the literature search, 74 were recognised as drought adapted, 29 were identified as drought sensitive and for 75 species no clear definition

could be assigned (Appendix 4.1). These species were found to be distributed across the plant phylogeny, occurring in all major evolutionary groups (Figure 4.1).

4.5.2 Maximum Likelihood approach to ancestral state reconstruction of drought adaptation

Using the likelihood approach in Mesquite, ancestral state reconstruction of drought adaptations was completed for all green plants (Fig. 4.3). The analysis suggested that the last common ancestor (LCA) of Streptophyta was likely to be drought adapted. Further investigation of the literature suggests that these plants were desiccation tolerant, which is the capacity to survive near complete dehydration (Oliver *et al.*, 2000), since the early diverging extant species are also desiccation tolerant (Fig. 4.3). This includes the streptophyte algae *Klebsormidium flaccidum* (Hori et al, 2014) and *Zygnema circumcarinatum* (Becker *et al.*, 2020) and the bryophytes *Physcomitrella patens* (Xiao *et al.*, 2018) and *Tortula ruralis* (Proctor *et al.*, 2007).

Furthermore, the analysis revealed that the LCA of vascular plants was likely to have been a drought adapted plant (Figure 4.3). This further suggests that tolerance to drought as an adaptation was acquired once which potentially occurred with the development of vascular tissue and a sporophyte dominated lifestyle (Harrison, 2017). Drought tolerance appears to have been highly retained which suggests that, for any drought sensitive species that appear later than this ancestor, the ability to adapt to drought has subsequently been lost. *Spirodela polyrhiza* (duckweed) (Wang *et al.*, 2014d)*, Zostera marina* (Olsen *et al.*, 2016) and *Zostera muelleri* (Lee *et al.*, 2016a) are in the order Alismatales and have all adapted to an aquatic lifestyle (i.e. *Zostera* are a genus of seagrasses). Thus, drought tolerance has been lost in these plants due to their transition back into aquatic environments.

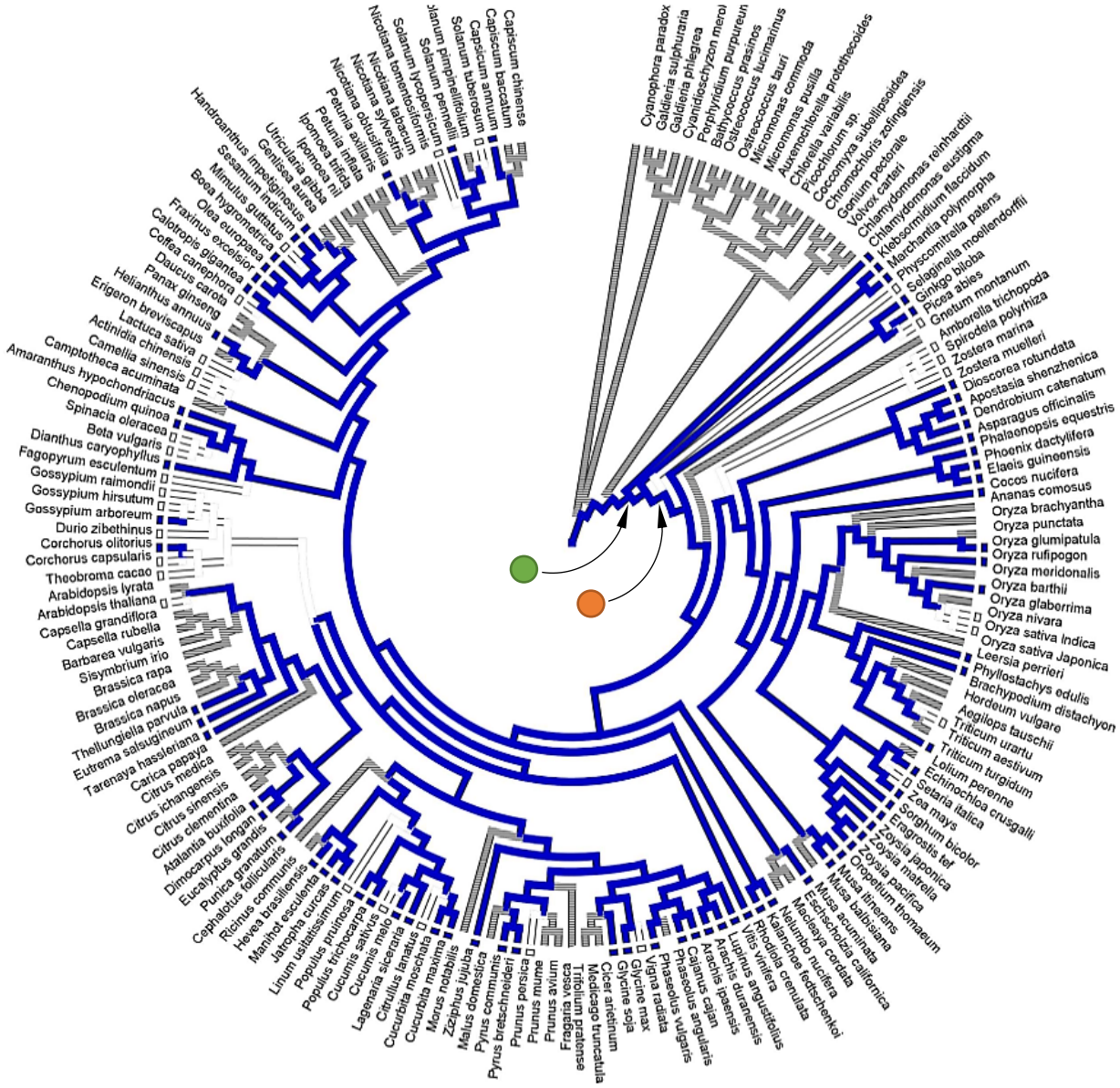


Figure 4.3. Distribution and Ancestral State Reconstruction of drought adaptation across the plant phylogeny. Blue terminal nodes indicate where a plant is drought adapted, white terminal nodes indicate drought sensitivity and grey striped nodes indicate that a drought status could not be assigned. Internal nodes indicate ancestral state reconstruction of drought adaptations completed using likelihood methods in Mesquite. The green dot denotes the last common ancestor of Streptophyta and the orange dot denotes the last common ancestor of Tracheophyta.

During the domestication process, artificial selection can lead to the loss of genetic diversity (Doebley *et al.*, 2006). For example, common signatures of selection include grain retention (for example, in rice, barley and wheat), reduction of lateral branching (for example in maize and sunflowers) and modifications to flowering-time (Ross-Ibarra *et al.*, 2007; Olsen *et al.*, 2013; Kantar *et al.*, 2017). Such indirect effects of domestication have also led to the loss of particular stress tolerance traits, including drought tolerance (Yu *et al.*, 2008; Zhu *et al.*, 2019a; Wang *et al.*, 2020b, 2020a).

In the same manner as drought adaptations above, domestication statuses for the same set of plant species were assigned. These statuses were exclusively sourced from the genome paper of each plant genome. To investigate the impact of domestication on the loss of drought tolerance, both the domestication history and plant drought statuses were mapped across the plant phylogeny (Fig. 4.4). Drought sensitivity appears to be common amongst many of the major crop species. In fact, drought sensitivity is predominantly found in crop species, suggesting domestication could potentially explain the loss of drought tolerance for many species. The only cases of loss of drought tolerance in wild species were found in the order Alismatales, as well as in the non-flowering plants, *Selaginella moellendorffii* and *Gnetum monatum*. The ancestral states of the last common ancestor of land plants was drought adapted and wild. Therefore, for any plants that are drought tolerant and cultivated, this likely represents the ancestral state.

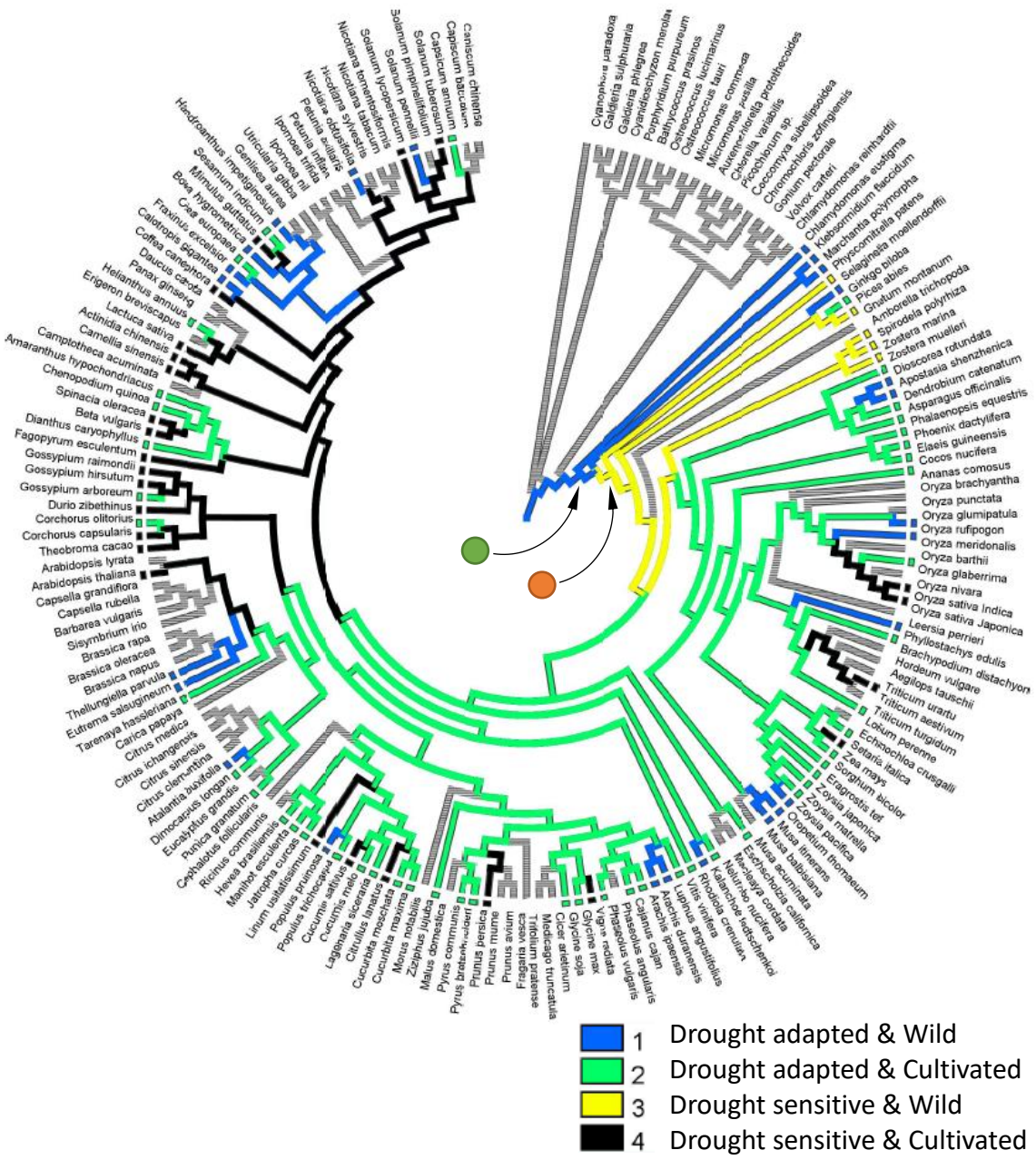


Figure 4.4. Distribution of domesticated and drought adapted species across the plant phylogeny. Grey striped nodes indicate that a drought status could not be assigned. Ancestral state reconstruction was completed using likelihood methods in Mesquite. The green dot denotes the last common ancestor of Streptophyta and the orange dot denotes the last common ancestor of Tracheophyta.

4.5.3 Bayesian approach to ancestral state reconstruction of drought tolerance

The species tree built to infer the evolutionary history of drought tolerance broadly supports the most up-to-date view of plant evolution (Figure 4.5). The analysis identifies that Rhodophyta are sister to a clade containing Glaucophyta and Viridiplantae (Figure 4.5), which is in agreement with a recent analysis of early Archaeplastida evolution from the 1000 plant transcriptomes project (Leebens-Mack *et al.*, 2019) and also a previous study investigating the *Cyanophora paradoxa* genome (Price *et al.*, 2012). The relationships of the remaining non-flowering plants are broadly in agreement with current evolutionary thinking. The charophyte, *Klebsormidium flaccidum*, is sister to land plants whilst gymnosperms are recovered as the sister group to flowering plants (Figure 4.5). The only discrepancy in the evolutionary tree is that *Selaginella moellendorffii*, a lycophyte, is placed in a clade with the two bryophytes in the dataset (Figure 4.5). This difference might be caused by the fact that the first gene from each Homology Group was chosen for building the species tree, as opposed to the longest gene which might have provided a greater amount of sequence information with less fragmentation. Alternatively, this could be caused by variation in sequencing quality across the *Selaginella moellendorffii* genome (Banks *et al.*, 2011). Broadly, the flowering relationships supported the current view of plant evolution, that is, that the ANA grade angiosperm, *Amborella trichopoda*, is a sister to the Mesangiospermae, which consists of the Eudicots and Monocots (Figure 4.5, Chase et al., 2016).

As the species tree supports the most current topology of plant evolution it could reliably be used for the ancestral state reconstruction of drought tolerance. A Bayesian approach to ancestral state reconstruction identifies similar patterns for the evolution of desiccation and drought tolerance to those identified using the maximum likelihood approach described above (Figures 4.3-4.5). For instance, the LCAs of land plants and seed plants were drought adapted, most likely desiccation tolerant and drought tolerant respectively (Figure 4.5). Equally, similar patterns can be identified that are associated with the loss of

drought tolerance with such losses identified in the order Alismatales and also across many crop species (Figure 4.5).

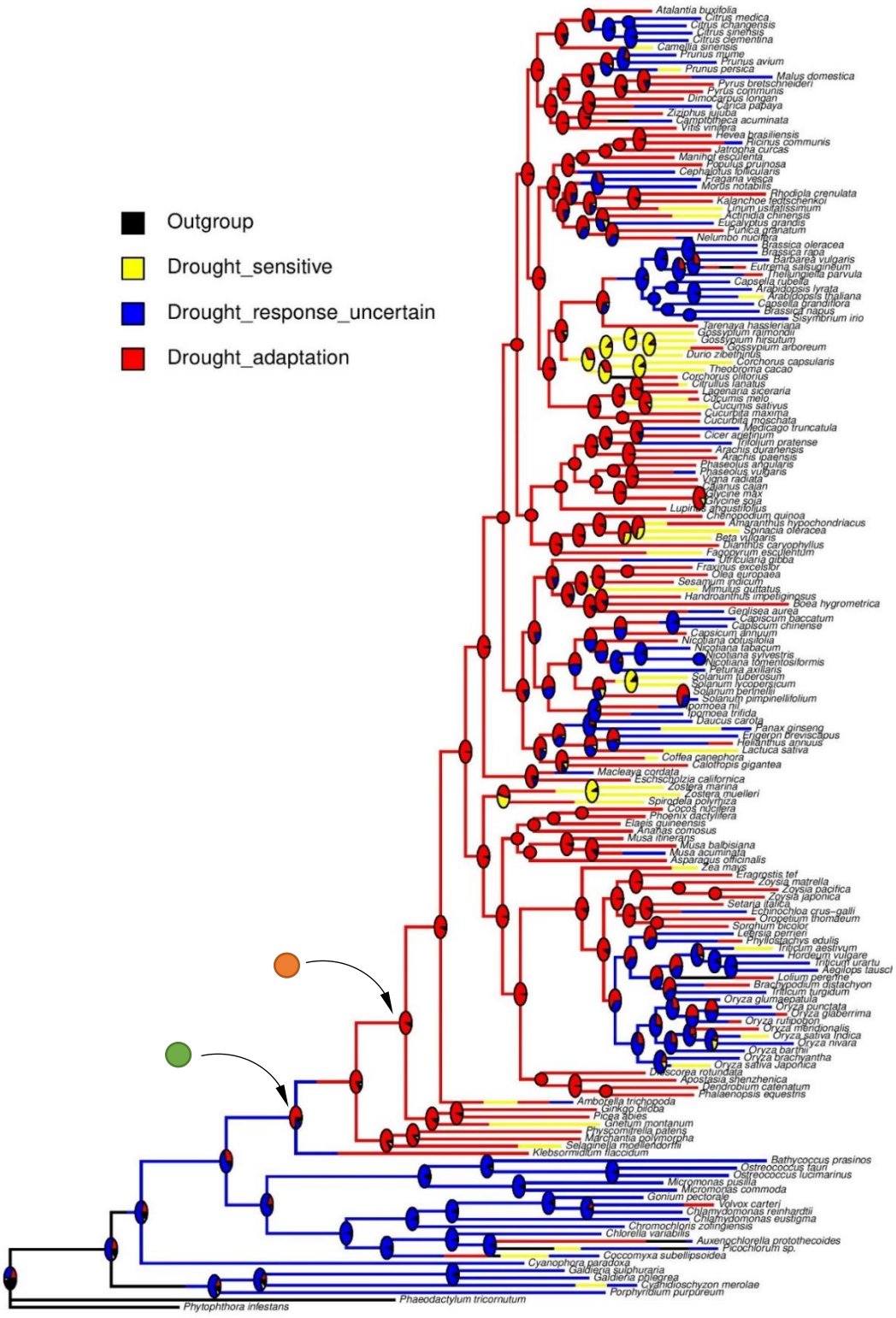


Figure 4.5. Ancestral state reconstruction of drought adaptation on a species tree of 178 plant species. The species tree was built from concatenation analysis of 315 universal genes. Branches are coloured by drought response status. Pie charts represent the support for the ancestral state at each internal node. The green dot denotes the last

common ancestor of Streptophyta and the orange dot denotes the last common ancestor of Spermatophyta.

4.6 Discussion and Conclusion

4.6.1 Ancestral state reconstruction of drought adaptations

In this chapter, a collective ‘drought adaptation’ trait was defined by querying the literature for drought response terms in reference to each species in the genomic dataset. Ancestral state reconstruction, using both maximum likelihood and Bayesian methods, was then used to map this trait onto the plant phylogeny. This revealed that the last common ancestor of Streptophyta was drought adapted. Based on a further search of the literature, this suggested that the ancestor of Streptophyta was desiccation tolerant and the ancestor of vascular plants was drought tolerant. The major occurrences of drought sensitivity were in crop species suggesting plant domestication as the selective pressure leading to the loss of drought adaptation.

An important point to note here, is that the majority of genomes are those of crop species and are mainly found within the flowering plants. It has previously been noted that a greater number of plant genomes from a diversity of plant species will be needed to understand the evolution of key traits and the diversity of plant life (Rensing, 2017). The inclusion of further genomic data will improve our understanding of a diverse number of adaptations and aid gene discovery for the improvement of crop stress tolerance. It would also provide further clarity on whether there are a greater number of instances in which drought sensitivity occurs in wild plant species and whether there are further factors that have driven the loss of drought adaptations other than the indirect effects of domestication.

4.6.2 Defining a collective drought tolerance trait

Defining a collective ‘drought adaptation’ trait is complex as there are a wide range of responses that plants adopt in response to drought. In this work, a collective ‘drought tolerance’ trait was defined by querying the literature for drought response terms in reference to each species in the genomic dataset. This collective ‘drought adaptation’ trait,

therefore, produces a binary outcome, where a plant is either drought adapted or drought sensitive. In reality, drought tolerance responses are far more diverse, but approaches that aim to capture this diversity are limited.

Work on a method to define drought tolerance has begun in the TRY trait database, a global database of curated plant traits (Kattge *et al.*, 2020). The database release in 2020 investigated the prevalence of species tolerance to drought which incorporated a low, medium and high level of tolerance. This approach aims to categorise a drought tolerance trait for a broader range of taxa. To date, the 'species tolerance to drought' trait has been categorised for 3324 species (Kattge *et al.*, 2020), although this does not cover all the species in the genomic dataset. There are limitations with this method, similar to the approach detailed in this study, that there are only three categories of drought tolerance. The TRY database is also still under construction, with trait representation only for euphyllophytes.

An additional approach for defining a collective 'drought tolerance' trait could be to include information about the geographical distribution of plant species. Some genome papers provide information about the geographical location of the plant material used to sequence the plant genome. For some species, this data is even listed as longitude and latitude coordinates, for example, plant material for the *Zostera marina* genome was sourced from Fårö Island, Sweden (latitude: 59° 55.234' N, longitude: 21° 47.766' E, Olsen *et al.*, 2016). Additionally, the global occurrence and severity of drought has been investigated (Sheffield *et al.*, 2008). With information about the geographical occurrence of drought, plant species could be defined based on their location in a drought prone region. However, there are limitations with this approach, for example, plant material sampled from a botanic garden or grown in a laboratory which are outside a plant's natural geographical range. Additionally, drought can be experienced at a highly localised level which may not be captured by broad-scale geographic data.

4.6.3 Potential impact for gene identification

Despite the caveats given above, it was demonstrated, in this body of work, that drought adapted plants are present across the plant phylogeny. This has highlighted how plant relationships with water have changed over the last billion years. Defining drought tolerance, understanding its evolution and the loss of this trait in certain lineages has important consequences for our mechanistic understanding of drought responses and could have implications for our ability to manipulate crop productivity. Drought tolerance is repeatedly referred to as one of the major constraints limiting crop production (Ramegowda *et al.*, 2017; Yu *et al.*, 2017; Roca Paixão *et al.*, 2019; Bao *et al.*, 2020). It therefore threatens global food security which has become increasingly important given the severe effects of global climate change accompanied by human population growth (Godfray *et al.*, 2010, 2014).

Despite the predicted increase in the frequency of droughts, there are still relatively few drought tolerant crops. Stress tolerant crops will be crucial for sustainably feeding future populations (Godfray *et al.*, 2010; Garnett *et al.*, 2013). Novel approaches that can be used to identify candidate genes could be valuable for improving drought tolerance in crops (Tuberosa *et al.*, 2006; Umezawa *et al.*, 2006; Cattivelli *et al.*, 2008; Ashraf, 2010). For example, crop wild relatives are considered to be a pool of genetic resources for engineering stress tolerant crops (Iseki *et al.*, 2018). With this in mind, examining the distribution of drought responses across the plant phylogeny may shed light on shared genes and their potential functions in drought tolerance. By exploring the genetic framework underlying these traits in the context of evolution, the genes and the associated changes in gene sequences responsible for diverse adaptations may be illuminated. Patterns of gene gain, loss, diversification or contraction can be indicative of gene function. For example, an analysis of transcriptome and genome data from across green plants found that the LCA of Streptophyta gained genes for fungal symbiosis before the transition to land. This algal ancestor was preadapted for this beneficial interaction, potentially aiding the colonisation of land. Subsequent genes emerged in the ancestor of land plants

enabling closer relationships with arbuscular mycorrhizae, which further promoted the diversification of land plants (Delaux *et al.*, 2015).

An example of gene loss being indicative of gene function comes from the analysis of 37 flowering plant genomes, of which, some species exhibit nitrogen-fixing root nodule symbiosis (Griesmann *et al.*, 2018). Predominantly found in legumes, this symbiosis enables plants to benefit from nitrogen produced by bacteria hosted in root nodules. Analysis found that the loss of a key symbiotic regulator gene *NODULE INCEPTION* could explain the loss of nitrogen-fixing root nodule symbiosis. This information suggests a single origin of this symbiosis in the ancestor of Fabales, Rosales, Cucurbitales and Fagales followed by multiple independent losses (Griesmann *et al.*, 2018).

Equally, lineage specific gene group expansion and contraction could relate to novel drought adaptations. Some gene groups, in reference to a plant lineage, can radiate more or less readily and this can act as an indicator of the biological adaptation they relate to (Brockington *et al.*, 2015; Yang *et al.*, 2015). For example, genome analysis of the orchid species, *Apostasia shenzhenica*, *Phalaenopsis equestris* and *Dendrobium catenatum*, found that distinct diversification patterns of MADS-box genes (MINICHROMOSOME MAINTENANCE FACTOR 1/ AGAMOUS/ DEFICIENS/ SERUM RESPONSE FACTOR) are responsible for orchid flower developmental evolution (Zhang *et al.*, 2016, 2017a). In *P. equestris*, B-AP3 Class and E Class MADS-box gene families have expanded leading to the development of a specialised labellum, the part of an orchid flower that attracts pollinating insects. In *A. shenzhenica*, AGL12 (AGAMOUS-LIKE 12) and ANR1 (ARABIDOPSIS NITRATE REGULATED 1) gene families have also expanded enabling adaptations to terrestrial habitats in this orchid; these genes have been lost in *P. equestris* which is an epiphytic orchid.

Therefore, as exemplified above, the origin and diversification of genes in relation to adaptations of interest can be investigated, in this case, to predict potential drought tolerance genes. By comparing gene content of the genomes analysed in the Chapter 2 in relation to the drought adaptations defined in this body of work, these patterns of marked

gene loss and gene family contraction can be elucidated and, as such, will be further investigated in Chapter 5.

Chapter 5 Identifying and characterising novel drought tolerance genes

5.1 Abstract

Feeding the world sustainably is becoming a major global challenge, particularly with the forecasted increase in world population growth and climate change. The development of stress tolerant crops will be crucial to enable us to attain this increase in yield. To achieve this, novel techniques need to be developed, for example those that facilitate the identification and characterisation of candidate genes that confer greater ability to tolerate abiotic stressors, including drought stress. Detailed here is a novel evolutionary approach to identify candidate drought tolerance genes that may function across a clade. Gene expression analysis provides preliminary evidence of the role of these identified candidate genes for plant adaptations to drought. Additional work is also described which would allow for the full characterisation of these candidate genes. It is hoped that the application of this novel gene identification approach could allow for the development of stress tolerant plants and would therefore contribute to future-proofing global food demands.

5.2 Introduction

Food security has become an increasingly important issue on the global agenda, as it is estimated that, by 2050, the world population will have risen to between 9 and 9.7 billion people (Godfray *et al.*, 2010). Feeding an additional 2 - 2.7 billion people in less than 30 years, but in a sustainable manner, represents an enormous challenge (United Nations DESA., 2015). Due to this pressure, there is an increasing emphasis on developing methods that enable the sustainable intensification of agriculture.

Drought is one of the major abiotic stressors which adversely affects crop plants, limiting their growth and yield potential. Developing crops that have a greater tolerance of such stressors is likely to play an important role in producing higher yields from the same area of land (Godfray *et al.*, 2014). With an emphasis on sustainably intensifying agriculture, identifying the natural variation associated with plant adaptations to drought has become a major focus for improving the food security of future generations (Garnett *et al.*, 2013).

A common method of improving crop adaptations to water shortages is to study the genetic underpinnings of drought tolerance.

Increasingly, novel approaches are needed to identify drought tolerance genes. Very few studies have used phylogenetic approaches which investigate the evolution of drought tolerance across a broad range of taxa. However, with the advent of modern sequencing technologies, genetic and genomic data are becoming increasingly available for a wide range of plant species, allowing for powerful insights into plant diversification (Li, 2018). By unravelling the evolutionary history of plants, the genomic consequences of plant adaptation to differing environmental conditions can be examined.

One way to understand the evolution of adaptations to stressors such as drought is to examine the origin and diversification of genes associated with these adaptations (Nagy *et al.*, 2020). In Chapter 2 of this thesis, the origin of land plants was identified to have been accompanied by the emergence of a large number of novel genes, more than have arisen at any other point in the history of the plant kingdom. In Chapter 3, distinct patterns of gene novelty and diversification were identified which were associated with specific anatomical innovations. In Chapter 4, drought adaptations were mapped onto the plant phylogeny to understand the emergence and evolution of this trait. By combining the rationale behind these three bodies of work, it may be possible to detect signatures of genome evolution related to the occupancy of traits, such as drought tolerance, in particular lineages.

Carrying this work forward, patterns of gene retention in relation to lineage specific adaptations could be indicative of gene function. For example, a recent comparative analysis of 72 streptophyte genomes, which focussed on the evolution of metabolic pathways, found that genes essential for the synthesis of selenocysteine were present in streptophyte algae but absent in land plants. Previously, selenocysteine, which is required for optimal growth, has been identified in bacteria, mammals and green algae but never in land plants (Novoselov *et al.*, 2002). Therefore the results from comparative genome

analysis confirm the metabolic loss of selenocysteine biosynthesis during land plant evolution (Cannell *et al.*, 2020).

A further example of patterns of gene loss in relation to lineage specific adaptations comes from an analysis of Alismatales genomes, an order of predominantly aquatic flowering plant species (Wang *et al.*, 2014d; Lee *et al.*, 2016a; Olsen *et al.*, 2016). Due to its aquatic lifestyle, the *Spirodela polyrhiza* (common duckweed) genome is characterised by the loss of genes that are associated with water transport as well as those involved in cell wall organization which is consistent with its specialized morphology (Wang *et al.*, 2014d). Similarly, an analysis of the genomes of two species of seagrass, *Zostera marina* and *Z. muelleri*, revealed major losses of genes related to UV light resistance, stomatal differentiation, volatile production (for airborne communication) and ethylene biosynthesis, consistent with its marine lifestyle (Lee *et al.*, 2016a; Olsen *et al.*, 2016).

In addition to gene losses, lineage specific gene group expansion and contraction could equally be related to novel drought adaptations. Some gene groups, in reference to a plant lineage, can radiate more readily and this can serve as an indicator of the biological adaptation they relate to (Brockington *et al.*, 2015; Yang *et al.*, 2015). For example, an analysis of the genomes of the Venus flytrap and other carnivorous plants found that the expansion of specific gene groups enabled the evolution of particular hunting strategies e.g. snap traps, pitfall traps, flypaper traps. The analysis also identified large scale gene loss and gene group contractions in relation to root development and nutrient acquisition (Palfalvi *et al.*, 2020).

In the previous chapter, it was identified that the last common ancestor of land plants was desiccation tolerant (tolerant of extreme water shortages) and the last common ancestor of vascular plants was likely to be drought tolerant (tolerant of intermittent water shortages). With this ancestor being drought tolerant, any cases of drought sensitivity therefore represent a loss of drought tolerance. These cases of drought tolerance loss are

likely to have been accompanied by marked patterns of gene loss and gene family contraction related to drought tolerance.

By comparing gene content in relation to the drought adaptations defined in Chapter 4, these patterns of marked gene loss and gene family contraction can be elucidated. Tackling this challenge from an evolutionary genomics perspective is a novel approach and aims to reveal undiscovered genes that aid plant adaptations to drought. The motivations for this research are twofold; first, to address fundamental questions about how the diversity of plant life arose and, secondly, to apply this evolutionary thinking to produce transgenic plants with a greater tolerance of drought. Applying this comparative genomic approach could allow the development of high yielding, sustainable crop varieties with greater water use efficiency. Additionally, applying this approach to other abiotic stressors and lineage specific adaptations could reveal genes that are relevant to other biological adaptations, such as salt tolerance and nitrogen fixation.

The work below describes a novel method to identify a list of candidate genes that are thought to be involved in drought tolerance. Subsequently, experimental analysis is detailed that aims to characterise the function of a selection of these candidate drought tolerance genes. The model organism *Arabidopsis thaliana* and a wild extremophile relative, *Thellungiella parvula,* were chosen for experimental analysis of putative candidate drought tolerant genes. This is based on the evidence, as shown in Chapter 4, that *A. thaliana* is a drought sensitive plant whilst *Thellungiella parvula* is a drought adapted plant. Several recent studies have used overexpression and gene knockout experiments to highlight the function of genes involved in plant drought tolerance (Ramegowda *et al.*, 2017; Yu *et al.*, 2017; Roca Paixão *et al.*, 2019; Bao *et al.*, 2020). Therefore, the proposed experimental design aimed to knock out candidate genes from *T. parvula* and overexpress them in *A. thaliana* to investigate the impacts of gain and loss of candidate genes for drought responses.

5.3 Methods

5.3.1 Methods overview

The methods detailed below described the evolutionary approach to identifying and characterising novel drought tolerance genes. A summary of this approach is provided below (Figure 5.1), which begins by analysing plant genome data in relation to drought adaptations, leading to experimental analysis that attempts to validate computational findings.

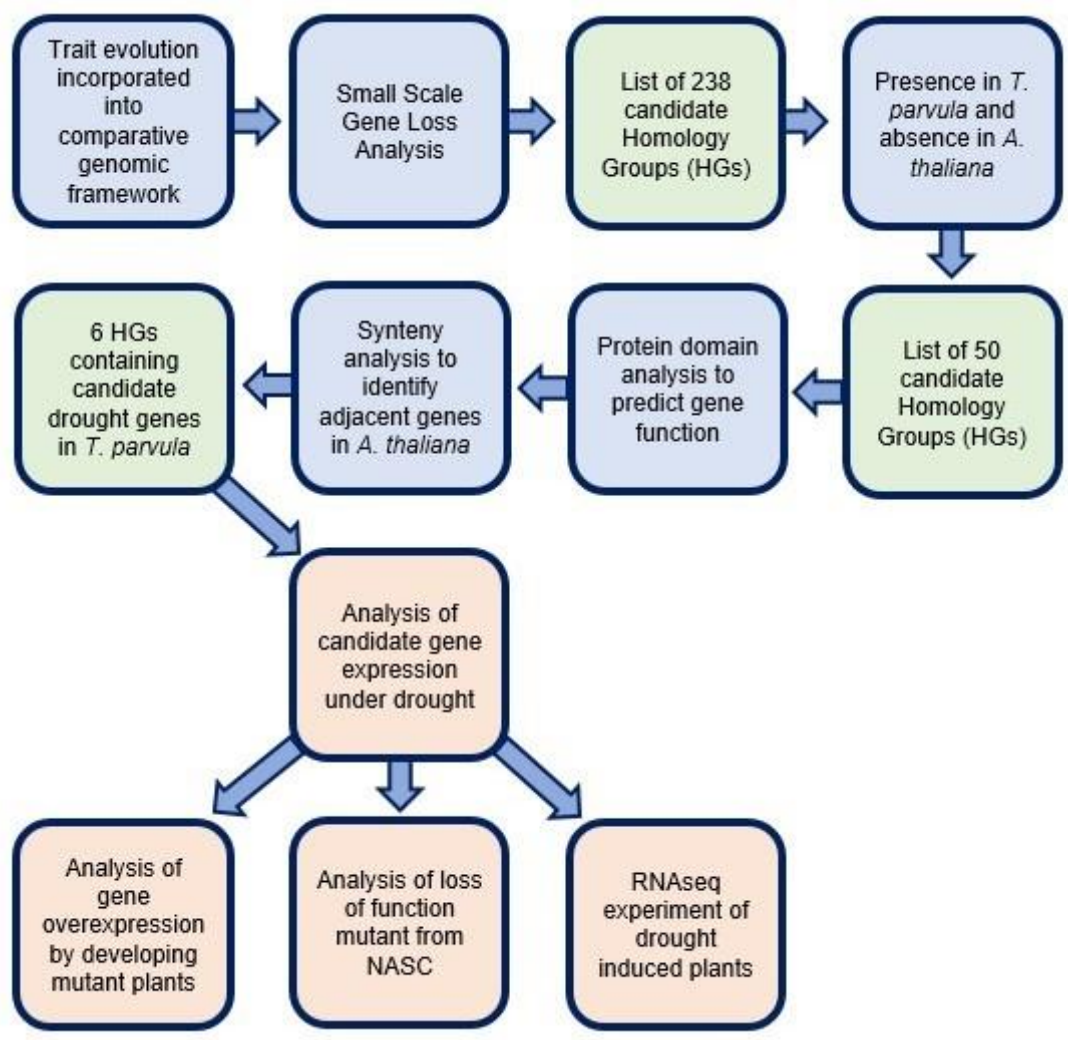


Figure 5.1 Overview of methods used in this chapter. Blue squares highlight computational analysis, orange squares highlight experimental analysis and green squares highlight outputs from computational analysis.

5.3.2 Identifying novel drought tolerance genes

The drought adaptation status for each plant within the genomic dataset described in Chapter 4 (containing 178 plant genomes) were incorporated into scripts of the computational pipeline. In Chapter 2, the pipeline was queried with an interest in taxonomic occupancy to identify gene novelties that accompanied the origin of land plants and other taxonomic groups (Supplementary Data 5.1). In the same manner, the pipeline was queried with an interest in the occupancy of drought tolerance. Preliminary queries investigated broad scale occupancy of genes across the plant tree of life in relation to drought adaptations. Secondary queries of the pipeline investigated gene absences in drought sensitive species to establish whether cases of drought sensitivity were linked to losses of the same genes. Further to this, the presence of these genes was quantified across the remaining land plant species to confirm that these genes were likely to be important in drought tolerance and were not simply lost in the majority of land plants.

5.3.3 Small Scale Gene Loss

To investigate gene loss within closely related taxa, clades with representatives for drought tolerant and drought sensitive species were identified. Gene loss at the clade level was next investigated by querying the pipeline using gene presence (for drought tolerant species) and gene absence (for drought sensitive species). To identify a list of candidate genes, a series of taxonomic queries of the computational pipeline were conducted, specifying gene absence in drought sensitive species and gene presence in drought tolerant species (Supplementary Data 5.2). As introduced above, these genes needed to be absent in *A. thaliana* (drought sensitive) and present in *T. parvula* (drought adapted). With this in mind, homology groups (HGs) were filtered to ensure absence in *A. thaliana* and presence in *T. parvula*. The outputs of these queries were then used in downstream analyses (Supplementary Data 5.3).

5.3.4 Analysis of protein domains

To predict the potential function of candidate genes and reduce the list of candidate genes to a number that was viable to experimentally validate, protein domains of each HG were analysed based on *T. parvula* genes. Specifically, pfam (Finn *et al.*, 2014) and interproscan (Jones *et al.*, 2014) analyses were conducted to provide insights into any known classification and function of protein domains of the potential drought responsive HGs identified in the small scale gene loss queries.

5.3.5 Synteny analysis of possible DT genes

Due to the taxonomic nature of the queries of the genomic dataset, that is ensuring gene absence in *A. thaliana* and gene presence in *T. parvula*, there are no genes in *A. thaliana* that are homologous to the candidate drought tolerance genes identified in *T. parvula*. As one of these homology groups was identified as a potential retrotransposon (see results) and could therefore potentially control the expression of surrounding genes, synteny analysis was used to identify any blocks of genes shared between *A. thaliana* and *T. parvula* adjacent to the focal genes for all HGs. The Gevo and SynFind function of CoGe (*CoGe: Comparative Genomics*, 2020) and Genomicus plants (Louis *et al.*, 2013, 2015) were used to view the syntenic regions surrounding the focal gene of interest. Outputs were illustrated in Inkscape (The Inkscape Project, 2019).

5.3.6 Primer design

Based on HG taxonomic occupancy, pfam and synteny analysis, a selection of candidate drought tolerance genes and their syntenic counterparts in *A. thaliana* were chosen and their expression was tested via qPCR in the laboratory. Forward and Reverse primers were designed for each gene using Primer 3 (Untergasser *et al.*, 2012), factoring in the likelihood of primer dimers forming during amplification, as well as other secondary structures (Table 5.1). Primers were also designed for two housekeeping genes (PP2AA2 and Actin) to enable gene expression data to be normalised for each plant species.

Table 5.1. List of primer sequences used in qPCR experiments to validate the expression of candidate drought tolerance genes. HG = homology group number. HK = housekeeping genes used for normalising gene expression data.

HG	Gene	Sequence	Annealing temperature
72	Tp4g06740 F	TCGTTCACCTTGTCTGAGCT	58.96
72	Tp4g06740 R	TAAACCGGCCCAAATCCTCC	60.03
72	Tp4g06680 F	AATACAGTTCAGCCCCGTGG	60.04
72	Tp4g06680 R	CTCTGTCCACCCACGTCAAA	59.89
72	Tp4g06700 F	TCTGGAGGAGAAAGGAGGGA	58.91
72	Tp4g06700 R	GCGGTCCTTTGCACACATAA	59.12
72	ARSK1 (AT2G26290) F	CACGGAGGGAACAAAGCCTA	59.68
72	ARSK1 (AT2G26290) R	ACCTGAGCCGCTTCTGTTTT	60.18
72	GPA1 (AT2G26300) F	GCAAGAGTTCGCACAACTGG	60.04
72	GPA1 (AT2G26300) R	ACCCACGTCAAACAATCGGT	60.18
72	Tp7g04180 F	ACGAGTGGCTCATCAAGGTG	60.04
72	Tp7g04180 R	AGGTCTTCTTGCATCGCCTC	60.11
72	Tp7g04210 F	TGGATGTGGAGTTGTGGTGG	59.89
72	Tp7g04210 R	ACCCCAAGTCACACATATCCC	59.44
72	NIP4;2 (AT5G37820) F	TCTCTGGATGTGGAGTAGTGGT	59.96
72	NIP4;2 (AT5G37820) R	CCCCAAGTCACACAGATCCC	60.04
72	Tp7g04200 F	CGGAAAGGAGGGGAGTTAGC	59.82
72	Tp7g04200 R	GCTGATCTTCGTTGGGACCA	60.04
72	OXP1 (AT5G37830) F	GCTAGTAGAGGTCACCACGC	59.9
72	OXP1 (AT5G37830) R	GCAGCTCCTTCCTCCCAAAT	60.03
72	Tp7g04170 F	GGTTGTTCCCCTGGCTTCTA	59.3
72	Tp7g04170 R	AAGCTTCCCTGCCATCTTCC	60.03
72	SOS4 (AT5G37850) F	AATGACGACGCCTCCAGTTC	60.39
72	SOS4 (AT5G37850) R	CCTGAACAGTGTGGGATTGGA	59.93
2909	Tp1g09090 F	GTGGACGGTGTGCTTCTG	58.36
2909	Tp1g09090 R	AACACGAATGCCTTACCCGG	60.68
2909	GLP7 (AT1G10460) F	GACCCGCTCCAAGACTACTG	59.83
2909	GLP7 (AT1G10460) R	GGCTTGTGTCGGATCTTTGC	59.83
5775	Tp2g19280 F	TCCCCTGCCTCTTTTGTTCT	58.85
5775	Tp2g19280 R	TCTTGTGGGCATTCTGGTGG	60.25

5775	AT5G48890 F	TCCCCTGCCTCTTTTGTTCT	58.85
5775	AT5G48890 R	CTCCTTCTTGTGGGCGTTCT	59.96
7522	Tp2g22420 F	TCTTTTGCACCACCAGAGCT	59.82
7522	Tp2g22420 R	CCTCACTGTTCCTCCTTCCAC	60
7522	Tp2g22410 F	TTCGCCAAAGTCGCTAGAGG	60.11
7522	Tp2g22410 R	TGTGACAGAAATCGACGGCT	59.68
7522	Tp2g22430 F	TGGAGTTTCGGAGCAGCTTT	59.89
7522	Tp2g22430 R	GAAAGAGTGAGCACCGTGGA	59.97
7522	CIPK25 (AT5G25110) F	GGGAGGAAAGGACAGATCGC	60.18
7522	CIPK25 (AT5G25110) R	CCGCCGACTTACACAACTCA	60.32
7522	AT5G25100 F	TCCCTCTTGTCTTTGTCGGC	59.97
7522	AT5G25100 R	TGTTGGTTTTCACGGGGTCA	60.03
9215	Tp6g08250 F	GGCAATCTCCCACCGTTGAA	60.61
9215	Tp6g08250 R	GGGTTACAGAAGGACAAACGC	59.47
9215	Tp6g08200 F	TCGTGAACTGCGGTCATTGA	59.97
9215	Tp6g08200 R	TTGCTCTTCAGAAGCCGGTT	59.89
9215	AT4G09340 F	AGTGGCTGGGTTTGAACTGT	59.74
9215	AT4G09340 R	ACTCCAACCCGTCTGTTTCA	59.17
10098	Tp7g15080 F	GGTACTGGGTGGAGTCGAGA	60.32
10098	Tp7g15080 R	GCCTTTTCGTTGTGGATGGG	59.76
10098	Tp7g15100 F	ATACTGCTGGCCACCTGAAC	60.04
10098	Tp7g15100 R	TTGCAACATTTCACCAGGCG	59.97
10098	Tp7g15110 F	GCTGGGAAAGGAGTGAAGCT	59.96
10098	Tp7g15110 R	ACAATCATCATCCTCCCCGC	59.89
10098	RGF6 (AT4G16515) F	AATGGTGGAGAAAGGAGGCG	60.04
10098	RGF6 (AT4G16515) R	TTGTGGATCGGAGGCTTACG	59.83
10098	AT4G16530 F	TGGCGTTGATGGAGATTTGGA	60
10098	AT4G16530 R	TCTTCAAGCCCTCGTACCAA	58.66
HK	AT3G18780 (Actin) F	ACAGCAGAGCGGGAAATTGT	60.25
HK	AT3G18780 (Actin) R	GGTTTCCATCTCCTGCTCGT	59.75
HK	AT3G25800 (PP2AA2) F	ATGCCGATGGTAAGGAGAGC	59.61
HK	AT3G25800 (PP2AA2) R	AACGTCGGTCTTCAAATGCG	59.49
HK	Tp2g15040 (Actin) F	TTCACCACAACAGCAGAACG	58.99
HK	Tp2g15040 (Actin) R	GAGGTCTCCATCTCCTGCTC	58.96
HK	Tp3g16870 (PP2AA2) F	TCCACATTGCATACCCAAGC	58.53

HK	Tp3g16870 (PP2AA2) R	CTCTCCGCACCATAGGCATA	59.03

5.3.7 Drought experiment

To test the expression levels of candidate drought genes (and syntenic genes in *A. thaliana*) between well-watered and drought stressed plants, a drought experiment was conducted. Seeds from *Arabidopsis thaliana Col-0* and *Thellungiella parvula* were sown on compost and kept at 4ºC for 3-4 days to break dormancy. Seeds were then placed in a growth cabinet at room temperature under short day conditions (8h-light, 16h-dark) for 7 days. Seedlings were then pricked out into individual pots. After 4 weeks of growth, *Arabidopsis thaliana Col-0* and *Thellungiella parvula* pots were saturated with water on Day 1 and then subjected to 12 days without watering. Control plants for *Arabidopsis thaliana Col-0* and *Thellungiella parvula* were grown alongside but remained well-watered throughout the experiment. Pots were weighed every two days to calculate the relative water content as a proxy for the severity of drought treatment. For each treatment, 15 plants were analysed.

5.3.8 RNA extraction, cDNA synthesis, qPCR

To test the expression of candidate drought genes, RNA was extracted from plants that had been grown under the drought and well-watered conditions in the experiment detailed above. This was completed using the method below, followed by cDNA synthesis and qPCR.

5.3.8.1 RNA extraction

Leaf material was taken from three samples per treatment for each plant species and placed immediately into liquid nitrogen. RNA was extracted using a Trizol extraction protocol which is described below (Rio *et al.*, 2010). A single leaf (~0.1g of material) was sampled per plant and ground with a pestle and mortar in liquid nitrogen. Samples were transferred into tubes that had been placed on dry ice. 1 ml of TRIzol reagent (Invitrogen) was added to the ground sample and vortexed for 30 seconds. This step allows for homogenisation of plant tissue without degradation of RNA. Samples were allowed to

stand for 3 minutes at room temperature. 200 µl of chloroform (Sigma Aldrich) was added and then vortexed for 15 seconds. Samples were centrifuged for 15 minutes at 4ºC at 13000 x g to separate the RNA from proteins and lipids. The aqueous phase, containing the RNA, was transferred into new tubes.

1 volume of isopropanol (~600 µl) was added, mixed by inversion and then placed at room temperature for 10 minutes. The isopropanol is used to precipitate the RNA from the solution. Samples were centrifuged for 15 minutes at 4ºC at 13000 x g to separate RNA from the surrounding solution. The supernatant was discarded leaving a white pellet containing the RNA. The pellet was washed with 1 ml of 75% ethanol (Sigma Aldrich) to purify the extraction. The ethanol was then discarded and the pellet air dried for 10 minutes to remove any remaining liquid.

The pellet was dissolved in 26 µl of RNase-free water. 1 µl of DNAse enzyme (Invitrogen) and 3 µl of DNA buffer (Invitrogen) were added and incubated at 37ºC for 1 hour to remove any genomic DNA from the RNA sample. 1.8 µl of Ethylenediaminetetraacetic acid (EDTA) (Invitrogen) was added and incubated at 65ºC for 10 minutes to deactivate the DNase enzyme. 60 µl of ethanol and 15 µl of Ammonium acetate were added to precipitate the RNA. These samples were placed in a -20ºC freezer overnight.

Samples were centrifuged for 15 minutes at 4ºC at 13000 x g to separate RNA from the surrounding solution. The supernatant was discarded leaving a white pellet containing the RNA. The pellet was washed with 1 ml of 75% ethanol. The ethanol was then discarded and the pellet air dried for 10 minutes to remove any remaining liquid. Finally, the RNA pellet was dissolved in 30ul of RNase-free water. RNA quality and quantity was then assessed using a nanodrop spectrophotometer.

5.3.8.2 cDNA synthesis

For cDNA synthesis, 1000 ng of RNA was added to RNA-free water equaling a total volume of 11 µl. 1 µl of random hexamer primers (Invitrogen) was added to the RNA/ water solution. This mix was then placed in a thermocycler at 65ºC for 10 minutes to denature

the sample and primers and then immediately placed on ice. To this reaction, 4 µl of 5X Reverse Transcriptase buffer (Invitrogen), 2 µl of 10mM dNTPs (Invitrogen), 1 µl of Reverse Transcriptase (Invitrogen) and 1 µl of RNA-free water were added to make a total volume of 20 µl. The 20 µl cDNA synthesis reaction was then incubated at 42ºC for 60 minutes in a thermocycler. Following this enzymes were inactivated by increasing the temperature to 70ºC for 5 minutes. Samples were then stored at –20ºC until further use.

5.3.8.3 qPCR protocol

To assess the expression of each candidate drought tolerance gene between well-watered and drought stressed plants, qPCR reactions were set up. A reaction mix containing 10 µl of SYBR Green master mix (Sigma Aldrich), 6 µl of RNA-free water, 0.4 µl of each primer (forward and reverse for each gene of interest or housekeeping genes, Table 5.1) and 0.2 µl of Taq polymerase per sample was created. For each well of a 96 well plate, 3 µl of cDNA and 17 µl of master mix were added to create a total volume of 20 µl. Gene expression was measured in a BioRad qPCR Detection System. The qPCR program consisted of an initial step of 95ºC for 5 minutes to denature the DNA, followed by 40 cycles of 15 seconds at 95ºC, 45 seconds at 60ºC and 30 seconds at 72ºC. These steps further denature the DNA, enable primer annealing and allow for primer extension before repeating. A melt curve analysis was performed after the qPCR run by ramping the temperature by 0.5ºC every 10 seconds from 55ºC to 95ºC. This step was used to confirm the synthesis of a single PCR product, in which case a single peak is observed.

5.3.8.4 Statistics and figures

All statistics were conducted in R version 3.6.3 (R Core Team, 2014). A two-sample t-test was used to test differences in gene expression between drought induced and well-watered plants. Figures were produced using ggplot2 (Wickham, 2016) and tidyr (Henry, 2018).

5.3.9 Mutant design

Based on the results of the qPCR experiments, a subset of syntenic genes were investigated further via loss-of-function *A. thaliana* mutants. To assess the impact of gene loss, *Arabidopsis thaliana* mutants were selected and ordered from the Nottingham Arabidopsis Stock Centre (NASC) (O'Malley *et al.*, 2015). For each syntenic *A. thaliana* gene, two mutants were ordered (Table 5.2). This was completed to ensure that any phenotype seen in the mutant plants was a result of a correct knockdown/knockout for the focal gene, rather than a knockout of any other polymorphisms associated with a mutant line.

Table 5.2. SALK lines selected for investigating the effects of syntenic gene loss on the drought response in *A. thaliana*. HG= homology group number. Line IDs A and B refer to the NASC IDs for the two lines of mutant plant used to investigate each syntenic gene.

Gene name	HG	Function of *A. thaliana* gene	Line ID A	Line ID B
AT2G26300	72	Negative and positive regulation of ABA (Pandey *et al.*, 2004; Chakraborty *et al.*, 2015, 2019; Jangam *et al.*, 2016)	N528135	N561522
AT5G37830	72	Glutathione catabolic process (Ohkama-Ohtsu *et al.*, 2008)/ Glutathione enhances plant abiotic tolerance (Hasanuzzaman *et al.*, 2017)	N590917	N668525
AT5G37850	72	Hypersalinity response/ Root hair development (Shi *et al.*, 2002)	N618343	N638992
AT1G10460	2909	Cold and osmotic stress tolerance (Kumar *et al.*, 2016)	N572453	N533384
AT5G48890	5775	Negative regulation of flower development (Weingartner *et al.*, 2011)/ Enhance salt tolerance in rice (Zhang *et al.*, 2018b)	N818436	N546014
AT5G25100	7522	Protein localisation to membrane (Parsons *et al.*, 2013)	N608055	N529679
AT4G09340	9215	Uncharacterised	N552021	N533044
AT4G16515	10098	Regulation of root growth (Moubayidin *et al.*, 2010; Sozzani *et al.*, 2014; Shinohara *et al.*, 2016)	N633489	N573605
AT4G16530	10098	Uncharacterised	N577075	N593165

5.3.10 Mutant confirmation

Once received from NASC, seeds from all mutant lines were sown on compost and kept at 4^{o}C for 3-4 days to break dormancy. Seeds were then placed in a growth cabinet at room temperature under short day conditions (8h-light, 16h-dark) for 7 days. Seedlings were then pricked out into individual pots. After 4 weeks, one leaf was sampled from each plant of every line for DNA extraction and placed on dry ice. To each sample, a small amount of sand and 200 µl of cetyl trimethylammonium bromide (CTAB) DNA extraction buffer were added to the leaf and ground with a micro pestle until the leaf had dissolved. Samples were then placed at 65^{o}C for 30 minutes to increase DNA yield by promoting the breakdown of cell and nuclear membranes. Solutions were cooled to room temperature before adding 200 µl of chloroform under a fume hood. Samples were vortexed and then centrifuged at 13000 x g for 15 minutes. The aqueous phase was extracted, added to a new tube and the solid phase discarded. 180 µl of isopropanol was added to the supernatant and mixed well by inversion. Samples were again centrifuged at 13000 x g for 15 minutes. The supernatant was then discarded, ensuring the pellet remained. The pellet was then washed with 500 µl of ice cold 70% ethanol. The ethanol was then discarded and the pellet was dried completely before eluting in 50 µl of RNA free water. Extracted DNA was quantified on a nanodrop spectrophotometer and then stored at -20^{o}C for later work.

Successful mutants all contain a tDNA insert with a known DNA sequence (Appendix 5.1). It is possible to isolate the genomic junction of this insert through additionally designing primers for the left border region of the tDNA insert; amplification of the tDNA insert confirms successful knockdown/ knockout of the gene of interest (O'Malley *et al.*, 2015). These primers were designed for each SALK line using http://signal.salk.edu/tdnaprimers.2.html (Table 3). These primers were ordered from Primer3 (Untergasser *et al.*, 2012) as described above. To confirm successful mutation, PCRs were run for every plant line using 3 primer mixes. Firstly, the left border primer

(Table 3) and the Forward primer (Table 1) of the focal gene were used to test insertion. The left border primer (Table 3) and the Reverse primer (Table 1) of the focal gene were also used to test successful insertion. Amplification when the Forward primer and Reverse primer (Table 1) of the focal gene were used together would confirm that the plant was in fact a wild type plant. For each PCR reaction, the following reaction mix was made: 2 µl DNA, 1.25 µl of each primer, 2.5 µl of 10x buffer, 0.5 µl 10mM dNTPs, 0.5 µl Taq polymerase and 17 µl PCR grade water. These reactions were run in a thermocycler with the following conditions: 4 minutes at 94ºC; 35 cycles of 94ºC for 30 seconds, 58ºC for 30 seconds and 72ºC for 60 seconds; 5 minutes at 72ºC and held at 12ºC. PCR products were run on a gel to confirm successful mutant lines (Appendix 5.2). Gels were made using 100 µl of 1x TE buffer, 0.8 g of agarose and 5 µl of SafeView Nucleic Acid Stain (NBS Biologicals). To each well, 15 µl of DNA mixed with 2 µl of loading dye were added alongside 5 µl of 1KB ladder.

Table 5.3. Primers sequences used to amplify the left border of the tDNA insert for the confirmation of SALK lines. HG = homology group number.

HG	Gene	SALK lines	Sequence	Annealing temperature
72	AT2G26300	N528135	GCAAAATCAGAACACCATTGG	60.36
72	AT2G26300	N561522	ACGCAGAAAACATCCTTCATG	60.13
72	AT5G37830	N590917	TGGGTGTCTCCACTCATTCTC	60.1
72	AT5G37830	N668525	AACACTCACAAAACCCATTGC	59.89
72	AT5G37850	N618343	CACTTTTCTTGCAGGGAACAG	59.9
72	AT5G37850	N638992	AGGACGACAACAATGACGATC	59.99
2909	AT1G10460	N572453	GAATTGAACTCGGGACCTCTC	60.07
2909	AT1G10460	N533384	ATGGAGAAATATGCCCCAAAG	60.16
5775	AT5G48890	N818436	TCCTCTGTGTACCACAAACCC	59.88

5775	AT5G48890	N546014	ATAAAATCATGTTTTCCCGCC	60.04
7522	AT5G25100	N608055	GACTTCGGGGTTTTCACTTTC	59.97
7522	AT5G25100	N539673	ACCAGAGGCGATTGAATCTTC	60.59
7522	AT5G25100	N529679	GAAAGTGAAAACCCCGAAGTC	59.97
9215	AT4G09340	N552021	TTTGCTGAGAAGCACTACGAAG	59.83
9215	AT4G09340	N533044	TCGATGCACATTTTGACTACG	59.74
10098	AT4G16515	N633489	ACCATTCATTGTTCCTTGCAG	59.99
10098	AT4G16515	N573605	CGCTCATGTCTTCTGTGTACG	59.52
10098	AT4G16530	N577075	AAACGCAAATCCCAAATTTTC	60.17
10098	AT4G16530	N593165	TGATTGGCTCAACCTTAATGC	60.09

5.3.11 Gene overexpression

5.3.11.1 Clone design and synthesis

Based on the results of the qPCR analysis (see 5.3.7-8) four HGs were chosen for gene overexpression analysis. For this, individual constructs were designed using the coding sequence of each of the focal genes for *T. parvula* or the corresponding syntenic gene of *A. thaliana* was extracted (Appendix 5.3). To the end of each sequences, attL1 and attL3 sequences were added. These are two flanking recombination sequences used to develop a gateway entry clone. Briefly, the gateway cloning approach allows the user to quickly insert the gene of interest into a destination vector using an LR (attL/attR) reaction (Chin *et al.*, 2015). This reaction takes place between attL sites of the entry clone and the attR sites of a destination vector and leads to the creation of an expression clone for downstream work. The constructs were synthesised and put into pUC57 by NBS biologicals. Upon arrival, entry clones were rehydrated in 50ul of RNA free water and stored at -20^{o}C.

5.3.11.2 Making stocks of entry plasmid

Entry clones were stocked via transformation into *Escherichia coli*. To make competent *E. coli* cells, a saturated overnight culture of TOP10 *E. coli* cells was inoculated (20 µl) into

20 ml of Lysogeny Broth (LB: Tryptone 10 g L^{-1}, NaCl 10 g L^{-1}, Yeast extract 5 g L^{-1}). This sub-culture was then grown, shaking, at 37ºC until the optical density (OD) was 0.4; OD was measured with using an OD600 DiluPhotometer. The culture was then placed on ice for 10 minutes and from here on, everything that came into contact with the competent cells was kept ice cold. The culture was separated into two 50 ml falcon tubes and centrifuged at 2700 x g for 10 minutes at 4ºC. The supernatant was removed and the cell pellet was resuspended in 100 µl of ice-cold 100 mM calcium chloride. The falcon tubes were then incubated on ice for 20 minutes. The cells were combined and 0.5 ml of ice cold 80% glycerol was added to prevent damage to the cells during freezing. Cells were then divided into 50 µl aliquots on dry ice and stored at -80ºC until use in transformation reactions.

For the transformation of each entry plasmid into the *E. coli* cells, 50 µl of the TOP10 competent cells were thawed on ice. 2 µl of DNA clone was added to the cells and then left on ice for 30 minutes. Cells were then heat shocked by placing them at 42ºC for 60 seconds to allow for the incorporation of the entry clone. Cells were immediately placed on ice for 5 minutes and then 900 µl of LB broth was added. They were then allowed to recover in a shaker at 37ºC for 1 hour. Following this, 100 µl of the cell solution was spread onto a petri dish consisting of LB agar, containing carbenicillin (10 µg μl^{-1}) which was used for antibiotic selection. The designed entry clones are resistant to ampicillin but carbenicillin was used as it is an analogous antibiotic with lower toxicity and longer half-life. Plates were grown overnight at 37ºC. A single colony was then picked off and grown in 20 ml of LB broth with 10 µg μl^{-1} carbenicillin overnight, shaking, at 37ºC. Following this, glycerol was added to the culture in a 50:50 (v/v) ratio and stored at -80ºC.

5.3.11.3 LR reaction

An LR reaction is the process of recombination between the attL sites of the entry clone and the attR sites of the expression clone to create the destination vector, as explained above. To complete this reaction, between 1 – 7 µl, equating to 50 – 150 ng of entry clone

DNA, was added to 2 µl of the destination vector (35 ng $^{µl-1}$) in a PCR tube. The destination vector was the Gateway Binary Vector (pGWB2) which contains a 35S promoter used for gene overexpression. This mixture was then made up to 8 µl with TE buffer and then 2 µl of Clonase II Enzyme Mix (Invitrogen) was added. PCR tubes were placed in a thermocycler at 25^{o}C for 60 minutes. After an hour, 1 µl of proteinase K (Invitrogen) was added and tubes were placed at 37^{o}C for 10 minutes to terminate the reaction. The resulting expression vector was then transformed into TOP10 E. coli as detailed above (section 5.3.11.2). Additionally, pENTR-gus is an entry vector used as a positive control to confirm the viability of *E. coli* competent cells. The pENTR-gus plasmid were also transformed into *E. coli.* Successfully transformed cells were then confirmed by growth overnight on LB agar plates containing kanamycin (10 ug µl^{-1}), the antibiotic resistance gene present in the destination vector. A single colony was then picked off the agar and placed in 20 ml of LB broth. This was incubated overnight at 37^{o}C whilst shaken constantly. 50% glycerol was added to create stocks (as above) which were stored at -80^{o}C for later use.

5.3.11.4 Agrobacterium transformation

The expression vector was prepared for transformation into *Agrobacterium tumefaciens* by inoculating TOP10 *E. coli* containing the completed expression vector (described above) into 20 ml of LB broth containing 10 µg µl^{-1} kanamycin. These were grown overnight, shaking, at 37^{o}C. Plasmid DNA was then purified from the overnight cultures using the GeneJet Plasmid Miniprep kit. Briefly, 2 ml of overnight culture was centrifuged at 13000 x g for 5 minutes to retrieve a pellet. Supernatant was removed and the pellet was resuspended in 250 µl of P1 buffer. Following this, 250 µl of lysis buffer was added and the solution was mixed by inversion. Next, 350 µl of Neutralisation buffer was then added and the solution was mixed by inversion, before centrifuging for 5 minutes at 12000 x g. 700 µl of the supernatant was then transferred to a spin column and centrifuged for 1 minute at 9000 x g. The resulting flow-through was discarded and 350 µl of wash buffer

was added, followed by another centrifugation step for 1 minute at 12000 x g. The resulting flow-through was again discarded and the spin column centrifuged again for 2 minutes to remove residual buffer. Plasmid DNA was then eluted from the column by adding 30 µl of RNA free water and centrifuging for 1 minute at 12000 x g. DNA concentration was assessed using a nanodrop spectrophotometer.

To transform the resulting expression plasmid into *Agrobacterium tumefaciens,* 2 µl of purified plasmid DNA was added to 50 µl of electrocompotent *A. tumefaciens* cells. Electroporation was then used to transform the *A. tumefaciens* cells. Following electroporation, 1 ml of LB broth was added and cells were recovered at 28^{o}C for 1 hour. These cells were then streaked onto agar plates containing kanamycin (10 µg µl^{-1}) to select successfully transformed agrobacterium. To stock the resulting cells, a single colony was picked off the agar and placed into 20 ml of LB broth. This was incubated for 3 days at 28^{o}C whilst shaken constantly. A 50% glycerol solution was then added (50:50 v/v ratio of glycerol to culture) to create stocks which were stored at -80^{o}C for later use.

5.3.11.5 Floral dip

5.3.11.5.1 Preparation of agrobacterium

An established protocol was used for *Arabidopsis thaliana* transformation as has previously been described (Clough *et al.*, 1998; Narusaka *et al.*, 2010). Briefly, transformed *Agrobacterium* cells were grown overnight at 28^{o}C in 2 ml of LB broth containing kanamycin (10 µg µl^{-1}) for selection purposes. These *Agrobacterium* cells were then pelleted via centrifugation and the supernatant was discarded. 1 ml of sucrose solution, which contains 5% sucrose and 95% RNA free water, was then added to the pellet. Before inoculation, 4 µl of Silwet L-77 was added to the *Agrobacterium*/sucrose solution. Silwet L-77 is a wetting agent that reduces the surface tension of the sucrose solution, improving the efficiency of transformation.

5.3.11.5.2 Arabidopsis plants

A. thaliana seeds were sown on compost and kept at 4ºC for 3-4 days to break dormancy. Seeds were then placed in a growth cabinet at room temperature under short day conditions (8h-light, 16h-dark) for 7 days after which seedlings were pricked out into individual pots. After 4 weeks, initial bolts were clipped to promote secondary bolts. 4-6 days after this, plants were ready for their first floral dip. For this, 5 µl of the sucrose/ *Agrobacterium* solution prepared in the steps above, was added to each flower bud. In total, approximately 50-100 µl of *Agrobacterium* inoculum was added to each plant. Plants were then placed under covers for 24 hours to maintain high humidity which improves the efficiency of transformation. To increase the rate of transformation, inoculation with *Agrobacterium* was repeated twice more at 7 day intervals. Plants were grown until seeds developed at which point watering was halted and, once the plants were dry, seeds were harvested.

5.4 Results

5.4.1 Broad scale evolutionary patterns of drought gene loss

Initial queries investigated the commonality and prevalence of gene losses across all drought sensitive species and the presence of these same genes in all drought tolerant species within a homology group. This first search revealed that no HGs matched this criterion suggesting there were no common patterns of loss across drought sensitive species within HGs. This is not surprising given the complexity of evolutionary relationships between plants and confirms the independent loss of drought tolerance. To understand the occupancy of HGs in relation to drought statuses, further queries were therefore conducted. Secondary queries quantified the presence of genes within homology groups across increasing numbers of land plant species, given their absence in all drought sensitive species. This was to confirm whether common genes were lost across drought sensitive plant species, which were likely to be important for drought tolerance, and to confirm these were not simply lost in the majority of land plants. Considering 178 land plant species were included in the genomic dataset, the maximum number of species

meeting this criterion was 16. This revealed that there are no broad scale evolutionary patterns of the loss of drought tolerance genes. Taxa matching this query are species of the genus *Oryza* that have good representation in the dataset suggesting this result may be an artefact of their close phylogenetic relationships rather than an indicator of drought tolerance.

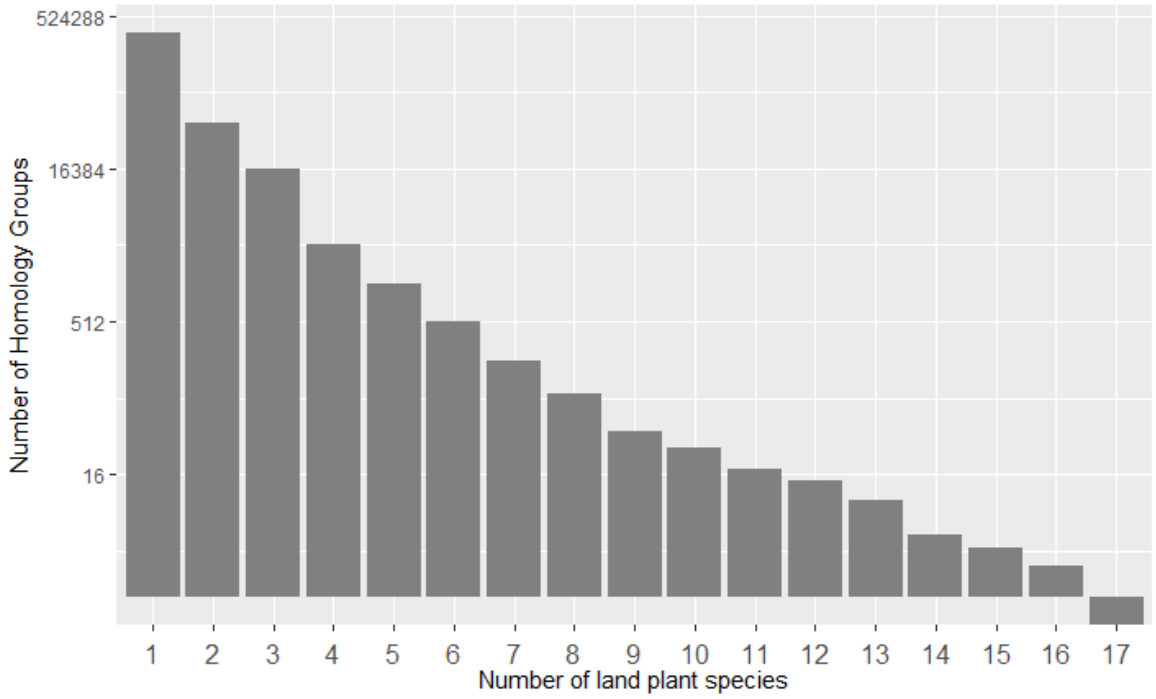


Figure 5.2. There is no kingdom wide patterns of drought gene loss. The graph shows the number of homology groups resulting from a query searching for absences in drought sensitive species and presence in increasing numbers of land plant species.

This data led to the conclusion that to identify candidate drought genes, individual cases of the loss of drought tolerance would need to be investigated. As there is no kingdom wide pattern of gene loss, then these differences in phenotypic response can be attributed to patterns of small scale gene loss (e.g. clade specific losses).

5.4.2 Small scale gene loss

Investigation of gene loss within specific clades of drought tolerant and drought sensitive species identified 238 HGs in total, however the number of HGs that were present in drought tolerant species but absent in drought sensitive species varied by clade (Figure 5.2). This differential retention between drought tolerant and drought sensitive species across the plant phylogeny is potentially indicative of gene function (Supplementary Data 5.2). Four of these queries focussed on gene occupancy in non-flowering plants, predominantly eliminating the lycophyte *Selaginella moellendorffii* and the gymnosperm *Gnetum monatum* (Figure 5.3). Several other queries investigated gene loss in the monocots, ensuring absence in the aquatic species in Alismatales and cultivated rice species, *Oryza sativa*. The remaining queries investigated gene loss in drought sensitive eudicot species. As detailed above, for experimental purposes, these HGs need to be present in *Thellungiella parvula* (a drought tolerant species) but absent in *Arabidopsis thaliana* (a drought sensitive species). Therefore, HGs were further selected based on these taxonomic occupancy criteria which led 50 HGs being chosen for downstream analysis (Supplementary Data 5.3).

5.4.3 Protein domain analysis of possible DT genes

To select a practical number of genes to experimentally test for their role in drought tolerance, protein domains were analysed by comparing protein sequences of genes from *T. parvula* against the pfam database for the selected 50 HGs (Supplementary Data 5.4). By analysing the protein domains, the predicted function of sequences could be identified. With this information, the differences in drought tolerance between the model organism *A. thaliana* and the wild relative *T. parvula* could be explored. Specifically, the underlying genes that contribute to this difference in drought phenotypes between the two species could be investigated, for example the role of retrotransposons which may influence the expression of drought tolerance genes.

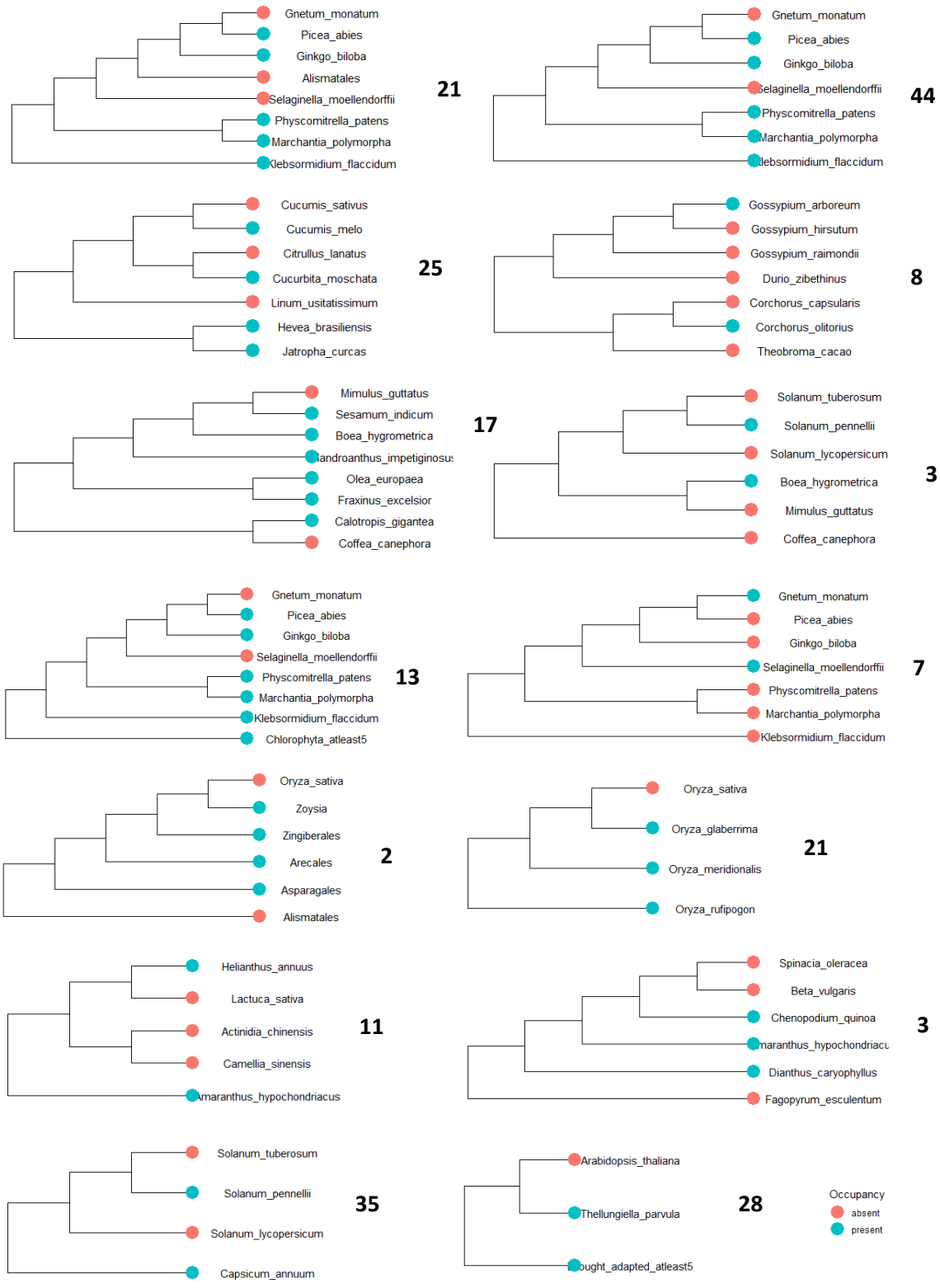


Figure 5.3. Query terms used in small scale gene loss searches of the genomic pipeline. Red dots at terminal branches denote gene absences in drought sensitive species. Blue dots at terminal branches denotes gene presences in drought adapted species. The number of homology groups (HGs) identified for each query are shown.

The protein sequences of many HGs were identified as proteins of unknown function so these were automatically discarded (Supplementary Data 5.4). Based on the protein domains identified by pfam analysis and their reported function in the literature, 6 HGs were chosen for the experimental phase (72, 2909, 5775, 7522, 9215, 10098) (Table 5.4).

Table 5.4. Protein domains predicted from pfam analysis of predicted drought response HGs based on *T. parvula* genes.

HG	Protein Domains	Role in drought tolerance
72	• Gag-polyprotein putative aspartyl protease • Reverse transcriptase (RNA-dependent DNA polymerase) • Retrotransposon gag protein /aspartyl protease • His(2)-Cys(2) zinc finger • Chromo (CHRromatin Organisation Modifier) domain • Integrase core domain	Aspartic proteases known to be involved in plant stress responses, for example to water deprivation (Simões *et al.*, 2004; Vicient *et al.*, 2020)
2909	• Cupin • S25 ribosomal protein	Cupins have roles in plant development and defense responses inc. biotic and abiotic defense (water deprivation) (Wang *et al.*, 2014f)
5775	• C2H2-type zinc finger • Zinc-finger double-stranded RNA-binding	C2H2 is involved in responses to abiotic and biotic stress (e.g. salt, drought, osmotic and oxidative stress) (Kiełbowicz-Matuk, 2012; Yuan *et al.*, 2018)
7522	• PLD-like domain • Enoyl-CoA hydratase/isomerase	Involved in Plant Growth, Development and Stress Responses (Wang, 2005)

9215	• Probable lipid transfer	Possible role in drought stress – no examples of a successful mutants (Salminen *et al.*, 2016)
10098	• HIT zinc finger • Zinc knuckle • Domain of unknown function (DUF4535)	Drought and salt stress tolerance (Li *et al.*, 2014d)

5.4.4 Analysis for genes within the six HGs

The *T. parvula* genes from the six HGs identified in the query above (HG 72, 2909, 5775, 7522, 9215, 10098) were chosen for further analysis. The following computational and experimental analyses apply to each Homology Group. As such, the results for these analyses are reported by Homology Group. Although homologs of candidate genes are not present in *A. thaliana*, blocks of genes may still have been preserved between species. The physical co-location of genes in genomes of different species is termed synteny (Tang *et al.*, 2008). This feature enables comparisons between the genomes of different species even if particular genes are absent across the 6 selected HGs.

5.4.4.1 HG_72

5.4.4.1.1 Synteny analysis and the function of syntenic genes

Protein domain analysis of HG 72 identified sequences containing retrotransposon and retroviral protein domains as well as reverse transcriptase domains (Table 5.4). Focussing on the latter of these domains, reverse transcriptases enable the copying of RNA into DNA which can then become integrated into eukaryotic genomes. Whole genome sequencing has revealed that a large proportion of eukaryotic genomes consist of reverse transcriptase genes, more than any other protein coding gene (Orozco-Arias *et al.*, 2019).

Genetic elements containing reverse transcriptase genes are termed retrotransposons. They are able to use the reverse transcriptase to move from location to location in the genome through an RNA intermediate (Finnegan, 2012). Transposable element

replication also includes the activity of integrases, which facilitate the insertion of retrotransposon DNA into the host genome, as well as aspartic proteases which process large transposon transcripts enabling their conversion into protein products. In addition to this, chromodomain enables the targeted integration of retrotransposon DNA into the genome (Orozco-Arias *et al.*, 2019). All of these domains were identified in protein domain analysis of HG 72 (Table 5.4).

Initially, it was thought that transposable elements only had negative impacts on the host genome (Kim, 2017). However recently, they have been shown to play key roles in chromosome organisation (Vicient *et al.*, 2017), genome size variations (Li *et al.*, 2017c) and genome stability after polyploidy events (Parisod *et al.*, 2010). Additionally, retrotransposons are able to influence genomic regulation whereby they can have effects on the expression of neighbouring genes (Elbarbary *et al.*, 2016; Mita *et al.*, 2016). Additional to the protein domain analysis of this HG, 42 copies of the same gene (with high sequence similarity) were identified in HG 72. With the hypothesis that these 42 sequence copies in *T. parvula* may act as a regulator to enhance the expression of drought and abiotic stress tolerance genes, the functions of syntenic genes in *A. thaliana* were investigated (Figure 5.4).

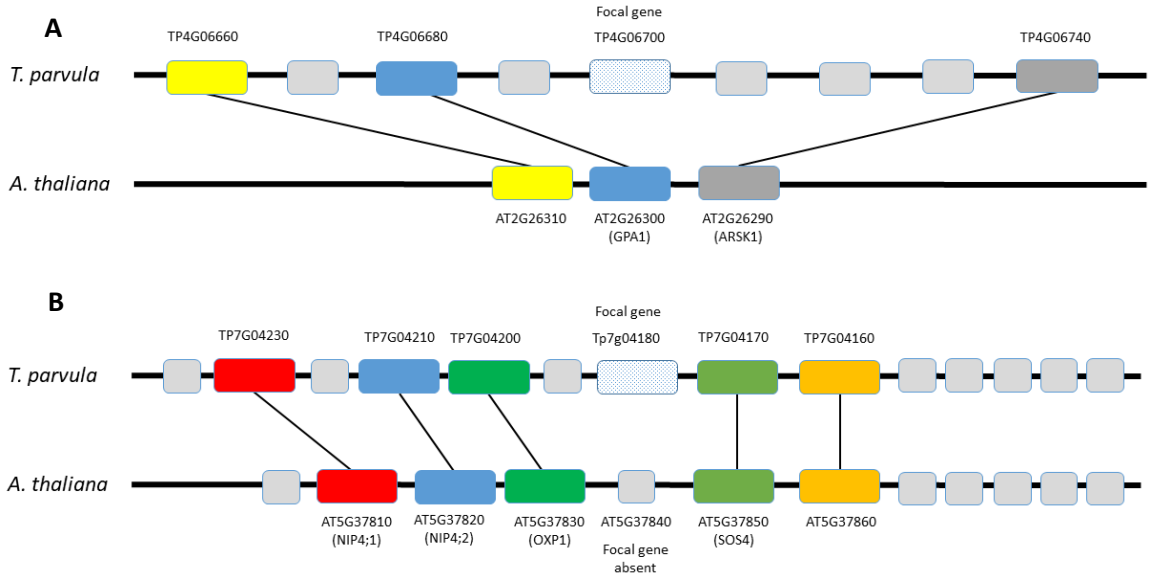


Figure 5.4. Synteny plots for the genes A) Tp4g06700 and B) Tp7g04180 in HG 72. Plots are based on outputs from Genomicus Plants. Genes that are linked with a line and that have matching coloration denote syntenic, homologous genes.

Two of the 42 representative *T. parvula* genes (Tp4g06700, Tp7g04180) were chosen for analysis based on the proposed function of adjacent genes for plant drought tolerance (Supplementary Data 5.5, Figure 5.4). In addition to being syntenic, these genes are homologous between *T. parvula* and *A. thaliana,* suggesting a conserved function. Close to the first focal gene in HG 72 (Tp4g06700) was Tp4g06680 which is syntenic to GPA1 (At2g26300) (Figure 5.4a). The GPA1 gene encodes a G Protein ALPHA Subunit 1, which is known to be involved in many important drought related functions. These include regulation of blue light signalling pathways, cell death, stomatal movement and seed germination (Warpeha *et al.*, 2007; Jeon *et al.*, 2019). GPA1 also acts during abscisic acid responses to guard cell opening (Jin *et al.*, 2013) and regulates transpiration efficiency and stomatal density by controlling epidermal cell size during stomatal formation (Nilson *et al.*, 2010).

Close to the second focal gene, Tp7g04180, in the *T. parvula* genome is the gene Tp7g04170, which is syntenic to SOS4 in *A. thaliana* (At5g37850) (Figure 5.4b). The

SOS4 gene is known to be involved in the plant salt stress response as well as root hair development (Shi *et al.*, 2002). In addition to this, the gene Tp7g04210 is also near to the candidate drought gene in the *T. parvula* genome; this is syntenic to the NIP4:2 gene in *A. thaliana* (At5g37820) (Figure 5.4b). NIP4:2 is an aquaporin channel protein, mainly involved in enabling the transport of water across the plant cell membrane (Di Giorgio *et al.*, 2016).

5.4.4.1.2 Gene expression of possible drought tolerance genes

Given the identified genes in *T. parvula* and the syntentic genes in *A. thaliana* could feasibly play a role in plant drought responses, gene expression analysis was then conducted to assess whether the genes were differentially expressed under watered versus drought conditions. Typically, *A. thaliana* genes had a higher fold change of gene expression between well-watered and drought conditions (Figure 5.5). However, the only significantly differentially expressed gene ($P<0.05$ in a two-sample t-test) is the adjacent gene Tp4g06680 which had a higher expression under watered conditions (Figure 5.5). This appears to be the opposite behaviour of the syntenic *A. thaliana* gene, GPA1 (At2g26300) which had higher expression under drought conditions, although this was not significant ($P>0.05$) (Figure 5.5). Unfortunately, not all primers were successfully confirmed due to laboratory closures and therefore expression analysis was not conducted for all syntenic genes (see Planned further work).

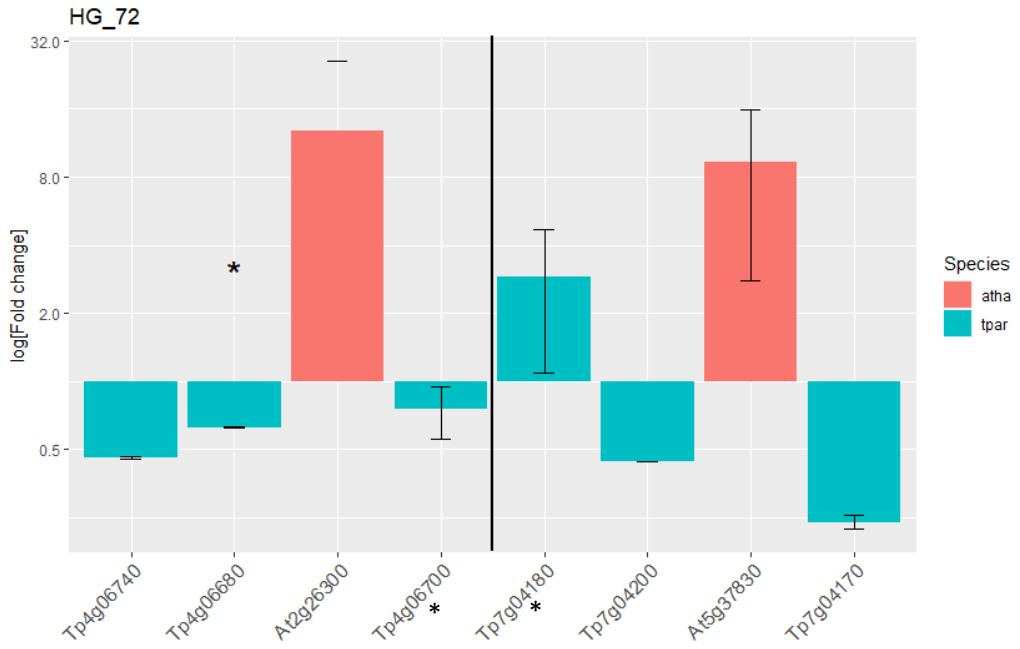


Figure 5.5. Average log fold change in gene expression of focal and syntenic genes of HG 72 for *T. parvula* and *A. thaliana* between drought and well-watered conditions. Gene expression was normalised against the housekeeping gene, Protein Phosphatase 2A Subunit A2 (At3g25800). A positive log fold change suggests greater expression under drought. Averages are based on 3 technical and 3 biological replicates in qPCR experiments. Error bars represent standard errors. Blue shows expression of *T. parvula* genes and red shows expression of the syntenic *A. thaliana* genes. Asterisk in the figure highlight genes that were significantly differentially expressed between drought and well-watered conditions (P <0.05 in a two-sample t-test). Asterisk next to gene IDs indicate the focal gene from HG 72.

5.4.4.2 HG_2909

5.4.4.2.1 Synteny analysis and function of syntenic genes

Protein domain analysis of HG 2909 identified domains of the cupin superfamily (Table 5.4). Cupins are known to play a role in plant development, as well as in plant defense responses to both biotic and abiotic pressures, including water deprivation (Wang *et al.*, 2014f). Although not homologous to an *A. thaliana* gene, the focal gene of HG 2909, Tp1g09090, was found to be syntenic to At1g10460 (Figure 5.6). At1g10460 has been characterised as a germin-like protein (GLP7) which have previously been linked to stress

responses in plants (Nakata *et al.*, 2004; Li *et al.*, 2016). This gene was originally identified in association with germination in wheat but has subsequently been found to be involved in plant resistance to heat treatment (Nakata *et al.*, 2004). More recently, overexpression of soybean GLP7 in *A. thaliana* improves abiotic stress tolerances most notably to drought, salt and oxidative tolerance (Li *et al.*, 2016).

Other genes identified in this region of the *A. thaliana* genome include Protein Phosphatase 2A 2 (PP2A2), Arabidopsis Response Regulator 4 (ARR4) and Zinc Finger Protein 5 (ZFP5). PP2A2, as discussed in Chapter 3, is involved in negative regulation of the abscisic acid pathway for stomatal closure. Additionally, ARR4 is involved in cytokinin signalling leading to the development of roots and ZFP5 is also involved in root and root hair development through cytokinin mediated signalling (both are also described in more detail in Chapter 3). This tight grouping of drought-related genes formed the hypothesis that this region might act as an operon-like gene cluster, whereby neighbouring genes have similar biological functions (Boycheva *et al.*, 2014). Similar gene clusters have been identified in plants, for example those involved in metabolism (Nützmann, Scazzocchio and Osbourn, 2018) or genomic hotspots found to be involved in the drought responses of wheat (Gálvez *et al.*, 2019). The operon-like gene cluster hypothesis was also applied to the remaining HGs (described in sections below).

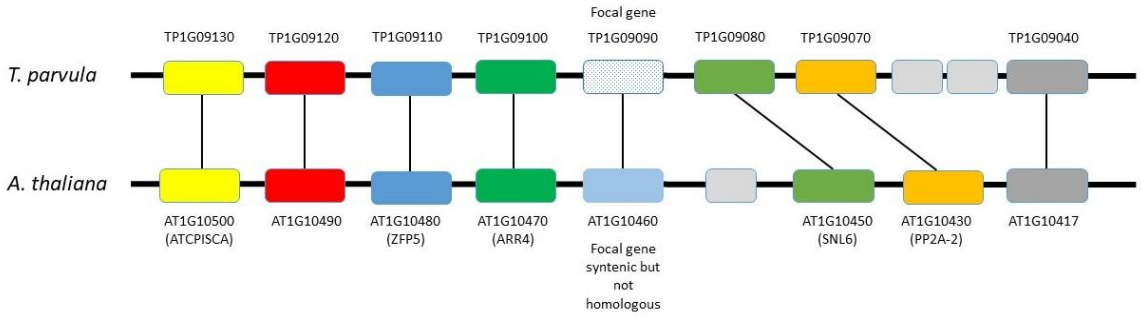


Figure 5.6. Synteny plots for genes in HG 2909. Plots are based on outputs from Genomicus Plants. Genes connected by lines denote syntenic genes whilst genes coloured in the same colour denote homologous genes.

5.4.4.2.2 Gene expression of possible drought tolerance genes

To further investigate whether Tp1g09090, the focal gene of HG 2909, and its syntenic counterpart, At1g10460 (GLP7) in *A. thaliana,* may play a role in the plant drought response, gene expression analysis was carried out under drought and well-watered conditions. The qPCR experiments demonstrated that At1g10460 (GLP7) was significantly differentially expressed (P<0.05) between drought and well-watered environments, with a greater level of expression under drought conditions (Figure 5.7).

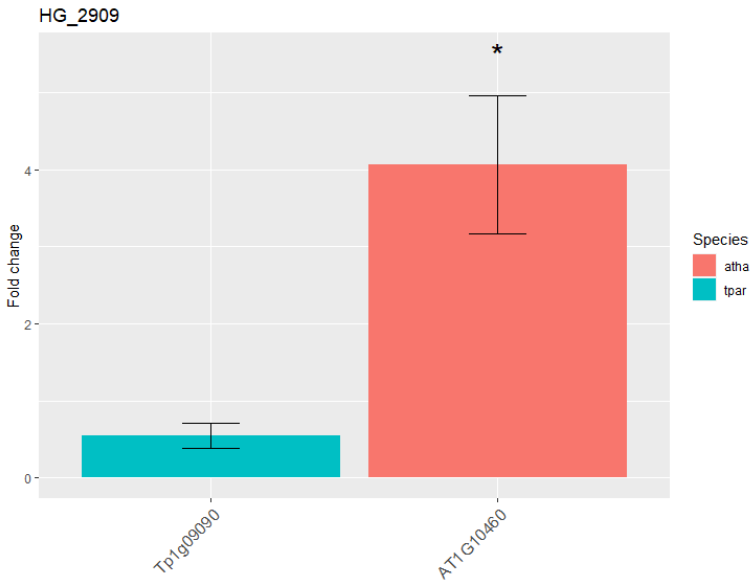


Figure 5.7. The average fold change in gene expression of the focal gene of HG 2909 for *T. parvula* (Tp1g09090) and the syntenic gene in *A. thaliana* (At1g10460) between drought and well-watered conditions. Gene expression was normalised against the housekeeping gene, Protein Phosphatase 2A Subunit A2 (At3g25800). Averages are based on 3 technical and 3 biological replicates in qPCR experiments and error bars represent standard errors. A positive fold change represents greater expression under drought. Blue shows expression of *T. parvula* genes and red shows expression of *A. thaliana* genes. Asterisks in the figure highlight genes that were significantly differentially expressed between drought and well-watered conditions (P <0.05 in a two-sample t-test).

5.4.4.3 HG_5775

5.4.4.3.1 Synteny analysis and function of syntenic genes

Protein domain analysis of HG 5775 identified a C2H2-type zinc finger domain (Table 5.4) which have previously been shown to be involved in plant stress responses to a range of abiotic and biotic pressures, including those related to high salinity, drought and osmotic stress (Kiełbowicz-Matuk, 2012; Yuan *et al.*, 2018). Although not homologous to an *A. thaliana* gene, the focal gene of HG 5775, Tp2g19280, was syntenic to At5g48890 (Figure 5.8). At5g48890 has been characterised as a LATE FLOWERING C2H2-type zinc-finger transcriptional regulator which acts as a floral repressor (Weingartner *et al.*, 2011). Overexpression mutants of this gene in *O. sativa* exhibit increased salt tolerance, which is thought to occur via an enhanced ability to scavenge reactive oxygen species (Zhang *et al.*, 2018b).

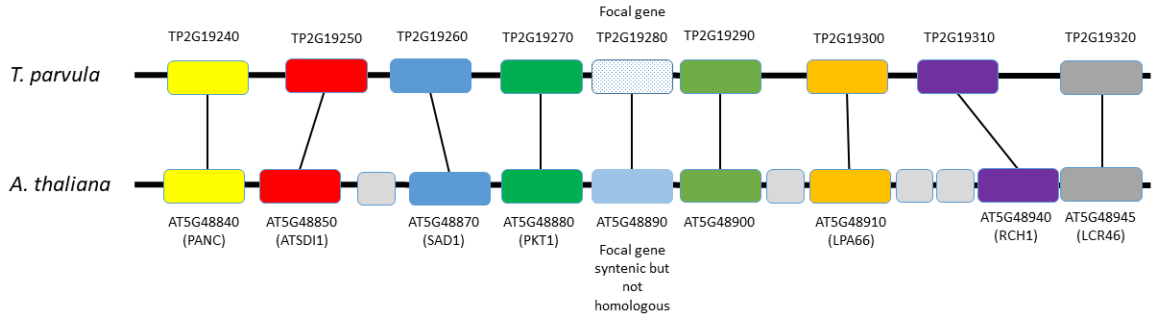


Figure 5.8. Synteny plots for genes in HG 5775. Plots are based on outputs from Genomicus Plants. Genes with matching colours denote homologous genes whilst genes connected by a line are syntenic.

5.4.4.3.2 Gene expression of possible drought tolerance genes

As above, gene expression experiments were conducted via qPCR, to identify whether Tp2g19280, the focal gene of HG 5775, and the syntenic *A. thaliana* gene, At5g48890, were drought responsive. Results showed that At5g48890 was highly expressed compared to the *T. parvula*, although this was highly variable and was not found to be significantly greater in a statistical test (P>0.05 in a two-sample t-test) (Figure 5.9).

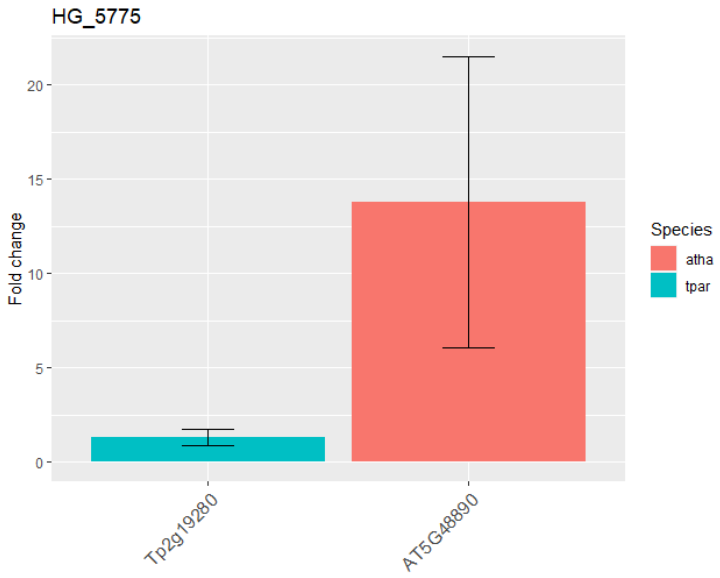


Figure 5.9. The average fold change in gene expression of the focal gene of HG 5775 for *T. parvula* (Tp2g19280) and the syntenic gene in *A. thaliana* (AT5G48890), between drought and well-watered conditions. Gene expression was normalised against the housekeeping gene, Protein Phosphatase 2A Subunit A2 (At3g25800). Averages are based on 3 technical and 3 biological plant replicates. Error bars represent standard errors. Positive fold changes represent greater expression under drought. Blue shows expression of *T. parvula* genes and red shows expression of *A. thaliana* genes.

5.4.4.4 HG_7522

5.4.4.4.1 Synteny analysis and function of syntenic genes

Protein domain analysis of HG 7522 identified a PLD-like and an Enoyl-CoA hydratase/isomerase domain (Table 5.4), which have both been shown to be involved in

plant development and stress responses (Wang, 2005). For HG 7522, there were no homologous or syntenic genes in *A. thaliana* (Figure 5.10). Instead, adjacent genes of the focal gene were identified based on the hypothesis that this is an operon-like gene cluster where neighbouring genes respond similarly to related biological functions (Boycheva *et al.*, 2014).

For HG 7522, Tp2g22410 and Tp2g22430, which are next to the focal gene (Tp2g22420) in the *T. parvula* genome, are syntenic to At5g25110 (CIPK25) and At5g25100 (TMN9), respectively. CIPK25, also known as SNRK3.25, is a member of the SRNK gene family (Colina *et al.*, 2019) which is known to be involved in conferring tolerance to drought stress and has also been shown to be crucial for plant adaptation to terrestrial environments (Shinozawa *et al.*, 2019). Specifically, CIPK25 has functions in the development of the root meristem (Meena *et al.*, 2019) and also confers hypoxic stress tolerance (Tagliani *et al.*, 2020).

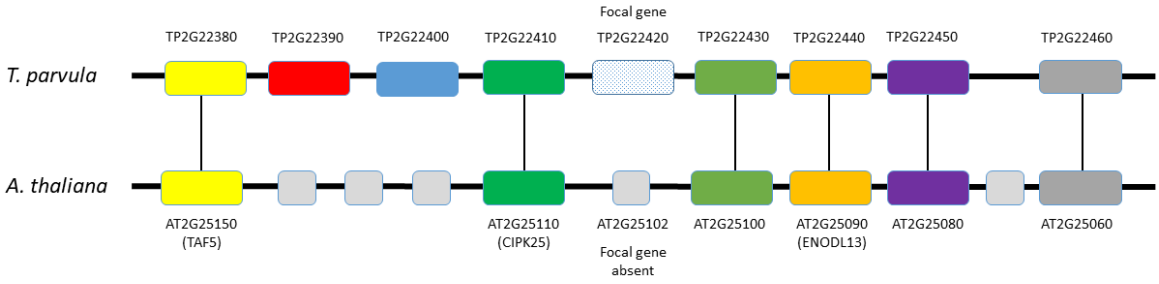


Figure 5.10. Synteny plots for genes in HG 7522. Plots are based on outputs from Genomicus Plants. Genes with matching colours and connecting lines denote syntenic, homologous genes.

5.4.4.4.2 Gene expression of possible drought tolerance genes

For HG 7522, patterns of gene expression under drought and well-watered conditions were less clear and there were no significant differences in expression between drought

and watered conditions (*P*>0.05 in two-sample t-tests). High levels of variation were seen between biological replicates, as shown by the error bars (Figure 5.11). This could be explained by the variability associated with responses to drought, although it could also be due to poor replicability across qPCR runs. Unfortunately, it was not possible to differentiate between these two competing explanations by running further qPCR experiments with different primer pairs, as this work was halted due to the coronavirus pandemic. Further work planned to complete the analysis of candidate drought genes, is detailed in the discussion.

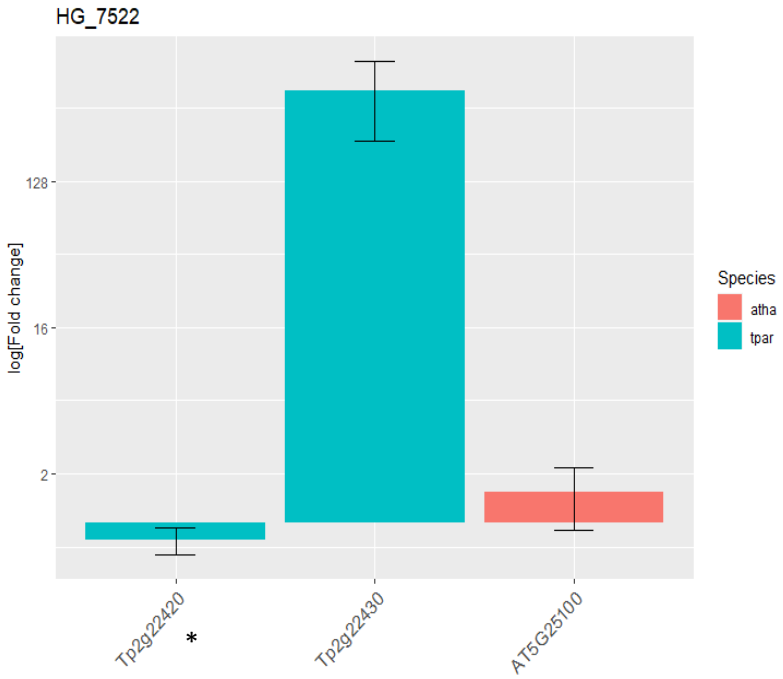


Figure 5.11. The average log fold change in gene expression of the focal gene of HG 7522 for *T. parvula* (Tp2g22420) and the syntenic gene in *A. thaliana* (AT5G25100), between drought and well-watered conditions. Gene expression was normalised against the housekeeping gene, Protein Phosphatase 2A Subunit A2 (At3g25800). Averages are calculated based on 3 technical and 3 biological plant replicates investigated in qPCR experiments. Error bars represent standard errors. Blue shows expression of T. parvula genes and red shows expression of A. thaliana genes. Asterisk next to gene IDs indicate the focal gene from HG 7522.

5.4.4.5 HG_9215

5.4.4.5.1 Synteny analysis and function of syntenic genes

Similar to HG 7522, there were no homologous or syntenic genes in HG 9215 to those in *A. thaliana* (Figure 5.10). Thus, as before, adjacent genes of the focal gene in HG 9215 were identified based on the operon-like gene cluster hypothesis (Boycheva *et al.*, 2014). For HG 9215, Tp6g08200 is the nearest gene to the focal gene (Tp6g08250) and is syntenic to At4g09340 in *A. thaliana* (Figure 5.12). The At4g09340 gene was characterised as a SPla/RYanodine receptor (SPRY) domain-containing protein, a member of the Trithorax gene group, which are developmental regulators that play a role in embryogenesis (Aquea *et al.*, 2010).

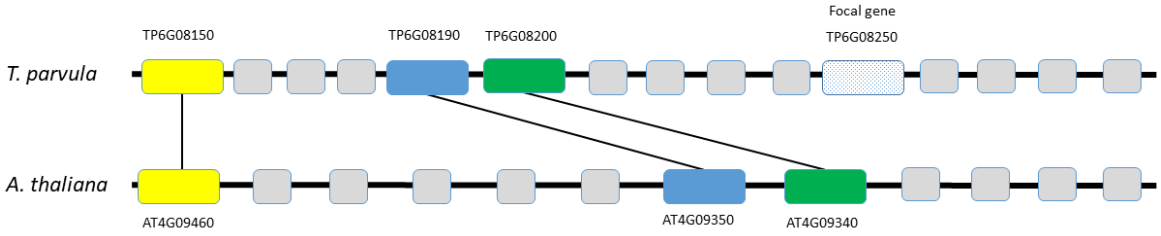


Figure 5.12. Synteny plots for genes in HG 9215. Plots are based on outputs from Genomicus Plants**.** Genes with matching colours and connecting lines denote syntenic, homologous genes.

5.4.4.5.2 Gene expression of possible drought tolerance genes

For HG 9215, the patterns of gene expression were also unclear (Figure 5.12), particularly because the primer pair for the *A. thaliana* homolog (At4g09340) was not successfully designed to avoid the amplification of off-target sequences and so could not be used for the PCR experiment. Unfortunately, further work to investigate these genes was halted by the coronavirus pandemic. Detailed below in the discussion is the work planned to comprehensively analyse the gene expression of these candidate drought genes. Although the focal (Tp6g08250) and adjacent (Tp6g08200) *T. parvula* genes showed an increase in expression under drought conditions (Figure 5.13), this difference was not significant (*P*>0.05).

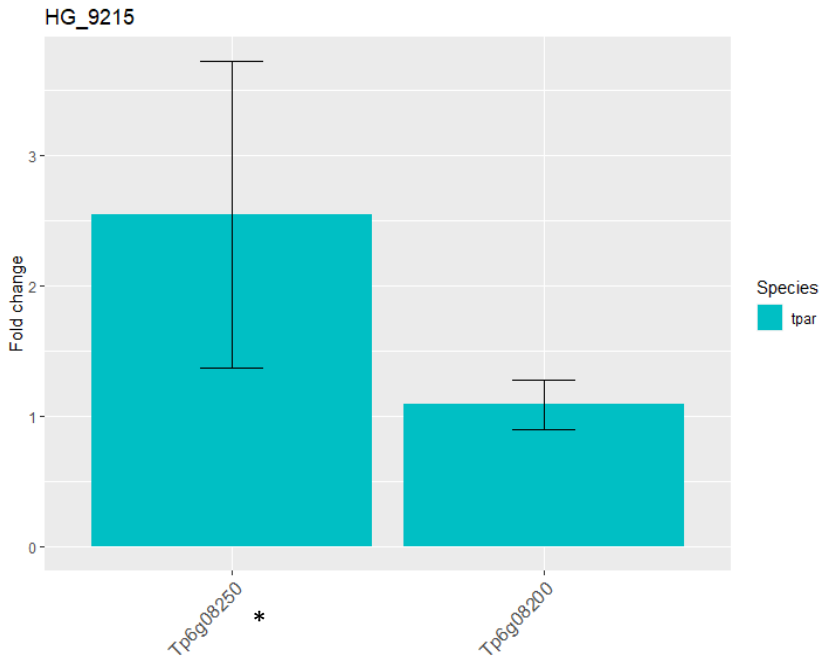


Figure 5.13. The average fold change in gene expression of the focal and adjacent genes of HG 9215 for *T. parvula* between drought and well-watered conditions. Gene expression was normalised against the housekeeping gene, Protein Phosphatase 2A Subunit A2 (At3g25800). Averages are based on 3 technical and 3 biological plant replicates. Error bars represent standard errors. Asterisk next to gene IDs indicate the focal gene from HG 9215.

5.4.4.6 HG_10098

5.4.4.6.1 Synteny analysis and function of syntenic genes

Protein domain analysis of HG 7522 revealed both HIT zinc finger and Zinc knuckle domains, which have previously been shown to play a role in plant drought and salt tolerance (Li *et al.*, 2014d). Similar to HG 7522 and 9215, genes adjacent to the focal gene (Tp7g15100) in HG 10098 were identified based on the operon-like gene cluster hypothesis. The Tp7g15080 gene, which is near to the focal gene Tp7g15100, was found to be syntenic to the At4g16515 gene in *A. thaliana* (Figure 5.14). The gene Tp7g15110 was also found next to Tp7g15100 in the *T. parvula* genome and was identified as syntenic to At4g16530 (Figure 5.13). At4g16515 (RGF6) is a root meristem growth factor, required for the maintenance of the root stem cell niche, root hair development and root gravitropism (Matsuzaki *et al.*, 2010; Fernandez *et al.*, 2013). At4g16530 is an uncharacterised protein, although it has been demonstrated that it is regulated by GDS1 (Growth, Development and Splicing 1) which is involved in abiotic and biotic stress responses (Kim *et al.*, 2016).

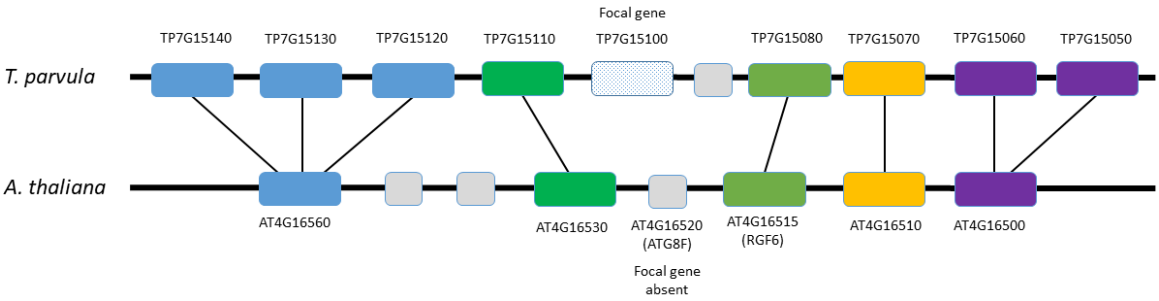


Figure 5.14. Synteny plots for genes in HG 10098 based on outputs from Genomicus Plants. Genes coloured in the same colour denote syntenic, homologous genes.

5.4.4.6.2 Gene expression of possible drought tolerance genes

Gene expression analysis identified that four of the five genes showed significantly greater levels of expression under drought than under well-watered conditions ($P<0.05$ in two-

sample t-tests) (Figure 5.14). This included the focal gene, Tp7g15100, as well as the adjacent gene Tp7g15080 and its syntenic counterpart, At4g16515. Although Tp7g15110 (which is adjacent to the focal gene of HG 10098) was not significantly differentially expressed (*P*>0.05), the syntenic gene At4g16530, which is an uncharacterised protein, also showed significantly greater levels of expression under drought (Figure 5.15).

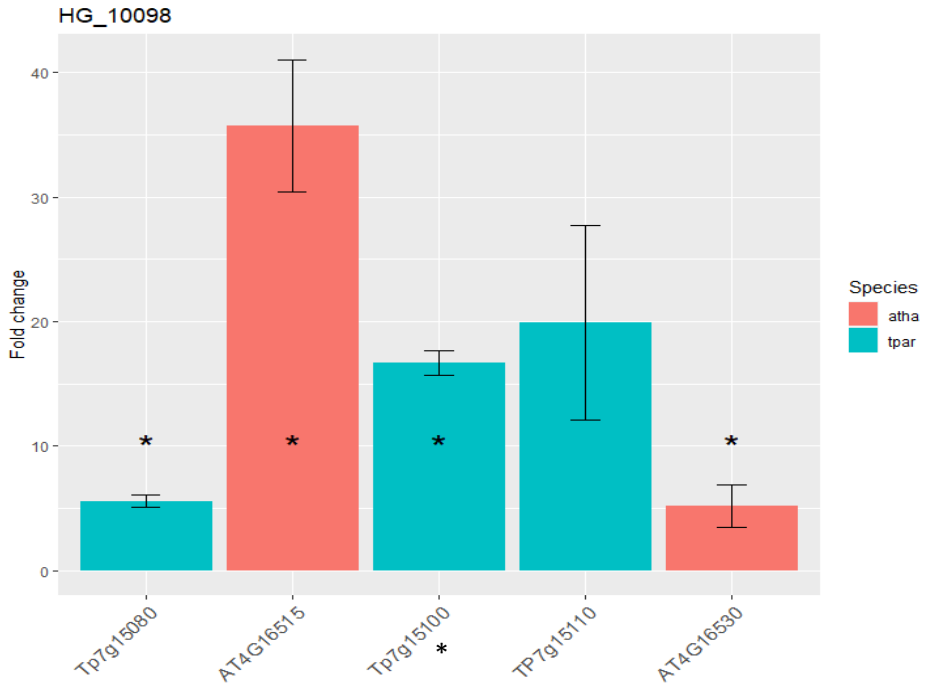


Figure 5.15. The average fold change in gene expression of the focal and adjacent genes of HG 10098 for T. parvula and the syntenic genes in A. thaliana, between drought and well-watered conditions. Gene expression was normalised against the housekeeping gene, Protein Phosphatase 2A Subunit A2 (At3g25800). Averages are based on 3 technical and 3 biological replicates. Error bars represent standard errors. Blue bars show expression of *T. parvula* genes and red shows expression of *A. thaliana* genes. Asterisks in the figure highlight genes that were significantly differentially expressed between drought and well-watered conditions (P<0.05 in two-sample t-tests). The asterisk next to gene IDs indicate the focal gene from HG 10098.

5.4.5 Loss of function mutants

To further investigate the impact of the candidate genes for all homology groups on plant drought responses, *A. thaliana* loss of function mutants were ordered from NASC (Table 5.5). For all genes of interests, two SALK lines were ordered which each contained T-DNA inserts at different locations within the same gene. These mutants were then confirmed via PCR and gel electrophoresis (Appendix 5.2). To date, this has only been completed for 13 of the 18 NASC mutant lines. Additionally, further characterisation of these mutants was halted by the coronavirus pandemic. Thus, future research to comprehensively analyse the impact of gene overexpression on drought responses is outlined in the discussion (section 5.5).

Table 5.5. A summary of SALK lines designed for each *A. thaliana* gene, the homology group (HG) they are associated with and whether mutants have been confirmed by PCR.

Gene name	HG	SALK ID	Confirmed
AT2G26300	72	N528135 N561522	Yes Yes
AT5G37830		N590917 N668525	No Yes
AT5G37850		N618343 N638992	Yes Yes
AT4G16515	10098	N633489 N573605	Yes No
AT4G16530		N577075 N593165	No Yes
AT5G48890	5775	N818436 N546014	Yes Yes
AT1G10460	2909	N572453 N533384	Yes No
AT5G25100	7522	N529679 N539673	Yes Yes
AT4G09340	9215	N552021 N533044	Yes No

5.4.6 Overexpression analysis of candidate genes

The gene expression analysis provided preliminary evidence that some of the candidate genes played a role in plant drought response pathways, particularly those that showed differential expression under drought conditions. To further understand the role of these candidate genes in plant drought responses, these focal genes were overexpressed in *A. thaliana*. To do this, gene constructs (Appendix 5.3) were designed and transformed into *Agrobacterium* and these were used to generate transgenic lines of *A. thaliana* by floral dip. Unfortunately, screening of primary transgenics could not be carried out due to the coronavirus pandemic. To date, seeds of *Agrobacterium* transformed *A. thaliana* have been collected and await selection on antibiotic media. Thus, planned future research designed to investigate the impact of gene overexpression on plant drought responses is outlined in the discussion (section 5.5).

5.5 Discussion

5.5.1 Identification and function of candidate genes

The work in this chapter aimed to identify a list of candidate genes that may be involved in plant drought tolerance, by using a novel comparative genomics approach, namely, comparing gene content in relation to plant drought adaptations. Other studies that have tried to link genes to traits, often have an a priori knowledge of what particular families of genes might be involved in the trait of interest (Nagy *et al.*, 2020). In doing this, these studies are only examining known unknowns and are not able to identify unknown unknowns. These are genes that have not previously been linked to the function of interest, for example responses to drought (Dunn *et al.*, 2016). The approach used in this chapter has the benefit that such genes can be identified as, by examining all genes associated with patterns of drought tolerance and sensitivity, there is no inherent bias towards a subset of already known genes.

Using comparative genomics to identify genes for traits is still in its infancy as a field of research. This is due to the fast evolving pace of genome sequencing and lack of

appropriate analytical tools to identify candidate genes from large scale genome datasets (Nagy *et al.*, 2020). There are several approaches that have been explored to investigate candidate genes linked to traits. These methods begin by reconstructing ancestral character states by mapping the gain and loss of a trait of interest onto a phylogeny (as completed in Chapter 4). Signatures of genome evolution in relation to trait evolution are next investigated by analysing gene groups from the outputs of comparative genomics. These approaches aim to identify patterns of gene duplication (Nagy *et al.*, 2014), sequence divergence rates (Chikina *et al.*, 2016; Partha *et al.*, 2019) and the frequency of gene gain and loss as an indicator of the biological function of genes. The latter of these, patterns of gene gain and loss, was the approach utilised in this chapter.

To this end, queries of the genomic pipeline revealed that there were no broad-scale patterns of drought gene loss associated with drought sensitivity across the plant tree of life, however, patterns of lineage specific gene loss could be identified. This enabled the identification of 238 homology groups that were differentially retained between drought tolerant and sensitive plants within plant clades. Further to this, protein domain analysis identified 50 homology groups of potential interest for drought tolerance, of which, 6 were chosen for further investigation. Based on taxonomic occupancy and protein domain analysis, as well as the subsequent preliminary experimental analysis, some of the selected HGs appeared to be suitable candidate drought genes. Therefore, the approach, incorporating trait evolution into a comparative genomics framework, is potentially effective for identifying candidate genes linked to specific adaptations. The effectiveness and potential applicability of this approach is discussed in further detail below.

5.5.2 Mechanisms of conferring drought tolerance

5.5.2.1 Retrotransposons and drought tolerance

Protein domain analysis revealed several interesting mechanisms through which the identified HGs could play a role in plant tolerance to drought. For example, the protein domains of *T. parvula* genes from HG 72 were identified as retrotransposons and retroviral

elements. Transposable elements (TEs) or jumping genes, are known to promote various chromosomal rearrangements which can, in turn, alter target gene expression (Elbarbary *et al.*, 2016; Mita *et al.*, 2016). TEs can operate thousands of nucleotides away from the genes that they regulate, however, in many instances they are also found to be proximal to their target gene. As a result of their ability to alter the expression of other genes, the activity of TEs has been linked to differential plant stress responses. For example, it has been shown in tomato plants (*Solanum lycopersicum*) that drought stress regulates the activation of *Rider* retrotransposons in the tomato plant genome which subsequently results in a greater level of drought tolerance (Benoit *et al.*, 2019). As a result of their potential impact on gene expression, it was hypothesised that the combination of the 42 retrotransposon sequences identified in HG 72 in this chapter, and their location in the *T. parvula* genome, may be responsible for altering drought phenotypes. The results of qPCR experiments suggested that for the focal gene of this HG, the majority of adjacent and syntenic genes were not differentially expressed under drought conditions. However, results of qPCR experiments were highly variable across replicates and so it is difficult to make firm conclusions about whether or not these genes are involved in plant drought tolerance. Further experiments that could help to validate the role of HGs in drought tolerance, such as RNA sequencing experiments and the characterisation of mutant plants, are discussed below (section 5.5.4).

5.5.2.2 Regulation of root development

Plants are known to alter their root and shoot growth rates in response to drought (Smith *et al.*, 2012) as highlighted in previous chapters. Under water stress, plants adjust their root system architecture by stimulating deeper primary and lateral root growth (Bao *et al.*, 2014; Orosa-Puente *et al.*, 2018; von Wangenheim *et al.*, 2020). Protein domain analysis revealed that several of the HGs and syntenic genes identified in the analysis above were involved in the regulation of root development. For example, CIPK25 was identified as being syntenic to Tp2g22410 (HG_7522) and is known to function in root meristem

development (Meena *et al.*, 2019). In loss of function CIPK25 mutants, plant roots have been shown to be shorter than in wild type *A. thaliana* (Meena *et al.*, 2015). Under drought, the differential regulation of CIPK25 may be involved in conferring greater tolerance (Meena *et al.*, 2015).

Other homology groups also contained genes that may contribute to plant drought tolerance via altered root development. The gene At4g16515 was identified as being syntenic to Tp7g15080 (in HG 10098) and has previously been characterised as a root meristem growth factor (RGF6). The RGF gene family have been shown to control the pattern of root growth and lateral root development (Meng *et al.*, 2012). Therefore, the regulation of these genes may also be altered under drought, leading to a difference in drought responses via root development. Indeed, this gene was found to be significantly upregulated under drought conditions relative to well-watered conditions in qPCR experiments.

5.5.2 Gene expression

Upon identification of candidate HGs and syntenic genes, attempts were made to validate the comparative genomics approach, by assessing the relative expression of identified candidate genes under drought and well-watered conditions. There were a couple of promising results from these experiments, for example, both the focal gene of HG 10098 as well as all but one of the adjacent and syntenic genes tested in qPCR experiments, showed significantly greater levels of expression under conditions of drought compared to well-watered conditions. This suggests that these genes are likely to be drought responsive and could play a role in the plant stress response. However, the gene expression analyses was not entirely clear cut for all homology groups, making it difficult to decisively conclude whether the comparative genomic approach taken in this chapter is a valid method of identifying unknown drought tolerance genes. For instance, there were cases (such as in HG 2909 and HG 72) where the focal and syntenic candidate genes showed different patterns of expression under drought. This could suggest that the genes

that were not significantly differentially expressed were not drought responsive. However, there was also a high degree of variability in the levels of gene expression seen across biological plant replicates. It is possible that this could reflect the high variability of responses to drought seen across individual plants (Guo *et al.*, 2014; Pabuayon *et al.*, 2016). Another important factor that could have impacted upon the levels of gene expression observed for individual plants is the degree to which each individual plant was stressed during the drought experiment. Multiple factors could have affected the severity of drought experienced, including the placement of plants in the growth cabinet (for example, near a fan). In future repeats of these experiments, these factors should be considered and mitigated. Further house-keeping genes could also be included as controls in qPCR experiments to further normalise the gene expression data and control for variability across individual plants (Guo *et al.*, 2014; Pabuayon *et al.*, 2016).

5.5.4 Planned further work

Due to the coronavirus pandemic, there are several components of this experimental chapter that could not be completed. The proposed future work to further validate the comparative genomics approach to finding novel drought genes is therefore discussed below.

Firstly, gene expression analysis (via qPCR) of the remaining genes that could not be completed prior to lockdown would be conducted. To date, 22 primer pairs have been successfully confirmed for both drought and well-watered samples. The primers for these 7 genes (72: At2g26290, 72: Tp7g04210, At5g37820, At5g37850, 5775: Tp2g22410, At5g25110, 9215: At4g09340) had been re-designed but further analysis would need to be completed to finalise conclusions about the expression of all candidate and syntenic genes (Table 1.1).

Secondly, there were many candidate genes that could not be investigated. 238 Homology Groups were originally identified to be differentially retained between drought tolerant and drought sensitive species, of which only 6 have been investigated in this body of work.

Instead of conducting further qPCR expeirments, an RNA-seq experiment would enable all of these 238 homology groups to be comprehensively characterised. Such an experiment would compare gene expression across all genes under drought and well-watered conditions for *T. parvula* and *A. thaliana*. This would also help to validate the findings of qPCR experiments presented in this chapter.

In addition to these experiments, the full characterisation the loss of function *A. thaliana* mutant lines (from NASC) would provide insight into the potential functions of the identified candidate syntenic genes and their potential role in drought tolerance. To do this, mutant lines would be subjected to drought and well-watered conditions. Key physiological traits would then be compared between wild type and mutant plants, such as leaf area, rosette weight and root growth. Finally, overexpression experiments would be completed, to assess whether the identified focal genes in *T. parvula* could confer drought tolerance on *A. thaliana* plants. As with loss of function mutants, drought experiments with these transgenic plants would provide insights into the function of these genes.

5.6 Conclusion

In this chapter, genes from six Homology Groups have been identified that are potentially involved in plant drought tolerance. These were identified based on their occupancy amongst drought tolerant and sensitive species. Protein domain analysis and gene expression experiments under drought conditions provide initially promising insights into the function of these genes. Further analysis, as described above, would reveal the extent to which these genes are able to produce plants with greater drought tolerance while maintaining growth and may help to validate the comparative genomics approach applied here as a method to identify novel drought tolerance genes. If valid, this approach could be used to identify genes involved in other key plant adaptations such as salt tolerance and nitrogen fixation.

Chapter 6 Discussion and Conclusions

6.1 Understanding plant evolution through genome analysis

The overarching goal of this thesis, as outlined in the introduction, was to explore the evolution of genes involved in the major transitions in plant evolution and the consequences of these events for some of the traits associated with these transitions, including drought tolerance. As detailed in chapters 2-5, plant evolution was investigated through the analysis of genes, genomes and traits. Specifically this was completed by addressing four aims below.

- Examine gene gains and losses across the plant tree of life
- Investigate gene group dynamics in relation to drought tolerance innovations
- Understand how drought tolerance as a trait has evolved
- Discover unknown drought tolerance genes by incorporating trait evolution into a comparative genomic framework

The progress made in addressing these aims is discussed below, placing the findings of each chapter in the context of the thesis as a whole.

6.2 Loss and gain of homology groups during plant diversification

Understanding the role of gene novelty and gene loss is intrinsically linked to our understanding of organismal evolution. It has previously been established that gene novelty was integral to the origin of animals (Paps *et al.*, 2018). In the first research chapter of this thesis (Chapter 2), it was identified that patterns of gene novelty were equally important for plant diversification. Specifically, the origin of Streptophyta (50 HGs) and Embryophyta (103 HGs) were associated with large bursts of gene novelty. These novel genes in Streptophyta were associated with increasing complexity of the cell wall, signalling pathways and stress responses which are indicators of multicellularity. In Embryophyta novel genes, functions were linked to UV light protection, environmental stress signalling and host microbe interactions which are hallmarks of terrestrialisation. Other major nodes within land plants also display genomic novelty although in a lesser degree, for example, 8, 55 and 16 Novel Core HGs were identified in the LCA of

Tracheophyta, Spermatophyta and Angiosperms respectively. This indicates that novelty was important for these groups and enabled re-wiring of genetic toolkits, increasing the adaptive plasticity of plants to changing environments.

These findings highlight a broader aspect of plant evolution. In "*On the Origin of Species"*, Charles Darwin postulated that species arose over long periods of time through a mechanism of slow and gradual change (Darwin, 1859). However, what could not be established at the time was the genetic basis of this gradual change and how it enabled evolution to occur. Since then, technological revolutions have transformed our understanding of the diversity of life at the DNA level (Li, 2018; Li *et al.*, 2018b). Through this study of plant DNA, the mechanisms by which plant species diversify are becoming better understood. It has since been suggested that organisms can also undergo rapid bursts of evolutionary change, termed punctuated equilibria (Gould *et al.*, 1993). This process is consistent with the two bursts of genomic novelty leading to the evolution of land plants, suggesting dramatic evolutionary change.

These findings alone can only offer partial insight into the genetic factors contributing to the evolutionary history of plants. There are distinct biological patterns that can be attributed to unbalanced gene novelty seen across the Archaeplastida phylogeny. Recent analysis of one thousand plant transcriptomes investigated patterns of gene birth, expansion and contraction for 23 gene families (Leebens-Mack *et al.*, 2019). Gene birth (or novel genes) and gene family expansion were common to Viridiplantae, Streptophyta and Embryophyta whilst gene family contraction was most common in the origin of flowering plants. Although only based on a small sampling of genes, this result is consistent with the findings in this chapter that there was a switch from gene novelty to more complex gene family dynamics in the evolutionary history of plants.

This highlights the complexity of plant genomes and the processes contributing to plant genome evolution. Plant genomes are characterised by multiple rounds of whole genome duplications (Leebens-Mack *et al.*, 2019) which is followed by genome regulatory

processes such as loss of duplicate genes and repetitive elements, chromosome restructuring and genome downsizing (Wendel *et al.*, 2016). This can lead to large expansions of gene family size through gene duplication and diversification, which can enable the radiation of diverse plant forms and functions (Harrison, 2017). For example, many gene families linked to flowering plant development are found in the ancestor of Embryophyta and have diversified throughout the evolutionary history of plants (Rensing *et al.*, 2008; Hori *et al.*, 2014). Due to the broad clustering of HGs, as opposed to orthology groups, patterns of gene duplication could remain undetected. Duplicated genes are likely to be placed within the same HG and as such further analysis, such as gene tree inference, would be needed to distinguish the orthology and paralogy relationships of genes within the same HG.

It has also been shown that many flowering plant transcription factors, known to be involved in plant development, have homologs in early diverging land plants (e.g. bHLH, Homeobox transcription factors) (Catarino *et al.*, 2016). This suggests there was an ancient repertoire of genes in the ancestral land plant and that many have been recruited to coordinate the development of new structures (Pires *et al.*, 2012). This process, of using old genes for novel functions, is termed co-option. To a certain degree, the role of these two other genetic factors forms the basis of research questions that Chapter 3 aims to begin to address. Other factors such as horizontal gene transfer (Yue *et al.*, 2012; Cheng *et al.*, 2019; Wickell *et al.*, 2019) and parasitism (Kado *et al.*, 2018; Vogel *et al.*, 2018), which are common to plants, can impact gene content. Furthermore, non-coding regions, such as the activity of transposable elements, can have consequences for genome architecture and function (Wendel *et al.*, 2016).

Additionally, the inclusion of new, taxonomically significant genomes may change the clustering of Homology Groups which is discussed in more detail below (section 6.5). These genomes would include the first two fern genomes (*Azolla filiculoides, Salvinia cucullata* (Li *et al.*, 2018a)), the first hornwort genomes (*Anthoceros sp.* (Li *et al.*, 2020a;

Zhang *et al.*, 2020b)) and a number of genomes from streptophyte algal lineages including *Chlorokybus atmophyticus, Mesostigma viride* (Wang *et al.*, 2019)*, Chara braunii* (Nishiyama *et al.*, 2018) and *Spirogloea muscicola* (Cheng *et al.*, 2019). However, to date, this is the largest published comparative genomic study of plants, incorporating genome data from a diverse range of plant taxa with comprehensive outgroup sampling using complete genomes, providing unprecedented insight into plant genome evolution.

6.3 Gene group dynamics for the evolution of plants on land

As highlighted above, other genetic factors, other than conserved gene novelty, are crucial for plant evolution. With an interest in investigating the evolution of drought tolerance, three innovations (stomata, vascular tissue, roots) were investigated that play an integral role in plant drought responses. The work in this chapter identified the role of gene novelty, gene duplication and gene co-option for the evolution of these innovations (Chapter 3). These finding suggest that distinct evolutionary mechanisms are responsible for the evolution of each of these innovations. They also emphasise the role of water during the course of plant evolutionary history for driving adaptations to novel environments. This is due to the stepwise nature of the emergence of stomata, vascular tissue, primary roots and lateral roots in the ancestors of land plants, vascular plants, Euphyllophyta and seed plants respectively.

Typically, across all innovations, the pattern of switching from gene novelty in early land plant evolution to more complex evolutionary dynamics of genes was observed. Co-option, the repurposing of old genes for new functions, appears to be a common feature of plant developmental pathways. Clear signatures of novel genes were generally not identified for the evolution of innovations that originated after the divergence of land plants (Figure 3.1). This finding could have important implications for the evolution of the entire gene repertoire of plants. Expanding sampling across all plant genes to analyse patterns of gene duplication and co-option could be an interesting future avenue.

There is strong debate around the origin of active stomatal closure in plants, in the ancestor of seed plants (Brodribb *et al.*, 2011; McAdam *et al.*, 2012, 2013) or earlier (Ruszala *et al.*, 2011; Cai *et al.*, 2017). The analysis presented here suggest that active stomatal closure originated in the ancestor of seed plants. However, this process occurred in a stepwise manner through duplication of core regulatory genes in the ABA signalling pathway in the euphyllophyte ancestor. Active control of seed plant stomata occurs in response to ABA under drought stress (Sussmilch *et al.*, 2017b). Carrying this work forward, it would be interesting to investigate the function of duplicate genes for drought induced stomatal closure.

As confirmed in this work, and in other studies, stomata were present in the ancestor of land plants and are present in every major lineage apart from liverworts (Chater *et al.*, 2017; Harris *et al.*, 2020). This loss of stomata in liverworts and some mosses occurred through a process of reductive evolution (Duckett *et al.*, 2018; Harris *et al.*, 2020). In liverworts, the air pore complex has independently evolved to enable gas exchange (Jones *et al.*, 2017). The evolutionary development of these analogous features likely required an individual genetic toolkit, facilitated by lineage specific gene group novelty and expansion. Future analysis of bryophyte genomes, which are now available, would enables us to ask questions about how the air pore complex independently evolved (Rensing *et al.*, 2008; Bowman *et al.*, 2017; Li *et al.*, 2020a).

These innovations are crucial for the evolution of plants on land, adapting to a myriad of environmental stresses. Root hairs, primary roots and lateral roots increasingly enabled fine-tuned responses of plants to water uptake. Vascular tissue enabled efficient water transport, promoting the development of plants with increased height, photosynthetic capacity and potential to colonise diverse habitats. Stomata and the evolution of stomatal control facilitated rapid responses to desiccation and drought stress which were common in terrestrial environments. Therefore, studying these developmental and signalling genes,

their function and their patterns of diversification is integral to understanding the evolution of plants on land.

6.4 The evolution of drought tolerance

In the previous chapter, the evolution of the genetic networks leading to the development of stomata, vascular tissue and roots was explored, as these innovations were intrinsically linked to drought tolerance. To understand the evolution of drought adaptations as a whole, the taxonomic occupancy and ancestral state reconstruction of drought adaptations as a collective trait was investigated (Chapter 4). This revealed that the first land plants were drought adapted and likely capable of desiccation tolerance whilst the first vascular plants were drought adapted and likely drought tolerant. The results from multiple ancestral state reconstruction methods incorporating species tree information for 178 Archaeplastida species found similar results providing robust support to the research findings. These findings highlight the changing relationship of plants with water, during the course of land plant evolution beginning ~500 million years ago.

In a recent essay on reconstructing trait evolution, the importance of understanding the evolutionary relationships of a study's organisms were emphasised (Delaux *et al.*, 2019). Although the species tree used in this analysis described the known relationships of Archaeplastida accurately, there were several branches that were misplaced (e.g. the lycophyte *Selaginella moellendorffii*, Figure 4.1). In future analyses, multiple species tree estimation approaches would be compared, most notably coalescent based analyses (e.g. ASTRAL Zhang *et al.*, 2018a) contrasted with concatenation based approaches. Concatenation based approaches concatenate multi-gene alignments and analyse these in a single analysis (Kubatko *et al.*, 2007). Coalescent-based approaches build a species tree by reconstructing multiple individual gene trees and then summarising the output into a single species tree (Springer *et al.*, 2014). This second approach has successfully been used to understand the evolutionary relationships of Archaeplastida (Leebens-Mack *et al.*, 2019), Viridiplantae (Wickett *et al.*, 2014) and Embryophyta (Li *et al.*, 2020a). Analytical

advances now enable species trees to be built from multiple copy genes, as opposed to single copy genes, which would be particularly advantageous given the broad clustering of Homology Groups (Zhang *et al.*, 2020a).

The species used to reconstruct the evolutionary history of drought tolerance were present in the genomic dataset. This formed the basis of the integration of drought tolerance characters into a comparative genomics framework detailed in Chapter 5. This collective drought tolerance trait was defined in the broadest terms for species across the tree of life. This was completed by searching species name in relation to a series of drought adaptation terms such as drought tolerance, drought resistant and drought sensitive (Table 4.1). Although, applicable for the phylogenetic breadth of the dataset, this does not capture the diversity of drought responses. For example, the tree *Populus pruinosa* survives in desert environments by accessing hypersaline underground water (Yang *et al.*, 2017b) and the resurrection plant *Boea hygrometrica* by altering the expression of dehydration responsive genes (Xiao *et al.*, 2015). The drought adaptation, crassulacean acid metabolism (CAM), has evolved multiple times and therefore is found across the plant phylogeny (Liu, 2015; Ming *et al.*, 2015; Yang *et al.*, 2017c). This drought avoidance mechanism enables the uptake of carbon dioxide at night when temperatures are lower, reducing water loss (Bräutigam *et al.*, 2017). To capture this diversity of drought responses, experimental evaluation of non-terminal and terminal drought stress could be considered, as seen for the genus *Vigna* (Iseki *et al.*, 2018) for a selection of plant species with genomic representation. Findings from this work found similar patterns of drought sensitivity in domesticated species with the potential for improvement of drought tolerance from crop wild relatives (Iseki *et al.*, 2018). The experimental approach would allow for quantification and comparison of drought responses across a broad range of taxa.

Additionally there are many more drought tolerant and drought sensitive species in the plant phylogeny. For example, the drought adapted species within the lycophyte order Isoetales (Li *et al.*, 2015b), the drought tolerant tree fern species in the order Cyatheales

(Volkova *et al.*, 2010) and the desiccation tolerant species in the genus *Xerophyta* (Gaff, 1971) to name but a few. As mentioned in the discussion of Chapter 4, the work of characterising drought response across a greater diversity of plants has begun with trait databases such as TRY (Kattge *et al.*, 2020). However, this data for species tolerance to drought is only available for euphyllophytes. To fully characterise the evolutionary history of drought tolerance, a broader taxonomic sampling of drought tolerant and sensitive species should be completed.

Taking into account these limitations of taxonomic coverage, the LCA of Embryophyta and Tracheophyta were likely identified as desiccation and drought tolerant respectively. With the ancestors of these plant groups likely adapted to variable water availability, any incidences of drought sensitivity represent loss of this trait. A major correlation for drought sensitivity in plants were crop species that have been domesticated (Figure 4.4). This has major implications for food security. This artificial selection process has reduced the genetic diversity of crop species which means that many crops have lost particular stress tolerance (Zhang *et al.*, 2017b). In an ever changing climate, crop yield will likely become less predictable. Therefore, crop wild relatives are being considered as a pool of genetic diversity to improve crop stress tolerance. Additionally, novel approaches to identify stress tolerant genes are required which was the aim of final research chapter of this thesis.

6.5 Application of evolutionary genomics approach for identifying drought tolerance genes

The results from the previous chapter provided the rationale that any cases of drought sensitivity will be accompanied by the loss of drought tolerance genes. In the final research chapter of this thesis, an evolutionary approach focussing on lineage specific gene loss to identify target drought tolerance genes is described, followed by preliminary functional evaluation to validate these computational findings (Chapter 5). Based on taxonomic occupancy of genes, protein domain analysis and synteny analysis, the candidate drought tolerance genes were suitable for experimental analysis. This suggests that the application of evolutionary thinking, integrating trait evolution with comparative genome analysis,

could be effective in determining uncharacterised genes linked to drought tolerance. Although initially promising, the experimental component of this work was halted by the coronavirus pandemic. Therefore, further experimental work is needed to fully characterise the function of candidate drought tolerance genes.

There are several important environmental stresses that limit crop productivity. Aside from drought stress, major limitations for securing crop yield include extremes of light availability, UV radiation, temperature, heavy metal and salinity (Pereira, 2016). Salt stress and plant adaptations to saline environments are well characterised. Additionally, the evolution of salt tolerance has been thoroughly investigated, finding multiple origins across euphyllophytes (Flowers *et al.*, 2010). As a well characterised trait whose evolution is also well understood, salt tolerant plants may be more easily identifiable than drought tolerant plants and, thus, could be mapped onto a phylogeny to facilitate the identification of novel candidate genes involved in salt tolerance.

This raises a broader point about the need for novel approaches to develop stress tolerant crop varieties. Food security is a major global agenda (Godfray *et al.*, 2010). Coupled with a rise in population size are issues associated with climate change which is predicted to alter crop yields through pollinator decline, impacts from pests, pathogens, weeds and abiotic stresses (Myers *et al.*, 2017). Considering these factors, there is an emphasis to sustainably intensify agriculture, by producing greater yield from the same area of land (Garnett *et al.*, 2013; Godfray *et al.*, 2014).

A component of this will be producing crops with greater stress tolerance, guided by evidence from plant genomes. Recent genome sequencing of barley (*Hordeum vulgare L.*) and wheat (*Triticum spp.*) lines revealed the genomic diversity amongst crop lines and wild relatives with an interest in improving breeding programs. This genome data will provide insights into future crop cultivars, with increased yield, stress tolerance and adaptation to diverse environments (Jayakodi *et al.*, 2020; Walkowiak *et al.*, 2020). Future sequencing of plant genomes is predicted to accelerate the development of stress tolerant

crops by further identifying the genetic variation within crops (Lewin *et al.*, 2018). This highlights the significance of genome data and the revolutionary impact it will have for crop science research (Michael *et al.*, 2013) as well as the field of plant sciences more broadly including everything from evolution (Soltis *et al.*, 2020) to developmental biology (Sinha, 2011).

6.6 Future research

The genome data that supports the main body of this thesis was sourced in January 2018. At the time, there were key phylogenetic positions where genomic representation was missing. Since then, as highlighted by Fig. 1.2, genome availability has increased rapidly. Additional to the genomes highlighted in Section 6.2, future analysis would include genome data from the chlorophyte *Prasinoderma coloniale* (Li *et al.*, 2020b), the moss *Calohypnum plumiforme* (Mao *et al.*, 2020), the lycophyte *Selaginella lepidophylla* (VanBuren *et al.*, 2018) and the gymnosperm *Sequoiadendron giganteum* (Scott *et al.*, 2020) to improve the evolutionary resolution of analysis for non-flowering plant genomes. Evolutionary significant flowering plant genomes to be incorporated into future analysis would include the ANA grade angiosperm *Nymphaea colorata* (Zhang *et al.*, 2019a), several magnoliid genomes (Chaw *et al.*, 2019; Chen *et al.*, 2019; Rendón-Anaya *et al.*, 2019), the first Ceratophyllales genome (Yang *et al.*, 2020) and the early-diverging eudicot *Aquilegia coerulea* (Filiault *et al.*, 2018).

In part, this genomic revolution is supported by genome sequencing projects that aim to sequence the diversity of life. These include the 10KP project (Cheng *et al.*, 2018), aiming to sequence 10,000 diverse plant genomes by 2023, the Darwin Tree of Life Project (Wellcome Sanger Institute, 2020) that aims to sequence all 60,000 eukaryotic species in Britain and Ireland and finally the Earth BioGenome Project (Lewin *et al.*, 2018), which aims to sequence the genomes of all of Earth's 15 million eukaryotic species by 2028. The inclusion of new data from more diverse and representative plant species from these

genome sequencing projects would improve the resolution and detail of comparative genomic analysis.

Below are two examples of potential new insights for the major transitions in the plant tree of life from genome analysis. Firstly, as identified in the work in Chapter 2, the origin of land plants and Streptophyta were associated with two large groups of gene novelty. The last common ancestor of Streptophyta emerged approximately 700 million years ago (Morris *et al.*, 2018). This period in Earth's history is classified as the Cryogenian, a period in which the Earth was almost completely frozen, known as a snowball Earth (Brocks *et al.*, 2017). The divergence of streptophyte algae in this environment had important implications for the evolution of the Earth's atmosphere, through global oxygenation events (Lyons *et al.*, 2014; Hoffman *et al.*, 2017). Recent molecular phylogenetic analysis has placed the streptophyte algal lineage Zygnematophyceae as the sister group to land plants (Figure 1.1) (Wickett *et al.*, 2014). It would be interesting to investigate whether the ancestor of Zygnematophyceae and land plants were present on the ice surface in the Cryogenian and what were the biological innovations required for life in these environments including extremes of heat, UV radiation and a lack of water. Comparative genomic analysis with the extra streptophyte algal genomes highlighted above would provide detailed analysis of the processes of plant terrestrialisation, the genes that enabled this transition and the consequences for other eukaryotic life.

Secondly, gene duplication is a common origin of biological novelty. For example, recently, in the animal kingdom, it has been identified that gene duplication played an important role in metazoan evolution (Fernández *et al.*, 2020; Guijarro-Clarke *et al.*, 2020). Whole genome duplication, or polyploidy, is the process that creates an organism with an additional copy of its entire genome and is a common phenomenon in the plant kingdom (Leebens-Mack *et al.*, 2019). However, the frequency of whole genome duplication in plants is widely contested (Jiao *et al.*, 2011; Ruprecht *et al.*, 2017). Another explanation for large numbers of duplicated genes is high rates of gene duplication (Panchy *et al.*,

2016). Regardless of the origin of duplicated genes, they are able to facilitate the acquisition of novel components in the genetic toolkit (Moriyama *et al.*, 2018). After gene (or genome) duplication, additional genes are free to evolve novel functions as they are not essential to the plant's biology. When doubled genes evolve an advantageous function, this can lead to phenotypic and adaptive evolution. With this new genome data, it would interesting to assess gene family dynamics for the major transitions of plants, specifically investigating the prevalence of gene family expansion and contraction and biological implications of gene diversification.

Including more data to comparative genome analysis would likely change the results of HGs identified to have emerged during the major transitions in the plant phylogeny. This may lead to a separation of HGs at different plant nodes. For example, some Novel Core HGs found in all seed plants may in fact be present in all Euphyllophyta. Alternatively, extra genomic data may improve the definition of distinct HGs, providing more support for the clustering of genes within groups. Inclusion of more data may also answer questions about the evolution of genes important for water regulatory innovations. For example, as discussed above, there is debate around the origin of active stomatal control. Genomic representation for fern species may clarify the origin of stomatal signalling genes or patterns of stomatal gene diversification. This classification could provide insights into stomatal physiology and activity in ancestral plants under different atmospheres which may be informative for future stomatal physiology in atmospheres with greater CO_2 concentrations.

Additionally, the predicted influx of genome data will require appropriate analytical tools for comparative genomics and phylogenetics. These tools need to balance analytical accuracy with computational speed and data storage requirements. This work is beginning to be produced by the scientific community, e.g. the latest version of OrthoFinder provides a higher accuracy of orthology inference compared to earlier versions (Emms *et al.*, 2015, 2019). DIAMOND BLAST (Buchfink *et al.*, 2015) has become widely used to compare

sequences and is significantly computationally faster than BLAST (Altschul *et al.*, 1990), which is particularly relevant when working with large datasets. These technological advances and others to come will improve the speed and accuracy of comparative genome analysis, providing greater insight into the plant evolution.

6.7 Conclusion

Overall, analyses reported in this thesis have investigated the evolutionary dynamics of genes in the course of the evolutionary history of plants. Plants exhibited striking patterns of gene, genome and trait evolution. With the increasing wealth and diversity of genome data, intricate questions can be asked about the diversification of plants over the last billion years, the factors that can explain this diversity and the ways this information can be used to address applied biological questions. Major priorities for future research will be to understand the complex patterns of plant genome evolution in extensive detail and the implications for the major transitions in the plant tree of life.

Chapter 7 Bibliography

Abdel-Ghany, S. E., Ullah, F., Ben-Hur, A. and Reddy, A. S. N. (2020) Transcriptome analysis of drought-resistant and drought-sensitive sorghum (Sorghum bicolor) genotypes in response to peg-induced drought stress, *International Journal of Molecular Sciences*, **21**(3).

Adams, M. D. *et al.* (2000) The genome sequence of Drosophila melanogaster, *Science*, **287**(5461), pp. 2185–2195.

Agarwal, P. K., Gupta, K., Lopato, S. and Agarwal, P. (2017) Dehydration responsive element binding transcription factors and their applications for the engineering of stress tolerance, *Journal of Experimental Botany*, **68**(9), pp. 2135–2148.

Agrama, H. A. S. and Moussa, M. E. (1996) Mapping QTLs in breeding for drought tolerance in maize (Zea mays L.), *Euphytica*, **91**(1), pp. 89–97.

Al-Mssallem, I. S. *et al.* (2013) Genome sequence of the date palm Phoenix dactylifera L, *Nature Communications*, **4**.

Al-Naggar, A., El-Salam, R., Badran, A. and El-Moghazi, M. (2017) Drought tolerance of Five Quinoa (Chenopodium quinoa Willd.) Genotypes and Its Association with Other Traits under Moderate and Severe Drought Stress, *Asian Journal of Advances in Agricultural Research*, **3**(3), pp. 1–13.

Albert, R., Acharya, B. R., Jeon, B. W., Zañudo, J. G. T., Zhu, M., Osman, K. and Assmann, S. M. (2017) A new discrete dynamic model of ABA-induced stomatal closure predicts key feedback loops, *PLoS Biology*, **15**(9).

Albert, V. A. *et al.* (2013) The Amborella Genome and the Evolution of Flowering Plants, *Science*, **342**(6165).

Alföldi, J. *et al.* (2011) The genome of the green anole lizard and a comparative analysis with birds and mammals, *Nature*, **477**(7366), pp. 587–591.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool, *Journal of Molecular Biology*, **215**(3), pp. 403–410.

Aquea, F., Johnston, A. J., Cañon, P., Grossniklaus, U. and Arce-Johnson, P. (2010) TRAUCO, a Trithorax-group gene homologue, is required for early embryogenesis in Arabidopsis thaliana, *Journal of Experimental Botany*, **61**(4), pp. 1215–1224.

Armbrust, E. V. *et al.* (2004) The genome of the diatom Thalassiosira Pseudonana: Ecology, evolution, and metabolism, *Science*, **306**(5693), pp. 79–86.

Artur, M. A. S., Zhao, T., Ligterink, W., Schranz, E. and Hilhorst, H. W. M. (2019) Dissecting the Genomic Diversification of Late Embryogenesis Abundant (LEA) Protein Gene Families in Plants, *Genome Biology and Evolution*, **11**(2), pp. 459–471.

Ashraf, M. (2010) Inducing drought tolerance in plants: Recent advances, *Biotechnology Advances*, **28**(1), pp. 169–183.

Aury, J. M. *et al.* (2006) Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia, *Nature*, **444**(7116), pp. 171–178.

Avni, R. *et al.* (2017) Wild emmer genome architecture and diversity elucidate wheat evolution and domestication, *Science*, **357**(6346), pp. 93–97.

Azevedo Neto, A. D., Nogueira, R. J. M. C., Melo Filho, P. A. and Santos, R. C. (2010) Physiological and biochemical responses of peanut genotypes to water deficit, *Journal of Plant Interactions*, **5**(1), pp. 1–10.

Azua-Bustos, A., González-Silva, C., Arenas-Fajardo, C. and Vicuña, R. (2012) Extreme environments as potential drivers of convergent evolution by exaptation: The Atacama Desert Coastal Range case, *Frontiers in Microbiology*, **3**(426).

Badouin, H. *et al.* (2017) The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution, *Nature*, **546**(7656), pp. 148–152.

Bae, H., Kim, S. H., Kim, M. S., Sicher, R. C., Lary, D., Strem, M. D., Natarajan, S. and Bailey, B. A. (2008) The drought response of Theobroma cacao (cacao) and the regulation of genes involved in polyamine biosynthesis by drought and other stresses, *Plant Physiology and Biochemistry*, **46**(2), pp. 174–188.

Bairoch, A. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000, *Nucleic Acids Research*, **28**(1), pp. 45–48.

Banks, J. A. J. A. *et al.* (2011) The Selaginella Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants, *Science*, **332**(6032), pp. 960–963.

Bao, Y., Aggarwal, P., Robbins, N. E., Sturrock, C. J., Thompson, M. C., Tan, H. Q., Tham, C., Duan, L., Rodriguez, P. L., Vernoux, T., Mooney, S. J., Bennett, M. J. and Dinneny, J. R. (2014) Plant roots use a patterning mechanism to position lateral root

branches toward available water, *Proceedings of the National Academy of Sciences of the United States of America*, **111**(25), pp. 9319–9324.

Bao, Y., Song, W. M., Wang, P., Yu, X., Li, B., Jiang, C., Shiu, S. H., Zhang, H. and Bassham, D. C. (2020) COST1 regulates autophagy to control plant drought tolerance, *Proceedings of the National Academy of Sciences of the United States of America*, **117**(13), pp. 7482–7493.

Barba-Montoya, J., dos Reis, M., Schneider, H., Donoghue, P. C. J. and Yang, Z. (2018) Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution, *New Phytologist*, **218**(2), pp. 819–834.

Baydoun, E. A. H. and T. Bretr, C. (1985) Comparison of cell wall compositions of a desert xerophyte and a related mesophyte, *Phytochemistry*, **24**(7), pp. 1595–1597.

Bechtold, U. (2018) Plant life in extreme environments: How do you improve drought tolerance?, *Frontiers in Plant Science*, **9**(543).

Becker, B., Feng, X., Yin, Y., Holzinger, A. and Buschmann, H. (2020) Desiccation tolerance in streptophyte algae and the algae to land plant transition: Evolution of LEA and MIP protein families within the Viridiplantae, *Journal of Experimental Botany*, **71**(11), pp. 3270–3278.

Becker, B. and Marin, B. (2009) Streptophyte algae and the origin of embryophytes, *Annals of Botany*, **103**(7), pp. 999–1004.

Benkova, E. and Bielach, A. (2010) Lateral root organogenesis - from cell to organ, *Current Opinion in Plant Biology*, **13**(6), pp. 677–683.

Benoit, M., Drost, H. G., Catoni, M., Gouil, Q., Lopez-Gomollon, S., Baulcombe, D. and Paszkowski, J. (2019) Environmental and epigenetic regulation of Rider retrotransposons in tomato, *PLoS Genetics*, **15**(9).

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E. and Huala, E. (2015) The arabidopsis information resource: Making and mining the ‘gold standard’ annotated reference plant genome, *Genesis*, **53**(8), pp. 474–485.

Bergonci, T., Ribeiro, B., Ceciliato, P. H. O., Guerrero-Abad, J. C., Silva-Filho, M. C. and Moura, D. S. (2014) Arabidopsis thaliana RALF1 opposes brassinosteroid effects on root cell elongation and lateral root formation, *Journal of Experimental Botany*, **65**(8), pp.

2219–2230.

Bertioli, D. J. *et al.* (2016) The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut, *Nature Genetics*, **48**(4), pp. 438–446.

Bertolini, E., Verelst, W., Horner, D. S., Gianfranceschi, L., Piccolo, V., Inzé, D., Pè, M. E. and Mica, E. (2013) Addressing the role of microrna s in reprogramming leaf growth during drought stress in brachypodium distachyon, *Molecular Plant*, **6**(2), pp. 423–443.

Betts, H. C., Puttick, M. N., Clark, J. W., Williams, T. A., Donoghue, P. C. J. and Pisani, D. (2018) Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin, *Nature Ecology and Evolution*, **2**(10), pp. 1556–1562.

Bhattacharya, D., Price, D. C., Xin Chan, C., Qiu, H., Rose, N., Ball, S., Weber, A. P. M., Cecilia Arias, M., Henrissat, B., Coutinho, P. M., Krishnan, A., Zäuner, S., Morath, S., Hilliou, F., Egizi, A., Perrineau, M. M. and Yoon, H. S. (2013) Genome of the red alga Porphyridium purpureum, *Nature Communications*, **4**(1941).

Biaolin, H., Shouwu, Y., Yong, W., Zheng, Z., Bingyu, Q. and Jiankun, X. (2010) Drought-resistance identification of dongxiang common wild rice Oryza rufipogon Griff. in whole growth period, *Acta Agronomica Sinica*, **33**(3), pp. 425–432.

Bimpong, I. K., Serraj, R., Chin, J. H., Ramos, J., Mendoza, E. M. T., Hernandez, J. E., Mendioro, M. S. and Brar, D. S. (2011) Identification of QTLs for Drought-Related Traits in Alien Introgression Lines Derived from Crosses of Rice (Oryza sativa cv. IR64) × O. glaberrima under Lowland Moisture Stress, *Journal of Plant Biology*, **54**(4), pp. 237–250.

Blanc, G. *et al.* (2012) The genome of the polar eukaryotic microalga Coccomyxa subellipsoidea reveals traits of cold adaptation, *Genome Biology*, **13**(5).

Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J., Salamov, A., Terry, A., Yamada, T., Dunigan, D. D., Grigoriev, I. V., Claverie, J. M. and van Etten, J. L. (2010) The Chlorella variabilis NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex, *Plant Cell*, **22**(9), pp. 2943–2955.

Bobrownyzky, J. (2016) Production of branched root hairs under progressive drought stress in Arabidopsis thaliana, *Cytology and Genetics*, **50**(5), pp. 324–329.

Boguszewska-Mańkowska, D., Pieczyński, M., Wyrzykowska, A., Kalaji, H. M., Sieczko, L., Szweykowska-Kulińska, Z. and Zagdańska, B. (2018) Divergent strategies displayed by potato (*Solanum tuberosum* L.) cultivars to cope with soil drought, *Journal of Agronomy and Crop Science*, **204**(1), pp. 13–30.

Bolger, A. *et al.* (2014) The genome of the stress-tolerant wild tomato species Solanum pennellii, *Nature Genetics*, **46**(9), pp. 1034–1038.

Bombarely, A. *et al.* (2016) Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida, *Nature Plants*, **2**(16074).

Boulc'h, P.-N., Caullireau, E., Faucher, E., Gouerou, M., Guérin, A., Miray, R. and Couée, I. (2020) Abiotic stress signalling in extremophile land plants, *Journal of Experimental Botany*, **71**(19), pp. 5771–5785.

Bowler, C. *et al.* (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes, *Nature*, **456**(7219), pp. 239–244.

Bowles, A. M. C., Bechtold, U. and Paps, J. (2020) The Origin of Land Plants Is Rooted in Two Bursts of Genomic Novelty, *Current Biology*, **30**(3), pp. 530–536.

Bowman, J. L. *et al.* (2017) Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome, *Cell*, **171**(2), pp. 287–304.

Bowman, J. L., Briginshaw, L. N., Fisher, T. J. and Flores-Sandoval, E. (2019) Something ancient and something neofunctionalized—evolution of land plant hormone signaling pathways, *Current Opinion in Plant Biology*, **47**, pp. 64–72.

Boyce, C. K., Brodribb, T. J., Feild, T. S. and Zwieniecki, M. A. (2009) Angiosperm leaf vein evolution was physiologically and environmentally transformative, *Proceedings of the Royal Society B: Biological Sciences*, **276**(1663), pp. 1771–1776.

Bräutigam, A., Schlüter, U., Eisenhut, M. and Gowik, U. (2017) On the evolutionary origin of CAM photosynthesis, *Plant Physiology*, **174**(2), pp. 473–477.

Brawley, S. H. *et al.* (2017) Insights into the red algae and eukaryotic evolution from the genome of Porphyra umbilicalis (Bangiophyceae, Rhodophyta), *Proceedings of the National Academy of Sciences of the United States of America*, **114**(31), pp. 6361–6370.

Bremer, B. *et al.* (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II, *Botanical Journal of the Linnean*

Society, **141**(4), pp. 399–436.

Bremer, B., Bremer, K., Chase, M. W., Fay, M. F., Reveal, J. L., Bailey, L. H., Soltis, D. E., Soltis, P. S., Stevens, P. F., Anderberg, A. A., Moore, M. J., Olmstead, R. G., Rudall, P. J., Sytsma, K. J., Tank, D. C., Wurdack, K., Xiang, J. Q. Y. and Zmarzty, S. (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III, *Botanical Journal of the Linnean Society*, **161**(2), pp. 105–121.

Bremer, K. (1998) An ordinal classification for the families of flowering plants the angiosperin phytogeny group, *Annals of the Missouri Botanical Garden*, **85**(4), pp. 531–533.

Briggs, G. C., Mouchel, C. F. and Hardtke, C. S. (2006) Characterization of the Plant-Specific BREVIS RADIX Gene Family Reveals Limited Genetic Redundancy Despite High Sequence Conservation, *Plant Physiology*, **140**(4), pp. 1306–1316.

Broad Institute (2009) *Fonticula alba str. ATCC 38817*. Available at: http://protists.ensembl.org/Fonticula_alba_gca_000388065/Info/Index.

Broad Institute (2010) *Allomyces macrogynus ATCC 38327*. Available at: https://www.ncbi.nlm.nih.gov/bioproject/20563.

Brockington, S. F., Yang, Y., Gandia-Herrero, F., Covshoff, S., Hibberd, J. M., Sage, R. F., Wong, G. K. S., Moore, M. J. and Smith, S. A. (2015) Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales, *New Phytologist*, **207**(4), pp. 1170–1180.

Brocks, J. J., Jarrett, A. J. M., Sirantoine, E., Hallmann, C., Hoshino, Y. and Liyanage, T. (2017) The rise of algae in Cryogenian oceans and the emergence of animals, *Nature*, **548**, pp. 578–581.

Brodribb, T. J., Carriquí, M., Delzon, S., McAdam, S. A. M. M. and Holbrook, N. M. (2020) Advanced vascular function discovered in a widespread moss, *Nature Plants*, **6**(3), pp. 273–279.

Brodribb, T. J. and McAdam, S. A. M. (2017) Evolution of the Stomatal Regulation of Plant Water Content, *Plant Physiology*, **174**(2), pp. 639–649.

Brodribb, T. J. and McAdam, S. A. M. M. (2011) Passive origins of stomatal control in vascular plants, *Science*, **331**(6017), pp. 582–585.

Buchfink, B., Xie, C. and Huson, D. H. (2015) Fast and sensitive protein alignment using DIAMOND, *Nature Methods*, **12**, pp. 59–60.

Burki, F., Roger, A. J., Brown, M. W. and Simpson, A. G. B. (2020) The New Tree of Eukaryotes, *Trends in Ecology and Evolution*, **35**(1), pp. 43–55.

Butterfield, N. J. (2000) Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes, *Paleobiology*, **26**(3), pp. 386–404.

Butterfield, N. J., Knoll, A. H. and Swett, K. (1994) Paleobiology of the Neoproterozoic Svanbergfjellet Formation, Spitsbergen, *Fossils and Strata*, **27**(1), pp. 76–76.

Byrne, S. L., Erthmann, P. Ø., Agerbirk, N., Bak, S., Hauser, T. P., Nagy, I., Paina, C. and Asp, T. (2017) The genome sequence of Barbarea vulgaris facilitates the study of ecological biochemistry, *Scientific Reports*, **7**(40728).

Byrne, S. L., Nagy, I., Pfeifer, M., Armstead, I., Swain, S., Studer, B., Mayer, K., Campbell, J. D., Czaban, A., Hentrup, S., Panitz, F., Bendixen, C., Hedegaard, J., Caccamo, M. and Asp, T. (2015) A synteny-based draft genome sequence of the forage grass *Lolium perenne*, *The Plant Journal*, **84**(4), pp. 816–826.

Caarls, L., Pieterse, C. M. J. and Van Wees, S. C. M. (2015) How salicylic acid takes transcriptional control over jasmonic acid signaling, *Frontiers in Plant Science*, **6**(170).

Cai, S. *et al.* (2017) Evolutionary Conservation of ABA Signaling for Stomatal Closure, *Plant Physiology*, **174**(2), pp. 732–747.

Cannarozzi, G., Plaza-Wüthrich, S., Esfeld, K., Larti, S., Wilson, Y., Girma, D., de Castro, E., Chanyalew, S., Blösch, R., Farinelli, L., Lyons, E., Schneider, M., Falquet, L., Kuhlemeier, C., Assefa, K. and Tadele, Z. (2014) Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (Eragrostis tef), *BMC Genomics*, **15**(1), p. 581.

Cannell, N., Emms, D. M., Hetherington, A. J., MacKay, J., Kelly, S., Dolan, L. and Sweetlove, L. J. (2020) Multiple Metabolic Innovations and Losses Are Associated with Major Transitions in Land Plant Evolution, *Current Biology*, **30**(10), pp. 1783–1800.

Cao, H., Wang, L., Nawaz, M. A., Niu, M., Sun, J., Xie, J., Kong, Q., Huang, Y., Cheng, F. and Bie, Z. (2017) Ectopic Expression of Pumpkin NAC Transcription Factor CmNAC1

Improves Multiple Abiotic Stress Tolerance in Arabidopsis, *Frontiers in Plant Science*, **8**(2052).

Cao, Y., Meng, D., Chen, Y., Abdullah, M., Jin, Q., Lin, Y. and Cai, Y. (2018) Comparative and Expression Analysis of Ubiquitin Conjugating Domain-Containing Genes in Two Pyrus Species, *Cells*, **7**(7).

Capella-Gutiérrez, S., Silla-Martínez, J. M. and Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, *Bioinformatics*, **25**(15), pp. 1972–3.

Catarino, B., Hetherington, A. J., Emms, D. M., Kelly, S. and Dolan, L. (2016) The Stepwise Increase in the Number of Transcription Factor Families in the Precambrian Predated the Diversification of Plants On Land, *Molecular Biology and Evolution*, **33**(11), pp. 2815–2819.

Catola, S., Marino, G., Emiliani, G., Huseynova, T., Musayev, M., Akparov, Z. and Maserti, B. E. (2016) Physiological and metabolomic analysis of Punica granatum (L.) under drought stress, *Planta*, **243**(2), pp. 441–449.

Cattivelli, L., Rizza, F., Badeck, F. W., Mazzucotelli, E., Mastrangelo, A. M., Francia, E., Marè, C., Tondelli, A. and Stanca, A. M. (2008) Drought tolerance improvement in crop plants: An integrated view from breeding to genomics, *Field Crops Research*, pp. 1–14.

Chagné, D. *et al.* (2014) The Draft Genome Sequence of European Pear (Pyrus communis L. 'Bartlett'), *PLoS ONE*, **9**(4).

Chakraborty, N., Kanyuka, K., Jaiswal, D. K., Kumar, A., Arora, V., Malik, A., Gupta, N., Hooley, R. and Raghuram, N. (2019) GCR1 and GPA1 coupling regulates nitrate, cell wall, immunity and light responses in Arabidopsis, *Scientific Reports*, **9**(1).

Chakraborty, N., Singh, N., Kaur, K. and Raghuram, N. (2015) G-protein Signaling Components GCR1 and GPA1 Mediate Responses to Multiple Abiotic Stresses in Arabidopsis, *Frontiers in Plant Science*, **6**(1000).

Chalhoub, B. *et al.* (2014) Early allopolyploid evolution in the post-neolithic Brassica napus oilseed genome, *Science*, **345**(6199), pp. 950–953.

Chan, A. P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., Melake-Berhan, A., Jones, K. M., Redman, J., Chen, G., Cahoon, E. B., Gedil, M., Stanke, M., Haas, B. J.,

Wortman, J. R., Fraser-Liggett, C. M., Ravel, J. and Rabinowicz, P. D. (2010) Draft genome sequence of the oilseed species Ricinus communis, *Nature Biotechnology*, **28**(9), pp. 951–956.

Channing, A. and Edwards, D. (2009) Yellowstone hot spring environments and the palaeo-ecophysiology of rhynie chert plants: Towards a synthesis, *Plant Ecology and Diversity*, **2**(2), pp. 111–143.

Chase, M. W. *et al.* (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV, *Botanical Journal of the Linnean Society*, **181**(1), pp. 1–20.

Chater, C. C., Caine, R. S., Fleming, A. J. and Gray, J. E. (2017) Origins and Evolution of Stomatal Development, *Plant Physiology*, **174**(2), pp. 624–638.

Chater, C. C., Caine, R. S., Tomek, M., Wallace, S., Kamisugi, Y., Cuming, A. C., Lang, D., MacAlister, C. A., Casson, S., Bergmann, D. C., Decker, E. L., Frank, W., Gray, J. E., Fleming, A., Reski, R. and Beerling, D. J. (2016) Origin and function of stomata in the moss Physcomitrella patens, *Nature Plants*, **2**(12), pp. 1–7.

Chaw, S.-M., Liu, Y.-C., Wu, Y.-W., Wang, H.-Y., Lin, C.-Y. I., Wu, C.-S., Ke, H.-M., Chang, L.-Y., Hsu, C.-Y., Yang, H.-T., Sudianto, E., Hsu, M.-H., Wu, K.-P., Wang, L.-N., Leebens-Mack, J. H. and Tsai, I. J. (2019) Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution, *Nature Plants*, **5**, pp. 63–73.

Chen, F., Dong, W., Zhang, J., Guo, X., Chen, J., Wang, Z., Lin, Z., Tang, H. and Zhang, L. (2018) The Sequenced Angiosperm Genomes and Genome Databases, *Frontiers in Plant Science*, **9**(418).

Chen, J. *et al.* (2013a) Whole-genome sequencing of Oryza brachyantha reveals mechanisms underlying Oryza genome evolution, *Nature Communications*, **4**(1595).

Chen, J. *et al.* (2019) Liriodendron genome sheds light on angiosperm phylogeny and species–pair differentiation, *Nature Plants*, **5**(1), pp. 18–25.

Chen, X. *et al.* (2016) Draft genome of the peanut A-genome progenitor (Arachis duranensis) provides insights into geocarpy, oil biosynthesis, and allergens, *Proceedings of the National Academy of Sciences of the United States of America*, **113**(24), pp. 6785–6790.

Chen, Y., Liu, Z. H., Feng, L., Zheng, Y., Li, D. Di and Li, X. B. (2013b) Genome-wide functional analysis of cotton (Gossypium hirsutum) in response to drought, *PLoS ONE*, **8**(11).

Cheng, S. *et al.* (2013) The Tarenaya hassleriana genome provides insight into reproductive trait and genome evolution of crucifers, *Plant Cell*, **25**(8), pp. 2813–2830.

Cheng, S. *et al.* (2019) Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution, *Cell*, **179**(5), pp. 1057–1067.

Cheng, S., Melkonian, M., Smith, S. A., Brockington, S., Archibald, J. M., Delaux, P. M., Li, F. W., Melkonian, B., Mavrodiev, E. V., Sun, W., Fu, Y., Yang, H., Soltis, D. E., Graham, S. W., Soltis, P. S., Liu, X., Xu, X. and Wong, G. K. S. (2018) 10KP: A phylodiverse genome sequencing plan, *GigaScience*, **7**(3), pp. 1–9.

Cheng, T. (2011) NaCl-induced responses in giant duckweed (Spirodela polyrhiza), *Journal of Aquatic Plant Management*, **49**(2), pp. 62–71.

Cheplick, G. P., Perera, A. and Koulouris, K. (2000) Effect of drought on the growth of Lolium perenne genotypes with and without fungal endophytes, *Functional Ecology*, **14**(6), pp. 657–667.

Cheserek, J. J. and Gichimu, B. M. (2012) Drought and heat tolerance in coffee: a review, *International Research Journal of Agricultural Science and Soil Science*, **2**(12), pp. 498–501.

Chikina, M., Robinson, J. D. and Clark, N. L. (2016) Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals, *Molecular Biology and Evolution*, **33**(9), pp. 2182–2192.

Chin, C. F. and Chee, J. Y. (2015) Gateway Cloning Technology: Advantages and Drawbacks, *Cloning & Transgenesis*, **4**(1).

Cho, S. M., Lee, H., Jo, H., Lee, H., Kang, Y., Park, H. and Lee, J. (2018) Comparative transcriptome analysis of field- and chamber-grown samples of Colobanthus quitensis (Kunth) Bartl, an Antarctic flowering plant, *Scientific Reports*, **8**(1), p. 11049.

Christenhusz, M. J. M. and Byng, J. W. (2016) The number of known plant species in the world and its annual increase, *Phytotaxa*, **261**(3), pp. 201–217.

Chung, Y. and Choe, S. (2013) The Regulation of Brassinosteroid Biosynthesis in

Arabidopsis, *Critical Reviews in Plant Sciences*, **32**(6), pp. 396–410.

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, J. W. and Donoghue, P. C. J. (2018) Whole-Genome Duplication and Plant Macroevolution, *Trends in Plant Science*, **23**(10), pp. 933–945.

Clarke, M. *et al.* (2013) Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling, *Genome Biology*, **14**(2).

De Clerck, O. *et al.* (2018) Insights into the Evolution of Multicellularity from the Sea Lettuce Genome, *Current Biology*, **28**(18), pp. 2921–2933.

Clough, S. J. and Bent, A. F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana, *The Plant Journal*, **16**(6), pp. 735–43.

Clouse, J. W., Adhikary, D., Page, J. T., Ramaraj, T., Deyholos, M. K., Udall, J. A., Fairbanks, D. J., Jellen, E. N. and Maughan, P. J. (2016) The Amaranth Genome: Genome, Transcriptome, and Physical Map Assembly, *The Plant Genome*, **9**(1).

Clouse, S. D. (2011) Brassinosteroid Signal Transduction: From Receptor Kinase Activation to Transcriptional Networks Regulating Plant Development, *Plant Cell*, **23**(4), pp. 1219–1230.

CoGe: Comparative Genomics (2020). Available at: https://genomevolution.org/coge/ (Accessed: 2 December 2019).

Colina, F., Amaral, J., Carbó, M., Pinto, G., Soares, A., Cañal, M. J. and Valledor, L. (2019) Genome-wide identification and characterization of CKIN/SnRK gene family in Chlamydomonas reinhardtii, *Scientific Reports*, **9**(1), pp. 1–16.

Collén, J. *et al.* (2013) Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida, *Proceedings of the National Academy of Sciences of the United States of America*, **110**(13), pp. 5247–5252.

Copetti, D., Búrquez, A., Bustamante, E., Charboneau, J. L. M., Childs, K. L., Eguiarte, L. E., Lee, S., Liu, T. L., McMahon, M. M., Whiteman, N. K., Wing, R. A., Wojciechowski, M. F. and Sanderson, M. J. (2017) Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti, *Proceedings of the National*

Academy of Sciences of the United States of America, **114**(45), pp. 12003–12008.

Corlett, R. T. (2016) Plant diversity in a changing world: Status, trends, and conservation needs, *Plant Diversity*, **38**(1), pp. 10–16.

Cortés, A. J., Monserrate, F. A., Ramírez-Villegas, J., Madriñán, S. and Blair, M. W. (2013) Drought Tolerance in Wild Plant Populations: The Case of Common Beans (Phaseolus vulgaris L.), *PLoS ONE*, **8**(5).

Costello, R., Emms, D. M. and Kelly, S. (2020) Gene duplication accelerates the pace of protein gain and loss from plant organelles, *Molecular Biology and Evolution*, **37**(4), pp. 969–981.

Cotelle, V. and Leonhardt, N. (2016) 14-3-3 proteins in guard cell signaling, *Frontiers in Plant Science*, **6**(1210).

Crane, P. R., Friis, E. M. and Pedersen, K. R. (1995) The origin and early diversification of angiosperms, *Nature*, **374**(6517), pp. 27–33.

Cruz, F., Julca, I., Gómez-Garrido, J., Loska, D., Marcet-Houben, M., Cano, E., Galán, B., Frias, L., Ribeca, P., Derdak, S., Gut, M., Sánchez-Fernández, M., García, J. L., Gut, I. G., Vargas, P., Alioto, T. S. and Gabaldón, T. (2016) Genome sequence of the olive tree, Olea europaea, *GigaScience*, **5**(29).

Cruz, Z. N., Rodríguez, P., Galindo, A., Torrecillas, E., Ondoño, S., Mellisho, C. D. and Torrecillas, A. (2012) Leaf mechanisms for drought resistance in Zizyphus jujuba trees, *Plant Science*, **197**, pp. 77–83.

Curtis, B. A. *et al.* (2012) Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs, *Nature*, **492**(7427), pp. 59–65.

D'hont, A. *et al.* (2012) The banana (Musa acuminata) genome and the evolution of monocotyledonous plants, *Nature*, **488**(7410), pp. 213–217.

Darwin, C. (1859) *On The Origin of Species by Means of Natural Selection, or The Preservation of Favoured Races in the Struggle for Life*. London, UK: John Murray.

Dash, P. K., Cao, Y., Jailani, A. K., Gupta, P., Venglat, P., Xiang, D., Rai, R., Sharma, R., Thirunavukkarasu, N., Abdin, M. Z., Yadava, D. K., Singh, N. K., Singh, J., Selvaraj, G., Deyholos, M., Kumar, P. A. nand. and Datla, R. (2014) Genome-wide analysis of drought induced gene expression changes in flax (Linum usitatissimum), *GM Crops &*

Food, **5**(2), pp. 106–119.

Dassanayake, M., Oh, D. H., Haas, J. S., Hernandez, A., Hong, H., Ali, S., Yun, D. J., Bressan, R. A., Zhu, J. K., Bohnert, H. J. and Cheeseman, J. M. (2011) The genome of the extremophile crucifer Thellungiella parvula, *Nature Genetics*, **43**(9), pp. 913–918.

Davey, M. W., Gudimella, R., Harikrishna, J. A., Sin, L. W., Khalid, N. and Keulemans, J. (2013) A draft Musa balbisiana genome sequence for molecular genetics in polyploid, inter- and intra-specific Musa hybrids, *BMC Genomics*, **14**(1), p. 683.

Davies, T. J., Barraclough, T. G., Chase, M. W., Soltis, P. S., Soltis, D. E. and Savolainen, V. (2004) Darwin's abominable mystery: Insights from a supertree of the angiosperms, *Proceedings of the National Academy of Sciences of the United States of America*, **101**(7), pp. 1904–1909.

Degu, H. D., Ohta, M. and Fujimura, T. (2008) Drought tolerance of Eragrostis tef and development of roots, *International Journal of Plant Sciences*, **169**(6), pp. 768–775.

Delahaie, J., Hundertmark, M., Bove, J., Leprince, O., Rogniaux, H. and Buitink, J. (2013) LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-regulated LEA protein abundance linked to desiccation tolerance, *Journal of Experimental Botany*, **64**(14), pp. 4559–4573.

Delaux, P. M. *et al.* (2015) Algal ancestor of land plants was preadapted for symbiosis, *Proceedings of the National Academy of Sciences of the United States of America*, **112**(43), pp. 13390–13395.

Delaux, P. M., Hetherington, A. J., Coudert, Y., Delwiche, C., Dunand, C., Gould, S., Kenrick, P., Li, F. W., Philippe, H., Rensing, S. A., Rich, M., Strullu-Derrien, C. and de Vries, J. (2019) Reconstructing trait evolution in plant evo–devo studies, *Current Biology*, **29**(21), pp. 1110–1118.

Delwiche, C. F. and Cooper, E. D. (2015) The evolutionary origin of a terrestrial flora, *Current Biology*, **25**(19), pp. 899–910.

Denoeud, F. *et al.* (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis, *Science*, **345**(6201), pp. 1181–1184.

DePamphilis, C. W. *et al.* (2013) The Amborella genome and the evolution of flowering plants, *Science*, **342**(6165).

Derelle, E. *et al.* (2006) Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features, *Proceedings of the National Academy of Sciences of the United States of America*, **103**(31), pp. 11647–11652.

Dietrich, D. *et al.* (2017) Root hydrotropism is controlled via a cortex-specific growth mechanism, *Nature Plants*, **3**(6), pp. 1–8.

Dietrich, D. (2018) Hydrotropism: how roots search for water, *Journal of Experimental Botany*, **69**(11), pp. 2759–2771.

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. and Thermes, C. (2018) The Third Revolution in Sequencing Technology, *Trends in Genetics*, **34**(9), pp. 666–681.

Dinakar, C. and Bartels, D. (2013) Desiccation tolerance in resurrection plants: New insights from transcriptome, proteome, and metabolome analysis, *Frontiers in Plant Science*, **4**(482).

Dobrowolska, D., Hein, S., Oosterbaan, A., Wagner, S., Clark, J. and Skovsgaard, J. P. (2011) A review of European ash (Fraxinus excelsior L.): implications for silviculture, *Forestry*, **84**(2), pp. 133–148.

Doebley, J. F., Gaut, B. S. and Smith, B. D. (2006) The Molecular Genetics of Crop Domestication, *Cell*, **127**(7), pp. 1309–1321.

Dohm, J. C. *et al.* (2014) The genome of the recently domesticated crop plant sugar beet (Beta vulgaris), *Nature*, **505**(7484), pp. 546–549.

Dombroski, J. L. D., Freitas, R. M. O. de, Tomczak, V. E., Pinto, J. R. de S. and Farias, R. M. de (2014) Ecophysiology of water stressed Handroanthus impetiginosus (Mart. Ex. DC) Mattos) Seedlings, *Scientia Forestalis/Forest Sciences*, **42**(101), pp. 155–163.

Donoghue, P. and Paps, J. (2020) Plant Evolution: Assembling Land Plants, *Current Biology*, **30**(2), pp. 81–83.

Dossa, K., Li, D., Wang, L., Zheng, X., Yu, J., Wei, X., Fonceka, D., Diouf, D., Liao, B., Cisse, N. and Zhang, X. (2017) Dynamic transcriptome landscape of sesame (Sesamum indicum L.) under progressive drought and after rewatering, *Genomics Data*, **11**, pp. 122–124.

Doyle, J. A. (2017) Phylogenetic Analyses and Morphological Innovations in Land Plants, in *Annual Plant Reviews online*. Chichester, UK: John Wiley & Sons, Ltd, pp. 1–

50.

Duckett, J. G. and Pressel, S. (2018) The evolution of the stomatal apparatus: Intercellular spaces and sporophyte water relations in bryophytes—two ignored dimensions, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **373**(1739).

Dunn, C. W. and Munro, C. (2016) Comparative genomics and the diversity of life, *Zoologica Scripta*, **45**(1), pp. 5–13.

Dunning, L. T., Olofsson, J. K., Parisod, C., Choudhury, R. R., Moreno-Villena, J. J., Yang, Y., Dionora, J., Paul Quick, W., Park, M., Bennetzen, J. L., Besnard, G., Nosil, P., Osborne, C. P. and Christin, P. A. (2019) Lateral transfers of large DNA fragments spread functional genes among grasses, *Proceedings of the National Academy of Sciences of the United States of America*, **116**(10), pp. 4416–4425.

Dunwell, T. L., Paps, J. and Holland, P. W. H. (2017) Novel and divergent genes in the evolution of placental mammals, *Proceedings of the Royal Society B: Biological Sciences*, **284**(1864).

Egea, I., Albaladejo, I., Meco, V., Morales, B., Sevilla, A., Bolarin, M. C. and Flores, F. B. (2018) The drought-tolerant Solanum pennellii regulates leaf water loss and induces genes involved in amino acid and ethylene/jasmonate metabolism under dehydration, *Scientific Reports*, **8**(1).

Eichinger, I. *et al.* (2005) The genome of the social amoeba Dictyostelium discoideum, *Nature*, **435**(7038), pp. 43–57.

Elbarbary, R. A., Lucas, B. A. and Maquat, L. E. (2016) Retrotransposons as regulators of gene expression, *Science*, **351**(6274).

Eldem, V., Çelikkol Akçay, U., Ozhuner, E., Bakır, Y., Uranbey, S. and Unver, T. (2012) Genome-Wide Identification of miRNAs Responsive to Drought in Peach (Prunus persica) by High-Throughput Deep Sequencing, *PLoS ONE*, **7**(12).

EMBL-EBI (2010) *Thecamonas trahens ATCC 50062*. Available at: https://protists.ensembl.org/Thecamonas_trahens_atcc_50062_gca_000142905/Info/Index.

Emms, D. M. and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole

genome comparisons dramatically improves orthogroup inference accuracy, *Genome Biology*, **16**(157).

Emms, D. M. and Kelly, S. (2019) OrthoFinder: Phylogenetic orthology inference for comparative genomics, *Genome Biology*, **20**(238).

Endress, P. K. (2011) Evolutionary diversification of the flowers in angiosperms, *American Journal of Botany*, **98**(3), pp. 370–396.

Enright, A. J., Van Dongen, S. and Ouzounis, C. A. (2002) An efficient algorithm for large-scale detection of protein families, *Nucleic Acids Research*, **30**(7), pp. 1575–84.

Eriksson, O. (2016) Evolution of angiosperm seed disperser mutualisms: The timing of origins and their consequences for coevolutionary interactions between angiosperms and frugivores, *Biological Reviews*, **91**(1), pp. 168–186.

Fahad, S., Bajwa, A. A., Nazir, U., Anjum, S. A., Farooq, A., Zohaib, A., Sadia, S., Nasim, W., Adkins, S., Saud, S., Ihsan, M. Z., Alharby, H., Wu, C., Wang, D. and Huang, J. (2017) Crop production under drought and heat stress: Plant responses and management options, *Frontiers in Plant Science*, **8**(1147).

Fairclough, S. R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H. M., Begovic, E., Richter, D. J., Russ, C., Westbrook, M. J., Manning, G., Lang, B. F., Haas, B., Nusbaum, C. and King, N. (2013) Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta, *Genome Biology*, **14**(2), pp. 1–15.

Fang, L., Leliaert, F., Zhang, Z. H., Penny, D. and Zhong, B. J. (2017) Evolution of the Chlorophyta: Insights from chloroplast phylogenomic analyses, *Journal of Systematics and Evolution*, **55**(4), pp. 322–332.

Federhen, S. (2012) The NCBI Taxonomy database, *Nucleic Acids Research*, **40**, pp. 136–143.

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach, *Journal of Molecular Evolution*, **17**(6), pp. 368–376.

Fernandez, A., Hilson, P. and Beeckman, T. (2013) Golven peptides as important regulatory signalling molecules of plant development, *Journal of Experimental Botany*, **64**(17), pp. 5263–5268.

Fernández, R. and Gabaldón, T. (2020) Gene gain and loss across the metazoan tree of life, *Nature Ecology and Evolution*, **4**, pp. 524–533.

Feuda, R., Dohrmann, M., Pett, W., Philippe, H., Rota-Stabelli, O., Lartillot, N., Wörheide, G. and Pisani, D. (2017) Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to All Other Animals, *Current Biology*, **27**(24), pp. 3864–3870.

Field, K. J., Pressel, S., Duckett, J. G., Rimington, W. R. and Bidartondo, M. I. (2015) Symbiotic options for the conquest of land, *Trends in Ecology & Evolution*, **30**(8), pp. 477–486.

Filiault, D. L. *et al.* (2018) The Aquilegia genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history, *eLife*, **7**.

Finet, C., Berne-Dedieu, A., Scutt, C. P. and Marlétaz, F. (2013) Evolution of the ARF Gene Family in Land Plants: Old Domains, New Tricks, *Molecular Biology and Evolution*, **30**(1), pp. 45–56.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J. and Punta, M. (2014) Pfam: The protein families database, *Nucleic Acids Research*, **42**(1).

Finnegan, D. J. (2012) Retrotransposons, *Current Biology*, **22**(11).

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

Flowers, T. J., Galal, H. K. and Bromham, L. (2010) Evolution of halophytes: Multiple origins of salt tolerance in land plants, *Functional Plant Biology*, **37**(7), pp. 604–612.

Foflonker, F., Price, D. C., Qiu, H., Palenik, B., Wang, S. and Bhattacharya, D. (2015) Genome of the halotolerant green alga *P icochlorum* sp. reveals strategies for thriving under fluctuating environmental conditions, *Environmental Microbiology*, **17**(2), pp. 412–426.

Fonouni-Farde, C., Miassod, A., Laffont, C., Morin, H., Bendahmane, A., Diet, A. and Frugier, F. (2019) Gibberellins negatively regulate the development of Medicago truncatula root system, *Scientific Reports*, **9**(1).

Franks, P. J. and Farquhar, G. D. (2007) The mechanical diversity of stomata and its

significance in gas-exchange control, *Plant Physiology*, **143**(1), pp. 78–87.

Freschi, L. (2013) Nitric oxide and phytohormone interactions: current status and perspectives, *Frontiers in Plant Science*, **4**(398).

Fritz-Laylin, L. K. *et al.* (2010) The Genome of Naegleria gruberi Illuminates Early Eukaryotic Versatility, *Cell*, **140**(5), pp. 631–642.

Fu, Y. *et al.* (2017) Draft genome sequence of the Tibetan medicinal herb Rhodiola crenulata, *GigaScience*, **6**(6), pp. 1–5.

Fukushima, K. *et al.* (2017) Genome of the pitcher plant Cephalotus reveals genetic changes associated with carnivory, *Nature Ecology and Evolution*, **1**(3).

Fürst-Jansen, J. M. R., de Vries, S. and Vries, J. de (2020) Evo-physio: on stress responses and the earliest land plants, *Journal of Experimental Botany*, **71**(11), pp. 3254–3269.

Gaff, D. F. (1971) Desiccation-tolerant flowering plants in Southern Africa, *Science*, **174**(4013), pp. 1033–1034.

Gaff, D. F. and Oliver, M. (2013) The evolution of desiccation tolerance in angiosperm plants: A rare yet common phenomenon, *Functional Plant Biology*, **40**(4), pp. 315–328.

Gambetta, G. A., Herrera, J. C., Dayer, S., Feng, Q., Hochberg, U. and Castellarin, S. D. (2020) The physiology of drought stress in grapevine: Towards an integrative definition of drought tolerance, *Journal of Experimental Botany*, **71**(16), pp. 4658–4676.

Gao, B., Li, X., Zhang, D., Liang, Y., Yang, H., Chen, M., Zhang, Y., Zhang, J. and Wood, A. J. (2017) Desiccation tolerance in bryophytes: The dehydration and rehydration transcriptomes in the desiccation-tolerant bryophyte Bryum argenteum, *Scientific Reports*, **7**(1).

Gao, C., Wang, Y., Shen, Y., Yan, D., He, X., Dai, J. and Wu, Q. (2014) Oil accumulation mechanisms of the oleaginous microalga Chlorella protothecoides revealed through its genome, transcriptomes, and proteomes, *BMC Genomics*, **15**(1), p. 582.

Garcia-Mas, J. *et al.* (2012) The genome of melon (Cucumis melo L.), *Proceedings of the National Academy of Sciences of the United States of America*, **109**(29), pp. 11872–11877.

Garnett, T., Appleby, M. C., Balmford, A., Bateman, I. J., Benton, T. G., Bloomer, P., Burlingame, B., Dawkins, M., Dolan, L., Fraser, D., Herrero, M., Hoffmann, I., Smith, P., Thornton, P. K., Toulmin, C., Vermeulen, S. J. and Godfray, H. C. J. (2013) Sustainable Intensification in Agriculture: Premises and Policies, *Science*, **341**(6141), pp. 33–34.

Gawryluk, R. M. R., Tikhonenkov, D. V., Hehenberger, E., Husnik, F., Mylnikov, A. P. and Keeling, P. J. (2019) Non-photosynthetic predators are sister to red algae, *Nature*, **572**(7768), pp. 240–243.

Geiger, D., Scherzer, S., Mumm, P., Stange, A., Marten, I., Bauer, H., Ache, P., Matschi, S., Liese, A., Al-Rasheid, K. A. S., Romeis, T. and Hedrich, R. (2009) Activity of guard cell anion channel SLAC1 is controlled by drought-stress signaling kinase-phosphatase pair, *Proceedings of the National Academy of Sciences of the United States of America*, **106**(50), pp. 21425–21430.

Ghosh, T. K., Kaneko, M., Akter, K., Murai, S., Komatsu, K., Ishizaki, K., Yamato, K. T., Kohchi, T. and Takezawa, D. (2016) Abscisic acid-induced gene expression in the liverwort Marchantia polymorpha is mediated by evolutionarily conserved promoter elements, *Physiologia Plantarum*, **156**(4), pp. 407–420.

Gibling, M. R. and Davies, N. S. (2012) Palaeozoic landscapes shaped by plant evolution, *Nature Geoscience*, **5**(2), pp. 99–105.

Gibson, T. M., Shih, P. M., Cumming, V. M., Fischer, W. W., Crockford, P. W., Hodgskiss, M. S. W., Wörndle, S., Creaser, R. A., Rainbird, R. H., Skulski, T. M. and Halverson, G. P. (2018) Precise age of Bangiomorpha pubescens dates the origin of eukaryotic photosynthesis, *Geology*, **46**(2), pp. 135–138.

Di Giorgio, J. A. P., Bienert, G. P., Ayub, N. D., Yaneff, A., Barberini, M. L., Mecchia, M. A., Amodeo, G., Soto, G. C. and Muschietti, J. P. (2016) Pollen-specific aquaporins NIP4;1 and NIP4;2 are required for pollen development and pollination in Arabidopsis thaliana, *Plant Cell*, **28**(5), pp. 1053–1077.

Glover, N., Dessimoz, C., Ebersberger, I., Forslund, S. K., Gabaldón, T., Huerta-Cepas, J., Martin, M. J., Muffato, M., Patricio, M., Pereira, C., Da Silva, A. S., Wang, Y., Sonnhammer, E., Thomas, P. D. and Rogers, R. (2019) Advances and Applications in the Quest for Orthologs, *Molecular Biology and Evolution*, **36**(10), pp. 2157–2164.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M. and Toulmin, C. (2010) Food Security: The

Challenge of Feeding 9 Billion People, *Science*, **327**(5967), pp. 812–818.

Godfray, H. C. J. and Garnett, T. (2014) Food security and sustainable intensification, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**(1639).

Goff, S. A. *et al.* (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica), *Science*, **296**(5565), pp. 92–100.

Goffeau, A., Barrell, G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S. G. (1996) Life with 6000 genes, *Science*, **274**(5287), pp. 546–567.

Goh, T., Kasahara, H., Mimura, T., Kamiya, Y. and Fukaki, H. (2012) Multiple AUX/IAA-ARF modules regulate lateral root formation: The role of Arabidopsis SHY2/IAA3-mediated auxin signalling, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**(1595), pp. 1461–1468.

Golestani, M. and Pakniyat, H. (2015) Evaluation of Traits Related to Drought Stress in Sesame (Sesamum Indicum L.) Genotypes, *Journal of Asian Scientific Research*, **5**(9), pp. 465–472.

Gomes, F. P. and Prado, C. H. B. A. (2007) Ecophysiology of coconut palm under water stress, *Brazilian Journal of Plant Physiology*, **19**(4), pp. 377–391.

Gonzalez-Guzman, M., Pizzio, G. A., Antoni, R., Vera-Sirera, F., Merilo, E., Bassel, G. W., Fernández, M. A., Holdsworth, M. J., Perez-Amador, M. A., 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.

Goodwin, S., McPherson, J. D. and McCombie, W. R. (2016) Coming of age: Ten years of next-generation sequencing technologies, *Nature Reviews Genetics*, **17**(6), pp. 333–351.

Gould, S. jay and Eldredge, N. (1993) Punctuated equilibrium comes of age, *Nature*, **366**, pp. 223–227.

van der Graaff, E., Laux, T. and Rensing, S. A. (2009) The WUS homeobox-containing (WOX) protein family, *Genome Biology*, **10**(12).

Griesmann, M. *et al.* (2018) Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis, *Science*, **361**(6398).

Griffith, M., Timonin, M., Wong, A. C. E., Gray, G. R., Akhter, S. R., Saldanha, M., Rogers, M. A., Weretilnyk, E. A. and Moffatt, B. (2007) Thellungiella: an Arabidopsis-related model plant adapted to cold temperatures, *Plant, Cell & Environment*, **30**(5), pp. 529–538.

Grondin, A., Rodrigues, O., Verdoucq, L., Merlot, S., Leonhardt, N. and Maurel, C. (2015) Aquaporins contribute to ABA-triggered stomatal closure through OST1-mediated phosphorylation, *Plant Cell*, **27**(7), pp. 1945–1954.

Guan, R. *et al.* (2016) Draft genome of the living fossil Ginkgo biloba, *GigaScience*, **5**(49).

Guijarro-Clarke, C., Holland, P. W. H. and Paps, J. (2020) Widespread patterns of gene loss in the evolution of the animal kingdom, *Nature Ecology & Evolution*, **4**, pp. 519–523.

Guimarães, P. M., Brasileiro, A. C. M., Morgante, C. V., Martins, A. C. Q., Pappas, G., Silva, O. B., Togawa, R., Leal-Bertioli, S. C. M., Araujo, A. C. G., Moretzsohn, M. C. and Bertioli, D. J. (2012) Global transcriptome analysis of two wild relatives of peanut under drought and fungi infection, *BMC Genomics*, **13**(387).

Guiry, M. D. (2012) How many species of algae are there?, *Journal of Phycology*, **48**(5), pp. 1057–1063.

Guo, J., Ling, H., Wu, Q., Xu, L. and Que, Y. (2014) The choice of reference genes for assessing gene expression in sugarcane under salinity and drought stresses, *Scientific Reports*, **4**(7042).

Guo, L. *et al.* (2017) Echinochloa crus-galli genome analysis provides insight into its adaptation and invasiveness as a weed, *Nature Communications*, **8**(1).

Haas, B. J. *et al.* (2009) Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans, *Nature*, **461**(7262), pp. 393–398.

Hahn, M. W. (2007) Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution, *Genome Biology*, **8**(7).

Hane, J. K. *et al.* (2017) A comprehensive draft genome sequence for lupin (Lupinus angustifolius), an emerging health food: insights into plant-microbe interactions and

legume evolution, *Plant Biotechnology Journal*, **15**(3), pp. 318–330.

Hanschen, E. R. *et al.* (2016) The Gonium pectorale genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity, *Nature Communications*, **7**, p. 11370.

Harkess, A. *et al.* (2017) The asparagus genome sheds light on the origin and evolution of a young y chromosome, *Nature Communications*, **8**(1).

Harris, B. J., Harrison, C. J., Hetherington, A. M. and Williams, T. A. (2020) Phylogenomic Evidence for the Monophyly of Bryophytes and the Reductive Evolution of Stomata, *Current Biology*, **30**(11), pp. 2001–2012.

Harrison, C. J. (2017) Development and genetics in the evolution of land plant body plans, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **372**(1713).

Hasanuzzaman, M., Nahar, K., Anee, T. I. and Fujita, M. (2017) Glutathione in plants: biosynthesis and physiological role in environmental stress tolerance, *Physiology and Molecular Biology of Plants*, **23**(2), pp. 249–268.

Haudry, A. *et al.* (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions, *Nature Genetics*, **45**(8), pp. 891–898.

He, L., Gao, Z. and Li, R. (2009) Pretreatment of seed with H 2 O 2 enhances drought tolerance of wheat (Triticum aestivum L.) seedlings, *African Journal of Biotechnology*, **8**(22), pp. 6151–6157.

He, N. *et al.* (2013) Draft genome sequence of the mulberry tree Morus notabilis, *Nature Communications*, **4**(2445).

Heidel, A. J., Lawal, H. M., Felder, M., Schilde, C., Helps, N. R., Tunggal, B., Rivero, F., John, U., Schleicher, M., Eichinger, L., Platzer, M., Noegel, A. A., Schaap, P. and Glöckner, G. (2011) Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication, *Genome Research*, **21**(11), pp. 1882–1891.

Hellsten, U. *et al.* (2010) The genome of the western clawed frog xenopus tropicalis, *Science*, **328**(5978), pp. 633–636.

Hellsten, U., Wright, K. M., Jenkins, J., Shu, S., Yuan, Y., Wessler, S. R., Schmutz, J.,

Willis, J. H. and Rokhsar, D. S. (2013) Fine-scale variation in meiotic recombination in Mimulus inferred from population shotgun sequencing, *Proceedings of the National Academy of Sciences of the United States of America*, **110**(48), pp. 19478–19482.

Henry, H. W. and L. (2018) tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions.

Hetherington, A. J., Berry, C. M. and Dolan, L. (2020) Multiple origins of dichotomous and lateral branching during root evolution, *Nature Plants*, **6**, pp. 454–459.

Hetherington, A. J. and Dolan, L. (2018) Stepwise and independent origins of roots among land plants, *Nature*, **561**(7722), pp. 235–238.

Hidalgo, O., Pellicer, J., Christenhusz, M. J. M. M., Schneider, H. and Leitch, I. J. (2017) Genomic gigantism in the whisk-fern family (Psilotaceae): Tmesipteris obliqua challenges record holder Paris japonica, *Botanical Journal of the Linnean Society*, **183**(4), pp. 509–514.

Hillier, L. W. *et al.* (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution, *Nature*, **432**(7018), pp. 695–716.

Hirakawa, H. *et al.* (2015) Survey of genome sequences in a wild sweet potato, Ipomoea trifida (H. B. K.) G. Don, *DNA Research*, **22**(2), pp. 171–179.

Hirooka, S., Hirose, Y., Kanesaki, Y., Higuchi, S., Fujiwara, T., Onuma, R., Era, A., Ohbayashi, R., Uzuka, A., Nozaki, H., Yoshikawa, H. and Miyagishima, S. Y. (2017) Acidophilic green algal genome provides insights into adaptation to an acidic environment, *Proceedings of the National Academy of Sciences of the United States of America*, **114**(39), pp. 8304–8313.

Hirose, N., Takei, K., Kuroha, T., Kamada-Nobusada, T., Hayashi, H. and Sakakibara, H. (2007) Regulation of cytokinin biosynthesis, compartmentalization and translocation, *Journal of Experimental Botany*, **59**(1), pp. 75–83.

Hoffman, P. F. *et al.* (2017) Snowball Earth climate dynamics and Cryogenian geology-geobiology, *Science Advances*, **3**(11).

Holland, P. W. H., Marlétaz, F., Maeso, I., Dunwell, T. L. and Paps, J. (2017) New genes from old: asymmetric divergence of gene duplicates and the evolution of development, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **372**(1713).

Holzinger, A. and Becker, B. (2015a) Desiccation tolerance in the streptophyte green alga Klebsormidium: The role of phytohormones, *Communicative and Integrative Biology*, **8**(4), pp. 1–4.

Holzinger, A., Herburger, K., Kaplan, F. and Lewis, L. A. (2015b) Desiccation tolerance in the chlorophyte green alga Ulva compressa: does cell wall architecture contribute to ecological success?, *Planta*, **242**(2), pp. 477–492.

Hoopes, G. M., Hamilton, J. P., Kim, J., Zhao, D., Wiegert-Rininger, K., Crisovan, E. and Buell, C. R. (2018) Genome assembly and annotation of the medicinal plant Calotropis gigantea, a producer of anticancer and antimalarial cardenolides, *G3: Genes, Genomes, Genetics*, **8**(2), pp. 385–391.

Hori, K. *et al.* (2014) Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation, *Nature Communications*, **5**(1), p. 3978.

Hori, K., Yamada, Y., Purwanto, R., Minakuchi, Y., Toyoda, A., Hirakawa, H. and Sato, F. (2018) Mining of the Uncharacterized Cytochrome P450 Genes Involved in Alkaloid Biosynthesis in California Poppy Using a Draft Genome Sequence, *Plant and Cell Physiology*, **59**(2), pp. 222–233.

Horodyski, R. J. and Knauth, L. P. (1994) Life on land in the Precambrian, *Science*, **263**(5146), pp. 494–498.

Hoshino, A. *et al.* (2016) Genome sequence and analysis of the Japanese morning glory Ipomoea nil, *Nature Communications*, **7**(13295).

Hosy, E., Vavasseur, A., Mouline, K., Dreyer, I., Gaymard, F., Porée, F., Boucherez, J., Lebaudy, A., Bouchez, D., Véry, A. A., Simonneau, T., Thibaud, J. B. and Sentenac, H. (2003) The Arabidopsis outward K+ channel GORK is involved in regulation of stomatal movements and plant transpiration, *Proceedings of the National Academy of Sciences of the United States of America*, **100**(9), pp. 5549–5554.

Howe, K. *et al.* (2013) The zebrafish reference genome sequence and its relationship to the human genome, *Nature*, **496**(7446), pp. 498–503.

Hu, T. T. *et al.* (2011) The Arabidopsis lyrata genome sequence and the basis of rapid genome size change, *Nature Genetics*, **43**(5), pp. 476–483.

Huang, J., Zhang, C., Zhao, X., Fei, Z., Wan, K., Zhang, Z., Pang, X., Yin, X., Bai, Y.,

Sun, X., Gao, L., Li, R., Zhang, J. and Li, X. (2016) The Jujube Genome Provides Insights into Genome Evolution and the Domestication of Sweetness/Acidity Taste in Fruit Trees, *PLOS Genetics*, **12**(12).

Huang, S. *et al.* (2009) The genome of the cucumber, Cucumis sativus L., *Nature Genetics*, **41**(12), pp. 1275–1281.

Huang, S. *et al.* (2013) Draft genome of the kiwifruit Actinidia chinensis, *Nature Communications*, **4**.

Huang, X., Hou, L., Meng, J., You, H., Li, Z., Gong, Z., Yang, S. and Shi, Y. (2018) The Antagonistic Action of Abscisic Acid and Cytokinin Signaling Mediates Drought Stress Response in Arabidopsis, *Molecular Plant*, **11**(7), pp. 970–982.

Huelsenbeck, J. P., Nielsen, R. and Bollback, J. P. (2003) Stochastic Mapping of Morphological Characters, *Systematic Biology*, **52**(2), pp. 131–158.

Hughes, R., Bachmann, K., Smirnoff, N. and Macnair, M. R. (2001) The role of drought tolerance in serpentine tolerance in the Mimulus guttatus Fischer ex DC. complex, *South African Journal of Science*, **97**(11–12), pp. 581–586.

Hundertmark, M. and Hincha, D. K. (2008) LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana, *BMC Genomics*, **9**(1), pp. 1–22.

Hutchison, C. E. and Kieber, J. J. (2002) Cytokinin signaling in Arabidopsis, *The Plant cell*, **14**(1), pp. 47–59.

Ibarra-Laclette, E. *et al.* (2013) Architecture and evolution of a minute plant genome, *Nature*, **498**(7452), pp. 94–98.

Imes, D., Mumm, P., Böhm, J., Al-Rasheid, K. A. S., Marten, I., Geiger, D. and Hedrich, R. (2013) Open stomata 1 (OST1) kinase controls R-type anion channel QUAC1 in Arabidopsis guard cells, *Plant Journal*, **74**(3), pp. 372–382.

Dello Ioio, R., Linhares, F. S., Scacchi, E., Casamitjana-Martinez, E., Heidstra, R., Costantino, P. and Sabatini, S. (2007) Cytokinins Determine Arabidopsis Root-Meristem Size by Controlling Cell Differentiation, *Current Biology*, **17**(8), pp. 678–682.

Iorizzo, M. *et al.* (2016) A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution, *Nature Genetics*, **48**(6), pp.

657–666.

Iseki, K., Takahashi, Y., Muto, C., Naito, K. and Tomooka, N. (2016) Diversity and Evolution of Salt Tolerance in the Genus Vigna, *PLoS ONE*, **11**(10).

Iseki, K., Takahashi, Y., Muto, C., Naito, K. and Tomooka, N. (2018) Diversity of Drought Tolerance in the Genus Vigna, *Frontiers in Plant Science*, **9**(729).

Islam, M. S. *et al.* (2017) Comparative genomics of two jute species and insight into fibre biogenesis, *Nature Plants*, **3**(16223).

Iwata, S., Miyazawa, Y., Fujii, N. and Takahashi, H. (2013) MIZ1-regulated hydrotropism functions in the growth and survival of Arabidopsis thaliana under natural conditions, *Annals of Botany*, **112**(1), pp. 103–114.

Jaenicke, L. and Gilles, R. (1982) Differentiation and Embryogenesis in Volvox carteri, in *Biochemistry of Differentiation and Morphogenesis*. Springer Berlin Heidelberg, pp. 288–294.

Jaillon, O. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla, *Nature*, **449**(7161), pp. 463–467.

Jamwal, A., Puri, S., Sharma, S., Bhattacharya, S. and Dhindsa, N. (2015) Polyethylene Glycol Induced Morphological Changes in Fagopyrum esculentum Moench of Indian Himalayan Region, *Asian Journal of Advanced Basci Sciences*, **3**(2), pp. 142–146.

Jangam, A. P., Pathak, R. R. and Raghuram, N. (2016) Microarray Analysis of Rice d1 (RGA1) Mutant Reveals the Potential Role of G-Protein Alpha Subunit in Regulating Multiple Abiotic Stresses Such as Drought, Salinity, Heat, and Cold, *Frontiers in Plant Science*, **7**(11).

Jarvis, D. E. *et al.* (2017) The genome of Chenopodium quinoa, *Nature*, **542**(7641), pp. 307–312.

Jayakodi, M. *et al.* (2020) The barley pan-genome reveals the hidden legacy of mutation breeding, *Nature*, **588**, pp. 284–289.

Jékely, G., Paps, J. and Nielsen, C. (2015) The phylogenetic position of ctenophores and the origin(s) of nervous systems, *EvoDevo*, **6**(1).

Jensen, C. R. and Henson, I. E. (1990) Leaf Water Relations Characteristics of Lupinus

angustifolius and L. cosentinii, *Oecologia*, **82**(1), pp. 114–121.

Jeon, B. W., Acharya, B. R. and Assmann, S. M. (2019) The Arabidopsis heterotrimeric G- protein β subunit, AGB1, is required for guard cell calcium sensing and calcium-induced calcium release, *The Plant Journal*, **99**(2).

Ji, W., Zhu, Y., Li, Y., Yang, L., Zhao, X., Cai, H. and Bai, X. (2010) Over-expression of a glutathione S-transferase gene, GsGST, from wild soybean (Glycine soja) enhances drought and salt tolerance in transgenic tobacco, *Biotechnology Letters*, **32**(8), pp. 1173–1179.

Jiao, C., Sørensen, I., Sun, X., Sun, H., Behar, H., Alseekh, S., Philippe, G., Palacio Lopez, K., Sun, L., Reed, R., Jeon, S., Kiyonami, R., Zhang, S., Fernie, A. R., Brumer, H., Domozych, D. S., Fei, Z. and Rose, J. K. C. (2020) The Penium margaritaceum Genome: Hallmarks of the Origins of Land Plants, *Cell*, **181**(5), pp. 1097–1111.

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J. and dePamphilis, C. W. (2011) Ancestral polyploidy in seed plants and angiosperms, *Nature*, **473**(7345), pp. 97–100.

Jin, X., Wang, R. S., Zhu, M., Jeon, B. W., Albert, R., Chen, S. and Assmann, S. M. (2013) Abscisic acid-responsive guard cell metabolomes of Arabidopsis wild-type and gpa1 G-protein mutants, *Plant Cell*, **25**(12), pp. 4789–4811.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R. and Hunter, S. (2014) InterProScan 5: Genome-scale protein function classification, *Bioinformatics*, **30**(9), pp. 1236–1240.

Jones, V. A. S. and Dolan, L. (2012) The evolution of root hairs and rhizoids, *Annals of Botany*, **110**(2), pp. 205–212.

Jones, V. A. S. and Dolan, L. (2017) MpWIP regulates air pore complex development in the liverwort Marchantia polymorpha, *Development*, **144**(8), pp. 1472–1476.

Ju, C. and Chang, C. (2015a) Mechanistic Insights in Ethylene Perception and Signal Transduction, *Plant Physiology*, **169**(1), pp. 85–95.

Ju, C., Van de Poel, B., Cooper, E. D., Thierer, J. H., Gibbons, T. R., Delwiche, C. F. and Chang, C. (2015b) Conservation of ethylene as a plant hormone over 450 million years of evolution, *Nature Plants*, **1**(14004).

Jung, J. K. H. and McCouch, S. (2013) Getting to the roots of it: Genetic and hormonal control of root architecture, *Frontiers in Plant Science*, **4**(186).

Kado, T. and Innan, H. (2018) Horizontal Gene Transfer in Five Parasite Plant Species in Orobanchaceae, *Genome Biology and Evolution*, **10**(12), pp. 3196–3210.

Kalandyk, A., Waligórski, P. and Dubert, F. (2017) Role of the maternal effect phenomena in improving water stress tolerance in narrow-leafed lupine (Lupinus angustifolius), *Plant Breeding*, **136**(2), pp. 167–173.

Kang, Y. J. *et al.* (2015) Draft genome sequence of adzuki bean, Vigna angularis, *Scientific Reports*, **5**(8069).

Kantar, M. B., Nashoba, A. R., Anderson, J. E., Blackman, B. K. and Rieseberg, L. H. (2017) The Genetics and Genomics of Plant Domestication, *BioScience*, **67**(11), pp. 971–982.

Kapli, P., Yang, Z. and Telford, M. J. (2020) Phylogenetic tree building in the genomic age, *Nature Reviews Genetics*, **21**(7), pp. 428–444.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, *Nucleic Acids Research*, **30**(14), pp. 3059–66.

Kattge, J. *et al.* (2020) TRY plant trait database – enhanced coverage and open access, *Global Change Biology*, **26**(1), pp. 119–188.

Kaul, S. *et al.* (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana, *Nature*, **408**(6814), pp. 796–815.

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

Kenrick, P. and Strullu-Derrien, C. (2014) The origin and early evolution of roots, *Plant Physiology*, **166**(2), pp. 570–580.

Kew Science (2020a) *Hevea brasiliensis (Willd. ex A.Juss.) Müll.Arg.*, *Plants of the*

World Online. Available at: http://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:349913-1#descriptions (Accessed: 8 December 2020).

Kew Science (2020b) *Musa itinerans Cheesman*, *Plants of the World Online*. Available at: http://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:584951-1 (Accessed: 8 December 2020).

Kiełbowicz-Matuk, A. (2012) Involvement of plant C2H2-type zinc finger transcription factors in stress responses, *Plant Science*, **185**–**186**, pp. 78–85.

Kim, D. W., Jeon, S. J., Hwang, S. M., Hong, J. C. and Bahk, J. D. (2016) The C3H-type zinc finger protein GDS1/C3H42 is a nuclear-speckle-localized protein that is essential for normal growth and development in Arabidopsis, *Plant Science*, **250**, pp. 141–153.

Kim, M. Y. *et al.* (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome, *Proceedings of the National Academy of Sciences of the United States of America*, **107**(51), pp. 22032–22037.

Kim, N. S. (2017) The genomes and transposable elements in plants: are they friends or foes?, *Genes and Genomics*, **39**(4), pp. 359–370.

Kim, S. *et al.* (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species, *Nature Genetics*, **46**(3), pp. 270–278.

Kim, S. *et al.* (2017) New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication, *Genome Biology*, **18**(1).

Kizil, Ü., Genc, L., İnalpulat, M., Şapolyo, D. and Mirik, M. (2012) Lettuce (Lactuca sativa L.) yield prediction under water stress using artificial neural network (ANN) model and vegetation indices, *Zemdirbyste*, **99**(4), pp. 409–418.

Kobayashi, A., Takahashi, A., Kakimoto, Y., Miyazawa, Y., Fujii, N., Higashitani, A. and Takahashi, H. (2007) A gene essential for hydrotropism in roots, *Proceedings of the National Academy of Sciences of the United States of America*, **104**(11), pp. 4724–4729.

Kocacinar, F. (2015) Photosynthetic, hydraulic and biomass properties in closely related C_3 and C_4 species, *Physiologia Plantarum*, **153**(3), pp. 454–466.

Kohler, M., Sohn, J., Nägele, G. and Bauhus, J. (2010) Can drought tolerance of Norway spruce (Picea abies (L.) Karst.) be increased through thinning?, *European Journal of Forest Research*, **129**(6), pp. 1109–1118.

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**(1), pp. 1–9.

van der Kooi, C. J. and Ollerton, J. (2020) The origins of flowering plants and pollinators, *Science*, **368**(6497), pp. 1306–1308.

Koonin, E. V. (2005) Orthologs, paralogs, and evolutionary genomics, *Annual Review of Genetics*, **39**, pp. 309–338.

Koonin, E. V, Aravind, L. and Kondrashov, A. S. (2000) The Impact of Comparative Genomics on Our Understanding of Evolution, *Cell*, **101**(6), pp. 573–576.

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A. and Zdobnov, E. M. (2019) OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs, *Nucleic Acids Research*, **47**(1), pp. 807–811.

Kubatko, L. S. and Degnan, J. H. (2007) Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence, *Systematic Biology*, **56**(1), pp. 17–24.

Kumar, D., Hazra, S., Datta, R. and Chattopadhyay, S. (2016) Transcriptome analysis of Arabidopsis mutants suggests a crosstalk between ABA, ethylene and GSH against combined cold and osmotic stress, *Scientific Reports*, **6**(36867).

Kumpf, R. P., Shi, C. L., Larrieu, A., Stø, I. M., Butenko, M. A., Péret, B., Riiser, E. S., Bennett, M. J. and Aalen, R. B. (2013) Floral organ abscission peptide IDA and its HAE/HSL2 receptors control cell separation during lateral root emergence, *Proceedings of the National Academy of Sciences of the United States of America*, **110**(13), pp. 5235–5240.

Kusvuran, S. (2012) Effects of drought and salt stresses on growth, stomatal conductance, leaf water and osmotic potentials of melon genotypes (Cucumis melo L.), *African Journal of Agricultural Research*, **7**(5).

Kuzniar, A., van Ham, R. C. H. J., Pongor, S. and Leunissen, J. A. M. (2008) The quest for orthologs: finding the corresponding gene across genomes, *Trends in Genetics*, **24**(11), pp. 539–551.

Labandeira, C. C. (2013) A paleobiologic perspective on plant-insect interactions, *Current Opinion in Plant Biology*, **16**(4), pp. 414–421.

Lander, E. S. *et al.* (2001) Initial sequencing and analysis of the human genome, *Nature*, **409**(6822), pp. 860–921.

Landis, J. B., Soltis, D. E., Li, Z., Marx, H. E., Barker, M. S., Tank, D. C. and Soltis, P. S. (2018) Impact of whole-genome duplication events on diversification rates in angiosperms, *American Journal of Botany*, **105**(3), pp. 348–363.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. and Calcott, B. (2017) Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses, *Molecular Biology and Evolution*, **34**(3), pp. 772–773.

Lau, O. S. and Bergmann, D. C. (2012) Stomatal development: A plant's perspective on cell polarity, cell fate transitions and intercellular communication, *Development*, **139**(20), pp. 3683–3692.

Lau, Y. L., Lee, W. C., Gudimella, R., Zhang, G. P., Ching, X. T., Razali, R., Aziz, F., Anwar, A. and Fong, M. Y. (2016) Deciphering the draft genome of Toxoplasma gondii RH strain, *PLoS ONE*, **11**(6).

Lavenus, J., Goh, T., Roberts, I., Guyomarc'h, S., Lucas, M., Smet, I. De, Fukaki, H., Beeckman, T., Bennett, M. and Laplaze, L. (2013) Lateral root development in Arabidopsis: fifty shades of auxin, *Trends in Plant Science*, **18**(8), pp. 450–458.

Le, J., Zou, J., Yang, K. and Wang, M. (2014) Signaling to stomatal initiation and cell division, *Frontiers in Plant Science*, **5**(297).

Lee, H. T., Golicz, A. A., Bayer, P. E., Jiao, Y., Tang, H., Paterson, A. H., Sablok, G., Krishnaraj, R. R., Chan, C.-K. K. K., Batley, J., Kendrick, G. A., Larkum, A. W. D. D., Ralph, P. J. and Edwards, D. (2016a) The Genome of a Southern Hemisphere Seagrass Species (Zostera muelleri), *Plant Physiology*, **172**(1), pp. 272–283.

Lee, J., Kang, Y., Shin, S. C., Park, H. and Lee, H. (2014) Combined Analysis of the

Chloroplast Genome and Transcriptome of the Antarctic Vascular Plant Deschampsia antarctica Desv, *PLoS ONE*, **9**(3).

Lee, J. M., Cho, C. H., Park, S. I., Choi, J. W., Song, H. S., West, J. A., Bhattacharya, D. and Yoon, H. S. (2016b) Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants, *BMC Biology*, **14**(75).

Lee, J., Noh, E. K., Choi, H. S., Shin, S. C., Park, H. and Lee, H. (2013) Transcriptome sequencing of the Antarctic vascular plant Deschampsia antarctica Desv. under abiotic stress, *Planta*, **237**(3), pp. 823–836.

Leebens-Mack, J. H. *et al.* (2019) One thousand plant transcriptomes and the phylogenomics of green plants, *Nature*, **574**, pp. 679–685.

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.

Leliaert, F., Verbruggen, H. and Zechman, F. W. (2011) Into the deep: New discoveries at the base of the green plant phylogeny, *BioEssays*, **33**(9), pp. 683–692.

Lenton, T. M., Crouch, M., Johnson, M., Pires, N. and Dolan, L. (2012) First plants cooled the Ordovician, *Nature Geoscience*, **5**(2), pp. 86–89.

Lenton, T. M., Dahl, T. W., Daines, S. J., Mills, B. J. W., Ozaki, K., Saltzman, M. R. and Porada, P. (2016) Earliest land plants created modern levels of atmospheric oxygen, *Proceedings of the National Academy of Sciences of the United States of America*, **113**(35), pp. 9704–9709.

Léran, S. *et al.* (2014) A unified nomenclature of nitrate transporter 1/peptide transporter family members in plants, *Trends in Plant Science*, **19**(1), pp. 5–9.

Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, *Nucleic Acids Research*, **44**(1), pp. 242–5.

Letunic, I. and Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments, *Nucleic Acids Research*, **47**(1), pp. 256-W259.

Leuschner, C. and Ellenberg, H. (2017) Ecology of Central European non-forest vegetation: Coastal to alpine, natural to man-made habitats, in *Ecology of Central*

European Non-Forest Vegetation: Coastal to Alpine, Natural to Man-Made Habitats. Springer International Publishing, pp. 1–1093.

Leushkin, E. V, Sutormin, R. A., Nabieva, E. R., Penin, A. A., Kondrashov, A. S. and Logacheva, M. D. (2013) The miniature genome of a carnivorous plant Genlisea aurea contains a low number of genes and short non-coding sequences, *BMC Genomics*, **14**(1), p. 476.

Lewin, H. A. *et al.* (2018) Earth BioGenome Project: Sequencing life for the future of life, *Proceedings of the National Academy of Sciences of the United States of America*, **115**(17), pp. 4325–4333.

Lewis, L. A. and McCourt, R. M. (2004) Green algae and the origin of land plants, *American Journal of Botany*, **91**(10), pp. 1535–1556.

Leyser, O. (2018) Auxin Signaling, *Plant Physiology*, **176**, pp. 465–479.

Li, C., Yue, J., Wu, X., Xu, C. and Yu, J. (2014a) An ABA-responsive DRE-binding protein gene from Setaria italica, SiARDP, the target gene of SiAREB, plays a critical role under drought stress, *Journal of Experimental Botany*, **65**(18), pp. 5415–5427.

Li, F.-W. *et al.* (2018a) Fern genomes elucidate land plant evolution and cyanobacterial symbioses, *Nature Plants*, **4**(7), pp. 460–472.

Li, F.-W. and Harkess, A. (2018b) A guide to sequence your favorite plant genomes, *Applications in Plant Sciences*, **6**(3).

Li, F. *et al.* (2014b) Genome sequence of the cultivated cotton gossypium arboreum, *Nature Genetics*, **46**(6), pp. 567–572.

Li, F. *et al.* (2015a) Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution, *Nature Biotechnology*, **33**(5), pp. 524–530.

Li, F. (2018) Surfing the genomic new wave, *Nature Plants*, **4**(393).

Li, F. *et al.* (2020a) Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts, *Nature Plants*, **6**(3), pp. 259–272.

Li, F., Li, M., Wang, P., Cox, K. L., Duan, L., Dever, J. K., Shan, L., Li, Z. and He, P. (2017a) Regulation of cotton (Gossypium hirsutum) drought responses by mitogen-

activated protein (MAP) kinase cascade-mediated phosphorylation of GhWRKY59, *New Phytologist*, **215**(4), pp. 1462–1475.

Li, F. W. *et al.* (2014c) 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, H. T. *et al.* (2019) Origin of angiosperms and the puzzle of the Jurassic gap, *Nature Plants*, **5**(5), pp. 461–470.

Li, J., Guan, Y., Fan, H., Liu, T. and Liu, B. (2015b) The physiological response of leaves of three kinds of endangered isoetes under drought stress, *Wetland Science*, **13**(2), pp. 217–222.

Li, L. *et al.* (2020b) The genome of Prasinoderma coloniale unveils the existence of a third phylum within green plants, *Nature Ecology and Evolution*, **4**, pp. 1220–1231.

Li, L., Stoeckert, C. J. and Roos, D. S. (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes, *Genome Research*, **13**(9), pp. 2178–2189.

Li, Q., Zhang, X., Lv, Q., Zhu, D., Qiu, T., Xu, Y., Bao, F., He, Y. and Hu, Y. (2017b) Physcomitrella patens dehydrins (PpDHNA and PpDHNC) confer salinity and drought tolerance to transgenic arabidopsis plants, *Frontiers in Plant Science*, **8**(1316).

Li, S. F., Su, T., Cheng, G. Q., Wang, B. X., Li, X., Deng, C. L. and Gao, W. J. (2017c) Chromosome evolution in connection with repetitive sequences and epigenetics in plants, *Genes*, **8**(10).

Li, X. L., Hu, Y. X., Yang, X., Yu, X. D. and Li, Q. L. (2014d) A Novel Zinc-Finger HIT Protein with an Additional PAPA-1-like Region from Suaeda liaotungensis K. Enhanced Transgenic Arabidopsis Drought and Salt Stresses Tolerance, *Molecular Biotechnology*, **56**(12), pp. 1089–1099.

Li, Y., Zhang, D., Li, W., Mallano, A. I., Zhang, Y., Wang, T., Lu, M., Qin, Z. and Li, W. (2016) Expression study of soybean germin-like gene family reveals a role of GLP7 gene in various abiotic stress tolerances, *Canadian Journal of Plant Science*, **96**(2), pp. 296–304.

Liang, Y., Mitchell, D. M. and Harris, J. M. (2007) Abscisic acid rescues the root meristem defects of the Medicago truncatula latd mutant, *Developmental Biology*,

304(1), pp. 297–307.

Liang, Z. *et al.* (2019) Mesostigma viride Genome and Transcriptome Provide Insights into the Origin and Evolution of Streptophyta, *Advanced Science*, **7**(1).

Ligrone, R., Duckett JG and Renzaglia KS (2012) Major transitions in the evolution of early land plants: a bryological perspective, *Annals of Botany*, **109**(5), pp. 851–871.

Lin, Y. *et al.* (2017) Genome-wide sequencing of longan (Dimocarpus longan Lour.) provides insights into molecular basis of its polyphenol-rich characteristics, *GigaScience*, **6**(5), pp. 1–14.

Lind, C., Dreyer, I., López-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.

Ling, H. Q. *et al.* (2013) Draft genome of the wheat A-genome progenitor Triticum urartu, *Nature*, **496**(7443), pp. 87–90.

Linkies, A., Graeber, K., Knight, C. and Leubner-Metzger, G. (2010) The evolution of seeds, *New Phytologist*, **186**(4), pp. 817–831.

Liu, P.-L., Du, L., Huang, Y., Gao, S.-M. and Yu, M. (2017a) Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants, *BMC Evolutionary Biology*, **17**(47).

Liu, S. *et al.* (2014) The brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes, *Nature Communications*, **5**(3930).

Liu, S. C., Jin, J. Q., Ma, J. Q., Yao, M. Z., Ma, C. L., Li, C. F., Ding, Z. T. and Chen, L. (2016) Transcriptomic analysis of tea plant responding to drought stress and recovery, *PLoS ONE*, **11**(1).

Liu, W. and Xu, L. (2018) Recruitment of IC-WOX Genes in Root Evolution, *Trends in Plant Science*, **23**(6), pp. 490–496.

Liu, X. *et al.* (2017b) The Genome of Medicinal Plant Macleaya cordata Provides New Insights into Benzylisoquinoline Alkaloids Metabolism, *Molecular Plant*, **10**(7), pp. 975–989.

Liu, Z. J. (2015) The genome sequence of the orchid Phalaenopsis equestris, *Nature genetics*, **47**(1), pp. 65–72.

Loko, Y. L., Adjatin, A., Dansi, A., Vodouhè, R. and Sanni, A. (2015) Participatory evaluation of Guinea yam (Dioscorea cayenensis Lam.–D. rotundata Poir. complex) landraces from Benin and agro-morphological characterization of cultivars tolerant to drought, high soil moisture and chips storage insects, *Genetic Resources and Crop Evolution*, **62**(8), pp. 1181–1192.

Lopez, P., Casane, D. and Philippe, H. (2002) Heterotachy, an important process of protein evolution, *Molecular Biology and Evolution*, **19**(1), pp. 1–7.

Louis, A., Muffato, M. and Crollius, H. R. (2013) Genomicus: Five genome browsers for comparative genomics in eukaryota, *Nucleic Acids Research*, **41**(1).

Louis, A., Nguyen, N. T. T., Muffato, M. and Crollius, H. R. (2015) Genomicus update 2015: KaryoView and MatrixView provide a genome-wide perspective to multispecies comparative genomics, *Nucleic Acids Research*, **43**(1), pp. 682–689.

Lu, K. J., van't Wout Hofland, N., Mor, E., Mutte, S., Abrahams, P., Kato, H., Vandepoele, K., Weijers, D. and de Rybel, B. (2020) Evolution of vascular plants through redeployment of ancient developmental regulators, *Proceedings of the National Academy of Sciences of the United States of America*, **117**(1), pp. 733–740.

Lu, Y., Ran, J.-H., Guo, D.-M., Yang, Z.-Y. and Wang, X.-Q. (2014) Phylogeny and Divergence Times of Gymnosperms Inferred from Single-Copy Nuclear Genes, *PLoS ONE*, **9**(9).

Lucas, M. *et al.* (2011) SHORT-ROOT regulates primary, lateral, and adventitious root development in Arabidopsis, *Plant Physiology*, **155**(1), pp. 384–398.

Lucas, W. J., Groover, A., Lichtenberger, R., Furuta, K., Yadav, S. R., Helariutta, Y., He, X. Q., Fukuda, H., Kang, J., Brady, S. M., Patrick, J. W., Sperry, J., Yoshida, A., López-Millán, A. F., Grusak, M. A. and Kachroo, P. (2013) The Plant Vascular System: Evolution, Development and Functions, *Journal of Integrative Plant Biology*, **55**(4), pp. 294–388.

Lughadha, E. N., Govaerts, R., Belyaeva, I., Black, N., Lindon, H., Allkin, R., Magill, R. E. and Nicolson, N. (2016) Counting counts: Revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of

other recent estimates, *Phytotaxa*, **272**(1), pp. 82–88.

Lukaszewski, A. J. *et al.* (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome, *Science*, **345**(6194).

Luo, M. C. *et al.* (2017) Genome sequence of the progenitor of the wheat D genome Aegilops tauschii, *Nature*, **551**(7681), pp. 498–502.

Lutzoni, F., Nowak, M. D., Alfaro, M. E., Reeb, V., Miadlikowska, J., Krug, M., Arnold, A. E., Lewis, L. A., Swofford, D. L., Hibbett, D., Hilu, K., James, T. Y., Quandt, D. and Magallón, S. (2018) Contemporaneous radiations of fungi and plants linked to symbiosis, *Nature Communications*, **9**(1), pp. 1–11.

Lyons, T., Reinhard, C. and Planavsky, N. (2014) The rise of oxygen in Earth's early ocean and atmosphere, *Nature*, **506**, pp. 307–315.

Macalister, C. A. and Bergmann, D. C. (2011) Sequence and function of basic helix-loop-helix proteins required for stomatal development in Arabidopsis are deeply conserved in land plants, *Evolution and Development*, **13**(2), pp. 182–192.

Magwanga, R. O., Lu, P., Kirungu, J. N., Lu, H., Wang, X., Cai, X., Zhou, Z., Zhang, Z., Salih, H., Wang, K. and Liu, F. (2018) Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton, *BMC Genetics*, **19**(1).

Mao, L. *et al.* (2020) Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants, *Proceedings of the National Academy of Sciences of the United States of America*, **117**(22), pp. 12472–12480.

Maqbool, A., Abbas, W., Rao, A. Q., Irfan, M., Zahur, M., Bakhsh, A., Riazuddin, S. and Husnain, T. (2009) Gossypium arboreum GHSP26 enhances drought tolerance in Gossypium hirsutum, *Biotechnology Progress*, **26**(1).

Marchant, D. B., Sessa, E. B., Wolf, P. G., Heo, K., Barbazuk, W. B., Soltis, P. S. and Soltis, D. E. (2019) The C-Fern (Ceratopteris richardii) genome: insights into plant genome evolution with the first partial homosporous fern genome assembly, *Scientific Reports*, **9**(1), pp. 1–14.

Margulis, L., Chapman, M., Guerrero, R. and Hall, J. (2006) The last eukaryotic common ancestor (LECA): Acquisition of cytoskeletal motility from aerotolerant spirochetes in the

Proterozoic Eon, *Proceedings of the National Academy of Sciences of the United States of America*, **103**(35), pp. 13080–13085.

Marín- de la Rosa, N., Lin, C., Kang, Y. J., Dhondt, S., Gonzalez, N., Inzé, D. and Falter- Braun, P. (2019) Drought resistance is mediated by divergent strategies in closely related Brassicaceae, *New Phytologist*, **223**(2), pp. 783–797.

Martin, F. *et al.* (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis, *Nature*, **464**(7291), pp. 1033–1038.

Mashilo, J., Odindo, A. O., Shimelis, H. A., Musenge, P., Tesfay, S. Z. and Magwaza, L. S. (2017) Drought tolerance of selected bottle gourd [Lagenaria siceraria (Molina) Standl.] landraces assessed by leaf gas exchange and photosynthetic efficiency, *Plant Physiology and Biochemistry*, **120**, pp. 75–87.

Matsuzaki, M. *et al.* (2004) Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D, *Nature*, **428**(6983), pp. 653–657.

Matsuzaki, Y., Ogawa-Ohnishi, M., Mori, A. and Matsubayashi, Y. (2010) Secreted peptide signals required for maintenance of root stem cell niche in Arabidopsis, *Science*, **329**(5995), pp. 1065–1067.

Mayer, K. F. X. *et al.* (2012) A physical, genetic and functional sequence assembly of the barley genome, *Nature*, **491**(7426), pp. 711–716.

McAdam, S. A. M. and Brodribb, T. J. (2012) Fern and lycophyte guard cells do not respond to endogenous abscisic acid, *Plant Cell*, **24**(4), pp. 1510–1521.

McAdam, S. A. M. and Brodribb, T. J. (2013) Ancestral stomatal control results in a canalization of fern and lycophyte adaptation to drought, *New Phytologist*, **198**(2), pp. 429–441.

McAdam, S. A. M., Brodribb, T. J., Banks, J. A., Hedrich, R., Atallah, N. M., Cai, C., Geringer, M. A., Lind, C., Nichols, D. S., Stachowski, K., Geiger, D. and Sussmilch, F. C. (2016) Abscisic acid controlled sex before transpiration in vascular plants, *Proceedings of the National Academy of Sciences of the United States of America*, **113**(45), pp. 12862–12867.

Meena, M. K., Ghawana, S., Dwivedi, V., Roy, A. and Chattopadhyay, D. (2015) Expression of chickpea CIPK25 enhances root growth and tolerance to dehydration and

salt stress in transgenic tobacco, *Frontiers in Plant Science*, **6**(683).

Meena, M. K., Vishwakarma, N. K., Tripathi, V. and Chattopadhyay, D. (2019) CBL-interacting protein kinase 25 contributes to root meristem development, *Journal of Experimental Botany*, **70**(1), pp. 133–147.

Meng, L., Buchanan, B. B., Feldman, L. J. and Luan, S. (2012) CLE-like (CLEL) peptides control the pattern of root growth and lateral root development in Arabidopsis, *Proceedings of the National Academy of Sciences of the United States of America*, **109**(5), pp. 1760–1765.

Merchant, S. S. *et al.* (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions, *Science*, **318**(5848), pp. 245–251.

Meyer-Berthaud, B., Scheckler, S. E. and Wendt, J. (1999) Archaeopteris is the earliest known modern tree, *Nature*, **398**(6729), pp. 700–701.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D. and Thomas, P. D. (2017) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements, *Nucleic Acids Research*, **45**(1), pp. 183–189.

Michael, T. P. and Jackson, S. (2013) The First 50 Plant Genomes, *The Plant Genome*, **6**(2).

Middleton, A. M., Úbeda-Tomás, S., Griffiths, J., Holman, T., Hedden, P., Thomas, S. G., Phillips, A. L., Holdsworth, M. J., Bennett, M. J., King, J. R. and Owen, M. R. (2012) Mathematical modeling elucidates the role of transcriptional feedback in gibberellin signaling., *Proceedings of the National Academy of Sciences of the United States of America*, **109**(19), pp. 7571–6.

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.

Mills, T. M., Li, J. and Behboudian, M. H. (2009) Physiological Responses of Gold Kiwifruit (Actinidia chinensis) to Reduced Irrigation, *Journal of the American Society for Horticultural Science*, **134**(6), pp. 677–683.

Ming, R. *et al.* (2008) The draft genome of the transgenic tropical fruit tree papaya

(Carica papaya Linnaeus), *Nature*, **452**(7190), pp. 991–996.

Ming, R. *et al.* (2013) Genome of the long-living sacred lotus (Nelumbo nucifera Gaertn.), *Genome Biology*, **14**(5).

Ming, R. *et al.* (2015) The pineapple genome and the evolution of CAM photosynthesis, *Nature Genetics*, **47**(12), pp. 1435–1442.

Mishra, U., Rai, A., Kumar, R., Singh, M. and Pandey, H. P. (2016) Gene expression analysis of Solanum lycopersicum and Solanum habrochaites under drought conditions, *Genomics Data*, **9**, pp. 40–41.

Mita, P. and Boeke, J. D. (2016) How retrotransposons shape genome regulation, *Current Opinion in Genetics and Development*, **37**, pp. 90–100.

Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., Van Bel, M., Poulain, J., Katinka, M., Hohmann-Marriott, M. F., Piganeau, G., Rouzé, P., Da Silva, C., Wincker, P., Van de Peer, Y. and Vandepoele, K. (2012) Gene functionalities and genome structure in Bathycoccus prasinos reflect cellular specializations at the base of the green lineage, *Genome Biology*, **13**(8).

Morffy, N., Faure, L. and Nelson, D. C. (2016) Smoke and Hormone Mirrors: Action and Evolution of Karrikin and Strigolactone Signaling, *Trends in Genetics*, **32**(3), pp. 176–188.

Moriyama, Y. and Koshiba-Takeuchi, K. (2018) Significance of whole-genome duplications on the emergence of evolutionary novelties, *Briefings in Functional Genomics*, **17**(5), pp. 329–338.

Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C. H., Yang, Z., Schneider, H. and Donoghue, P. C. J. (2018) The timescale of early land plant evolution, *Proceedings of the National Academy of Sciences of the United States of America*, **115**(10), pp. 2274–2283.

Motte, H. and Beeckman, T. (2019) The evolution of root branching: increasing the level of plasticity, *Journal of Experimental Botany*, **70**(3), pp. 785–793.

Moubayidin, L., Di Mambro, R. and Sabatini, S. (2009) Cytokinin-auxin crosstalk, *Trends in Plant Science*, **14**(10), pp. 557–562.

Moubayidin, L., Perilli, S., Dello Ioio, R., Di Mambro, R., Costantino, P. and Sabatini, S.

(2010) The rate of cell differentiation controls the arabidopsis root meristem growth phase, *Current Biology*, **20**(12), pp. 1138–1143.

Multicellgenome Lab (2019) *Creolimax fragrantissima*. Available at: http://multicellgenome.com/research/meet-our-organisms/creolimax-fragrantissima.

Murphy, E. and De Smet, I. (2014) Understanding the RALF family: A tale of many species, *Trends in Plant Science*, **19**(10), pp. 664–671.

Murugesan, P., Aswathy, G. M., Kumar, K. S., Masilamani, P., Kumar, V. and Ravi, V. (2017) Oil palm (Elaeis guineensis) genetic resources for abiotic stress tolerance: A review, *Indian Journal of Agricultural Sciences*, **87**(5), pp. 571–579.

Mutte, S. K., Kato, H., Rothfels, C., Melkonian, M., Wong, G. K. S. and Weijers, D. (2018) Origin and evolution of the nuclear auxin response system, *eLife*, **7**.

Mutwakil, M. Z., Hajrah, N. H., Atef, A., Edris, S., Sabir, M. J., Al-Ghamdi, A. K., Sabir, M. J. S. M., Nelson, C., Makki, R. M., Ali, H. M., El-Domyati, F. M., Al-Hajar, A. S. M., Gloaguen, Y., Al-Zahrani, H. S., Sabir, J. S. M., Jansen, R. K., Bahieldin, A. and Hall, N. (2017) Transcriptomic and metabolic responses of Calotropis procera to salt and drought stress, *BMC Plant Biology*, **17**(231).

Myburg, A. A. *et al.* (2014) The genome of Eucalyptus grandis, *Nature*, **510**(7505), pp. 356–362.

Myers, S. S., Smith, M. R., Guth, S., Golden, C. D., Vaitla, B., Mueller, N. D., Dangour, A. D. and Huybers, P. (2017) Climate Change and Global Food Systems: Potential Impacts on Food Security and Undernutrition, *Annual Review of Public Health*, **38**, pp. 259–277.

Nagy, L. G., Merényi, Z., Hegedüs, B. and Bálint, B. (2020) Novel phylogenetic methods are needed for understanding gene function in the era of mega-scale genome sequencing, *Nucleic Acids Research*, **48**(5), pp. 2209–2219.

Nagy, L. G., Ohm, R. A., Kovács, G. M., Floudas, D., Riley, R., Gácser, A., Sipiczki, M., Davis, J. M., Doty, S. L., Hoog, G. S. de, Lang, B. F., Spatafora, J. W., Martin, F. M., Grigoriev, I. V. and Hibbett, D. S. (2014) Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts, *Nature Communications*, **5**(4471).

Nakata, M., Watanabe, Y., Sakurai, Y., Hashimoto, Y., Matsuzaki, M., Takahashi, Y. and Satoh, T. (2004) Germin-like protein gene family of a moss, Physcomitrella patens, phylogenetically falls into two characteristic new clades, *Plant Molecular Biology*, **56**(3), pp. 381–395.

Nansamba, M., Sibiya, J., Tumuhimbise, R., Karamura, D., Kubiriba, J. and Karamura, E. (2020) Breeding banana (Musa spp.) for drought tolerance: A review, *Plant Breeding*, **139**(4), pp. 685–696.

Narusaka, M., Shiraishi, T., Iwabuchi, M. and Narusaka, Y. (2010) The floral inoculating protocol: a simplified Arabidopsis thaliana transformation method modified from floral dipping, *Plant Biotechnology*, **27**(4), pp. 349–351.

Newton, R. J. and Goodin, J. R. (1989) Moisture Stress Adaptation in Shrubs, in McKell, C. (ed.) *The Biology and Utilization of Shrubs*. Academic Press, pp. 365–383.

Nguyen, K. H., Ha, C. Van, Nishiyama, R., Watanabe, Y., Leyva-González, M. A., Fujita, Y., Tran, U. T., Li, W., Tanaka, M., Seki, M., Schaller, G. E., Herrera-Estrella, L. and Tran, L. S. P. (2016) Arabidopsis type B cytokinin response regulators ARR1, ARR10, and ARR12 negatively regulate plant responses to drought, *Proceedings of the National Academy of Sciences of the United States of America*, **113**(11), pp. 3090–3095.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. and Minh, B. Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, *Molecular Biology and Evolution*, **32**(1), pp. 268–74.

Nibau, C., Gibbs, D. J. and Coates, J. C. (2008) Branching out in new directions: the control of root architecture by lateral root formation, *New Phytologist*, **179**(3), pp. 595–614.

Nilson, S. E. and Assmann, S. M. (2010) The α-subunit of the Arabidopsis heterotrimeric G protein, GPA1, is a regulator of transpiration efficiency, *Plant Physiology*, **152**(4), pp. 2067–2077.

Nishiyama, T. *et al.* (2018) The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization, *Cell*, **174**(2), pp. 448–464.

Van Norman, J. M., Xuan, W., Beeckman, T. and Benfey, P. N. (2013) To branch or not to branch: the role of pre-patterning in lateral root formation, *Development*, **140**(21), pp. 4301–4310.

Novoselov, S. V., Rao, M., Onoshko, N. V., Zhi, H., Kryukov, G. V., Xiang, Y., Weeks, D. P., Hatfield, D. L. and Gladyshev, V. N. (2002) Selenoproteins and selenocysteine insertion system in the model plant cell system, Chlamydomonas reinhardtii, *EMBO Journal*, **21**(14), pp. 3681–3693.

Ntoulas, N., Nektarios, P. A., Spaneas, K. and Kadoglou, N. (2012) Semi-extensive green roof substrate type and depth effects on Zoysia matrella 'Zeon' growth and drought tolerance under different irrigation regimes, *Acta Agriculturae Scandinavica Section B: Soil and Plant Science*, **62**, pp. 165–173.

Nystedt, B. *et al.* (2013) The Norway spruce genome sequence and conifer genome evolution, *Nature*, **497**(7451), pp. 579–584.

O'Malley, R. C., Barragan, C. C. and Ecker, J. R. (2015) A user's guide to the arabidopsis T-DNA insertion mutant collections, *Methods in Molecular Biology*, **1284**, pp. 323–342.

Ogura, T., Goeschl, C., Filiault, D., Mirea, M., Slovak, R., Wolhrab, B., Satbhai, S. B. and Busch, W. (2019) Root System Depth in Arabidopsis Is Shaped by EXOCYST70A3 via the Dynamic Modulation of Auxin Transport, *Cell*, **178**(2), pp. 400–412.

Oh, D.-H., Dassanayake, M., Bohnert, H. J. and Cheeseman, J. M. (2013) Life at the extreme: lessons from the genome, *Genome Biology*, **13**(3).

Oh, E., Seo, P. J. and Kim, J. (2018) Signaling Peptides and Receptors Coordinating Plant Root Development, *Trends in Plant Science*, **23**(4), pp. 337–351.

Ohkama-Ohtsu, N., Oikawa, A., Zhao, P., Xiang, C., Saito, K. and Oliver, D. J. (2008) A γ-glutamyl transpeptidase-independent pathway of glutathione catabolism to glutamate via 5-oxoproline in Arabidopsis, *Plant Physiology*, **148**(3), pp. 1603–1613.

Okogbenin, E., Setter, T. L., Ferguson, M., Mutegi, R., Ceballos, H., Olasanmi, B. and Fregene, M. (2013) Phenotypic approaches to drought in cassava: review, *Frontiers in Physiology*, **4**(93).

Okushima, Y., Fukaki, H., Onoda, M., Theologis, A. and Tasaka, M. (2007) ARF7 and ARF19 regulate lateral root formation via direct activation of LBD/ASL genes in Arabidopsis, *Plant Cell*, **19**(1), pp. 118–130.

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.

Oliver, M., Tuba, Z. and Mishler, B. D. (2000) The evolution of vegetative desiccation tolerance in land plants, *Plant Ecology*, **151**(1), pp. 85–100.

Olsen, J. L. *et al.* (2016) The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea, *Nature*, **530**(7590), pp. 331–335.

Olsen, K. M. and Wendel, J. F. (2013) A Bountiful Harvest: Genomic Insights into Crop Domestication Phenotypes, *Annual Review of Plant Biology*, **64**(1), pp. 47–70.

Orosa-Puente, B. *et al.* (2018) Root branching toward water involves posttranslational modification of transcription factor ARF7, *Science*, **362**(6421), pp. 1407–1410.

Orozco-Arias, S., Isaza, G. and Guyot, R. (2019) Retrotransposons in plant genomes: Structure, identification, and classification through bioinformatics and machine learning, *International Journal of Molecular Sciences*, **20**(15).

Pabuayon, I. M., Yamamoto, N., Trinidad, J. L., Longkumer, T., Raorane, M. L. and Kohli, A. (2016) Reference genes for accurate gene expression analyses across different tissues, developmental stages and genotypes in rice for drought tolerance, *Rice*, **9**(1), pp. 1–8.

Palenik, B. *et al.* (2007) The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation, *Proceedings of the National Academy of Sciences of the United States of America*, **104**(18), pp. 7705–7710.

Palfalvi, G. *et al.* (2020) Genomes of the Venus Flytrap and Close Relatives Unveil the Roots of Plant Carnivory, *Current Biology*, **30**(12), pp. 2312–2320.

Palmer, J. D., Soltis, D. E. and Chase, M. W. (2004) The plant tree of life: an overview and some points of view, *American Journal of Botany*, **91**(10), pp. 1437–45.

Panchy, N., Lehti-Shiu, M. and Shiu, S. H. (2016) Evolution of gene duplication in plants, *Plant Physiology*, **171**(4), pp. 2294–2316.

Pandey, S. and Assmann, S. M. (2004) The Arabidopsis putative G protein-coupled receptor GCR1 interacts with the G protein α subunit GPA1 and regulates abscisic acid signaling, *Plant Cell*, **16**(6), pp. 1616–1632.

Paps, J. and Holland, P. W. H. (2018) Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty, *Nature Communications*, **9**(1730).

Parfrey, L. W. and Lahr, D. J. G. (2013) Multicellularity arose several times in the evolution of eukaryotes, *BioEssays*, **35**(4), pp. 339–347.

Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., Ainouche, M., Chalhoub, B. and Grandbastien, M. A. (2010) Impact of transposable elements on the organization and function of allopolyploid genomes, *New Phytologist*, **186**(1), pp. 37–45.

Park, S.-Y. *et al.* (2009) Abscisic Acid Inhibits Type 2C Protein Phosphatases via the PYR/PYL Family of START Proteins, *Science*, **324**(5930), pp. 1068–71.

Parsons, H. T., Drakakaki, G. and Heazlewood, J. L. (2013) Proteomic dissection of the Arabidopsis Golgi and trans-Golgi network, *Frontiers in Plant Science*, **3**(298).

Partha, R., Kowalczyk, A., Clark, N. L. and Chikina, M. (2019) Robust Method for Detecting Convergent Shifts in Evolutionary Rates, *Molecular Biology and Evolution*, **36**(8), pp. 1817–1830.

Paterson, A. H. *et al.* (2009) The Sorghum bicolor genome and the diversification of grasses, *Nature*, **457**(7229), pp. 551–556.

Patton, A. J., Schwartz, B. M. and Kenworthy, K. E. (2017) Zoysiagrass (Zoysia spp.) history, utilization, and improvement in the United States: A review, *Crop Science*, **57**(1).

Paudel, I., Gerbi, H., Zisovich, A., Sapir, G., Ben-Dor, S., Brumfeld, V. and Klein, T. (2019) Drought tolerance mechanisms and aquaporin expression of wild vs. cultivated pear tree species in the field, *Environmental and Experimental Botany*, **167**.

Van de Peer, Y., Mizrachi, E. and Marchal, K. (2017) The evolutionary significance of polyploidy, *Nature Reviews Genetics*, **18**(7), pp. 411–424.

Pellicer, J., Fay, M. F. and Leitch, I. J. (2010) The largest eukaryotic genome of them all?, *Botanical Journal of the Linnean Society*, **164**(1), pp. 10–15.

Pellicer, J., Hidalgo, O., Dodsworth, S. and Leitch, I. J. (2018) Genome Size Diversity and Its Impact on the Evolution of Land Plants, *Genes*, **9**(2).

Peng, Z. *et al.* (2013) The draft genome of the fast-growing non-timber forest species moso bamboo (Phyllostachys heterocycla), *Nature Genetics*, **45**(4), pp. 456–461.

Pereira, A. (2016) Plant Abiotic Stress Challenges from the Changing Environment, *Frontiers in Plant Science*, **7**(1123).

Péret, B., Li, G., Zhao, J., Band, L. R., Voß, U., Postaire, O., Luu, D.-T., Ines, O. Da, Casimiro, I., Lucas, M., Wells, D. M., Lazzerini, L., Nacry, P., King, J. R., Jensen, O. E., Schäffner, A. R., Maure, C. and Bennett, M. J. (2012) Auxin regulates aquaporin function to facilitate lateral root emergence, *Nature Cell Biology*, **14**(10), pp. 991–998.

Pett, W., Adamski, M., Adamska, M., Francis, W. R., Eitel, M., Pisani, D., Woerheide, G. and Wörheide, G. (2019) The Role of Homology and Orthology in the Phylogenomic Analysis of Metazoan Gene Content, *Molecular Biology and Evolution*, **36**(4), pp. 643–649.

Pidgeon, J. D., Ober, E. S., Qi, A., Clark, C. J. A., Royal, A. and Jaggard, K. W. (2006) Using multi-environment sugar beet variety trials to screen for drought tolerance, *Field Crops Research*, **95**(2–3), pp. 268–279.

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., Raven, P. H., Roberts, C. M. and Sexton, J. O. (2014) The biodiversity of species and their rates of extinction, distribution, and protection, *Science*, **344**(6187).

Pimm, S. L. and Joppa, L. N. (2015) How Many Plant Species are There, Where are They, and at What Rate are They Going Extinct?, *Annals of the Missouri Botanical Garden*, **100**(3), pp. 170–176.

Pires, N. and Dolan, L. (2012) Morphological evolution in land plants: new designs with old genes, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, pp. 508–518.

Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., Lartillot, N. and Wörheide, G. (2015) Genomic data do not support comb jellies as the sister group to all other animals, *Proceedings of the National Academy of Sciences of the United States of America*, **112**(50), pp. 15402–15407.

Plants for a Future, A. (2020) *Hevea brasiliensis - (A.Juss.) Meull.* Available at: https://pfaf.org/user/Plant.aspx?LatinName=Hevea+brasiliensis.

Popper, Z. A., Michel, G., Hervé, C., Domozych, D. S., Willats, W. G. T., Tuohy, M. G., Kloareg, B. and Stengel, D. B. (2011) Evolution and Diversity of Plant Cell Walls: From Algae to Flowering Plants, *Annual Review of Plant Biology*, **62**(1), pp. 567–590.

Price, D. C. *et al.* (2012) Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants, *Science*, **335**(6070), pp. 843–847.

Price, D. C., Goodenough, U. W., Roth, R., Lee, J.-H., Kariyawasam, T., Mutwil, M., Ferrari, C., Facchinelli, F., Ball, S. G., Cenci, U., Chan, C. X., Wagner, N. E., Yoon, H. S., Weber, A. P. M. and Bhattacharya, D. (2019) Analysis of an improved Cyanophora paradoxa genome assembly, *DNA Research*, **26**(4), pp. 287–299.

Prochnik, S. E. *et al.* (2010) Genomic analysis of organismal complexity in the multicellular green alga volvox carteri, *Science*, **329**(5988), pp. 223–226.

Proctor, M. C. F., Oliver, M. J., Wood, A. J., Alpert, P., Stark, L. R., Cleavitt, N. L. and Mishler, B. D. (2007) Desiccation-tolerance in bryophytes: a review, *The Bryologist*, **110**(4), pp. 595–621.

Proctor, M. C. F. and Tuba, Z. (2002) Poikilohydry and homoihydry: antithesis or spectrum of possibilities?, *New Phytologist*, **156**(3), pp. 327–349.

Puttick, M. N., Morris, J. L., Williams, T. A., Cox, C. J., Edwards, D., Kenrick, P., Pressel, S., Wellman, C. H., Schneider, H., Pisani, D. and Donoghue, P. C. J. (2018) The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte, *Current Biology*, **28**(5), pp. 733–745.

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S. and Paterson, A. H. (2019) Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants, *Genome Biology*, **20**(38).

Qin, G., Xu, C., Ming, R., Tang, H., Guyot, R., Kramer, E. M., Hu, Y., Yi, X., Qi, Y., Xu, X., Gao, Z., Pan, H., Jian, J., Tian, Y., Yue, Z. and Xu, Y. (2017) The pomegranate (Punica granatum L.) genome and the genomics of punicalagin biosynthesis, *The Plant Journal*, **91**(6), pp. 1108–1128.

Qiu, H., Price, D. C., Weber, A. P. M., Reeb, V., Chan Yang, E., Lee, J. M., Kim, S. Y., Yoon, H. S. and Bhattacharya, D. (2013) Adaptation through horizontal gene transfer in the cryptoendolithic red alga Galdieria phlegrea, *Current Biology*, **23**(19), pp. 865–866.

Qiu, H., Price, D. C., Yang, E. C., Yoon, H. S. and Bhattacharya, D. (2015) Evidence of ancient genome reduction in red algae (Rhodophyta), *Journal of Phycology*, **51**(4), pp. 624–636.

R Core Team, A. (2014) *A language and environment for statistical computing*.

Raffaele, S., Mongrand, S., Gamas, P., Niebel, A. and Ott, T. (2007) Genome-Wide Annotation of Remorins, a Plant-Specific Protein Family: Evolutionary and Functional Perspectives, *Plant Physiology*, **145**(3), pp. 593–600.

Rambaut A, D. A. (2012) *FigTree version 1.4.* Available at: http://tree.bio.ed.ac.uk/software/figtree.

Ramegowda, V., Gill, U. S., Sivalingam, P. N., Gupta, A., Gupta, C., Govind, G., Nataraja, K. N., Pereira, A., Udayakumar, M., Mysore, K. S. and Senthil-Kumar, M. (2017) GBF3 transcription factor imparts drought tolerance in Arabidopsis thaliana, *Scientific Reports*, **7**(1), pp. 1–13.

Rannala, B. and Yang, Z. (1996) Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference, *Journal of Molecular Evolution*, **43**(3), pp. 304–311.

Raven, J. A. and Edwards, D. (2001) Roots: evolutionary origins and biogeochemical significance, *Journal of Experimental Botany*, **52**, pp. 381–401.

Read, B. A. *et al.* (2013) Pan genome of the phytoplankton Emiliania underpins its global distribution, *Nature*, **499**(7457), pp. 209–213.

Rebolleda- Gómez, M., Forrester, N. J., Russell, A. L., Wei, N., Fetters, A. M., Stephens, J. D. and Ashman, T. (2019) Gazing into the anthosphere: considering how microbes influence floral evolution, *New Phytologist*, **224**(3), pp. 1012–1020.

Rendón-Anaya, M. *et al.* (2019) The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation, *Proceedings of the National Academy of Sciences of the United States of America*, **116**(34), pp. 17081–17089.

Rensing, S. A. *et al.* (2008) The Physcomitrella Genome Reveals Evolutionary Insights into the Conquest of Land by Plants, *Science*, **319**(5859), pp. 64–69.

Rensing, S. A. (2017) Why we need more non-seed plant models, *New Phytologist*, **216**(2), pp. 355–360.

Renzaglia, K. S., Villarreal, J. C., Piatkowski, B. T., Lucas, J. R. and Merced, A. (2017) Hornwort stomata: Architecture and fate shared with 400-million-year-old fossil plants

without leaves, *Plant Physiology*, **174**(2), pp. 788–797.

Revell, L. J. (2012) phytools: An R package for phylogenetic comparative biology (and other things), *Methods in Ecology and Evolution*, **3**(2), pp. 217–223.

Reyes-Chin-Wo, S. *et al.* (2017) Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce, *Nature Communications*, **8**(14953).

Rio, D. C., Ares, M., Hannon, G. J. and Nilsen, T. W. (2010) Purification of RNA using TRIzol (TRI Reagent), *Cold Spring Harbor Protocols*, **5**(6).

Roca Paixão, J. F., Gillet, F. X., Ribeiro, T. P., Bournaud, C., Lourenço-Tessutti, I. T., Noriega, D. D., Melo, B. P. de, de Almeida-Engler, J. and Grossi-de-Sa, M. F. (2019) Improved drought stress tolerance in Arabidopsis by CRISPR/dCas9 fusion with a Histone AcetylTransferase, *Scientific Reports*, **9**(1), pp. 1–9.

Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S. C., Roure, B., Burger, G., Löffelhardt, W., Bohnert, H. J., Philippe, H. and Lang, B. F. (2005) Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes, *Current Biology*, **15**(14), pp. 1325–1330.

Ross-Ibarra, J., Morrell, P. L. and Gaut, B. S. (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation, *Proceedings of the National Academy of Sciences of the United States of America*, **104**, pp. 8641–8648.

Roth, M. S., Cokus, S. J., Gallaher, S. D., Walter, A., Lopez, D., Erickson, E., Endelman, B., Westcott, D., Larabell, C. A., Merchant, S. S., Pellegrini, M. and Niyogi, K. K. (2017) Chromosome-level genome assembly and transcriptome of the green alga Chromochloris zofingiensis illuminates astaxanthin production, *Proceedings of the National Academy of Sciences of the United States of America*, **114**(21), pp. 4296–4305.

Rubinstein, C. V., Gerrienne, P., de la Puente, G. S., Astini, R. A. and Steemans, P. (2010) Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana), *New Phytologist*, **188**(2), pp. 365–369.

Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E. and Burleigh, J. G. (2014) From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes, *BMC Evolutionary Biology*, **14**(23).

Ruonala, R., Ko, D. and Helariutta, Y. (2017) Genetic Networks in Plant Vascular

Development, *Annual Review of Genetics*, **51**(1), pp. 335–359.

Ruprecht, C., Lohaus, R., Vanneste, K., Mutwil, M., Nikoloski, Z., Van de Peer, Y. and Persson, S. (2017) Revisiting ancestral polyploidy in plants, *Science Advances*, **3**(7).

Russ, C., Lang, B. F., Chen, Z., Gujja, S., Shea, T., Zeng, Q., Young, S., Cuomo, C. A. and Nusbaum, C. (2016) Genome sequence of Spizellomyces punctatus, *Genome Announcements*, **4**(4).

Ruszala, E. M., Beerling, D. J., Franks, P. J., Chater, C., Casson, S. A., Gray, J. E. and Hetherington, A. M. (2011) Land Plants Acquired Active Stomatal Control Early in Their Evolutionary History, *Current Biology*, **21**(12), pp. 1030–1035.

Ryan, J. F., Pang, K., Schnitzler, C. E., Nguyen, A. D., Moreland, R. T., Simmons, D. K., Koch, B. J., Francis, W. R., Havlak, P., Smith, S. A., Putnam, N. H., Haddock, S. H. D., Dunn, C. W., Wolfsberg, T. G., C.Mullikin, J., Martindale, M. Q. and Baxevanis, A. D. (2013) The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution, *Science*, **342**(6164).

De Rybel, B. *et al.* (2010) A novel Aux/IAA28 signaling cascade activates GATA23-dependent specification of lateral root founder cell identity, *Current Biology*, **20**(19), pp. 1697–1706.

Saeed, W., Naseem, S. and Ali, Z. (2017) Strigolactones Biosynthesis and Their Role in Abiotic Stress Resilience in Plants: A Critical Review, *Frontiers in Plant Science*, **8**(1487).

Safronov, O., Kreuzwieser, J., Haberer, G., Alyousif, M. S., Schulze, W., Al-Harbi, N., Arab, L., Ache, P., Stempfl, T., Kruse, J., Mayer, K. X., Hedrich, R., Rennenberg, H., Salojärvi, J. and Kangasjärvi, J. (2017) Detecting early signs of heat and drought stress in Phoenix dactylifera (date palm), *PLOS One*, **12**(6).

Sahitya, U. L., Krishna, M. S. R. and Suneetha, P. (2019) Integrated approaches to study the drought tolerance mechanism in hot pepper (Capsicum annuum L.), *Physiology and Molecular Biology of Plants*, **25**(3), pp. 637–647.

Salminen, T. A., Blomqvist, K. and Edqvist, J. (2016) Lipid transfer proteins: classification, nomenclature, structure, and function, *Planta*, **244**(5), pp. 971–997.

Salvi, E., Rutten, J. P., Di Mambro, R., Polverari, L., Licursi, V., Negri, R., Dello Ioio, R.,

Sabatini, S. and Ten Tusscher, K. (2020) A Self-Organized PLT/Auxin/ARR-B Network Controls the Dynamics of Root Zonation Development in Arabidopsis thaliana, *Developmental Cell*, **53**(4), pp. 431–443.

Santos Teixeira, J. A. and ten Tusscher, K. H. (2019) The Systems Biology of Lateral Root Formation: Connecting the Dots, *Molecular Plant*, **12**(6), pp. 784–803.

Sapeta, H., Lourenço, T., Lorenz, S., Grumaz, C., Kirstahler, P., Barros, P. M., Costa, J. M., Sohn, K. and Oliveira, M. M. (2016) Transcriptomics and physiological analyses reveal co-ordinated alteration of metabolic pathways in Jatropha curcas drought tolerance, *Journal of Experimental Botany*, **67**(3), pp. 845–860.

Sarkar, D. *et al.* (2017) The draft genome of Corchorus olitorius cv. JRO-524 (Navin), *Genomics Data*, **12**, pp. 151–154.

Sato, A., Sato, Y., Fukao, Y., Fujiwara, M., Umezawa, T., Shinozaki, K., Hibi, T., Taniguchi, M., Miyake, H., Goto, D. B. and Uozumi, N. (2009) Threonine at position 306 of the KAT1 potassium channel is essential for channel activity and is a target site for ABA-activated SnRK2/OST1/SnRK2.6 protein kinase, *Biochemical Journal*, **424**(3), pp. 439–448.

Sato, S. *et al.* (2012) The tomato genome sequence provides insights into fleshy fruit evolution, *Nature*, **485**(7400), pp. 635–641.

Sauquet, H. *et al.* (2017) The ancestral flower of angiosperms and its early diversification, *Nature Communications*, **8**(16047).

Saxena, R. K., Edwards, D. and Varshney, R. K. (2014) Structural variations in plant genomes, *Briefings in Functional Genomics*, **13**(4), pp. 296–307.

Schmutz, J. *et al.* (2010) Genome sequence of the palaeopolyploid soybean, *Nature*, **463**(7278), pp. 178–183.

Schmutz, J. *et al.* (2014) A reference genome for common bean and genome-wide analysis of dual domestications, *Nature Genetics*, **46**(7), pp. 707–713.

Schnable, P. S. *et al.* (2009) The B73 maize genome: Complexity, diversity, and dynamics, *Science*, **326**(5956), pp. 1112–1115.

Schönknecht, G., Chen, W. H., Ternes, C. M., Barbier, G. G., Shrestha, R. P., Stanke, M., Bräutigam, A., Baker, B. J., Banfield, J. F., Garavito, R. M., Carr, K., Wilkerson, C.,

Rensing, S. A., Gagneul, D., Dickenson, N. E., Oesterhelt, C., Lercher, M. J. and Weber, A. P. M. (2013) Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote, *Science*, **339**(6124), pp. 1207–1210.

Schuettpelz, E. *et al.* (2016) A community-derived classification for extant lycophytes and ferns, *Journal of Systematics and Evolution*, **54**(6), pp. 563–603.

Schwab, K. B. and Heber, U. (1984) Thylakoid membrane stability in drought-tolerant and drought-sensitive plants, *Planta*, **161**(1), pp. 37–45.

Scott, A. D., Zimin, A. V., Puiu, D., Workman, R., Britton, M., Zaman, S., Caballero, M., Read, A. C., Bogdanove, A. J., Burns, E., Wegrzyn, J., Timp, W., Salzberg, S. L. and Neale, D. B. (2020) A Reference Genome Sequence for Giant Sequoia, *G3: Genes, Genomes, Genetics*, **10**(11), pp. 3907–3919.

Serbet, R. and Rothwell, G. W. (1992) Characterizing the Most Primitive Seed Ferns. I. A Reconstruction of Elkinsia polymorpha, *International Journal of Plant Sciences*, **153**(4), pp. 602–621.

Sessa, E. B. *et al.* (2014) Between Two Fern Genomes, *GigaScience*, **3**(15).

Sfriso, A. A., Gallo, M. and Baldi, F. (2018) Phycoerythrin productivity and diversity from five red macroalgae, *Journal of Applied Phycology*, **30**(4), pp. 2523–2531.

Shan-An, H., Gu, Y. and Zi-Jie, P. (1997) Resources and Prospects of Ginkgo biloba in China, in *Ginkgo Biloba A Global Treasure*. Springer Japan, pp. 373–383.

Sheffield, J. and Wood, E. F. (2008) Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations, *Climate Dynamics*, **31**(1), pp. 79–105.

Shi, H. and Zhu, J. K. (2002) SOS4, a pyridoxal kinase gene, is required for root hair development in Arabidopsis, *Plant Physiology*, **129**(2), pp. 585–593.

Shinohara, H., Mori, A., Yasue, N., Sumida, K. and Matsubayashi, Y. (2016) Identification of three LRR-RKs involved in perception of root meristem growth factor in Arabidopsis, *Proceedings of the National Academy of Sciences of the United States of America*, **113**(14), pp. 3897–3902.

Shinozawa, A., Otake, R., Takezawa, D., Umezawa, T., Komatsu, K., Tanaka, K., Amagai, A., Ishikawa, S., Hara, Y., Kamisugi, Y., Cuming, A. C., Hori, K., Ohta, H.,

Takahashi, F., Shinozaki, K., Hayashi, T., Taji, T. and Sakata, Y. (2019) SnRK2 protein kinases represent an ancient system in plants for adaptation to a terrestrial environment, *Communications biology*, **2**(30).

Shirasawa, K., Isuzugawa, K., Ikenaga, M., Saito, Y., Yamamoto, T., Hirakawa, H. and Isobe, S. (2017) The genome sequence of sweet cherry (Prunus avium) for use in genomics-assisted breeding, *DNA Research*, **24**(5), pp. 499–508.

Shkolnik-Inbar, D. and Bar-Zvi, D. (2010) ABI4 mediates abscisic acid and cytokinin inhibition of lateral root formation by reducing polar auxin transport in Arabidopsis, *Plant Cell*, **22**(11), pp. 3560–3573.

Shkolnik, D., Nuriel, R., Bonza, M. C., Costa, A. and Fromm, H. (2018) MIZ1 regulates ECA1 to generate a slow, long-distance phloem-transmitted Ca2+ signal essential for root water tracking in Arabidopsis, *Proceedings of the National Academy of Sciences of the United States of America*, **115**(31), pp. 8031–8036.

Shulaev, V. *et al.* (2011) The genome of woodland strawberry (Fragaria vesca), *Nature Genetics*, **43**(2), pp. 109–116.

Sierro, N., Battey, J. N. D., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., Goepfert, S., Peitsch, M. C. and Ivanov, N. V. (2014) The tobacco genome sequence and its comparison with those of tomato and potato, *Nature Communications*, **5**(3833).

Sierro, N., Battey, J. N., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N., Peitsch, M. C. and Ivanov, N. V (2013) Reference genomes and transcriptomes of Nicotiana sylvestris and Nicotiana tomentosiformis, *Genome Biology*, **14**(6).

Silva-Junior, O. B., Grattapaglia, D., Novaes, E. and Collevatti, R. G. (2018) Genome assembly of the Pink Ipê (Handroanthus impetiginosus, Bignoniaceae), a highly valued, ecologically keystone Neotropical timber forest tree, *GigaScience*, **7**(1).

Simakov, O. *et al.* (2013) Insights into bilaterian evolution from three spiralian genomes, *Nature*, **493**(7433), pp. 526–531.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E. M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics*, **31**(19), pp. 3210–3212.

Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D. J., Di Franco, A., Roure, B.,

Satoh, N., Quéinnec, É., Ereskovsky, A., Lapébie, P., Corre, E., Delsuc, F., King, N., Wörheide, G. and Manuel, M. (2017) A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals, *Current Biology*, **27**(7), pp. 958–967.

Simões, I. and Faro, C. (2004) Structure and function of plant aspartic proteinases, *European Journal of Biochemistry*, **271**(11), pp. 2067–2075.

Simonin, K. A. and Roddy, A. B. (2018) Genome downsizing, physiological novelty, and the global dominance of flowering plants, *PLOS Biology*, **16**(1).

Singh, R. *et al.* (2013) Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds, *Nature*, **500**(7462), pp. 335–339.

Sinha, N. R. (2011) Plant Developmental Biology in the Post-Genomic Era, *Frontiers in Plant Science*, **2**(11).

Slotte, T. *et al.* (2013) The Capsella rubella genome and the genomic consequences of rapid mating system evolution, *Nature Genetics*, **45**(7), pp. 831–835.

De Smet, I. *et al.* (2010) Bimodular auxin response controls organogenesis in Arabidopsis, *Proceedings of the National Academy of Sciences of the United States of America*, **107**(6), pp. 2705–2710.

Smet, W., Sevilem, I., Balaguer, M. A. de L., Wybouw, B., Mor, E., Miyashima, S., Blob, B., Roszak, P., Jacobs, T. B., Boekschoten, M., Hooiveld, G., Sozzani, R., Helariutta, Y. and Rybel, B. De (2019) DOF2.1 Controls Cytokinin-Dependent Vascular Cell Proliferation Downstream of TMO5/LHW, *Current Biology*, **29**(3), pp. 520–529.

Smith, S. A. and Dunn, C. W. (2008) Phyutility: A phyloinformatics tool for trees, alignments and molecular data, *Bioinformatics*, **24**(5), pp. 715–716.

Smith, S. and de Smet, I. (2012) Root system architecture: Insights from Arabidopsis and cereal crops, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**(1595), pp. 1441–1452.

Sofo, A. (2011) Drought stress tolerance and photoprotection in two varieties of olive tree, *Acta Agriculturae Scandinavica Section B: Soil and Plant Science*, **61**(8), pp. 711–720.

Sollars, E. S. A. *et al.* (2017) Genome sequence and genetic diversity of European ash

trees, *Nature*, **541**(7636), pp. 212–216.

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., Sankoff, D., Depamphilis, C. W., Wall, P. K. and Soltis, P. S. (2009) Polyploidy and angiosperm diversification, *American Journal of Botany*, **96**(1), pp. 336–48.

Soltis, P. S. and Soltis, D. E. (2016) Ancient WGD events as drivers of key innovations in angiosperms, *Current Opinion in Plant Biology*, **30**, pp. 159–165.

Soltis, P. S. and Soltis, D. E. (2020) Plant genomes: Markers of evolutionary history and drivers of evolutionary change, *Plants, People, Planet*, **3**(1), pp. 74–82.

Sozzani, R. and Iyer-Pascuzzi, A. (2014) Postembryonic control of root meristem growth and development, *Current Opinion in Plant Biology*, **17**(1), pp. 7–12.

Spicer, R. and Groover, A. (2010) Evolution of development of vascular cambia and secondary growth, *New Phytologist*, **186**(3), pp. 577–592.

Springer, M. S. and Gatesy, J. (2014) Land plant origins and coalescence confusion, *Trends in Plant Science*, **19**(5), pp. 267–269.

Stein, J. C. *et al.* (2018) Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza, *Nature Genetics*, **50**(2).

Stein, W. E., Mannolini, F., Hernick, L. V., Landing, E. and Berry, C. M. (2007) Giant cladoxylopsid trees resolve the enigma of the Earth's earliest forest stumps at Gilboa, *Nature*, **446**(7138), pp. 904–907.

Stevenson, S. R., Kamisugi, Y., Trinh, C. H., Schmutz, J., Jenkins, J. W., Grimwood, J., Muchero, W., Tuskan, G. A., Rensing, S. A., Lang, D., Reski, R., Melkonian, M., Rothfels, C. J., Li, F. W., Larsson, A., Wong, G. K. S., Edwards, T. A. and Cuming, A. C. (2016) Genetic analysis of Physcomitrella patens identifies ABSCISIC ACID NON-RESPONSIVE, a regulator of ABA responses unique to basal land plants and required for desiccation tolerance, *Plant Cell*, **28**(6), pp. 1310–1327.

Stoeckle, D., Thellmann, M. and Vermeer, J. E. (2018) Breakout — lateral root emergence in Arabidopsis thaliana, *Current Opinion in Plant Biology*, **41**, pp. 67–72.

Su, X., Wei, F., Huo, Y. and Xia, Z. (2017) Comparative Physiological and Molecular Analyses of Two Contrasting Flue-Cured Tobacco Genotypes under Progressive

Drought Stress, *Frontiers in Plant Science*, **8**(827).

Sugano, S. S., Shimada, T., Imai, Y., Okawa, K., Tamai, A., Mori, M. and Hara-Nishimura, I. (2010) Stomagen positively regulates stomatal density in Arabidopsis, *Nature*, **463**(7278), pp. 241–244.

Sun, G., Dilcher, D. L., Zheng, S. and Zhou, Z. (1998) In search of the first flower: A Jurassic angiosperm, Archaefructus, from Northeast China, *Science*, **282**(5394), pp. 1692–1695.

Sun, G., Ji, Q., Dilcher, D. L., Zheng, S., Nixon, K. C. and Wang, X. (2002) Archaefructaceae, a new basal angiosperm family, *Science*, **296**(5569), pp. 899–904.

Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., Zhang, J., Zhang, H., Gong, G., Jia, Z., Zhang, F., Tian, J., Lucas, W. J., Doyle, J. J., Li, H., Fei, Z. and Xu, Y. (2017) Karyotype Stability and Unbiased Fractionation in the Paleo-Allotetraploid Cucurbita Genomes, *Molecular Plant*, **10**(10), pp. 1293–1306.

Sunil, M., Hariharan, A. K., Nayak, S., Gupta, S., Nambisan, S. R., Gupta, R. P., Panda, B., Choudhary, B. and Srinivasan, S. (2014) The draft genome and transcriptome of amaranthus hypochondriacus: A C4 dicot producing high-lysine edible pseudo-cereal, *DNA Research*, **21**(6), pp. 585–602.

Sussmilch, F. C., Atallah, N. M., Brodribb, T. J., Banks, J. A. and McAdam, S. A. M. (2017a) Abscisic acid (ABA) and key proteins in its perception and signaling pathways are ancient, but their roles have changed through time, *Plant Signaling & Behavior*, **12**(9).

Sussmilch, F. C., Schultz, J., Hedrich, R., Rob, M. and Roelfsema, G. (2019) Acquiring Control: The Evolution of Stomatal Signalling Pathways, *Trends in Plant Science*, **24**(4), pp. 342–351.

Sussmilch, F. and McAdam, S. (2017b) Surviving a Dry Future: Abscisic Acid (ABA)-Mediated Plant Mechanisms for Conserving Water under Low Humidity, *Plants*, **6**(4).

Svennblad, B., Erixon, P., Oxelman, B. and Britton, T. (2006) Fundamental Differences Between the Methods of Maximum Likelihood and Maximum Posterior Probability in Phylogenetics, *Systematic Biology*. Edited by R. Page, **55**(1), pp. 116–121.

Swarup, K. *et al.* (2008) The auxin influx carrier LAX3 promotes lateral root emergence,

Nature Cell Biology, **10**(8), pp. 946–954.

Szathmáry, E. and Smith, J. M. (1995) The major evolutionary transitions, *Nature*, **374**(6519), pp. 227–232.

Tagliani, A., Tran, A. N., Novi, G., Di Mambro, R., Pesenti, M., Sacchi, G. A., Perata, P. and Pucciariello, C. (2020) The calcineurin β-like interacting protein kinase CIPK25 regulates potassium homeostasis under low oxygen in Arabidopsis, *Journal of Experimental Botany*, **71**(9), pp. 2678–2689.

Tamaoki, M. (2008) The role of phytohormone signaling in ozone-induced cell death in plants, *Plant Signaling & Behavior*, **3**(3), pp. 166–74.

Tamiru, M. *et al.* (2017) Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination, *BMC Biology*, **15**(86).

Tanaka, H., Hirakawa, H., Kosugi, S., Nakayama, S., Ono, A., Watanabe, A., Hashiguchi, M., Gondo, T., Ishigaki, G., Muguerza, M., Shimizu, K., Sawamura, N., Inoue, T., Shigeki, Y., Ohno, N., Tabata, S., Akashi, R. and Sato, S. (2016) Sequencing and comparative analyses of the genomes of zoysiagrasses, *DNA Research*, **23**(2), pp. 171–180.

Tang, C. *et al.* (2016) The rubber tree genome reveals new insights into rubber production and species adaptation, *Nature Plants*, **2**(6).

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M. and Paterson, A. H. (2008) Synteny and collinearity in plant genomes, *Science*, **320**(5875), pp. 486–488.

Tang, Q., Pang, K., Yuan, X. and Xiao, S. (2020) A one-billion-year-old multicellular chlorophyte, *Nature Ecology and Evolution*, **4**(4), pp. 543–549.

Tang, S., Dong, Y., Liang, D., Zhang, Z., Ye, C. Y., Shuai, P., Han, X., Zhao, Y., Yin, W. and Xia, X. (2015) Analysis of the Drought Stress-Responsive Transcriptome of Black Cottonwood (Populus trichocarpa) Using Deep RNA Sequencing, *Plant Molecular Biology Reporter*, **33**(3), pp. 424–438.

Teh, B. T., Lim, K., Yong, C. H., Ng, C. C. Y., Rao, S. R., Rajasegaran, V., Lim, W. K., Ong, C. K., Chan, K., Cheng, V. K. Y., Soh, P. S., Swarup, S., Rozen, S. G., Nagarajan, N. and Tan, P. (2017) The draft genome of tropical fruit durian (Durio zibethinus), *Nature*

Genetics, **49**(11), pp. 1633–1641.

Tezara, W., Colombo, R., Coronel, I. and Marín, O. (2011) Water relations and photosynthetic capacity of two species of Calotropis in a tropical semi-arid ecosystem, *Annals of Botany*, **107**(3), pp. 397–405.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana, *Nature*, **408**(6814), pp. 796–815.

The Brassica rapa Genome Sequencing Project Consortium *et al.* (2011) The genome of the mesopolyploid crop species Brassica rapa, *Nature Genetics*, **43**(10), pp. 1035–1040.

The C. elegans Sequencing Consortium (1998) Genome sequence of the nematode C. elegans: A platform for investigating biology, *Science*, **282**(5396), pp. 2012–2018.

The Inkscape Project (2019) *Inkscape v0.92.4*. Available at: https://inkscape.org/.

Tomescu, A. M. F. (2009) Megaphylls, microphylls and the evolution of leaf development, *Trends in Plant Science*, **14**(1), pp. 5–12.

Tominaga, R., Iwata, M., Sano, R., Inoue, K., Okada, K. and Wada, T. (2008) Arabidopsis CAPRICE-LIKE MYB 3 (CPL3) controls endoreduplication and flowering development in addition to trichome and root hair formation, *Development*, **135**(7), pp. 1335–1345.

Tomitani, A., Okada, K., Miyashita, H., Matthijs, H. C. P., Ohno, T. and Tanaka, A. (1999) Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts, *Nature*, **400**(6740), pp. 159–162.

Trueba, S., Delzon, S., Isnard, S. and Lens, F. (2019) Similar hydraulic efficiency and safety across vesselless angiosperms and vessel-bearing species with scalariform perforation plates, *Journal of Experimental Botany*, **70**(12), pp. 3227–3240.

Tuberosa, R. and Salvi, S. (2006) Genomics-based approaches to improve drought tolerance of crops, *Trends in Plant Science*, **11**(8), pp. 405–412.

Tuskan, G. A. *et al.* (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray), *Science*, **313**(5793), pp. 1596–1604.

Ubeda-Tomás, S., Swarup, R., Coates, J., Swarup, K., Laplaze, L., Beemster, G. T. S., Hedden, P., Bhalerao, R. and Bennett, M. J. (2008) Root growth in Arabidopsis requires

gibberellin/DELLA signalling in the endodermis, *Nature Cell Biology*, **10**(5), pp. 625–628.

Uga, Y. *et al.* (2013) Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions, *Nature Genetics*, **45**(9), pp. 1097–1102.

Umen, J. G. (2014) Green algae and the origins of multicellularity in the plant kingdom, *Cold Spring Harbor Perspectives in Biology*, **6**(11).

Umezawa, T., Fujita, M., Fujita, Y., Yamaguchi-Shinozaki, K. and Shinozaki, K. (2006) Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future, *Current Opinion in Biotechnology*, **17**(2), pp. 113–122.

UniProt Consortium, T. (2018) UniProt: the universal protein knowledgebase, *Nucleic Acids Research*, **46**(5), pp. 2699–2699.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. and Rozen, S. G. (2012) Primer3 - new capabilities and interfaces, *Nucleic Acids Research*, **40**(15).

Vanburen, R., Bryant, D., Edger, P. P., Tang, H., Burgess, D., Challabathula, D., Spittle, K., Hall, R., Gu, J., Lyons, E., Freeling, M., Bartels, D., Ten Hallers, B., Hastie, A., Michael, T. P. and Mockler, T. C. (2015) Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum, *Nature*, **527**(7579), pp. 508–511.

VanBuren, R., Wai, C. M., Ou, S., Pardo, J., Bryant, D., Jiang, N., Mockler, T. C., Edger, P. and Michael, T. P. (2018) Extreme haplotype variation in the desiccation-tolerant clubmoss Selaginella lepidophylla, *Nature Communications*, **9**(13).

Vanneste, K., Baele, G., Maere, S., Van de Peer, Y. and Peer, Y. Van De (2014) Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous − Paleogene boundary, *Genome Research*, **32**(9052), pp. 1334–1347.

Varshney, R. K. *et al.* (2012) Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers, *Nature Biotechnology*, **30**(1), pp. 83–89.

Varshney, R. K. *et al.* (2013) Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement, *Nature Biotechnology*, **31**(3), pp. 240–246.

Varshney, R. K., Thudi, M., Nayak, S. N., Gaur, P. M., Kashiwagi, J., Krishnamurthy, L., Jaganathan, D., Koppolu, J., Bohra, A., Tripathi, S., Rathore, A., Jukanti, A. K.,

Jayalakshmi, V., Vemula, A., Singh, S. J., Yasin, M., Sheshshayee, M. S. and Viswanatha, K. P. (2014) Genetic dissection of drought tolerance in chickpea (Cicer arietinum L.), *Theoretical and Applied Genetics*, **127**(2), pp. 445–462.

Vaughan-Hirsch, J., Goodall, B. and Bishopp, A. (2018) North, East, South, West: mapping vascular tissues onto the Arabidopsis root, *Current Opinion in Plant Biology*, **41**, pp. 16–22.

Vaughan, D. A., Morishima, H. and Kadowaki, K. (2003) Diversity in the Oryza genus, *Current Opinion in Plant Biology*, **6**(2), pp. 139–146.

De Vega, J. J. *et al.* (2015) Red clover (Trifolium pratense L.) draft genome provides a platform for trait improvement, *Scientific Reports*, **5**(17394).

Velasco, R. *et al.* (2010) The genome of the domesticated apple (Malus × domestica Borkh.), *Nature Genetics*, **42**(10), pp. 833–839.

Vera-Sirera, F., De Rybel, B., Úrbez, C., Kouklas, E., Pesquera, M., Álvarez-Mahecha, J. C., Minguet, E. G., Tuominen, H., Carbonell, J., Borst, J. W., Weijers, D. and Blázquez, M. A. (2015) A bHLH-Based Feedback Loop Restricts Vascular Cell Proliferation in Plants, *Developmental Cell*, **35**(4), pp. 432–443.

Verde, I. *et al.* (2013) The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution, *Nature Genetics*, **45**(5), pp. 487–494.

Verstraeten, I., Schotte, S. and Geelen, D. (2014) Hypocotyl adventitious root organogenesis differs from lateral root development, *Frontiers in Plant Science*, **5**, p. 495.

Vicient, C. M. and Casacuberta, J. M. (2017) Impact of transposable elements on polyploid plant genomes, *Annals of Botany*, **120**(2), pp. 195–207.

Vicient, C. M. and Casacuberta, J. M. (2020) Additional ORFs in Plant LTR-Retrotransposons, *Frontiers in Plant Science*, **11**(555).

Vissenberg, K., Claeijs, N., Balcerowicz, D. and Schoenaers, S. (2020) Hormonal regulation of root hair growth and responses to the environment in Arabidopsis, *Journal of Experimental Botany*, **71**(8), pp. 2412–2427.

Vogel, A., Schwacke, R., Denton, A. K., Usadel, B., Hollmann, J., Fischer, K., Bolger, A.,

Schmidt, M. H.-W., Bolger, M. E., Gundlach, H., Mayer, K. F. X., Weiss-Schneeweiss, H., Temsch, E. M. and Krause, K. (2018) Footprints of parasitism in the genome of the parasitic flowering plant Cuscuta campestris, *Nature Communications*, **9**(2515).

Vogel, J. P. *et al.* (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon, *Nature*, **463**(7282), pp. 763–768.

Volkova, L., Bennett, L. T., Merchant, A. and Tausz, M. (2010) Shade does not ameliorate drought effects on the tree fern species Dicksonia antarctica and Cyathea australis, *Trees - Structure and Function*, **24**(2), pp. 351–362.

Vosolsobě, S., Skokan, R. and Petrášek, J. (2020) The evolutionary origins of auxin transport: what we know and what we need to know, *Journal of Experimental Botany*, **71**(11), pp. 3287–3295.

de Vries, J. and Archibald, J. M. (2018a) Plant evolution: landmarks on the path to terrestrial life, *New Phytologist*, **217**(4), pp. 1428–1434.

de Vries, J., Curtis, B. A., Gould, S. B. and Archibald, J. M. (2018b) Embryophyte stress signaling evolved in the algal progenitors of land plants, *Proceedings of the National Academy of Sciences of the United States of America*, **115**(15), pp. 3471–3480.

de Vries, J., Stanton, A., Archibald, J. M. and Gould, S. B. (2016) Streptophyte Terrestrialization in Light of Plastid Evolution, *Trends in Plant Science*, **21**(6), pp. 467–476.

Walkowiak, S. *et al.* (2020) Multiple wheat genomes reveal global variation in modern breeding, *Nature*, **588**, pp. 277–283.

Wan Nazri, W. B., Ezdiani, Z. N., Romainor, M. M., Erma, K. S., Jurina, J. and Noor Fadzlina, I. Z. A. (2014) Effect of fibre loading on mechanical properties of durian skin fibre composite, *Journal of Tropical Agriculture and Food Science*, **42**(2), pp. 169–174.

Wan, T. *et al.* (2018a) A genome for gnetophytes and early evolution of seed plants, *Nature Plants*, **4**(2), pp. 82–89.

Wan, X. L., Zhou, Q., Wang, Y. Y., Wang, W. E., Bao, M. Z. and Zhang, J. W. (2015) Identification of heat-responsive genes in carnation (Dianthus caryophyllus L.) by RNA-seq, *Frontiers in Plant Science*, **6**(519).

Wan, X., Zou, L. H., Zheng, B. Q., Tian, Y. Q. and Wang, Y. (2018b) Transcriptomic

profiling for prolonged drought in dendrobium catenatum, *Scientific Data*, **5**(1), pp. 1–9.

Wang, C. J., Yang, W., Wang, C., Gu, C., Niu, D. D., Liu, H. X., Wang, Y. P. and Guo, J. H. (2012a) Induction of Drought Tolerance in Cucumber Plants by a Consortium of Three Plant Growth-Promoting Rhizobacterium Strains, *PLoS ONE*, **7**(12).

Wang, C., Liu, Y., Li, S.-S. S. and Han, G.-Z. Z. (2015) Insights into the origin and evolution of the plant hormone signaling machinery, *Plant Physiology*, **167**(3), pp. 872–886.

Wang, J. *et al.* (2012b) Genome sequence of foxtail millet (Setaria italica) provides insights into grass evolution and biofuel potential, *Nature Biotechnology*, **30**(6), pp. 549–554.

Wang, K. *et al.* (2012c) The draft genome of a diploid cotton Gossypium raimondii, *Nature Genetics*, **44**(10), pp. 1098–1103.

Wang, L. *et al.* (2014a) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis, *Genome Biology*, **15**(2).

Wang, M. *et al.* (2014b) The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication, *Nature Genetics*, **46**(9), pp. 982–988.

Wang, N., Zhang, W., Qin, M., Li, S., Qiao, M., Liu, Z. and Xiang, F. (2017a) Drought tolerance conferred in soybean (Glycine max. L) by GmMYB84, a novel R2R3-MYB transcription factor, *Plant and Cell Physiology*, **58**(10), pp. 1764–1776.

Wang, S. *et al.* (2019) Genomes of early-diverging streptophyte algae shed light on plant terrestrialization, *Nature Plants*, **6**, pp. 95–106.

Wang, W. *et al.* (2014c) Cassava genome from a wild ancestor to cultivated varieties, *Nature Communications*, **5**(5110).

Wang, W. *et al.* (2014d) The Spirodela polyrhiza genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle, *Nature Communications*, **5**(3311).

Wang, X.-Q. and Ran, J.-H. (2014e) Evolution and biogeography of gymnosperms, *Molecular Phylogenetics and Evolution*, **75**, pp. 24–40.

Wang, X. (2005) Regulatory functions of phospholipase D and phosphatidic acid in plant growth, development, and stress responses, *Plant Physiology*, **139**(2), pp. 566–573.

Wang, X. *et al.* (2017b) Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction, *Nature Genetics*, **49**(5), pp. 765–772.

Wang, X., Chen, S., Zhang, H., Shi, L., Cao, F., Guo, L., Xie, Y., Wang, T., Yan, X. and Dai, S. (2010) Desiccation Tolerance Mechanism in Resurrection Fern-Ally *Selaginella tamariscina* Revealed by Physiological and Proteomic Analysis, *Journal of Proteome Research*, **9**(12), pp. 6561–6577.

Wang, X., Zhang, H., Gao, Y., Sun, G., Zhang, W. and Qiu, L. (2014f) A Comprehensive Analysis of the Cupin Gene Family in Soybean (Glycine max), *PLoS ONE*, **9**(10).

Wang, Z. *et al.* (2012d) The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads, *The Plant Journal*, **72**(3), pp. 461–473.

Wang, Z., Hong, Y., Li, Y., Shi, H., Yao, J., Liu, X., Wang, F., Huang, S., Zhu, G. and Zhu, J. (2020a) Natural variations in SlSOS1 contribute to the loss of salt tolerance during tomato domestication, *Plant Biotechnology Journal*.

Wang, Z., Hong, Y., Zhu, G., Li, Y., Niu, Q., Yao, J., Hua, K., Bai, J., Zhu, Y., Shi, H., Huang, S. and Zhu, J. (2020b) Loss of salt tolerance during tomato domestication conferred by variation in a Na^+/K^+ transporter, *The EMBO Journal*, **39**(10).

von Wangenheim, D., Banda, J., Schmitz, A., Boland, J., Bishopp, A., Maizel, A., Stelzer, E. H. K. and Bennett, M. (2020) Early developmental plasticity of lateral roots in response to asymmetric water availability, *Nature Plants*, **6**(2), pp. 73–77.

Warpeha, K. M., Upadhyay, S., Yeh, J., Adamiak, J., Hawkins, S. I., Lapik, Y. R., Anderson, M. B. and Kaufman, L. S. (2007) The GCR1, GPA1, PRN1, NF-Y signal chain mediates both blue light and abscisic acid responses in arabidopsis, *Plant Physiology*, **143**(4), pp. 1590–1600.

Wasternack, C. and Hause, B. (2013) Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany, *Annals of Botany*, **111**(6), pp. 1021–58.

Wei, H., Feng, F., Lou, Q., Xia, H., Ma, X., Liu, Y., Xu, K., Yu, X., Mei, H. and Luo, L. (2016) Genetic determination of the enhanced drought resistance of rice maintainer HuHan2B by pedigree breeding, *Scientific Reports*, **6**(1), pp. 1–11.

Weingartner, M., Subert, C. and Sauer, N. (2011) LATE, a C2H2 zinc-finger protein that

acts as floral repressor, *The Plant Journal*, **68**(4), pp. 681–692.

Wellcome Sanger Institute (2020) *Darwin Tree Of Life*. Available at: https://www.darwintreeoflife.org/ (Accessed: 17 August 2020).

Wendel, J. F., Jackson, S. A., Meyers, B. C. and Wing, R. A. (2016) Evolution of plant genome architecture, *Genome Biology*, **17**(37).

Whelan, N. V., Kocot, K. M., Moroz, T. P., Mukherjee, K., Williams, P., Paulay, G., Moroz, L. L. and Halanych, K. M. (2017) Ctenophore relationships and their placement as the sister group to all other animals, *Nature Ecology and Evolution*, **1**(11), pp. 1737–1746.

Whitmore, J. S. (2000) Vegetables for Drought-Prone Areas, in *Drought Management on Farmland*. Springer, Dordrecht, pp. 106–111.

Wickell, D. A. and Li, F. (2019) On the evolutionary significance of horizontal gene transfers in plants, *New Phytologist*, **225**(1), pp. 113–117.

Wickett, N. J. *et al.* (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants, *Proceedings of the National Academy of Sciences*, **111**(45), pp. 4859–4868.

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.

Wilhelmsson, P. K. I., Mühlich, C., Ullrich, K. K. and Rensing, S. A. (2017) Comprehensive Genome-Wide Classification Reveals That Many Plant-Specific Transcription Factors Evolved in Streptophyte Algae, *Genome Biology and Evolution*, **9**(12), pp. 3384–3397.

Willis, K. J. (2017) *State of the World's Plants 2017*. Royal Botanic Gardens Kew.

Wilts, B. D., Rudall, P. J., Moyroud, E., Gregory, T., Ogawa, Y., Vignolini, S., Steiner, U. and Glover, B. J. (2018) Ultrastructure and optics of the prism-like petal epidermal cells of *Eschscholzia californica* (California poppy), *New Phytologist*, **219**(3), pp. 1124–1133.

Wiriya-Alongkorn, W., Spreer, W., Ongprasert, S., Spohrer, K., Pankasemsuk, T. and Müller, J. (2013) Detecting drought stress in longan tree using thermal imaging, *Maejo International Journal of Science and Technology*, **7**(1), pp. 166–180.

Wood, A. J. (2007a) Eco-physiological Adaptations to Limited Water Environments, in *Plant Abiotic Stress*. Blackwell Publishing Ltd, pp. 1–13.

Wood, A. J. (2007b) The nature and distribution of vegetative desiccation-tolerance in hornworts, liverworts and mosses, *The Bryologist*, **110**(2), pp. 163–177.

Worden, A. Z. *et al.* (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes micromonas, *Science*, **324**(5924), pp. 268–272.

Wu, G. A. *et al.* (2014) Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication, *Nature Biotechnology*, **32**(7), pp. 656–662.

Wu, J. *et al.* (2013) The genome of the pear (Pyrus bretschneideri Rehd.), *Genome Research*, **23**(2), pp. 396–408.

Wu, P. *et al.* (2015) Integrated genome sequence and linkage map of physic nut (Jatropha curcas L.), a biodiesel plant, *The Plant Journal*, **81**(5), pp. 810–821.

Wu, S., Shamimuzzaman, M., Sun, H., Salse, J., Sui, X., Wilder, A., Wu, Z., Levi, A., Xu, Y., Ling, K.-S. and Fei, Z. (2017) The bottle gourd genome provides insights into Cucurbitaceae evolution and facilitates mapping of a *Papaya ring-spot virus* resistance locus, *The Plant Journal*, **92**(5), pp. 963–975.

Wu, W., Ng, W. L., Yang, J. X., Li, W. M. and Ge, X. J. (2018a) High cryptic species diversity is revealed by genome-wide polymorphisms in a wild relative of banana, Musa itinerans, and implications for its conservation in subtropical China, *BMC Plant Biology*, **18**(1), pp. 1–11.

Wu, W., Yang, Y. L., He, W. M., Rouard, M., Li, W. M., Xu, M., Roux, N. and Ge, X. J. (2016) Whole genome sequencing of a banana wild relative Musa itinerans provides insights into lineage-specific diversification of the Musa genus, *Scientific Reports*, **6**(31586).

Wu, Z. Z., Ying, Y. Q., Zhang, Y. Bin, Bi, Y. F., Wang, A. K. and Du, X. H. (2018b) Alleviation of drought stress in Phyllostachys edulis by N and P application, *Scientific Reports*, **8**(228).

Xi, J.-J., Chen, H.-Y., Bai, W.-P., Yang, R.-C., Yang, P.-Z., Chen, R.-J., Hu, T.-M. and Wang, S.-M. (2018) Sodium-Related Adaptations to Drought: New Insights From the

Xerophyte Plant Zygophyllum xanthoxylum, *Frontiers in Plant Science*, **9**(1678).

Xia, E. H. *et al.* (2017) The Tea Tree Genome Provides Insights into Tea Flavor and Independent Evolution of Caffeine Biosynthesis, *Molecular Plant*, **10**(6), pp. 866–877.

Xiao, G. and Zhang, Y. (2020) Adaptive Growth: Shaping Auxin-Mediated Root System Architecture, *Trends in Plant Science*, **25**(2), pp. 121–123.

Xiao, L. *et al.* (2015) The resurrection genome of Boea hygrometrica: A blueprint for survival of dehydration, *Proceedings of the National Academy of Sciences of the United States of America*, **112**(18), pp. 5833–5837.

Xiao, L., Yobi, A., Koster, K. L., He, Y. and Oliver, M. (2018) Desiccation tolerance in Physcomitrella patens: Rate of dehydration and the involvement of endogenous abscisic acid (ABA), *Plant Cell and Environment*, **41**(1), pp. 275–284.

Xiao, Y., Xu, P., Fan, H., Baudouin, L., Xia, W., Bocs, S., Xu, J., Li, Q., Guo, A., Zhou, L., Li, J., Wu, Y., Ma, Z., Armero, A., Issali, A. E., Liu, N., Peng, M. and Yang, Y. (2017) The genome draft of coconut (Cocos nucifera), *GigaScience*, **6**(11).

Xie, M., Chen, H., Huang, L., O'Neil, R. C., Shokhirev, M. N. and Ecker, J. R. (2018) A B-ARR-mediated cytokinin transcriptional network directs hormone cross-regulation and shoot development, *Nature Communications*, **9**(1), pp. 1–13.

Xie, Z., Nolan, T. M., Jiang, H. and Yin, Y. (2019) AP2/ERF transcription factor regulatory networks in hormone and abiotic stress responses in Arabidopsis, *Frontiers in Plant Science*, **10**(228).

Xu, C., Jiao, C., Sun, H., Cai, X., Wang, X., Ge, C., Zheng, Y., Liu, W., Sun, X., Xu, Y., Deng, J., Zhang, Z., Huang, S., Dai, S., Mou, B., Wang, Q., Fei, Z. and Wang, Q. (2017a) Draft genome of spinach and transcriptome diversity of 120 Spinacia accessions, *Nature Communications*, **8**(15275).

Xu, J. *et al.* (2017b) Panax ginseng genome examination for ginsenoside biosynthesis, *GigaScience*, **6**(11).

Xu, Q. *et al.* (2013a) The draft genome of sweet orange (Citrus sinensis), *Nature Genetics*, **45**(1), pp. 59–66.

Xu, S., Brockmöller, T., Navarro-Quezada, A., Kuhl, H., Gase, K., Ling, Z., Zhou, W., Kreitzer, C., Stanke, M., Tang, H., Lyons, E., Pandey, P., Pandey, S. P., Timmermann,

B., Gaquerel, E. and Baldwin, I. T. (2017c) Wild tobacco genomes reveal the evolution of nicotine biosynthesis, *Proceedings of the National Academy of Sciences of the United States of America*, **114**(23), pp. 6133–6138.

Xu, W., Jia, L., Shi, W., Liang, J., Zhou, F., Li, Q. and Zhang, J. (2013b) Abscisic acid accumulation modulates auxin transport in the root tip to enhance proton secretion for maintaining root growth under moderate water stress, *New Phytologist*, **197**(1), pp. 139–150.

Xu, X. *et al.* (2011) Genome sequence and analysis of the tuber crop potato, *Nature*, **475**(7355), pp. 189–195.

Xu, Y. *et al.* (2013c) The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions, *Nature Genetics*, **45**(1), pp. 51–58.

Xu, Z., Xin, T., Bartels, D., Li, Y., Gu, W., Yao, H., Liu, S., Yu, H., Pu, X., Zhou, J., Xu, J., Xi, C., Lei, H., Song, J. and Chen, S. (2018) Genome Analysis of the Ancient Tracheophyte Selaginella tamariscina Reveals Evolutionary Features Relevant to the Acquisition of Desiccation Tolerance, *Molecular Plant*, **11**(7), pp. 983–994.

Yagi, M. *et al.* (2014) Sequence Analysis of the Genome of Carnation (Dianthus caryophyllus L.), *DNA Research*, **21**(3), pp. 231–241.

Yamauchi, Y., Ogawa, M., Kuwahara, A., Hanada, A., Kamiya, Y. and Yamaguchi, S. (2004) Activation of gibberellin biosynthesis and response pathways by low temperature during imbibition of Arabidopsis thaliana seeds, *The Plant Cell*, **16**(2), pp. 367–78.

Yang, J., Zhang, G., Zhang, J., Liu, H., Chen, W., Wang, X., Li, Y., Dong, Y. and Yang, S. (2017a) Hybrid de novo genome assembly of the Chinese herbal fleabane Erigeron breviscapus, *GigaScience*, **6**(6).

Yang, R., Jarvis, D. E., Chen, H., Beilstein, M. A., Grimwood, J., Jenkins, J., Shu, S., Prochnik, S., Xin, M., Ma, C., Schmutz, J., Wing, R. A., Mitchell-Olds, T., Schumaker, K. S. and Wang, X. (2013) The Reference Genome of the Halophytic Plant Eutrema salsugineum, *Frontiers in Plant Science*, **4**.

Yang, W., Wang, K., Zhang, J., Ma, J., Liu, J. and Ma, T. (2017b) The draft genome sequence of a desert tree Populus pruinosa, *GigaScience*, **6**(9).

Yang, X. *et al.* (2017c) The Kalanchoë genome provides insights into convergent

evolution and building blocks of crassulacean acid metabolism, *Nature Communications*, **8**(1).

Yang, Y. *et al.* (2020) Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution, *Nature Plants*, **6**, pp. 215–222.

Yang, Y., Moore, M. J., Brockington, S. F., Soltis, D. E., Wong, G. K. S., Carpenter, E. J., Zhang, Y., Chen, L., Yan, Z., Xie, Y., Sage, R. F., Covshoff, S., Hibberd, J. M., Nelson, M. N. and Smith, S. A. (2015) Dissecting molecular evolution in the highly diverse plant clade caryophyllales using transcriptome sequencing, *Molecular Biology and Evolution*, **32**(8), pp. 2001–2014.

Yang, Z., Dai, Z., Lu, R., Wu, B., Tang, Q., Xu, Y., Cheng, C. and Su, J. (2017d) Transcriptome Analysis of Two Species of Jute in Response to Polyethylene Glycol (PEG)- induced Drought Stress, *Scientific Reports*, **7**(16565).

Yasar, F., Uzal, O., Kose, S., Yasar, O. and Ellialtioglu, S. (2014) Enzyme activities of certain pumpkin (Cucurbita spp) species under drought stress, *Fresenius Environmental Bulletin*, **23**(4), pp. 1093–1099.

Yasui, Y., Hirakawa, H., Ueno, M., Matsui, K., Katsube-Tanaka, T., Yang, S. J., Aii, J., Sato, S. and Mori, M. (2016) Assembly of the draft genome of buckwheat and its applications in identifying agronomically useful genes, *DNA Research*, **23**(3), pp. 215–224.

Yobi, A., Wone, B. W. M., Xu, W., Alexander, D. C., Guo, L., Ryals, J. A., Oliver, M. J. and Cushman, J. C. (2013) Metabolomic profiling in selaginella lepidophylla at various hydration states provides new insights into the mechanistic basis of desiccation tolerance, *Molecular Plant*, **6**(2), pp. 369–385.

Yokoyama, A., Yamashino, T., Amano, Y.-I., Tajima, Y., Imamura, A., Sakakibara, H. and Mizuno, T. (2007) Type-B ARR Transcription Factors, ARR10 and ARR12, are Implicated in Cytokinin-Mediated Regulation of Protoxylem Differentiation in Roots of Arabidopsis thaliana, *Plant and Cell Physiology*, **48**(1), pp. 84–96.

Young, N. D. *et al.* (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses, *Nature*, **480**(7378), pp. 520–524.

Yu, J. *et al.* (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica), *Science*, **296**(5565), pp. 79–92.

Yu, J., Zhang, Y., Liu, W., Wang, H., Wen, S., Zhang, Y. and Xu, L. (2020) Molecular Evolution of Auxin-Mediated Root Initiation in Plants, *Molecular Biology Evolution*, **37**(5), pp. 1387–1393.

Yu, T. F., Xu, Z. S., Guo, J. K., Wang, Y. X., Abernathy, B., Fu, J. D., Chen, X., Zhou, Y. Bin, Chen, M., Ye, X. G. and Ma, Y. Z. (2017) Improved drought tolerance in wheat plants overexpressing a synthetic bacterial cold shock protein gene SeCspA, *Scientific Reports*, **7**(1), pp. 1–14.

Yu, Y., Tang, T., Qian, Wang, Y., Yan, M., Zeng, D., Han, B., Wu, C. I., Shi, S. and Li, J. (2008) Independent Losses of Function in a Polyphenol Oxidase in Rice: Differentiation in Grain Discoloration between Subspecies and the Role of Positive Selection under Domestication, *Plant Cell*, **20**(11), pp. 2946–2959.

Yuan, X., Huang, P., Wang, R., Li, H., Lv, X., Duan, M., Tang, H., Zhang, H. and Huang, J. (2018) A zinc finger transcriptional repressor confers pleiotropic effects on rice growth and drought tolerance by down-regulating stress-responsive genes, *Plant and Cell Physiology*, **59**(10), pp. 2129–2142.

Yue, J., Hu, X., Sun, H., Yang, Y. and Huang, J. (2012) Widespread impact of horizontal gene transfer on plant colonization of land, *Nature Communications*, **3**(1152).

Zhang, C., Bousquet, A. and Harris, J. M. (2014a) Abscisic acid and lateral root organ defective/numerous infections and polyphenolics modulate root elongation via reactive oxygen species in Medicago truncatula, *Plant Physiology*, **166**(2), pp. 644–658.

Zhang, C., Rabiee, M., Sayyari, E. and Mirarab, S. (2018a) ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees, *BMC Bioinformatics*, **19**(153).

Zhang, C., Scornavacca, C., Molloy, E. K. and Mirarab, S. (2020a) ASTRAL-Pro: Quartet-Based Species-Tree Inference despite Paralogy, *Molecular Biology and Evolution*, **37**(11), pp. 3292–3307.

Zhang, G. Q. *et al.* (2016) The Dendrobium catenatum Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution, *Scientific Reports*, **6**(19029).

Zhang, G. Q. *et al.* (2017a) The Apostasia genome and the evolution of orchids, *Nature*, **549**(7672), pp. 379–383.

Zhang, H., Gong, G., Guo, S., Ren, Y., Xu, Y. and Ling, K.-S. (2011) Screening the USDA Watermelon Germplasm Collection for Drought Tolerance at the Seedling Stage, *HortScience*, **46**(9), pp. 1245–1248.

Zhang, H., Mittal, N., Leamy, L. J., Barazani, O. and Song, B. H. (2017b) Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement, *Evolutionary Applications*, **10**(1), pp. 5–24.

Zhang, J. *et al.* (2020b) The hornwort genome and early land plant evolution, *Nature Plants*, **6**(2), pp. 107–118.

Zhang, L. *et al.* (2019a) The water lily genome and the early evolution of flowering plants, *Nature*, **577**, pp. 79–84.

Zhang, L., Wu, M., Teng, Y., Jia, S., Yu, D., Wei, T., Chen, C. and Song, W. (2019b) Overexpression of the Glutathione Peroxidase 5 (RcGPX5) Gene From Rhodiola crenulata Increases Drought Tolerance in Salvia miltiorrhiza, *Frontiers in Plant Science*, **9**(1950).

Zhang, Q. *et al.* (2012) The genome of Prunus mume, *Nature Communications*, **3**(1318).

Zhang, Q. J. *et al.* (2014b) Rapid diversification of five Oryza AA genomes associated with rice adaptation, *Proceedings of the National Academy of Sciences of the United States of America*, **111**(46), pp. 4954–4962.

Zhang, T., Qiao, Q., Novikova, P. Y., Wang, Q., Yue, J., Guan, Y., Ming, S., Liu, T., De, J., Liu, Y., Al-Shehbaz, I. A., Sun, H., Van Montagu, M., Huang, J., Van de Peer, Y. and Qiong, L. (2019c) Genome of Crucihimalaya himalaica, a close relative of Arabidopsis, shows ecological adaptation to high altitude, *Proceedings of the National Academy of Sciences of the United States of America*, **116**(14), pp. 7137–7146.

Zhang, Z., Liu, H., Sun, C., Ma, Q., Bu, H., Chong, K. and Xu, Y. (2018b) A C2H2 zinc-finger protein OsZFP213 interacts with OsMAPK3 to enhance salt tolerance in rice, *Journal of Plant Physiology*, **229**, pp. 100–110.

Zhao, D., Hamilton, J. P., Pham, G. M., Crisovan, E., Wiegert-Rininger, K., Vaillancourt, B., DellaPenna, D. and Buell, C. R. (2017) De novo genome assembly of Camptotheca acuminata, a natural source of the anti-cancer compound camptothecin, *GigaScience*, **6**(9).

Zhong, B., Yonezawa, T., Zhong, Y. and Hasegawa, M. (2010) The Position of Gnetales among Seed Plants: Overcoming Pitfalls of Chloroplast Phylogenomics, *Molecular Biology and Evolution*, **27**(12), pp. 2855–2863.

Zhu, G., Li, W., Wang, G., Li, L., Si, Q., Cai, C. and Guo, W. (2019a) Genetic Basis of Fiber Improvement and Decreased Stress Tolerance in Cultivated Versus Semi-Domesticated Upland Cotton, *Frontiers in Plant Science*, **10**(1572).

Zhu, J.-Y., Sae-Seaw, J. and Wang, Z.-Y. (2013) Brassinosteroid signalling, *Development*, **140**(8), pp. 1615–20.

Zhu, Q., Shao, Y., Ge, S., Zhang, M., Zhang, T., Hu, X., Liu, Y., Walker, J., Zhang, S. and Xu, J. (2019b) A MAPK cascade downstream of IDA–HAE/HSL2 ligand–receptor pair in lateral root emergence, *Nature Plants*, **5**(4), pp. 414–423.

Zhu, S., Estévez, J. M., Liao, H., Zhu, Y., Yang, T., Li, C., Wang, Y., Li, L., Liu, X., Pacheco, J. M., Guo, H. and Yu, F. (2020) The RALF1–FERONIA Complex Phosphorylates eIF4E1 to Promote Protein Synthesis and Polar Root Hair Growth, *Molecular Plant*, **13**(5), pp. 698–716.

Zolla, G., Heimer, Y. M. and Barak, S. (2010) Mild salinity stimulates a stress-induced morphogenic response in Arabidopsis thaliana roots, *Journal of Experimental Botany*, **61**(1), pp. 211–224.

Zwaenepoel, A. and Van de Peer, Y. (2019) Inference of Ancient Whole-Genome Duplications and the Evolution of Gene Duplication and Loss Rates, *Molecular Biology and Evolution*, **36**(7), pp. 1384–1404.

Chapter 8 Appendices

Appendix 1 The Origin of Land Plants Is Rooted in Two Bursts of Genomic Novelty.

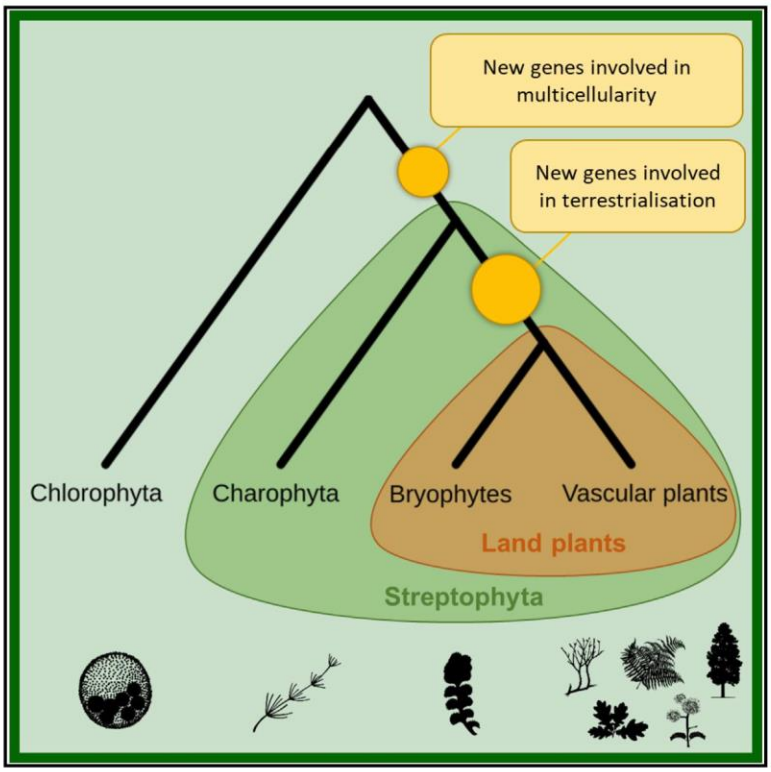
Bowles, A.M.C., Bechtold, U., and Paps, J. (2020). The Origin of Land Plants Is Rooted in Two Bursts of Genomic Novelty. Current Biology. 30. 530–536.

Report

Current Biology

The Origin of Land Plants Is Rooted in Two Bursts of Genomic Novelty

Graphical Abstract



Authors

Alexander M.C. Bowles, Ulrike Bechtold, Jordi Paps

Correspondence

ubech@essex.ac.uk (U.B.), jordi.paps@bristol.ac.uk (J.P.)

In Brief

Bowles et al. show that two consecutive bursts of genomic novelty predate the origin of land plants. Identifying these events provides insights into the evolution of flora that has defined modern ecosystems.

Highlights

- Comparing 208 genomes gives insight into the role of gene novelty in plant evolution
- Two bursts of genomic novelty played a major role in the evolution of land plants
- Functions linked to these novelties are multicellularity and terrestrialization
- The backbone of hormone signaling either predates or accompanies this transition

Bowles et al., 2020, Current Biology *30*, 530–536
February 3, 2020
https://doi.org/10.1016/j.cub.2019.11.090

CellPress

CellPress

Current Biology
Report

The Origin of Land Plants Is Rooted in Two Bursts of Genomic Novelty

Alexander M.C. Bowles,[1] Ulrike Bechtold,[1,*] and Jordi Paps[1,2,3,4,*]
[1]School of Life Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK
[2]School of Biological Sciences, University of Bristol, 24 Tyndall Avenue, Bristol BS8 1TQ, UK
[3]Department of Zoology, University of Oxford, 11a Mansfield Road, Oxford OX1 3SZ, UK
[4]Lead Contact
*Correspondence: ubech@essex.ac.uk (U.B.), jordi.paps@bristol.ac.uk (J.P.)
https://doi.org/10.1016/j.cub.2019.11.090

SUMMARY

Over the last 470 Ma, plant evolution has seen major evolutionary transitions, such as the move from water to land and the origins of vascular tissues, seeds, and flowers [1]. These have resulted in the evolution of terrestrial flora that has shaped modern ecosystems and the diversification of the Plant Kingdom, Viridiplantae, into over 374,000 described species [2]. Each of these transitions was accompanied by the gain and loss of genes in plant genomes. For example, whole-genome duplications are known to be fundamental to the origins of both seed and flowering plants [3, 4]. With the ever-increasing quality and quantity of whole-genome data, evolutionary insight into origins of distinct plant groups using comparative genomic techniques is now feasible. Here, using an evolutionary genomics pipeline to compare 208 complete genomes, we analyze the gene content of the ancestral genomes of the last common ancestor of land plants and all other major groups of plant. This approach reveals an unprecedented level of fundamental genomic novelties in two nodes related to the origin of land plants: the first in the origin of streptophytes during the Ediacaran and another in the ancestor of land plants in the Ordovician. Our findings highlight the biological processes that evolved with the origin of land plants and emphasize the importance of conserved gene novelties in plant diversification. Comparisons to other eukaryotic studies suggest a separation of the genomic origins of multicellularity and terrestrialization in plants.

RESULTS AND DISCUSSION

Analyzing the Ancestral Plant Gene Content

Understanding the diversification of plant life on Earth is still one of the major challenges in evolutionary biology. Defining the genomic changes accompanying plant evolution is key to unraveling the molecular basis of biological innovations. Recent studies have used comprehensive taxonomic transcriptome data to understand angiosperm diversification rates and gene family expansion in the major plant groups [5, 6]. Furthermore, reduced genomic datasets have been used to investigate whole-genome duplications as well as gene family gains and losses associated with plant diversification [4, 7, 8]. However, the role of genomic novelty in the origins of distinct plant groups using an extensive sampling of complete genomes with a phylogenetically broad outgroup has not been fully evaluated.

Adapting a previously described [9, 10] comparative genomics pipeline, we compared 208 eukaryotic genomes, including a broad representation of animal (10), other unikont (11), and non-embryophyte bikont (29) genomes (STAR Methods; Data S1; Figure S1). Genome quality was assessed with BUSCO, discarding genomes with more than 15% of BUSCO missing genes, and protein sequences were compared using BLAST and MCL to identify homology groups (HGs). To reduce the error produced by the complex evolutionary dynamics of genes involved in these transitions, further dissection of HGs was not conducted [10, 11]. Therefore, a single HG is defined as a set of proteins that have distinctly diverged from others. The 208 eukaryotic genomes contain ~9 million proteins, which were clustered into ~650,000 HGs. Using scripts incorporating a phylogenetic framework to inform comparative genomics, five evolutionarily distinct classifications of HG (ancestral, ancestral core, novel, novel core, and lost) were extracted (Data S2; Figure S2). Based on these outputs, patterns of large gene gains and losses were identified across the plant phylogeny (Figure 1).

The HG categorization juxtaposes between the traditional gene classification (e.g., gene families and classes) and their evolutionary dynamics. Therefore, a HG can either contain genes traditionally designated as subfamilies (e.g., GA3ox), gene families (e.g., allene oxide cyclase), or gene superfamilies. This recovery of traditional gene classifications demonstrates the reliability of this clustering approach (Data S3). There are limitations shared with other BLAST-based analyses, such as the impact of gene fusion, fission, and lateral gene transfer. However, genes in broad HGs are less likely to be misassigned than orthologs and paralogs (e.g., OrthoMCL) [12]. The pipeline approach also tackles biases seen in tree reconciliation methods, which are prone to inaccurate assignments of gene gains and losses [13].

The Role of Highly Conserved Gene Groups in Plant Evolution

The evolutions of Embryophyta (land plants) and Streptophyta (land plants and their closest algal relatives, Charophyta) are arguably the most dramatic transitions in the history of plants.

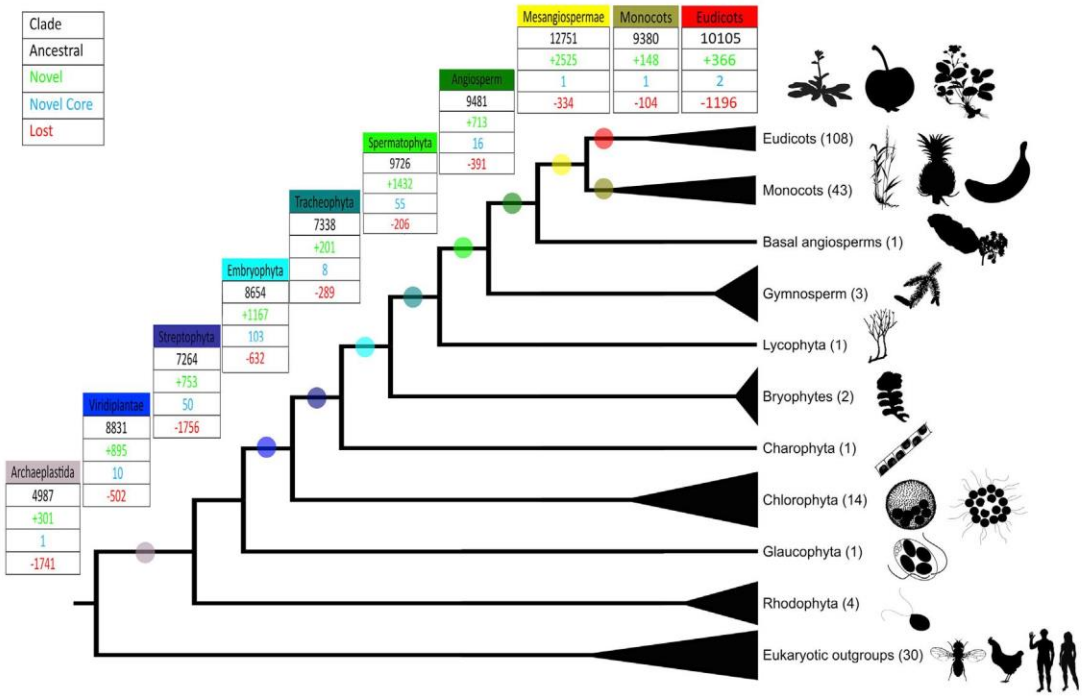


Figure 1. Analysis of the Gene Content of Ancestral Plant Genomes

The number of HGs of different categories indicated at each node for all major plant groups. Evolutionary relationships of these groups can be found in Data S1. Organism silhouettes were sourced from http://phylopic.org. See also Figures S1 and S2 and Table S1.

These events have previously been linked with the expansion of many processes and developmental traits, including embryogenesis [14], plant hormones [15], and symbiotic interactions with arbuscular mycorrhizae and rhizobacteria [16]. Our analyses revealed that there was a substantial increase in the number of highly retained gene novelties in the last common ancestor (LCA) of Streptophyta and the LCA of Embryophyta with 50 and 103 novel core HGs identified, respectively (Figure 1). Gene Ontology (GO) analyses using *Arabidopsis thaliana*, which has comprehensive GO annotations, were used to explore the modern functions of descendants of genes from novel core HGs (Data S4; Figure 2). The protein class category was used, as this classification is less prone to false assignments and biases [10]. All other GO categories, including molecular function, biological process, and pathway were produced (Data S4). HGs present in the LCA of embryophytes are abundant in classes involved in protein modification (e.g., transferase, oxidoreductase, and ligase) and protein transport (e.g., transporter proteins and membrane traffic proteins), whereas HGs present in the LCA of streptophytes are abundant in gene regulation (e.g., transcription factor) and cell structure, movement, and division (e.g., cytoskeletal proteins). The origins of Streptophyta were accompanied by the evolution of many plant-specific transcription factors (e.g., HD-ZIP) and an increasingly complex cell wall corresponding to the high number of the protein class hits seen in the Streptophyta novel core (NC) HGs [8, 14, 17].

It is possible that the bursts of conserved genomic novelty could be explained by the presence of one or multiple whole-genome duplications (WGDs). Inferring WGDs in these ancestral nodes is difficult with no events currently identified in the LCA of these groups [18, 19]. Analysis of over 1,000 transcriptomes has identified 244 WGDs across the green plant phylogeny [6]. These mostly occur after the origin of vascular plants and do not appear to coincide with the bursts of novelty seen in this study. This supports the theory that there was a change in strategy from gene family birth and expansion to WGD along the backbone of the plant phylogeny. Another contributing factor that might explain the origins of some novel core HGs is the presence of horizontal gene transfer (HGT). BLAST searches against the Swissprot database confirmed the absence of all novel core HGs in outgroup taxa, validating the outputs of the pipeline approach (BLAST outputs on Github: https://github.com/AlexanderBowles/Plant-Evomics/tree/master/Extended%20Data). Queries using the pipeline approach revealed that 323 HGs were present in fungal and land plant genomes but absent in all other taxa in this study's dataset (Data S1), suggesting widespread HGT in plants [20, 21]. The last eukaryotic common ancestor (LECA) is the ancestor that connects all eukaryotes, including plants and fungi. Either these HGs were in LECA and lost from all eukaryotic representatives aside from fungi and land plants or they are the product of HGT [22]. GO analysis of 25 of the HGs that contained at least 100 embryophyte taxa revealed that they were associated with gene regulation and protein modification (Data S5). Other possible HGT events that could explain the marked distribution of these novel core HGs include parasitism by other plants, symbiosis with other plants (e.g., transfer of a photoreceptor gene from bryophytes to ferns), and symbiosis with rhizobacteria [21, 23].

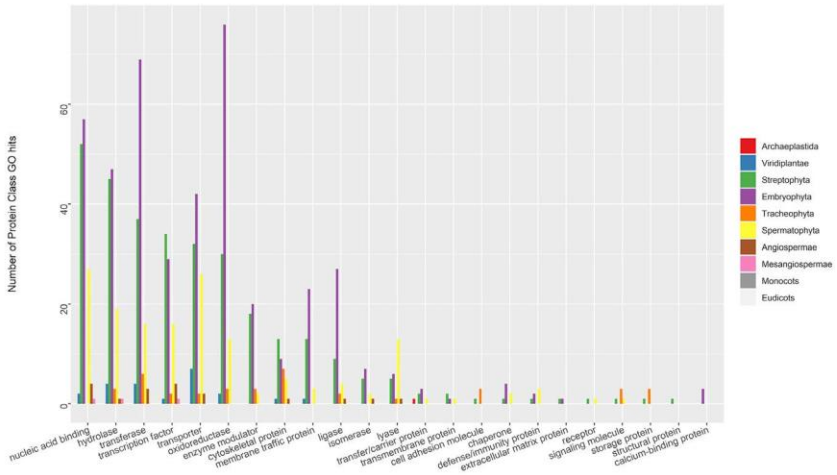


Figure 2. Gene Ontology Annotations of Novel Core HGs

Using *Arabidopsis thaliana* genes as an extant representative, protein classes were assigned for all novel core HGs. All other GO annotations (e.g., molecular function, biological process, cellular component, and pathways) were produced. See also Data S4.

The Functions of Highly Conserved Gene Groups

In streptophytes, novel core HGs were implicated in root, multicellular, and lateral organ development (Data S6; Figure 3). These terms were assigned based on the functions in extant *Arabidopsis thaliana* genes. In some cases, the evolutionary emergence of HGs predates the origin of the function with which they are often associated. For example, there is no evidence of roots outside Tracheophyta, yet genes associated with root development are found in older nodes [24, 25]. Therefore, these HGs are potential examples of co-option of old genes for new processes (Figure 3).

Other key functions include the increased complexity of the cell wall, which is crucial for multidimensional cell growth [26]. Further indicators of multicellularity in the predecessor of land plants are HGs involved in the regulation of transcription, cell adhesion, and division. The findings here also support an expansion of cellular signal transduction pathways associated with growth, development, and stress responses in streptophytes.

Many of the novel core HGs identified in our study have not previously been associated with the origin of land plants. These include proteins involved in plant organ development, cell wall construction, and host microbe interactions [27]. Other HGs are related to terrestrialization, with functions related to the synthesis of lignin, UV light protection, and cell signaling. The latter comprise plant hormones (phytohormones) linked with growth, such as auxin (body plan definition) [28], brassinosteroids (photomorphogenesis) [29], and gibberellins, as well as those associated with environmental responses, such as abscisic acid (ABA), salicylic acid, and jasmonic acid (primordial root growth) [30]. Several novel core HGs, including basic-helix-loop-helix (bHLH) transcription factors, receptor like kinases (LRR-RLKs), and three families of heavy-metal-associated isoprenylated plant proteins (HIPPs), have been previously linked to the origin of embryophytes, further validating our results (Data S6) [31].

The Evolution of Phytohormone Signaling

Some of these innovations have evolved in an incremental fashion. For example, phytohormone signaling genes identified as novel core to Streptophyta include ethylene-overproduction protein 1 (ETO1) and ethylene insensitive 3 (EIN3) (Figure S3). However, genes involved in ethylene signaling have been shown to originate before (1-aminocyclopropane-1-carboxylate synthase [ACS]) and after (1-aminocyclopropane-1-carboxylate oxidase [ACO]) this point in the evolutionary history of plants [14]. Therefore, these assigned functions do not demonstrate an establishment of these features but the additive developments contributing to their origin and evolution.

Using the same comparative genomics approach, we infer the evolutionary origins and conservation of phytohormone pathways in plants (Figure S3). The fundamental backbone of the biosynthesis and signaling pathways of all phytohormones either predates or accompanies the land plant transition [14, 32–34]. Genes involved in gibberellic acid production and signaling originate with plant terrestrialization (Figure 4). However, the role of hormones may have changed during land plant evolution, as recently highlighted for ABA signaling [39]. Important innovations in land plants include tightly controlled responses to drought and salt stresses, which require the production and perception of ABA. Our results show that ABA biosynthesis and perception evolved earlier than previously thought and are highly conserved across the plant phylogeny (Figure 4). The ABA receptor, PYL, has recently been identified in *Zygnema circumcarinatum* but is absent in other streptophyte algae [32]. In combination with the analysis presented here, this confirms that PYLs are conserved across Zygnematophyceae and Embryophyta. PP2Cs and SnRK2s, known to be present across Viridiplantae, are here supported as an Archaeplastida novelty [33]. Identifying these HGs is a significant step in understanding the evolution of phytohormones and their implications for plant diversification.

Other Evolutionarily Distinct Gene Groups of Ancestral Plant Genomes

Genomic novelty is considered to have an important role in the establishment of new features during the origins of land plants and other taxa. Genomic novelty in the LCA of distinct plant groups was substantial (Figure 1). In the LCAs of Streptophyta and Embryophyta, 753 and 1,167 novel HGs were identified, respectively, similar to values found in other studies (Data S4) [7, 14]. In contrast to other plant nodes, these values are relatively low compared to the 2,525 HGs identified in the origin of Mesangiospermae. As mentioned, WGD in plants is common and multiple events have been identified across the angiosperm phylogeny [3]. Two WGD events have been established in the

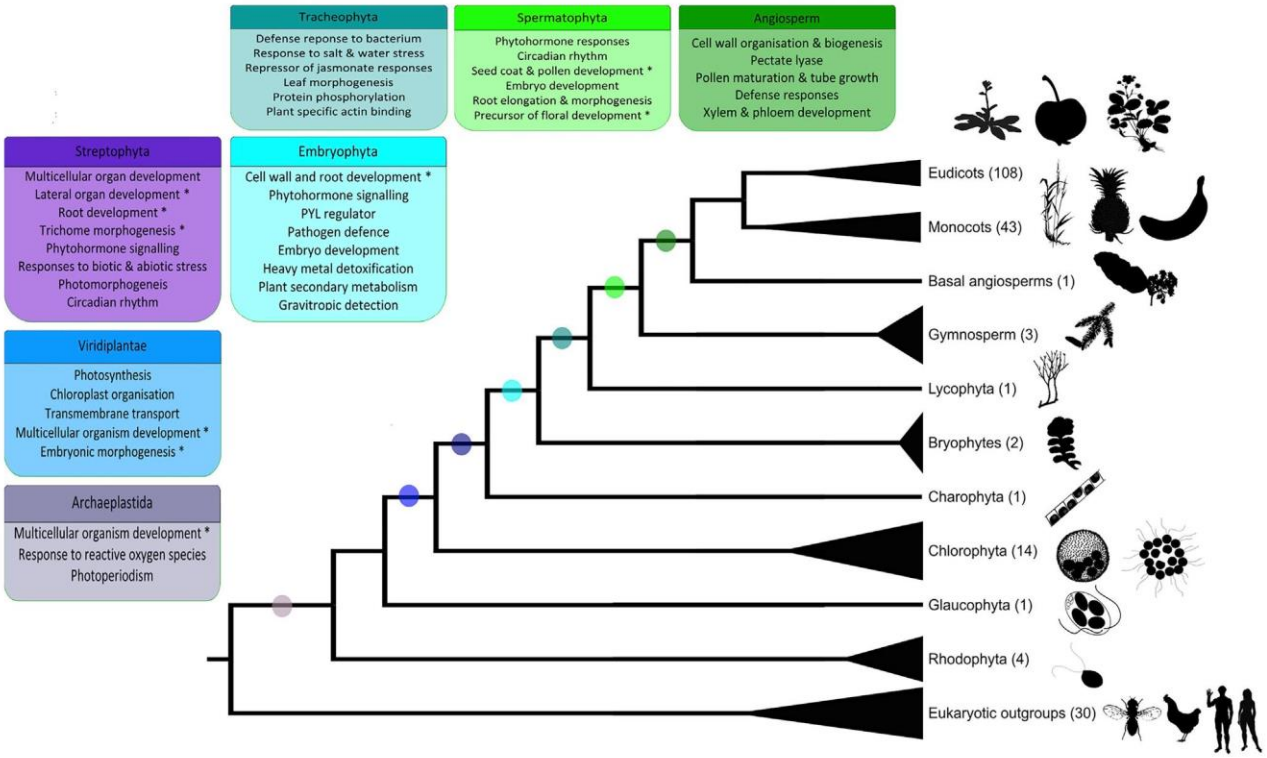


Figure 3. Biological Functions of Novel Core HGs based on *A. thaliana* Genes

Each box, color coded by phylogenetic group, is a summary of the modern day biological processes that are associated with each set of novel core HGs. An asterisk denotes an assigned biological term that is known to predate the origin of the function. Organism silhouettes were sourced from http://phylopic.org.

ancestors of seed plants (Spermatophyta) and flowering plants (Angiospermae), which could explain the 1,432 and 713 novel HGs identified in these nodes [40, 41].

Our analyses also identify that the LCA of extant land plants (Embryophyta) contained at least 8,654 ancestral HGs (Data S4). This number is likely lower than the total number of gene families present in the ancestral Embryophyta gene content because a HG can contain multiple genes, and HGs and genes can be lost from all extant representatives. *Arabidopsis thaliana* and *Brachypodium distachyon* genomes contain 27,655 and 34,310 genes clustered into 13,345 and 14,235 HGs, respectively, with 60%–70% of their genes present in the LCA of land plants. 2,254 of these ancestral HGs were retained (ancestral core) by at least 157 of the embryophyte genomes, demonstrating extensive gene loss has occurred across land plant evolution (Data S4). GO analysis revealed genes derived from HGs present in the LCA of embryophytes are abundant in gene regulation (e.g., nucleic acid binding and transcription factors) and protein modification (e.g., hydrolase and transferase; Data S4).

Furthermore, our analyses recognize HG losses (Data S4). *Drosophila melanogaster* was used as a representative of a well-annotated non-plant genome in the GO analyses of HGs lost in plant evolution. A total of 1,756 HGs were absent in the LCA of Streptophyta comprising protein classes involved in gene regulation (e.g., nucleic acid binding and transcription factor), cell signaling (e.g., enzyme modulator and signaling molecules), and catalytic activity (e.g., hydrolase and oxidoreductase). Lost HGs were also identified in Embryophyta, suggesting that gene turnover was prolific during the evolution of the ancestors of streptophytes and land plants (Figure 1). Large losses were also identified in branches leading to the LCA of eudicots and Archaeplastida with 1,196 and 1,741 HGs, respectively.

Comparisons with Animal Evolution

A previous study using the same comparative approach used in our study revealed an increase of genomic novelty during the origin of the animal kingdom, with an increase of conserved genomic novelty (novel core HGs) in a single node: the LCA of metazoans, which comprises 25 novel core HGs associated with multicellular processes; this represents a 5-fold increase from previous ancestors [10]. The origin of land plants shows two nodes with an increase of conserved genomic novelty: one in the LCA of streptophytes (in the Ediacaran; 629 mya) [1] and another in the LCA of land plants (Ordovician; 473 mya) [1]. Moreover, plants show higher numbers of conserved gene novelties than animals, representing a 10-fold increase compared to older ancestors (e.g., novel core HGs originating in the respective ancestors of Viridiplantae and Archaeplastida). In green plants, multicellularity has multiple independent evolutionary origins, with chlorophycean and charophycean algae showing a patchy

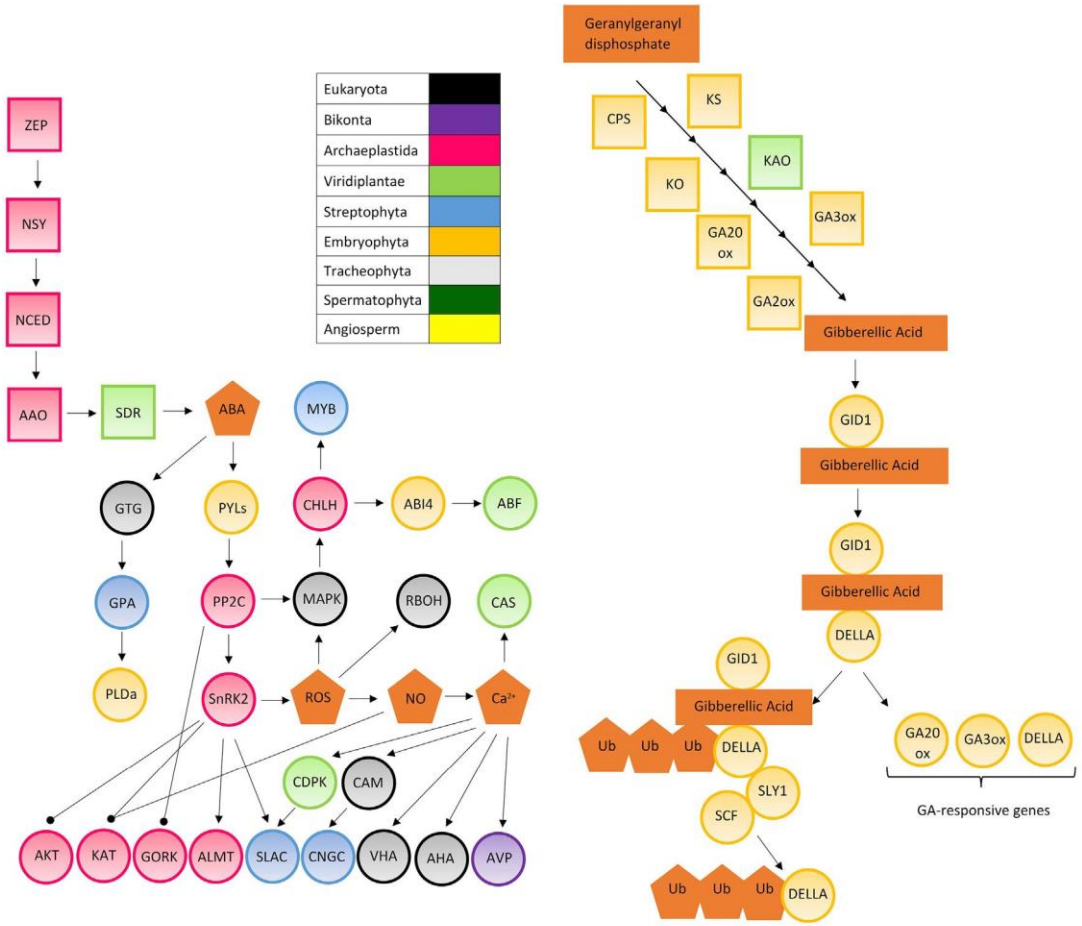


Figure 4. Evolution of Abscisic Acid (ABA) and Gibberellic Acid (GA) Biosynthesis and Signaling

Squares indicate genes that are involved in biosynthesis and circles indicate genes involved in signaling. Dark orange shapes indicate non-genetic elements. Color coding demonstrates that a gene was present in at least the last common ancestor of a clade. Arrows indicate positive regulation, and circle ended lines indicate negative regulation. Acronyms for genes: ABA biosynthesis: *AAO*, *ABA-ALDEHYDE OXIDASE*; *NCED*, *9-CIS-EPOXYCAROTENOID DIOXYGENASE*; *NSY*, *NEOXANTHIN SYNTHASE*; *SDR*, *SHORT-CHAIN ALCOHOL DEHYDROGENASE/REDUCTASE*; *ZEP*, *ZEAXANTHIN EPOXIDASE*. ABA signaling: *ABF*, *ABA RESPONSIVE ELEMENT-BINDING FACTOR*; *ABI4*, *ABA INSENSTIVE4*; *AHA*, *ARABIDOPSIS PLASMA MEMBRANE H^+-ATPASE*; *AKT*, *SER/THR KINASE1*; *ALMT*, *ALUMINUM-ACTIVATED MALATE TRANSPORTER*; *AVP*, *ARABIDOPSIS VACUOLAR H^+-PYROPHOSPHATASE*; *CAS*, *CALCIUM SENSING RECEPTOR*; *CHLH*, *PROTOPORPHYRIN IX MAGNESIUM CHELATASE, SUBUNIT H*; *CNGC*, *CYCLIC NUCLEOTIDE GATED CHANNEL*; *GORK*, *GATED OUTWARDLY RECTIFYING K^+ CHANNEL*; *KAT*, *GUARD CELL INWARDLY RECTIFYING K^+ CHANNEL*; *MAPK*, *MITOGEN ACTIVATED KINASE-LIKE PROTEIN*; *MYB*, *MYB DOMAIN PROTEIN*; *PLDa1*, *PHOSPHLIPASE Dα1*; *PP2C*, *PROTEIN PHOSPHATASE 2C*; *RBOH*, *RESPIRATORY BURST OXIDASE HOMOLOG PROTEIN*; *SLAC*, *SLOW ANION CHANNEL*; *VHA*, *VACUOLAR H^+-ATPASE*. GA biosynthesis: *CPS*, *ENT-COPALYL DIPHOSPHATE SYNTHASE*; *KS*, *KAURENE SYNTHASE*; *KO*, *ENT-KAURENE OXIDASE*; *KAO*, *ENT-KAURENOIC ACID OXIDASE*; *GA20ox*, *GIBBERELLIN 20 OXIDASE 1*; *GA3ox*, *GIBBERELLIN 3-BETA-DIOXYGENASE*; *GA2ox*, *GIBBERELLIN 2-BETA-DIOXYGENASE*. GA signaling: *GID1*, *GIBBERELLIN-INSENSITIVE DWARD PROTEIN 1*, *DELLA*; *SLY1*, *SLEEPY1*; *SCF*, *SKP1-CULLIN-F-BOX*. This figure has been adapted from previous publications for ABA [35, 36] and GA [37, 38]. See also Figure S3.

distribution, but is a trait that is conserved in all embryophytes [42, 43]. Here, we speculate that analysis of the gene content of the ancestral genomes of the plant kingdom (Viridiplantae) supports a decoupling between the emergence of multicellularity (streptophytes) and terrestrialization (embryophytes), which is in contrast to a single burst of novelty in the animal kingdom (Metazoa), whose origins did not involve a change of environment. In the future, the inclusion of new genomes may change the reconstruction of HGs at each node. Specifically, recent sequencing of the first two fern genomes and a second charophyte genome would help to fill phylogenetic gaps [7, 14]. Results from BLAST searches of novel core HGs against these phylogenetically important genomes supported the pipeline outputs, further validating our analyses (BLAST outputs available on Github: https://github.com/AlexanderBowles/Plant-Evomics/tree/master/Extended%20Data). In addition, this study solely focuses on protein-coding genes; however, non-coding genes, regulatory regions, and epigenetic modifications most likely contributed to the diversification of plant life. The analysis presented here, which incorporates genomic data for 208 taxa from across the tree of life, provides new insight into the composition of ancestral

plant genomes and emphasizes the role of genome evolution in the emergence of terrestrial flora.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- METHOD DETAILS
  - Compiling genomic dataset
  - Homology assignment
  - Phylogenetically Aware Parsing Script
  - Novel Core HG validation
  - Functional annotation
  - Inferring Horizontal Gene Transfer
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.11.090.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Phillip Mullineaux, Prof. Leonard Schalkwyk, Prof. Peter W.H. Holland, Prof. Phillip Donoghue, Dr. Nacho Maeso, Dr. Ferdinand Marlétaz, and Dr. Sarah F. Worsley for their comments on the manuscript. We also would like to thank Stuart Newman for his support of the Genomics HPC server. A.M.C.B., U.B., and J.P. received funding from the School of Life Sciences (University of Essex). U.B. was in part funded by BBSRC grant BB/N016831/1.

AUTHOR CONTRIBUTIONS

Conceptualization, A.M.C.B., U.B., and J.P.; Formal Analysis, A.M.C.B.; Visualization, A.M.C.B.; Writing – Original Draft, A.M.C.B., U.B., and J.P.; Writing – Review & Editing, A.M.C.B., U.B., and J.P.; Supervision, U.B. and J.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 13, 2019
Revised: October 15, 2019
Accepted: November 29, 2019
Published: January 16, 2020

REFERENCES

1. Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H., and Donoghue, P.C.J. (2018). The timescale of early land plant evolution. Proc. Natl. Acad. Sci. USA *115*, E2274–E2283.
2. Christenhusz, M.J.M., and Byng, J.W. (2016). The number of known plant species in the world and its annual increase. Phytotaxa *261*, 201–217.
3. Clark, J.W., and Donoghue, P.C.J. (2018). Whole-genome duplication and plant macroevolution. Trends Plant Sci. *23*, 933–945.
4. Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. Genome Res. *24*, 1334–1347.
5. Landis, J.B., Soltis, D.E., Li, Z., Marx, H.E., Barker, M.S., Tank, D.C., and Soltis, P.S. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. Am. J. Bot. *105*, 348–363.
6. Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzendanner, M.A., Graham, S.W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., et al.; One Thousand Plant Transcriptomes Initiative (2019). One thousand plant transcriptomes and the phylogenomics of green plants. Nature *574*, 679–685.
7. Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M., Eily, A., Koppers, N., Kuo, L.-Y., Li, Z., et al. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. Nat. Plants *4*, 460–472.
8. Wilhelmsson, P.K.I., Mühlich, C., Ullrich, K.K., and Rensing, S.A. (2017). Comprehensive genome-wide classification reveals that many plant-specific transcription factors evolved in streptophyte algae. Genome Biol. Evol. *9*, 3384–3397.
9. Dunwell, T.L., Paps, J., and Holland, P.W.H. (2017). Novel and divergent genes in the evolution of placental mammals. Proc. Biol. Sci. *284*, 20171357.
10. Paps, J., and Holland, P.W.H. (2018). Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. Nat. Commun. *9*, 1730.
11. Holland, P.W.H., Marlétaz, F., Maeso, I., Dunwell, T.L., and Paps, J. (2017). New genes from old: asymmetric divergence of gene duplicates and the evolution of development. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*, 20150480.
12. Pett, W., Adamski, M., Adamska, M., Francis, W.R., Eitel, M., Pisani, D., and Wörheide, G. (2019). The role of homology and orthology in the phylogenomic analysis of metazoan gene content. Mol. Biol. Evol. *36*, 643–649.
13. Hahn, M.W. (2007). Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution. Genome Biol. *8*, R141.
14. Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K., Haas, F.B., Vanderstraeten, L., Becker, D., Lang, D., et al. (2018). The Chara genome: secondary complexity and implications for plant terrestrialization. Cell *174*, 448–464.e24.
15. Wang, C., Liu, Y., Li, S.-S., and Han, G.-Z. (2015). Insights into the origin and evolution of the plant hormone signaling machinery. Plant Physiol. *167*, 872–886.
16. Field, K.J., Pressel, S., Duckett, J.G., Rimington, W.R., and Bidartondo, M.I. (2015). Symbiotic options for the conquest of land. Trends Ecol. Evol. *30*, 477–486.
17. Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., et al. (2014). Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. Nat. Commun. *5*, 3978.
18. Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. Nat. Rev. Genet. *18*, 411–424.
19. Zwaenepoel, A., and Van de Peer, Y. (2019). Inference of ancient whole-genome duplications and the evolution of gene duplication and loss rates. Mol. Biol. Evol. *36*, 1384–1404.
20. Lutzoni, F., Nowak, M.D., Alfaro, M.E., Reeb, V., Miadlikowska, J., Krug, M., Arnold, A.E., Lewis, L.A., Swofford, D.L., Hibbett, D., et al. (2018). Contemporaneous radiations of fungi and plants linked to symbiosis. Nat. Commun. *9*, 5451.
21. Yue, J., Hu, X., Sun, H., Yang, Y., and Huang, J. (2012). Widespread impact of horizontal gene transfer on plant colonization of land. Nat. Commun. *3*, 1152.
22. Margulis, L., Chapman, M., Guerrero, R., and Hall, J. (2006). The last eukaryotic common ancestor (LECA): acquisition of cytoskeletal motility from aerotolerant spirochetes in the Proterozoic Eon. Proc. Natl. Acad. Sci. USA *103*, 13080–13085.
23. Wickell, D.A., and Li, F. (2020). On the evolutionary significance of horizontal gene transfers in plants. New Phytol. *225*, 113–117.

24. Raven, J.A., and Edwards, D. (2001). Roots: evolutionary origins and biogeochemical significance. J. Exp. Bot. *52*, 381–401.
25. Hetherington, A.J., and Dolan, L. (2018). Stepwise and independent origins of roots among land plants. Nature *561*, 235–238.
26. Becker, B., and Marin, B. (2009). Streptophyte algae and the origin of embryophytes. Ann. Bot. *103*, 999–1004.
27. Raffaele, S., Mongrand, S., Gamas, P., Niebel, A., and Ott, T. (2007). Genome-wide annotation of remorins, a plant-specific protein family: evolutionary and functional perspectives. Plant Physiol. *145*, 593–600.
28. Finet, C., Berne-Dedieu, A., Scutt, C.P., and Marlétaz, F. (2013). Evolution of the ARF gene family in land plants: old domains, new tricks. Mol. Biol. Evol. *30*, 45–56.
29. Zhu, J.-Y., Sae-Seaw, J., and Wang, Z.-Y. (2013). Brassinosteroid signalling. Development *140*, 1615–1620.
30. Briggs, G.C., Mouchel, C.F., and Hardtke, C.S. (2006). Characterization of the plant-specific BREVIS RADIX gene family reveals limited genetic redundancy despite high sequence conservation. Plant Physiol. *140*, 1306–1316.
31. Liu, P.L., Du, L., Huang, Y., Gao, S.M., and Yu, M. (2017). Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants. BMC Evol. Biol. *17*, 47.
32. de Vries, J., Curtis, B.A., Gould, S.B., and Archibald, J.M. (2018). Embryophyte stress signaling evolved in the algal progenitors of land plants. Proc. Natl. Acad. Sci. USA *115*, E3471–E3480.
33. Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., et al. (2017). Insights into land plant evolution garnered from the Marchantia polymorpha genome. Cell *171*, 287–304.e15.
34. Ju, C., Van de Poel, B., Cooper, E.D., Thierer, J.H., Gibbons, T.R., Delwiche, C.F., and Chang, C. (2015). Conservation of ethylene as a plant hormone over 450 million years of evolution. Nat. Plants *1*, 14004.
35. Cai, S., Chen, G., Wang, Y., Huang, Y., Marchant, D.B., Wang, Y., Yang, Q., Dai, F., Hills, A., Franks, P.J., et al. (2017). Evolutionary conservation of ABA signaling for stomatal closure. Plant Physiol. *174*, 732–747.
36. Yamauchi, Y., Ogawa, M., Kuwahara, A., Hanada, A., Kamiya, Y., and Yamaguchi, S. (2004). Activation of gibberellin biosynthesis and response pathways by low temperature during imbibition of Arabidopsis thaliana seeds. Plant Cell *16*, 367–378.
37. Freschi, L. (2013). Nitric oxide and phytohormone interactions: current status and perspectives. Front. Plant Sci. *4*, 398.
38. Middleton, A.M., Úbeda-Tomás, S., Griffiths, J., Holman, T., Hedden, P., Thomas, S.G., Phillips, A.L., Holdsworth, M.J., Bennett, M.J., King, J.R., and Owen, M.R. (2012). Mathematical modeling elucidates the role of transcriptional feedback in gibberellin signaling. Proc. Natl. Acad. Sci. USA *109*, 7571–7576.
39. McAdam, S.A.M., Brodribb, T.J., Banks, J.A., Hedrich, R., Atallah, N.M., Cai, C., Geringer, M.A., Lind, C., Nichols, D.S., Stachowski, K., et al. (2016). Abscisic acid controlled sex before transpiration in vascular plants. Proc. Natl. Acad. Sci. USA *113*, 12862–12867.
40. Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature *473*, 97–100.
41. Ruprecht, C., Lohaus, R., Vanneste, K., Mutwil, M., Nikoloski, Z., Van de Peer, Y., and Persson, S. (2017). Revisiting ancestral polyploidy in plants. Sci. Adv. *3*, e1603195.
42. De Clerck, O., Kao, S.-M., Bogaert, K.A., Blomme, J., Foflonker, F., Kwantes, M., Vancaester, E., Vanderstraeten, L., Aydogdu, E., Boesger, J., et al. (2018). Insights into the evolution of multicellularity from the sea lettuce genome. Curr. Biol. *28*, 2921–2933.e5.
43. Umen, J.G. (2014). Green algae and the origins of multicellularity in the plant kingdom. Cold Spring Harb. Perspect. Biol. *6*, a016170.
44. Catarino, B., Hetherington, A.J., Emms, D.M., Kelly, S., and Dolan, L. (2016). The stepwise increase in the number of transcription factor families in the Precambrian predated the diversification of plants on land. Mol. Biol. Evol. *33*, 2815–2819.
45. Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A., et al. (2011). The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science *332*, 960–963.
46. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
47. Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. *30*, 1575–1584.
48. Rambaut, A.D.A. (2012). FigTree version 1.4. http://tree.bio.ed.ac.uk/software/figtree.
49. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.
50. The Inkscape Project (2019). Inkscape v0.92.4. https://inkscape.org/.
51. Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. *28*, 45–48.
52. Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. *45* (D1), D183–D189.
53. R Development Core Team. (2014). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
54. Wickham, H., and Henry, L. (2018). tidyr: easily tidy data with "spread()" and "gather()" functions. https://cran.r-project.org/package=tidyr.
55. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
SWISSPROT	[46]	https://www.uniprot.org/
Genome Data	Data S1	N/A
Software and Algorithms		
BUSCO v3	[35]	https://busco.ezlab.org/
BLAST 2.7	[36]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
mcl-14-137	[37]	https://micans.org/mcl/
Phylogenetic Aware Parsing Script	[10]	https://github.com/PapsLab/Phylogenetic_Aware_Parsing_Script
Panther GO v11	[47]	http://www.pantherdb.org/
R 3.4.2; R - tidyr; R - GGplot2	[44, 45, 48]	https://www.r-project.org/
PAPS Plant-Evomics	https://github.com/AlexanderBowles/Plant-Evomics	N/A

LEAD CONTACT AND MATERIALS AVAILABILITY

Genome sources and software (e.g., BLAST) are listed (Data S1) and referenced (Figure S1) and all scripts used are available on Github listed below. Further information about the study and intermediary files (BLAST and MCL outputs) should be directed to the Lead Contact, Jordi Paps (jordi.paps@bristol.ac.uk). This study did not generate any new, or unique reagents.

METHOD DETAILS

Compiling genomic dataset

A detailed description of the pipeline utilized here can be found elsewhere [10]. Briefly, the pipeline uses the protein coding genes of whole genome sequences to identify homologous groups of proteins within and between species (Figure S1). Broad taxonomic sampling of genomic data was implemented to be able to accurately infer the phylogenetic origin of different HGs (Data S1). 208 eukaryotic genomes were downloaded equating to 9,204,593 predicted proteins including 178 Archaeplastida genomes (including 158 land plant genomes) and 30 from a diverse representation of eukaryotic outgroups (Data S1). BUSCO analysis was used to assess the quality of the genome annotation, using a < 15% of missing genes in the BUSCO Eukaryota dataset as a benchmark to accept a genome for further analysis (Data S1) [49].

Homology assignment

Sequence similarity for all predicted proteins was identified with an all-versus-all BLAST [46] (version 2.7.1) using an e-value of 10^{e-5}, resulting in 84,724,532,295,649 comparisons with 3,680,714,880 significant BLAST hits. The BLAST search was launched on 7th February 2018 and therefore any genomes published after this date were not included in the analysis. Within the MCL protocols, it is recommended to assess the effects of changing of the granularity score which is the fineness of the clusters produced [47]. Outputs for granularity scores 1.2, 2, 4 and 6 were used to compare the phylogenetic appearance and clustering of plant gene families against published datasets of Banks et al. [45] and the transcription factor families from Catarino et al. [44] (Data S3). After testing the impact of altering this inflation value, BLAST outputs were clustered using MCL with the default granularity score (*I* = 2.0, Data S3) [47]. This approach identified 661,545 groups of homologous genes across all proteins.

Phylogenetically Aware Parsing Script

The MCL output was processed by modifying the Perl scripts described in Paps and Holland [10] with Perl version 5. In the form of three Perl scripts, the pipeline can be used to identify the origin or loss of homologous groups of proteins (HGs) based upon their taxonomic occupancy (Data S2). Different sets of HGs can be analyzed (initially defined in Paps and Holland [10]);

- Ancestral (HGs present in the Last Common Ancestor of a clade),
- Ancestral Core (HGs present in every representative species within a clade or absent only in one genome),
- Novel (HGs present in the Last Common Ancestor of a clade and absent in all outgroup taxa),

- Novel Core (HGs present in every representative species within a clade or absent only once and absent in all outgroup taxa),
- Lost (HGs lost in the Last Common Ancestor of a clade).

A more detailed explanation of these query terms with examples is available (Figure S2). The main tree figures were made in FigTree [48] and edited in Inkscape [50].

Novel Core HG validation

To confirm accurate identification of conserved gene novelties, *Arabidopsis thaliana* (and *Brachypodium distachyon* for Liliopsida novelties) genes for each HG were tested, by performing BLASTP searches against the Swissprot database [51] (25th July 2018) excluding in-group sequences with the option negative_gilist [46]. This offers the maximum breadth of taxonomic sampling possible. Based on sequence similarity, e-value, and taxonomic occupancy, BLAST searches further validated the identification of novel core Homology Groups.

Three evolutionarily significant genomes have recently been published, the first two fern genomes [7] and the second charophyte genome [14]. Novel Core HGs from all groups were BLASTP searched against the protein coding genes of these genomes (Data S4). Based on sequence similarity, e-value, and taxonomic occupancy, these BLAST searches refined the number of Novel Core HGs identified (Table S1).

Functional annotation

To obtain a functional description for all types of HG for every Archaeplastida node, their *Arabidopsis thaliana* genes were assessed using Panther GO [52] (Version 11). The number of Gene Ontology hits for all GO classifications were collated: Protein Class, Molecular Function, Biological Process, Cellular Component, Pathways (Data S4). A literature search further revealed the functions of the Novel Core Homology Groups (Data S6). Graphics were produced in R [53] using packages tidyr [54] and GGplot2 [55].

Inferring Horizontal Gene Transfer

Inferences about potential HGT were made. Based on the taxon sampling in the dataset, the pipeline was used to produce the query: Atleast1-fungi present, Atleast1-Embryophyta present and Outgroups absent. 323 HGs were identified which were subsequently whittled down to 25 HGs by stipulating that at least 100 land plant taxa must be present. Similar to the above, GO analysis was used to reveal the functions of these HGs (Data S5).

DATA AND CODE AVAILABILITY

All genomic data used in the study is publically available with sources listed in Data S1. The code used to process the outputs of MCL and extract the 5 evolutionarily distinct Homology Groups is available on Github at https://github.com/AlexanderBowles/Plant-Evomics. Also available on Github are the BLASTs of all Novel Core HGs against the SwissProt database and the results of the BLASTs against the protein coding genes of *Chara braunii, Azolla filiculoides* and *Salvinia cucullata* (https://github.com/AlexanderBowles/Plant-Evomics/tree/master/Extended%20Data).

Appendix 2: related to work presented in Chapter 2

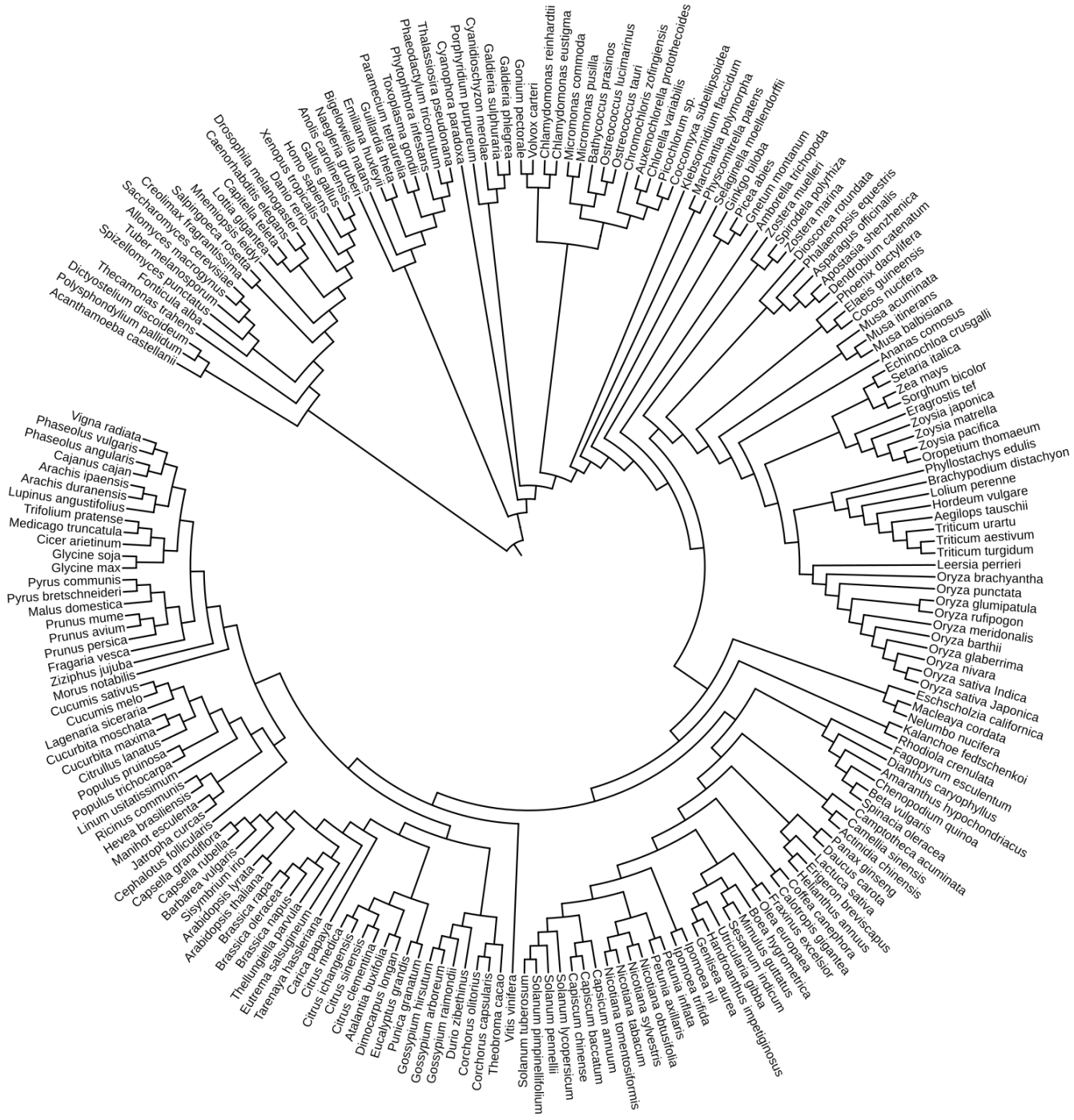


Figure S8.1. Phylogeny of species sampled in this study.

Evolutionary relationships of species with whole genomes sequenced (Goffeau *et al.*, 1996; The C. elegans Sequencing Consortium, 1998; Adams *et al.*, 2000; Kaul *et al.*, 2000; Lander *et al.*, 2001; Yu *et al.*, 2002; Goff *et al.*, 2002; Armbrust *et al.*, 2004; Hillier *et al.*,

2004; Matsuzaki *et al.*, 2004; Eichinger *et al.*, 2005; Tuskan *et al.*, 2006; Derelle *et al.*, 2006; Aury *et al.*, 2006; Palenik *et al.*, 2007; Jaillon *et al.*, 2007; Merchant *et al.*, 2007; Ming *et al.*, 2008, 2013, 2015; Rensing *et al.*, 2008; Bowler *et al.*, 2008; Paterson *et al.*, 2009; Schnable *et al.*, 2009; Worden *et al.*, 2009; Broad Institute, 2009, 2010; Haas *et al.*, 2009; Huang *et al.*, 2009, 2013, 2016; Blanc *et al.*, 2010; Prochnik *et al.*, 2010; Blanc *et al.*, 2012; Schmutz *et al.*, 2010, 2014; Velasco *et al.*, 2010; Vogel *et al.*, 2010; Chan *et al.*, 2010; EMBL-EBI, 2010; Fritz-Laylin *et al.*, 2010; Hellsten *et al.*, 2010, 2013; Kim *et al.*, 2010, 2014, 2017; Martin *et al.*, 2010; Shulaev *et al.*, 2011; The Brassica rapa Genome Sequencing Project Consortium *et al.*, 2011; Xu *et al.*, 2011, 2013c, 2013a, 2017b, 2017a, 2017c; Young *et al.*, 2011; Alföldi *et al.*, 2011; Dassanayake *et al.*, 2011; Heidel *et al.*, 2011; Hu *et al.*, 2011; Banks *et al.*, 2011; Moreau *et al.*, 2012; Price *et al.*, 2012; Sato *et al.*, 2012; Varshney *et al.*, 2012, 2013; Wang *et al.*, 2012d, 2012b, 2012c, 2014a, 2014c, 2014d, 2014b, 2017b; Zhang *et al.*, 2012, 2014b, 2016, 2017a; Curtis *et al.*, 2012; D'hont *et al.*, 2012; Garcia-Mas *et al.*, 2012; Mayer *et al.*, 2012; Al-Mssallem *et al.*, 2013; Bhattacharya *et al.*, 2013; Nystedt *et al.*, 2013; Peng *et al.*, 2013; Qiu *et al.*, 2013; Read *et al.*, 2013; Ryan *et al.*, 2013; Schönknecht *et al.*, 2013; Sierro *et al.*, 2013, 2014; Simakov *et al.*, 2013; Singh *et al.*, 2013; Slotte *et al.*, 2013; Verde *et al.*, 2013; Wu *et al.*, 2013, 2014, 2015, 2016, 2017; Yang *et al.*, 2013, 2017a, 2017b, 2017c; Albert *et al.*, 2013; Chen *et al.*, 2013a, 2016; Cheng *et al.*, 2013; Clarke *et al.*, 2013; Davey *et al.*, 2013; Fairclough *et al.*, 2013; Haudry *et al.*, 2013; He *et al.*, 2013; Howe *et al.*, 2013; Ibarra-Laclette *et al.*, 2013; Leushkin *et al.*, 2013; Ling *et al.*, 2013; Myburg *et al.*, 2014; Bolger *et al.*, 2014; Yagi *et al.*, 2014; Cannarozzi *et al.*, 2014; Chagné *et al.*, 2014; Chalhoub *et al.*, 2014; Denoeud *et al.*, 2014; Dohm *et al.*, 2014; Gao *et al.*, 2014; Hori *et al.*, 2014, 2018; Li *et al.*, 2014b, 2015a; Liu *et al.*, 2014, 2017b; Lukaszewski *et al.*, 2014; Vanburen *et al.*, 2015; Xiao *et al.*, 2015, 2017; Byrne *et al.*, 2015, 2017; De Vega *et al.*, 2015; Foflonker *et al.*, 2015; Hirakawa *et al.*, 2015; Kang *et al.*, 2015; Liu, 2015; Olsen *et al.*, 2016; Russ *et al.*, 2016; Tanaka *et al.*, 2016; Bombarely *et al.*, 2016; Tang *et al.*, 2016; Yasui *et al.*, 2016; Clouse *et al.*, 2016; Cruz *et al.*, 2016; Guan *et al.*, 2016; Hanschen *et al.*, 2016; Hoshino *et al.*,

2016; Iorizzo *et al.*, 2016; Lau *et al.*, 2016; Lee *et al.*, 2016a; Bertioli *et al.*, 2016; Qin *et al.*, 2017; Reyes-Chin-Wo *et al.*, 2017; Roth *et al.*, 2017; Sarkar *et al.*, 2017; Shirasawa *et al.*, 2017; Sollars *et al.*, 2017; Sun *et al.*, 2017; Tamiru *et al.*, 2017; Teh *et al.*, 2017; Bowman *et al.*, 2017; Xia *et al.*, 2017; Zhao *et al.*, 2017; Fu *et al.*, 2017; Fukushima *et al.*, 2017; Guo *et al.*, 2017; Hane *et al.*, 2017; Harkess *et al.*, 2017; Avni *et al.*, 2017; Hirooka *et al.*, 2017; Badouin *et al.*, 2017; Islam *et al.*, 2017; Jarvis *et al.*, 2017; Lin *et al.*, 2017; Luo *et al.*, 2017; Silva-Junior *et al.*, 2018; Stein *et al.*, 2018; Wan *et al.*, 2018a; Hoopes *et al.*, 2018; Multicellgenome Lab, 2019). The tree was manually curated and visualised in iTOL (Letunic *et al.*, 2016).

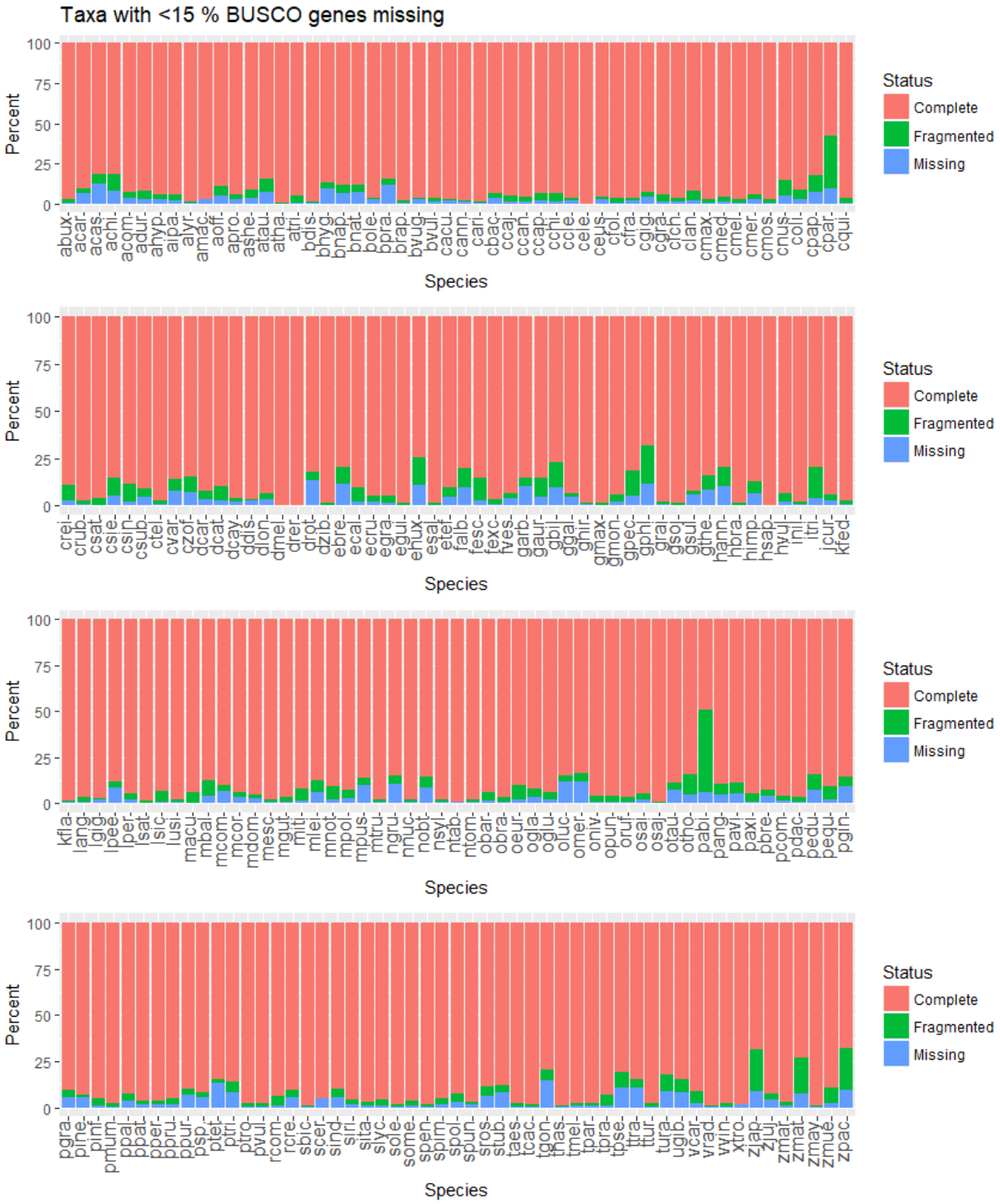


Figure S8.2. BUSCO results for genomes incorporated into pipeline.

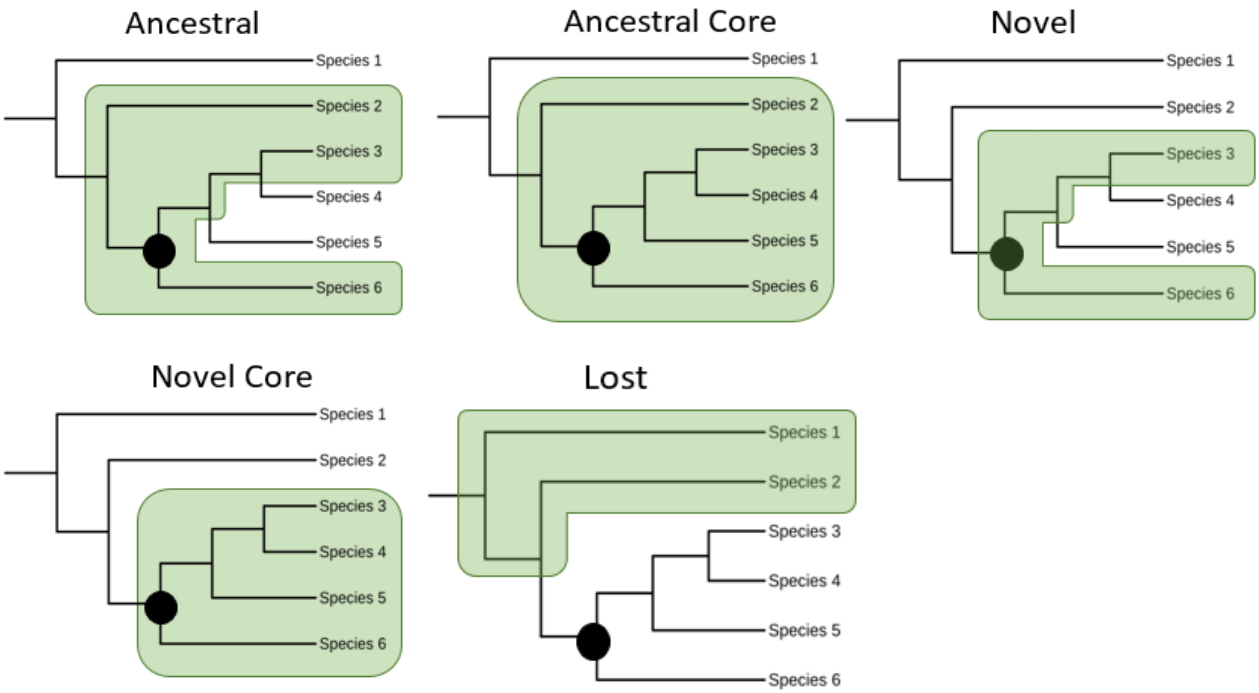
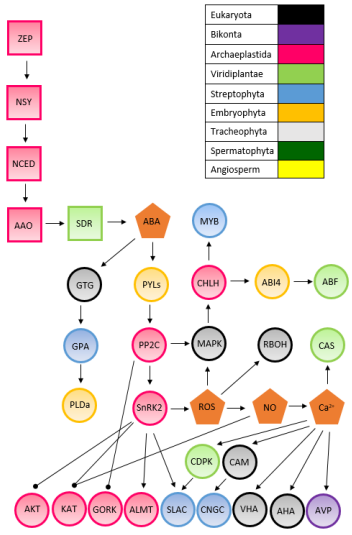


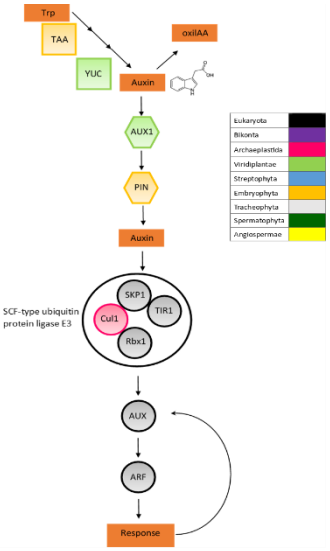
Figure S8.3. Five hypothetical species trees demonstrating the classification of the different evolutionary significant HG classifications.

All definitions are with reference to the clade of interest indicated by the dot in the tree. **Ancestral HGs** are genes present in the first splitting lineage of clade (Species_6) and at least one other lineage within a clade. They may also be present in older lineages (Species_2). **Ancestral Core HGs** are genes present in every single lineage of a clade (or absent only once). They may also be present in older lineages (Species_2). **Novel HGs** are genes present in the first splitting lineage of clade (Species_6) and at least one other lineage within a clade. However they are absent outside the clade of interest. **Novel Core HGs** are genes present in every single lineage of a clade (or absent only once) and are absent in taxa outside the clade of interest**. Lost HGs** are genes absent in the clade of interest but present in the closest splitting sister lineage (Species_2).

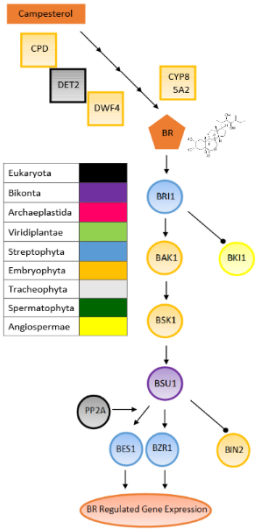
Abscisic Acid



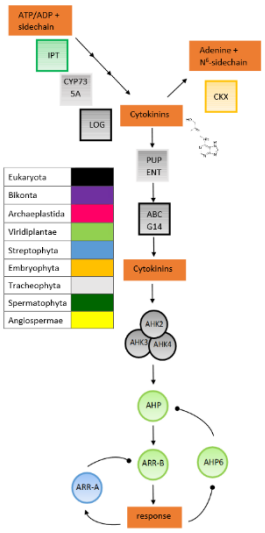
Auxin



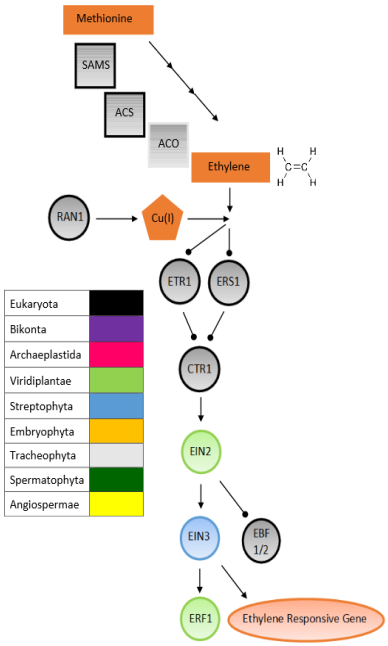
Brassinosteroids



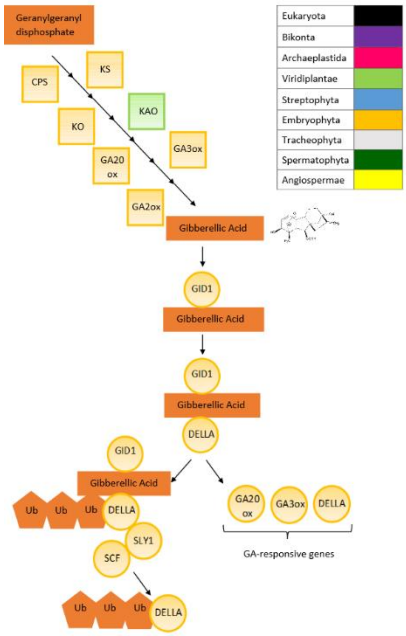
Cytokinins



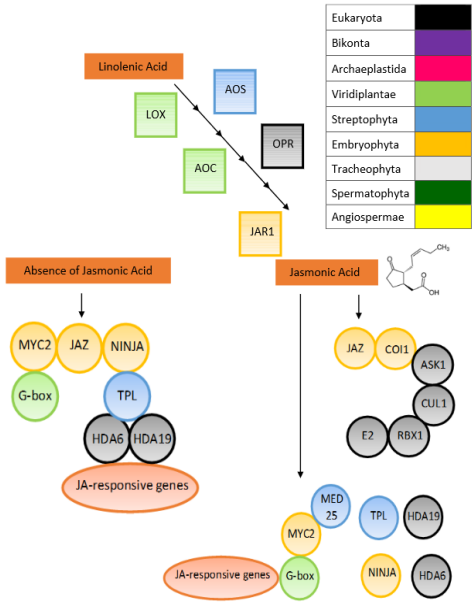
Ethylene



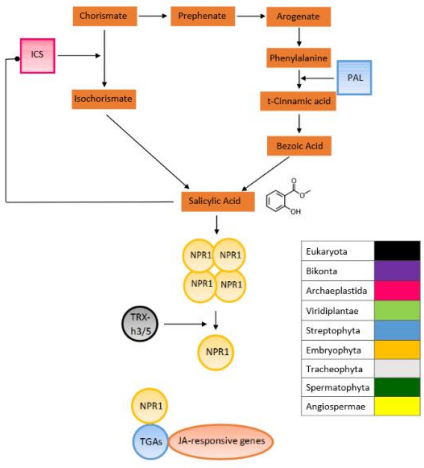
Gibberellic Acid



Jasmonic Acid



Salicylic Acid



Strigolactones and Karrakins

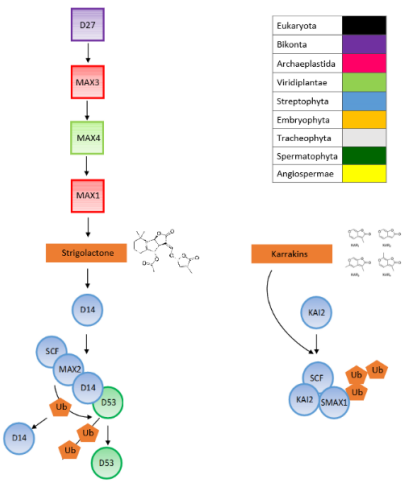


Figure S8.4. The evolution of phytohormone biosynthesis and signalling.

Shapes represent genes or steps in phytohormone pathways. Block orange rectangles and pentagons indicate non-genetic elements in phytohormone pathways whilst faded shapes indicates genes involved. Squares are genes that are involved in biosynthesis, hexagons are genes involved transport and circles are genes involved signalling. Each gene is colour coded by its phylogenetic appearance; that is, this gene is found in at least the Last Common Ancestor of each retrospective clade. For each of the major phytohormones, a recent paper was sourced to inform its known genetic pathway. Figures from the following papers were adapted to understand the evolution of each retrospective phytohormone: Abscisic Acid (Cai *et al.*, 2017); Auxin (Leyser, 2018); Brassinosteroid (Clouse, 2011; Chung *et al.*, 2013); Cytokinins biosynthesis (Hirose *et al.*, 2007) & signalling (Hutchison *et al.*, 2002); Ethylene (Ju *et al.*, 2015a); Gibberellic Acid biosynthesis (Yamauchi *et al.*, 2004) & signalling (Middleton *et al.*, 2012; Freschi, 2013); Jasmonic Acid (Wasternack *et al.*, 2013); Salicylic Acid biosynthesis (Tamaoki, 2008) & signalling (Caarls *et al.*, 2015); Strigolactone signalling (Saeed *et al.*, 2017) & biosynthesis and Karrakin signalling (Morffy *et al.*, 2016).

Table S8.1. Increasing phylogenetic coverage of sampled species. Novel Core Homology Groups removed based on BLAST against newly published genomes of *Chara braunii* (Nishiyama et al., 2018) and the ferns, *Azolla filiculoides* and *Salvinia cucullata* (Li et al., 2018a). Based on a 50% sequence similarity and a bitscore of over 150 (or over 250 if sequence similarity was below 50%), the categorisation of Novel Core Homology Groups was assessed. The LCA of Phragmoplastophyta and Euphyllophyta could not previously be assigned any Novel Core HGs due to genomic taxonomic sampling. Here these HGs are assigned to these nodes in the plant phylogeny. 50 Streptophyta Novel Core HGs supported by high sequence similarity and bitscore in *Chara braunii*. A further 45 HGs were assigned as Novel Core to Streptophyta but homologous sequences were not found in *Chara braunii* which has extensive gene loss (Nishiyama et al., 2018). These searches were taken as an indicator of the likelihood of homology. However, BLAST searches alone are not entirely accurate for homology assignment (Paps et al., 2018). To categorically confirm the inclusion of a sequence within a HG, the pipeline approach would have to re-run.

Classification	**Before BLAST**	**After BLAST**
Archaeplastida	1	1
Viridiplantae	14	10
Streptophyta	127	50
Phragmoplastophyta	NA	22
Embryophyta	168	103
Tracheophyta	19	8
Euphyllophyta	NA	8
Spermatophyta	65	55
Angiosperm	19	16
Mesangiospermae	1	1
Monocots	1	1
Eudicots	2	2

Appendix 3: related to work in Chapter 3

Table S8.2. Summary of novel, duplicated and coopted HGs linked each innovations. The origin of each innovation (root hairs, roots, lateral roots, vascular tissue, stomatal development and signalling) is highlighted in bold.

Root hairs						
Gene	Strepto.	**Embryo.**	Tracheo.	Euphyllo.	Spermato.	Angio.
Novel	4	**7**	0	0	1	1
Duplicated	0	**1**	0	0	8	4
Co-opted	0	**0**	0	0	0	0
Roots						
Gene	Strepto.	Embryo.	Tracheo.	**Euphyllo.**	Spermato.	Angio.
Novel	6	7	1	**3**	2	3
Duplicated	1	2	5	**0**	8	14
Co-opted	0	0	0	**0**	0	0
Lateral roots						
Gene	Strepto.	Embryo.	Tracheo.	Euphyllo.	**Spermato.**	Angio.
Novel	4	9	1	0	**4**	4
Duplicated	0	2	1	0	**13**	9
Co-opted	0	0	0	0	**0**	0
Vascular tissue						
Gene	Strepto.	Embryo.	**Tracheo.**	Euphyllo.	Spermato.	Angio.
Novel	5	15	**3**	0	4	3
Duplicated	0	5	**2**	0	17	10
Co-opted	0	0	**9**	0	0	0
Stomatal development						
Gene	Strepto.	**Embryo.**	Tracheo.	Euphyllo.	Spermato.	Angio.
Novel	1	**6**	1	0	0	1
Duplicated	1	**2**	2	0	4	6
Co-opted	0	**0**	0	0	0	0
Stomatal signalling						
Gene	Strepto.	Embryo.	Tracheo.	Euphyllo.	Spermato.	Angio.
Novel	4	5	0	0	2	0
Duplicated	3	2	2	6	15	7
Co-opted	0	0	0	0	0	0

Table S8.3. Evolutionary dynamics of genes linked to root hairs, roots and lateral roots. The origin of each innovation is highlighted in bold.

Gene	Uniprot ID	Strepto phyta	Embryo phyta	Trache ophyta	Euphyll ophyta	Spermat ophyta	Angios perms
Root hairs							
ZFP5	Q39264	Novel				Dup.	
CPC/TRY/ETC1	O22059/Q8GV05/Q9LNI5					Novel	
GL3/EGL3	Q9FN69		**Novel**			Dup.	
TTG1	Q9XGN1					Dup.	
EIN3/EIL1	O24606	Novel				Dup.	Dup.
GL2	P46607	Novel				Dup.	Dup.
WER	Q9SEI0	Novel				Dup.	
BIN2	F4JRM5						Novel
RHD6/RSL1	Q9C707		**Novel**				
RSL4/RSL2	Q8LEG1/Q84WK0		**Novel**				Dup.
ARF5	P93024		**Novel**				
OBP4	Q0WUB6		**Novel**			Dup.	
ARF7	P93022		**Novel**				
ARF19	Q8RYC8		**Novel**				
ERU	Q9FLJ8		**Dup.**			Dup.	Dup.
Roots							
CTR1	Q0WUI6	Dup.					
TTL1	Q9MAH1	Novel					Dup.
AXR2(3)	Q38825						Novel
BRX	Q17TI5		Novel			Dup.	
ARR12	P62598				**Novel**		
LATD	D2IU94				**Novel**		Dup.
RGA	Q9SLH3		Novel/ Dup.				Dup.
BRL3	Q9LJF3	Novel				Dup.	
ASA1	F4K0T5						Dup.
PLS	Q8LLV8						Novel
TAA1	Q9S7N2		Novel				
ABI8	Q9C9Z9						
ACS5(9)	Q37001					Dup.	Dup.
SCR	Q9M384		Novel			Dup.	
CKX	O22213		Novel	Dup.		Dup.	Dup.
ARR7	Q9ZWS7	Novel		Dup.		Dup.	
ARR22	Q9M8Y4	Novel		Dup.			
EIN2	Q9S814		Novel				
ARR1	Q940D0						Dup.
ASB1	F4IAW5						
SHR	Q9SZF7		Novel				

IPT2	Q9ZUX7					Novel	
TIR1	Q570C0	Novel		Dup.		Dup.	Dup.
PLT1	Q5YGP8			Novel		Dup.	Dup.
IPT9	Q9C5J6						
AHK3	Q9C5U1						Dup.
EIN3	O24606	Novel	Dup.				Dup.
IPT5	Q94ID2					Novel	Dup.
AUX1	Q96247			Dup.			Dup.
B450/BRox62	Q8VZC2				**Novel**		Dup.
SNE	Q9LUB6						Novel
Lateral roots							
BRI1	O22476	Novel				**Dup.**	
BAK1	Q94F62		Novel				
BIN2	F4JRM5						Novel
BZR1/BES1	Q8S307/F4HP45	Novel					
RALF	Q9SRY3					**Novel**	
PIN1,2,3	Q9C6B8/Q9LU77/Q9S7Z8	Novel					
MIZ1	O22227		Novel/ Dup.				
GNOM/MIZ2	Q42510					**Dup.**	
TIR1/AFB	Q570C0/Q9LW29	Novel		Dup.		**Dup.**	Dup.
IAA12/28	Q38830/Q9XFM0					**Novel**	
IAA8	Q38826		Novel				Dup.
IAA19	O24409						Novel
IAA14	Q38832		Novel			**Dup.**	
IAA3/SHY2	Q38822						Novel
ARF5,6,8	P93024/Q9ZTX8/Q9FGV1		Novel			**Dup.**	Dup.
ARF7,19	P93022/Q8RYC8		Novel				Dup.
GATA23	Q8LC59					**Dup.**	
TOSL2	F4JRC5						Novel
RLK7	F4I2N7						
PUCHI	A0A178URC3					**Dup.**	Dup.
LBD16/18	Q9SLB7/O22131					**Novel/Dup.**	Dup.
E2Fa	Q9FNY0						
CDKA1/CYCB1	P24100						
PLT3/4/5	Q9ZNS0/Q0WUU6/Q8VZ80						Dup.
PLT1/2/7	Q5YGP8/Q5YGP7/A0A1P8BA69					**Dup.**	
SHR	Q9SZF7		Novel			**Dup.**	
SCR	Q9M384		Novel			**Dup.**	
WOX5	Q8H1D2			Novel		**Dup.**	

IDA/AMG	Q8LAD7					**Novel**	
HAE/HSL2	P47735/C0LGX3						
MKK4/5	O80397/Q8RXG3		Novel				Dup.
MPK3/6	Q39023/Q39026		Dup.			**Dup.**	Dup.

Table S8.4. Evolutionary dynamics of genes linked to vascular tissue. The origin of each innovation is highlighted in bold.

Gene	Uniprot ID	Strepto phyta	Embryo phyta	Tracheo phyta	Euphyllo phyta	Spermat ophyta	Angiosp erms
Vascular tissue							
MP	P93024		Novel			Dup.	Dup.
ATHB8	Q39123	Novel					
TMO5	Q9LS08	Novel		**Dup.**			Dup.
LHW	Q9XIN0		Novel	**Co-opted**			Dup.
LOGs	Q8RUN2					Dup.	
BDL	Q38830					Novel	
SHR	Q9SZF7		Novel	**Co-opted**			
SCR	Q9M384		Novel	**Co-opted**			
PIN1	Q9C6B8	Novel				Dup.	Dup.
AHP6	Q9SSC9					Dup.	Dup.
SACL1	Q9FY69			**Novel**			
ACL5	Q9S7X6		Dup.				
BUD2	Q3E9D5		Dup.				
IAA 20/30	O24410/Q9M1R4						Novel
CLE 41/44	Q84W98/Q941C5						Novel
CLV1	Q9SYQ8		Novel			Dup.	
CLV2	O80809		Novel	**Co-opted**			
PXY	Q9FII5			**Dup.**		Dup.	
WOX1	Q6X7K0			**Novel**		Dup.	
WOX10	Q9LM83		Dup.			Dup.	
ERF_018	Q9S7L5		Novel	**Co-opted**		Dup.	
CYCD3;1	P42753		Novel	**Co-opted**			
ANT	Q38914		Dup.			Dup.	
BIN2	F4JRM5						Novel
BES1	F4HP45	Novel					
VND6	Q9FHC2						Dup.
VND7	Q9C8W9		Novel				Dup.
E2Fc	Q9FV70						Dup.
REV/PHB	Q9SE43/O04291	Novel				Dup.	Dup.
NST1,2/SND1	Q84WP6/Q9M274/Q9LPI7		Novel	**Co-opted**		Dup.	
KNAT1/STM	P46639/Q38874					Dup.	
BOP1	F4IH25						

BOP2	Q9ZVC2		Novel	**Co-opted**			
APL	Q9SAK5		Novel	**Co-opted**		Dup.	
NAC20/45/86	Q67Z40/A4VCM0/Q9FFI5		Novel			Dup.	Dup.
NEN1	Q9FLR0			**Novel**			
OPS	Q9SS80					Novel	
CVP2	Q9LR47						
BRX	Q17TI5		Novel			Dup.	
BAM3	O65440		Novel/ Dup.			Dup.	
CLE45	Q6IWA9					Novel	

Table S8.5. Evolutionary dynamics of genes linked to stomata signalling and development. The origin of each innovation is highlighted in bold.

Gene	Uniprot ID	Streptophyta	Embryophyta	Tracheophyta	Euphyllophyta	Spermatophyta	Angiosperms
Stomatal Development							
BAK1	Q94F62		**Novel**				
YODA	Q9CAD5		**Dup.**				
EPF1/2	Q8S8I4		**Novel**				Dup.
TMM	Q9SSD1			Dup.			
COP1	P43254						
MPK3/6	Q39023		**Dup.**			Dup.	Dup.
Erf	Q42371					Dup.	
BIN2	F4JRM5						Novel
BRI	O22476	Novel				Dup.	
MKK4/5	O80397		**Novel**				Dup.
BSU1f	Q9LR78	Dup.					
Stomagen	Q9SV72			Novel			
FLP	Q94FL6						
RBR	Q9LKZ3						
CDKA1	P24100						
AGO1	O04379					Dup.	Dup.
SCRM	Q9LSE2		**Novel**	Dup.			Dup.
MUTE/Speechless	Q9M8K6/ Q700C7		**Novel**				Dup.
FAMA	Q56YJ8		**Novel**				
Stomatal Signalling							
GEF1/4	Q93ZY2				Dup.		Dup.
GEF10	Q1KS66					**Novel**	
ROP11	O82481				Dup.	**Dup.**	
ABI1	O04719						
GHR1	C0LGQ9		Novel			**Dup.**	
GTG	Q9XIP7						
GPA	TRIDC1BG074450	Novel					
PLDa	Q38882		Novel			**Dup.**	
PYL1-3	Q8VZS8					**Novel**	Dup.
PYL4-6,11-13	O80920		Novel				
PYL7-10	Q1ECF1		Novel				
PP2C	P49598				Dup.	**Dup.**	
SnRK2	P43291		Dup.	Dup.	Dup.	**Dup.**	Dup.
MAPK	Q39023	Dup.				**Dup.**	
CHLH	Q9FNB0						
MYB	Q9SPG6						
ABI4	A0MES8		Novel		Dup.		

ABF	Q9M7Q5						
RBOH	Q9FIJ0	Novel				**Dup.**	Dup.
CAS	Q9FN48						
CDPK	Q06850	Dup.				**Dup.**	
CAM	P0DH95						
QUAC1	O49696					**Dup.**	
PIP2;1	P43286						Dup.
AKT1/KAT2	Q38998/Q38849					**Dup.**	
GORK	Q94A76					**Dup.**	
ALMT	Q9SJE9						Dup.
CLC-C	Q96282	Dup.				**Dup.**	
SLAC	Q9LD83	Novel		Dup.	Dup.		
CNGC	O65717	Novel				**Dup.**	Dup.
VHA	Q570K4						
AHA	P20649		Dup.			**Dup.**	
AVP	P31414					**Dup.**	

Appendix 3.1: Phylogenetic trees discussed in the text of Chapter 3. For all trees analysed in Chapter 3, see Supplementary Data 3.2. Values on branches indicate bootstrap support values. For each gene, a tree with and without branch lengths is provided for ease of viewing.

Appendix 3.1.1. Phylogenetic trees for stomatal signalling genes discussed in Chapter 3.

Appendix 3.1.1.1 Gene duplications in the ancestor of Spermatophyta

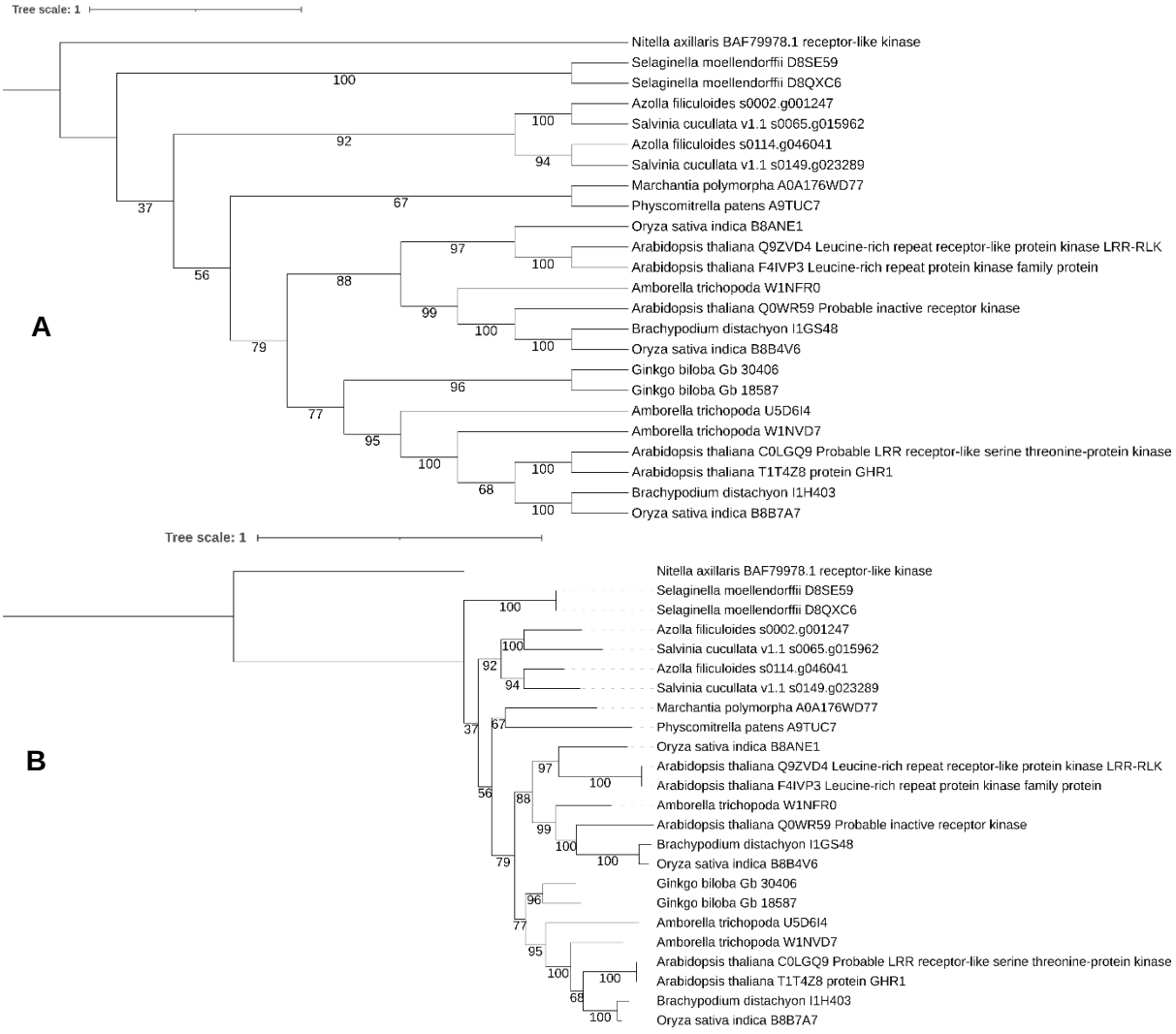
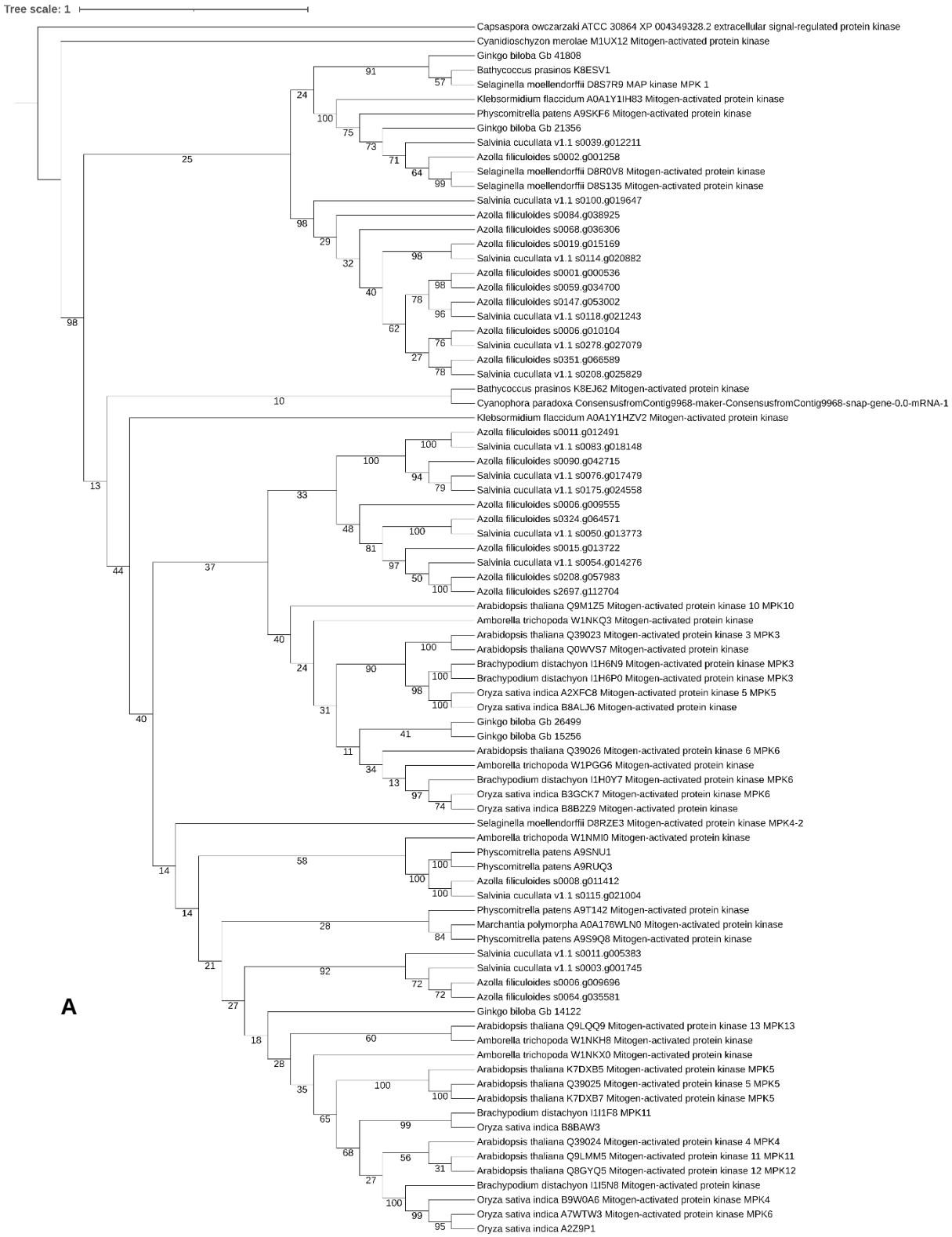


Figure S8.5. Maximum likelihood gene tree for GUARD CELL HYDROGEN PEROXIDE-RESISTANT 1 (GHR1, Uniprot ID: C0LGQ9) without (A) and with branch lengths (B).


Tree scale: 1
Capsaspora owczarzaki ATCC 30864 XP 004349328.2 extracellular signal-regulated protein kinase
Cyanidioschyzon merolae M1UX12 Mitogen-activated protein kinase
Ginkgo biloba Gb 41808
Bathycoccus prasinos K8ESV1
Selaginella moellendorffii D8S7R9 MAP kinase MPK 1
Klebsormidium flaccidum A0A1Y1IH83 Mitogen-activated protein kinase
Physcomitrella patens A9SKF6 Mitogen-activated protein kinase
Ginkgo biloba Gb 21356
Salvinia cucullata v1.1 s0039.g012211
Azolla filiculoides s0002.g001258
Selaginella moellendorffii D8R0V8 Mitogen-activated protein kinase
Selaginella moellendorffii D8S135 Mitogen-activated protein kinase
Salvinia cucullata v1.1 s0100.g019647
Azolla filiculoides s0084.g038925
Azolla filiculoides s0068.g036306
Azolla filiculoides s0019.g015169
Salvinia cucullata v1.1 s0114.g020882
Azolla filiculoides s0001.g000536
Azolla filiculoides s0059.g034700
Azolla filiculoides s0147.g053002
Salvinia cucullata v1.1 s0118.g021243
Azolla filiculoides s0006.g010104
Salvinia cucullata v1.1 s0278.g027079
Azolla filiculoides s0351.g066589
Salvinia cucullata v1.1 s0208.g025829
Bathycoccus prasinos K8EJ62 Mitogen-activated protein kinase
Cyanophora paradoxa ConsensusfromContig9968-maker-ConsensusfromContig9968-snap-gene-0.0-mRNA-1
Klebsormidium flaccidum A0A1Y1HZV2 Mitogen-activated protein kinase
Azolla filiculoides s0011.g012491
Salvinia cucullata v1.1 s0083.g018148
Azolla filiculoides s0090.g042715
Salvinia cucullata v1.1 s0076.g017479
Salvinia cucullata v1.1 s0175.g024558
Azolla filiculoides s0006.g009555
Azolla filiculoides s0324.g064571
Salvinia cucullata v1.1 s0050.g013773
Azolla filiculoides s0015.g013722
Salvinia cucullata v1.1 s0054.g014276
Azolla filiculoides s0208.g057983
Azolla filiculoides s2697.g112704
Arabidopsis thaliana Q9M1Z5 Mitogen-activated protein kinase 10 MPK10
Amborella trichopoda W1NKQ3 Mitogen-activated protein kinase
Arabidopsis thaliana Q39023 Mitogen-activated protein kinase 3 MPK3
Arabidopsis thaliana Q0WVS7 Mitogen-activated protein kinase
Brachypodium distachyon I1H6N9 Mitogen-activated protein kinase MPK3
Brachypodium distachyon I1H6P0 Mitogen-activated protein kinase MPK3
Oryza sativa indica A2XFC8 Mitogen-activated protein kinase 5 MPK5
Oryza sativa indica B8ALJ6 Mitogen-activated protein kinase
Ginkgo biloba Gb 26499
Ginkgo biloba Gb 15256
Arabidopsis thaliana Q39026 Mitogen-activated protein kinase 6 MPK6
Amborella trichopoda W1PGG6 Mitogen-activated protein kinase
Brachypodium distachyon I1H0Y7 Mitogen-activated protein kinase MPK6
Oryza sativa indica B3GCK7 Mitogen-activated protein kinase MPK6
Oryza sativa indica B8B2Z9 Mitogen-activated protein kinase
Selaginella moellendorffii D8RZE3 Mitogen-activated protein kinase MPK4-2
Amborella trichopoda W1NMI0 Mitogen-activated protein kinase
Physcomitrella patens A9SNU1
Physcomitrella patens A9RUQ3
Azolla filiculoides s0008.g011412
Salvinia cucullata v1.1 s0115.g021004
Physcomitrella patens A9T142 Mitogen-activated protein kinase
Marchantia polymorpha A0A176WLN0 Mitogen-activated protein kinase
Physcomitrella patens A9S9Q8 Mitogen-activated protein kinase
Salvinia cucullata v1.1 s0011.g005383
Salvinia cucullata v1.1 s0003.g001745
Azolla filiculoides s0006.g009696
Azolla filiculoides s0064.g035581
Ginkgo biloba Gb 14122
Arabidopsis thaliana Q9LQQ9 Mitogen-activated protein kinase 13 MPK13
Amborella trichopoda W1NKH8 Mitogen-activated protein kinase
Amborella trichopoda W1NKX0 Mitogen-activated protein kinase
Arabidopsis thaliana K7DXB5 Mitogen-activated protein kinase MPK5
Arabidopsis thaliana Q39025 Mitogen-activated protein kinase 5 MPK5
Arabidopsis thaliana K7DXB7 Mitogen-activated protein kinase MPK5
Brachypodium distachyon I1I1F8 MPK11
Oryza sativa indica B8BAW3
Arabidopsis thaliana Q39024 Mitogen-activated protein kinase 4 MPK4
Arabidopsis thaliana Q9LMM5 Mitogen-activated protein kinase 11 MPK11
Arabidopsis thaliana Q8GYQ5 Mitogen-activated protein kinase 12 MPK12
Brachypodium distachyon I1I5N8 Mitogen-activated protein kinase
Oryza sativa indica B9W0A6 Mitogen-activated protein kinase MPK4
Oryza sativa indica A7WTW3 Mitogen-activated protein kinase MPK6
Oryza sativa indica A2Z9P1

A

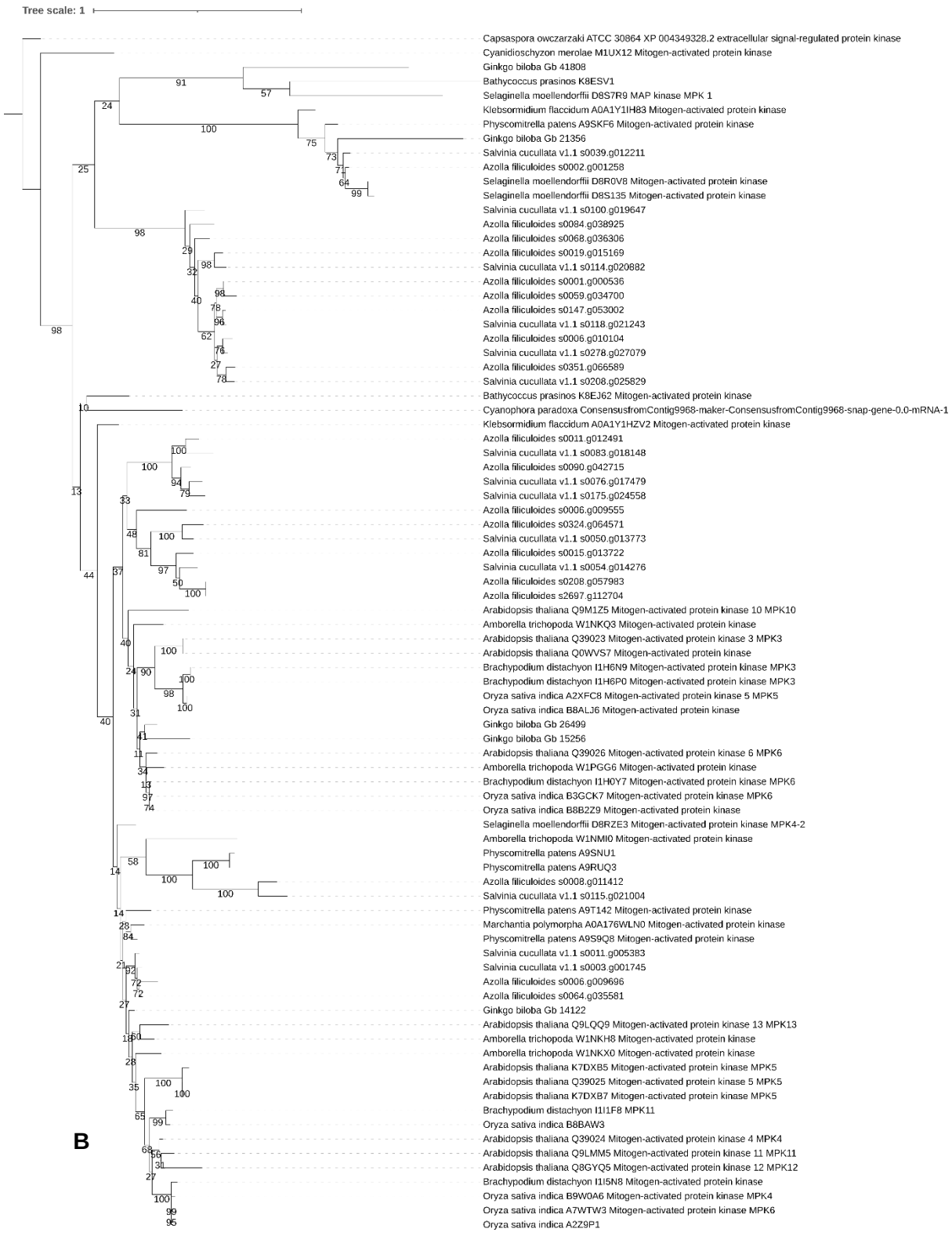
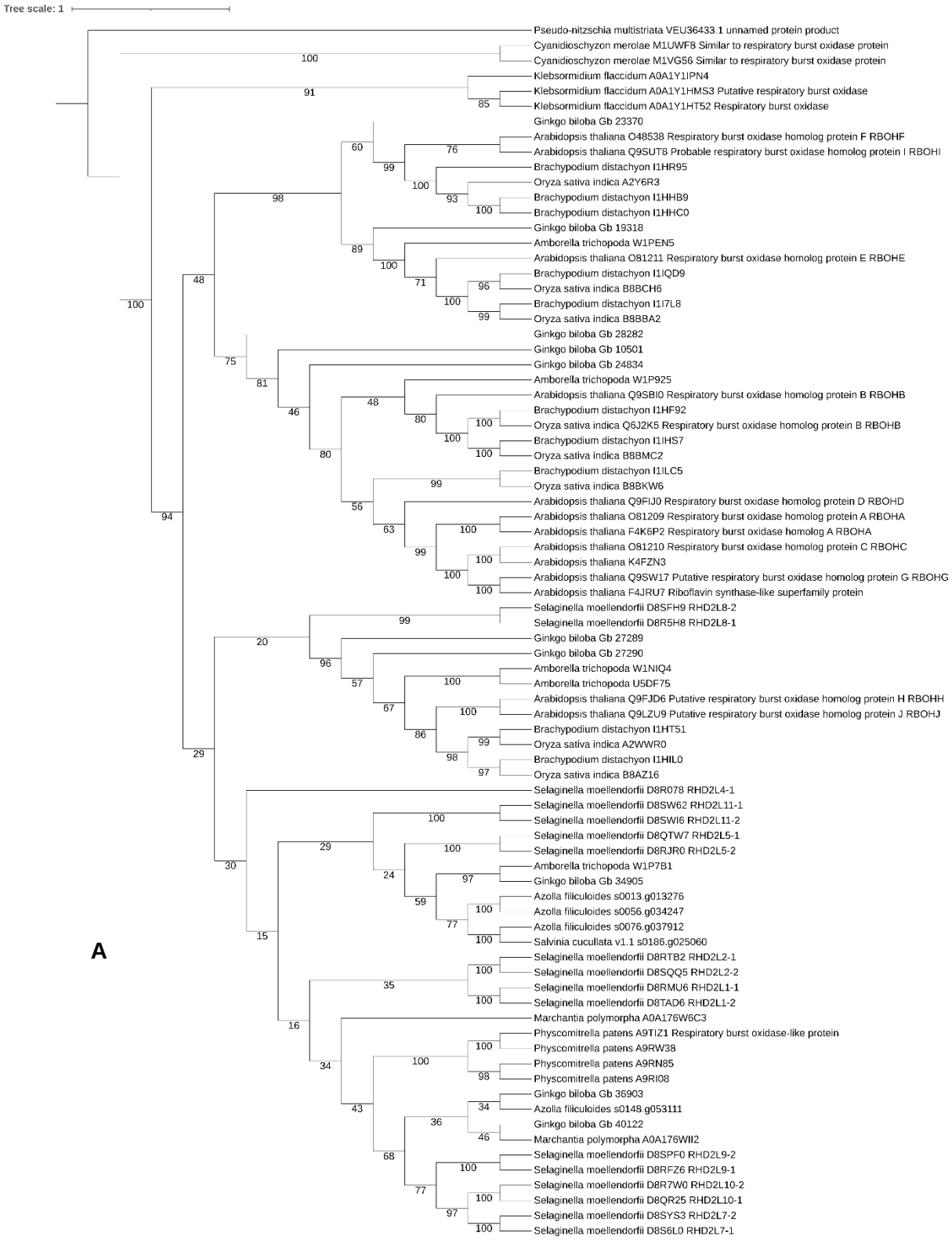


Figure S8.6. Maximum likelihood gene tree for MITOGEN-ACTIVATED PROTEIN KINASE (MAPK, Uniprot ID: Q39023) without (A) and with branch lengths (B).


Tree scale: 1
Pseudo-nitzschia multistriata VEU36433.1 unnamed protein product
Cyanidioschyzon merolae M1UWF8 Similar to respiratory burst oxidase protein
Cyanidioschyzon merolae M1VG56 Similar to respiratory burst oxidase protein
Klebsormidium flaccidum A0A1Y1IPN4
Klebsormidium flaccidum A0A1Y1HMS3 Putative respiratory burst oxidase
Klebsormidium flaccidum A0A1Y1HT52 Respiratory burst oxidase
Ginkgo biloba Gb 23370
Arabidopsis thaliana O48538 Respiratory burst oxidase homolog protein F RBOHF
Arabidopsis thaliana Q9SUT8 Probable respiratory burst oxidase homolog protein I RBOHI
Brachypodium distachyon I1HR95
Oryza sativa indica A2Y6R3
Brachypodium distachyon I1HHB9
Brachypodium distachyon I1HHC0
Ginkgo biloba Gb 19318
Amborella trichopoda W1PEN5
Arabidopsis thaliana O81211 Respiratory burst oxidase homolog protein E RBOHE
Brachypodium distachyon I1IQD9
Oryza sativa indica B8BCH6
Brachypodium distachyon I1I7L8
Oryza sativa indica B8BBA2
Ginkgo biloba Gb 28282
Ginkgo biloba Gb 10501
Ginkgo biloba Gb 24834
Amborella trichopoda W1P925
Arabidopsis thaliana Q9SBI0 Respiratory burst oxidase homolog protein B RBOHB
Brachypodium distachyon I1HF92
Oryza sativa indica Q6J2K5 Respiratory burst oxidase homolog protein B RBOHB
Brachypodium distachyon I1IHS7
Oryza sativa indica B8BMC2
Brachypodium distachyon I1ILC5
Oryza sativa indica B8BKW6
Arabidopsis thaliana Q9FIJ0 Respiratory burst oxidase homolog protein D RBOHD
Arabidopsis thaliana O81209 Respiratory burst oxidase homolog protein A RBOHA
Arabidopsis thaliana F4K6P2 Respiratory burst oxidase homolog A RBOHA
Arabidopsis thaliana O81210 Respiratory burst oxidase homolog protein C RBOHC
Arabidopsis thaliana K4FZN3
Arabidopsis thaliana Q9SW17 Putative respiratory burst oxidase homolog protein G RBOHG
Arabidopsis thaliana F4JRU7 Riboflavin synthase-like superfamily protein
Selaginella moellendorfii D8SFH9 RHD2L8-2
Selaginella moellendorfii D8R5H8 RHD2L8-1
Ginkgo biloba Gb 27289
Ginkgo biloba Gb 27290
Amborella trichopoda W1NIQ4
Amborella trichopoda U5DF75
Arabidopsis thaliana Q9FJD6 Putative respiratory burst oxidase homolog protein H RBOHH
Arabidopsis thaliana Q9LZU9 Putative respiratory burst oxidase homolog protein J RBOHJ
Brachypodium distachyon I1HT51
Oryza sativa indica A2WWR0
Brachypodium distachyon I1HIL0
Oryza sativa indica B8AZ16
Selaginella moellendorfii D8R078 RHD2L4-1
Selaginella moellendorfii D8SW62 RHD2L11-1
Selaginella moellendorfii D8SWI6 RHD2L11-2
Selaginella moellendorfii D8QTW7 RHD2L5-1
Selaginella moellendorfii D8RJR0 RHD2L5-2
Amborella trichopoda W1P7B1
Ginkgo biloba Gb 34905
Azolla filiculoides s0013.g013276
Azolla filiculoides s0056.g034247
Azolla filiculoides s0076.g037912
Salvinia cucullata v1.1 s0186.g025060
Selaginella moellendorfii D8RTB2 RHD2L2-1
Selaginella moellendorfii D8SQQ5 RHD2L2-2
Selaginella moellendorfii D8RMU6 RHD2L1-1
Selaginella moellendorfii D8TAD6 RHD2L1-2
Marchantia polymorpha A0A176W6C3
Physcomitrella patens A9TIZ1 Respiratory burst oxidase-like protein
Physcomitrella patens A9RW38
Physcomitrella patens A9RN85
Physcomitrella patens A9RI08
Ginkgo biloba Gb 36903
Azolla filiculoides s0148.g053111
Ginkgo biloba Gb 40122
Marchantia polymorpha A0A176WII2
Selaginella moellendorfii D8SPF0 RHD2L9-2
Selaginella moellendorfii D8RFZ6 RHD2L9-1
Selaginella moellendorfii D8R7W0 RHD2L10-2
Selaginella moellendorfii D8QR25 RHD2L10-1
Selaginella moellendorfii D8SYS3 RHD2L7-2
Selaginella moellendorfii D8S6L0 RHD2L7-1

A

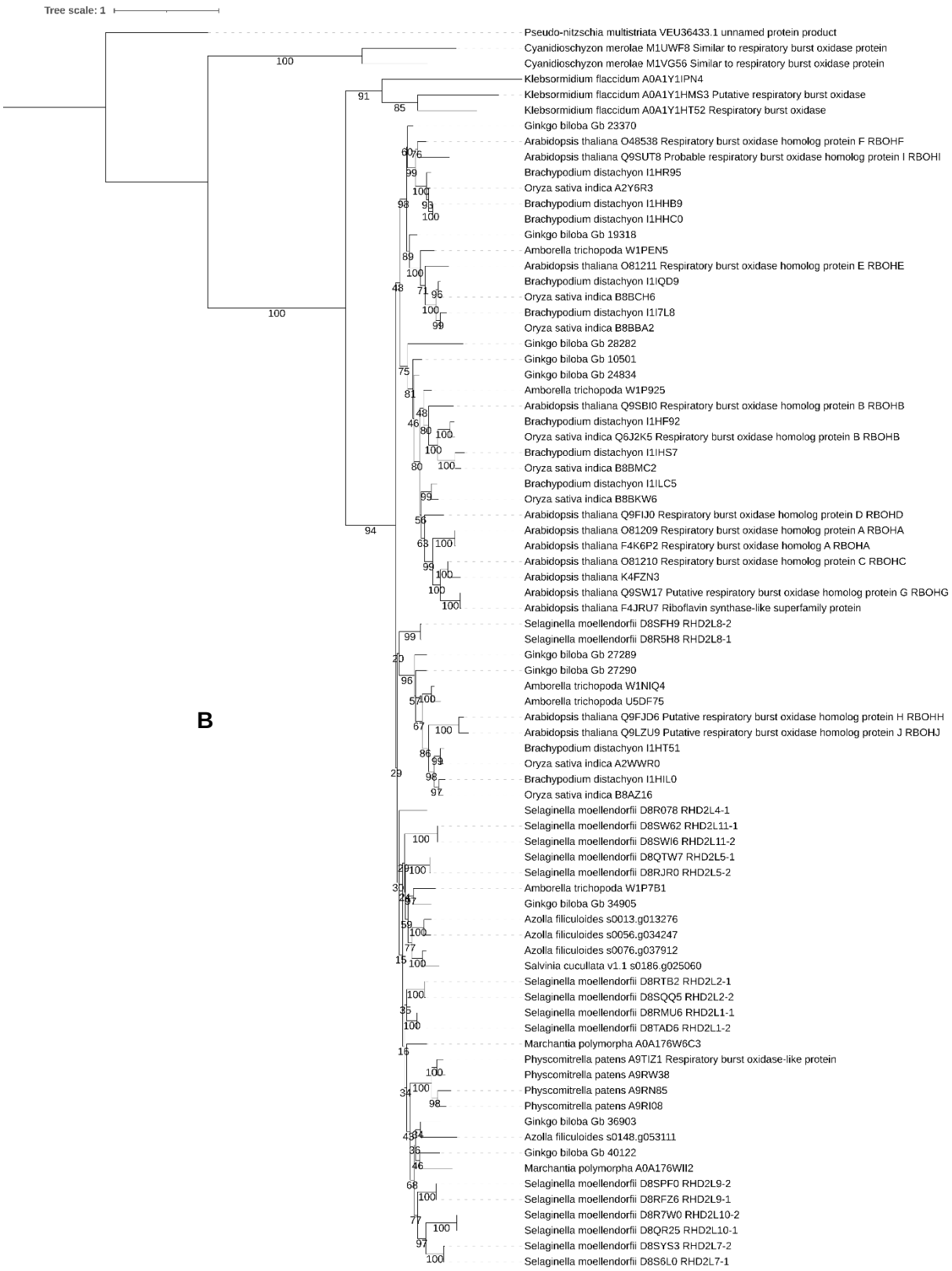


Figure S8.7. Maximum likelihood gene tree for RESPIRATORY BURST OXIDASE HOMOLOGY PROTEIN (RBOH, Uniprot ID: O81210) without (A) and with branch lengths (B).

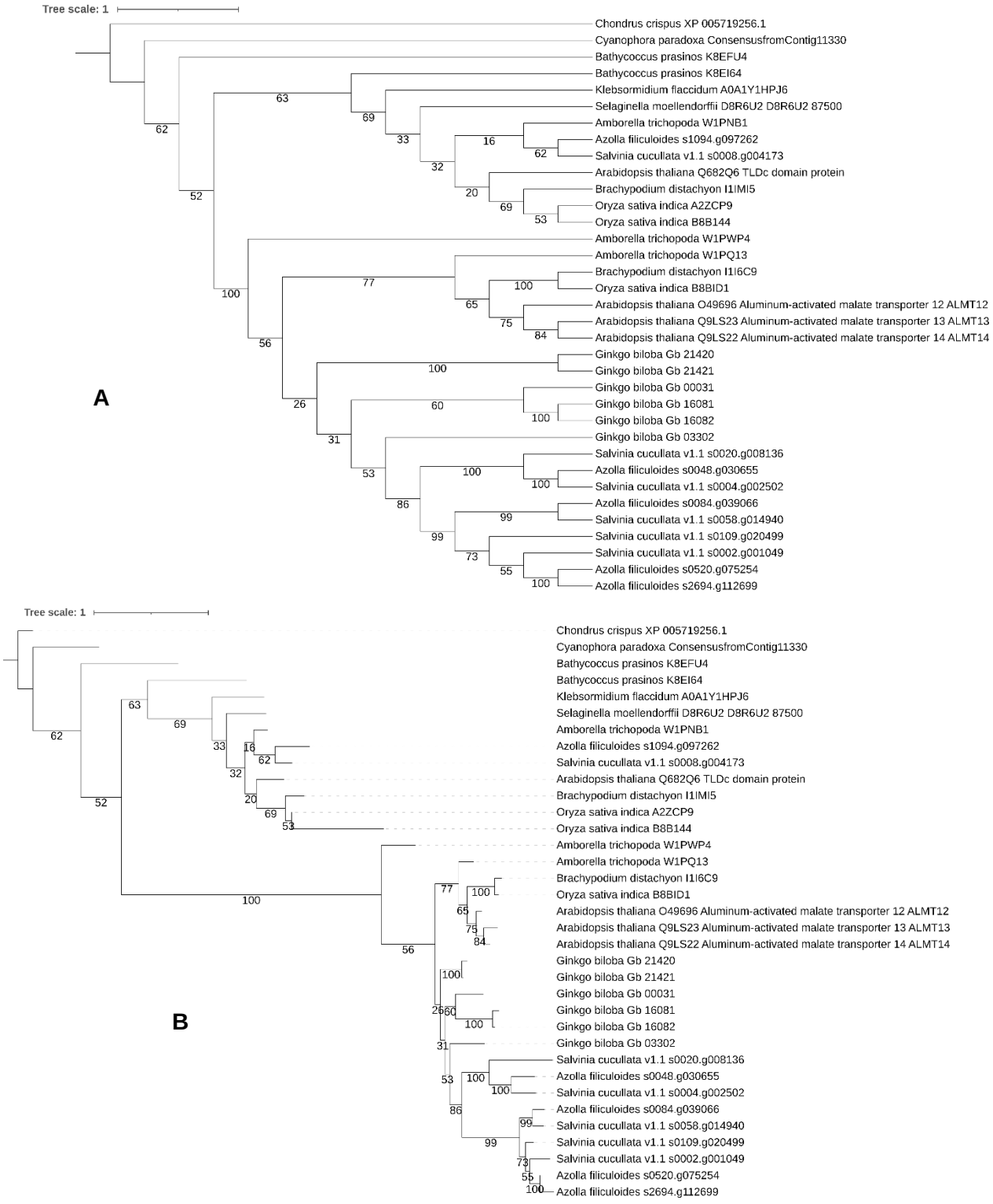
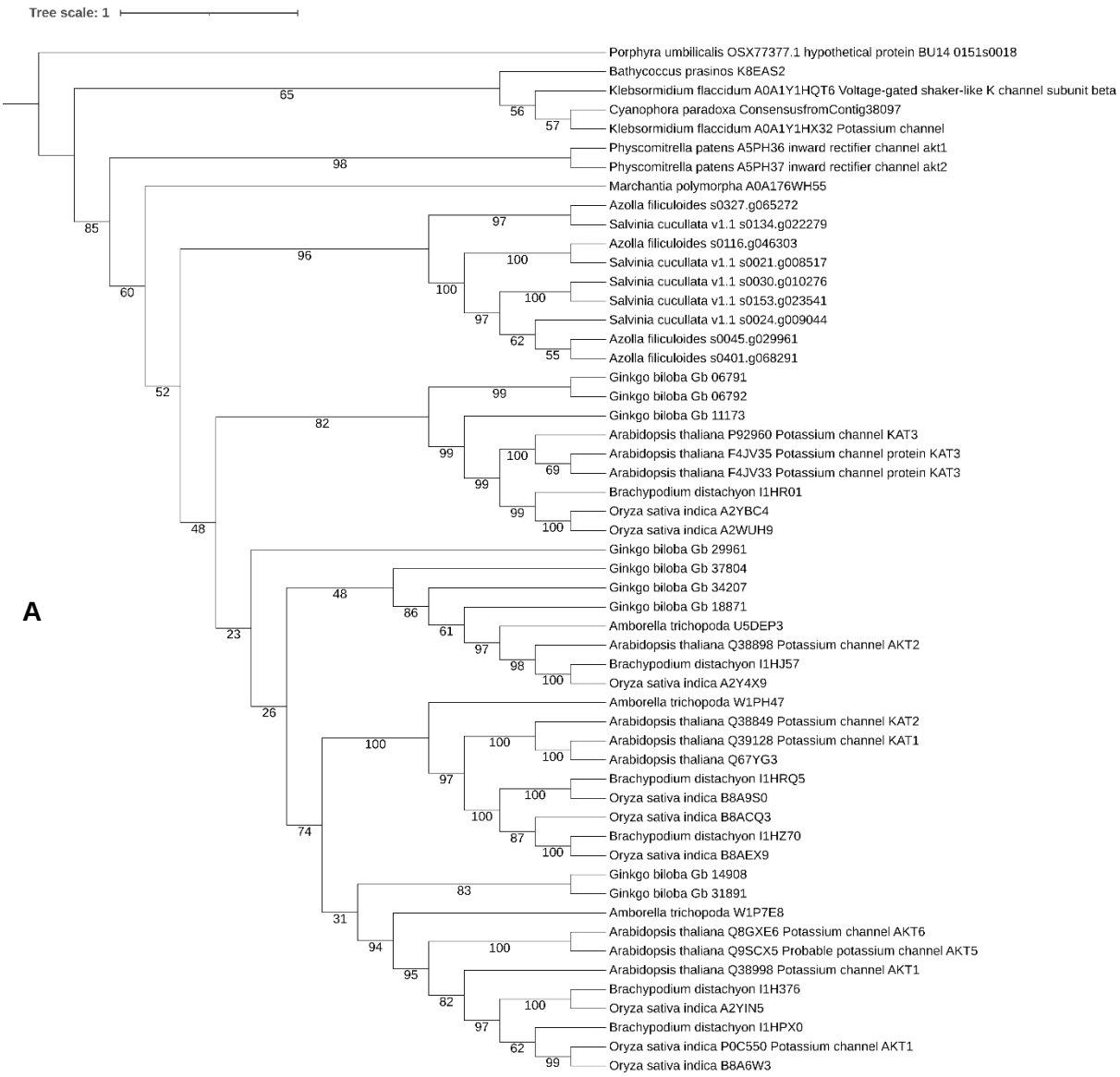


Figure S8.8. Maximum likelihood gene tree for QUICK ANION CHANNEL 1 (QUAC1, Uniprot ID: O49696) without (A) and with branch lengths (B).


Tree scale: 1
Porphyra umbilicalis OSX77377.1 hypothetical protein BU14 0151s0018
Bathycoccus prasinos K8EAS2
Klebsormidium flaccidum A0A1Y1HQT6 Voltage-gated shaker-like K channel subunit beta
Cyanophora paradoxa ConsensusfromContig38097
Klebsormidium flaccidum A0A1Y1HX32 Potassium channel
Physcomitrella patens A5PH36 inward rectifier channel akt1
Physcomitrella patens A5PH37 inward rectifier channel akt2
Marchantia polymorpha A0A176WH55
Azolla filiculoides s0327.g065272
Salvinia cucullata v1.1 s0134.g022279
Azolla filiculoides s0116.g046303
Salvinia cucullata v1.1 s0021.g008517
Salvinia cucullata v1.1 s0030.g010276
Salvinia cucullata v1.1 s0153.g023541
Salvinia cucullata v1.1 s0024.g009044
Azolla filiculoides s0045.g029961
Azolla filiculoides s0401.g068291
Ginkgo biloba Gb 06791
Ginkgo biloba Gb 06792
Ginkgo biloba Gb 11173
Arabidopsis thaliana P92960 Potassium channel KAT3
Arabidopsis thaliana F4JV35 Potassium channel protein KAT3
Arabidopsis thaliana F4JV33 Potassium channel protein KAT3
Brachypodium distachyon I1HR01
Oryza sativa indica A2YBC4
Oryza sativa indica A2WUH9
Ginkgo biloba Gb 29961
Ginkgo biloba Gb 37804
Ginkgo biloba Gb 34207
Ginkgo biloba Gb 18871
Amborella trichopoda U5DEP3
Arabidopsis thaliana Q38898 Potassium channel AKT2
Brachypodium distachyon I1HJ57
Oryza sativa indica A2Y4X9
Amborella trichopoda W1PH47
Arabidopsis thaliana Q38849 Potassium channel KAT2
Arabidopsis thaliana Q39128 Potassium channel KAT1
Arabidopsis thaliana Q67YG3
Brachypodium distachyon I1HRQ5
Oryza sativa indica B8A9S0
Oryza sativa indica B8ACQ3
Brachypodium distachyon I1HZ70
Oryza sativa indica B8AEX9
Ginkgo biloba Gb 14908
Ginkgo biloba Gb 31891
Amborella trichopoda W1P7E8
Arabidopsis thaliana Q8GXE6 Potassium channel AKT6
Arabidopsis thaliana Q9SCX5 Probable potassium channel AKT5
Arabidopsis thaliana Q38998 Potassium channel AKT1
Brachypodium distachyon I1H376
Oryza sativa indica A2YIN5
Brachypodium distachyon I1HPX0
Oryza sativa indica P0C550 Potassium channel AKT1
Oryza sativa indica B8A6W3

A

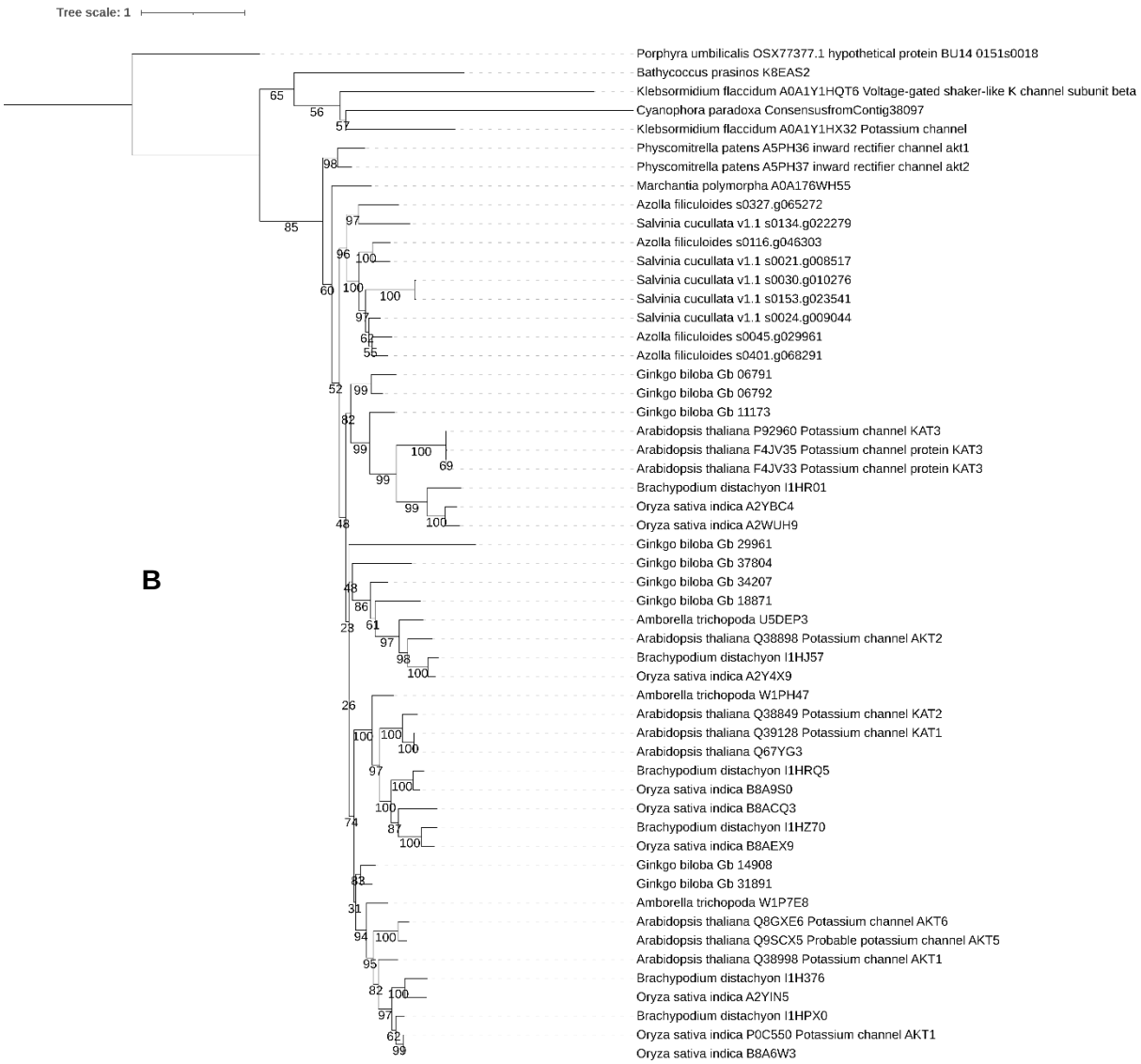


Figure S8.9. Maximum likelihood gene tree for Potassium channels AKT1/ KAT2 (Uniprot: Q38998/ Q38849) without (A) and with branch lengths (B).

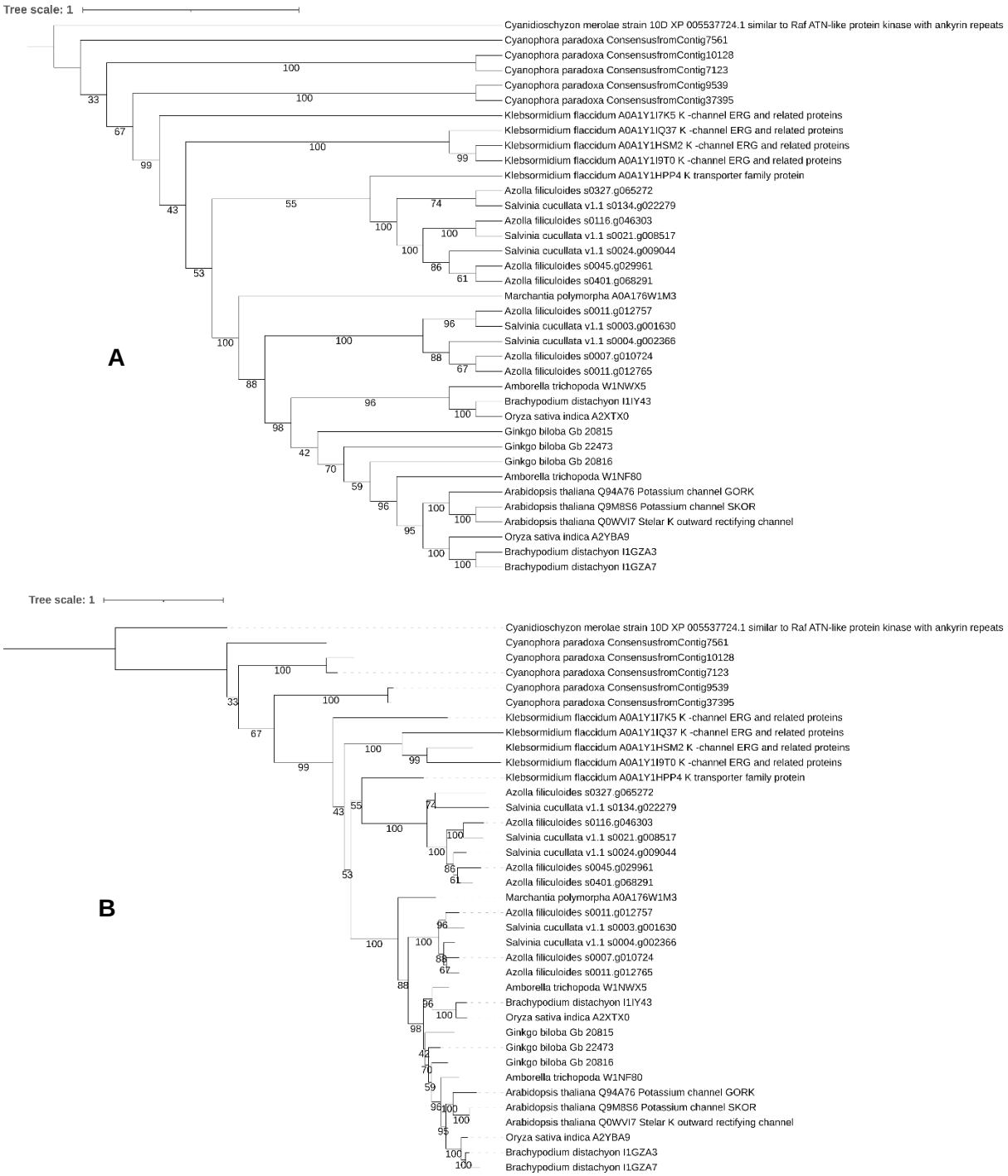
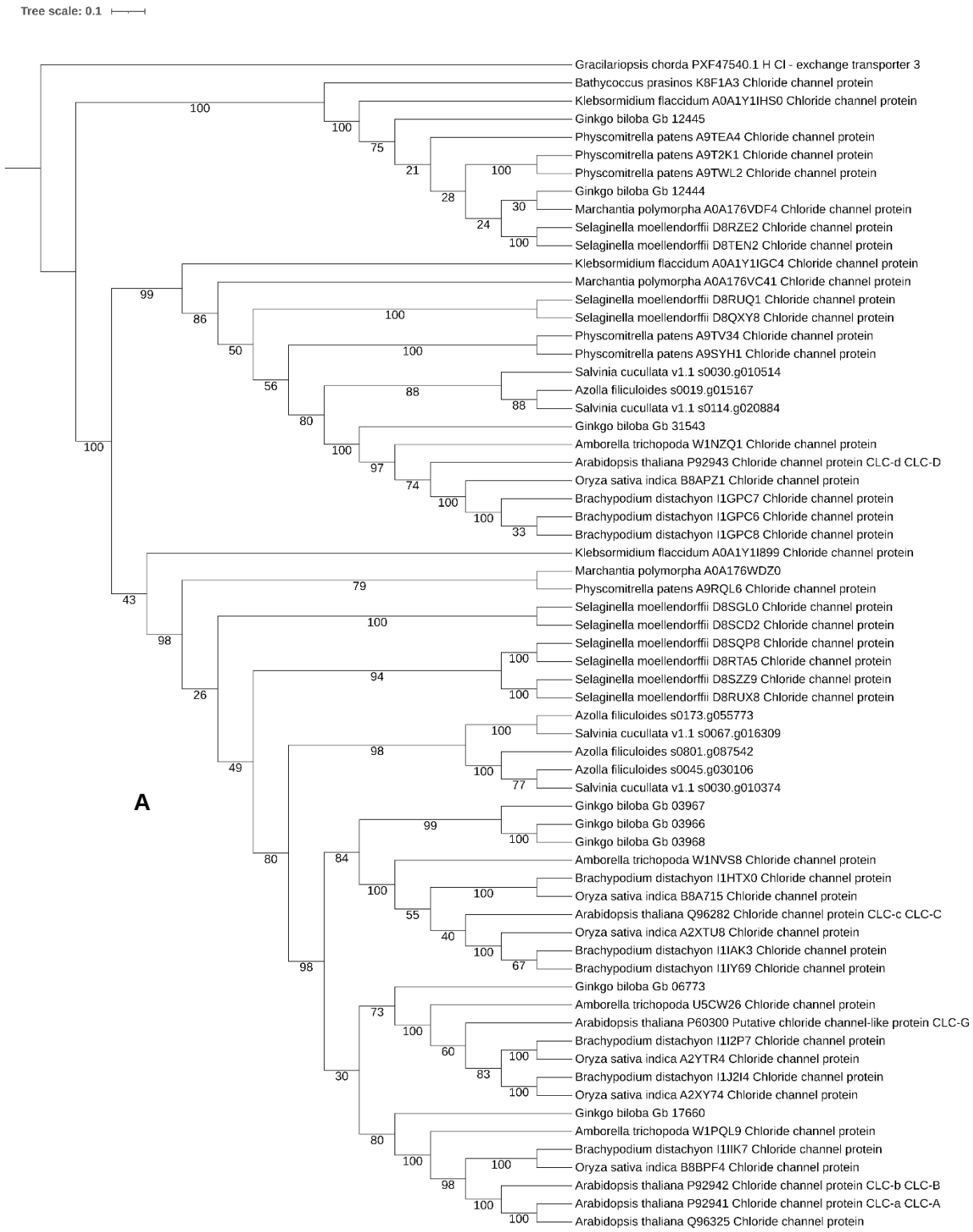


Figure S8.10. Maximum likelihood gene tree for GUARD CELL OUTWARD RECTIFYING K(+) CHANNEL (GORK, Uniprot ID: Q94A76) without (A) and with branch lengths (B).


Tree scale: 0.1
Gracilariopsis chorda PXF47540.1 H Cl - exchange transporter 3
Bathycoccus prasinos K8F1A3 Chloride channel protein
Klebsormidium flaccidum A0A1Y1IHS0 Chloride channel protein
Ginkgo biloba Gb 12445
Physcomitrella patens A9TEA4 Chloride channel protein
Physcomitrella patens A9T2K1 Chloride channel protein
Physcomitrella patens A9TWL2 Chloride channel protein
Ginkgo biloba Gb 12444
Marchantia polymorpha A0A176VDF4 Chloride channel protein
Selaginella moellendorffii D8RZE2 Chloride channel protein
Selaginella moellendorffii D8TEN2 Chloride channel protein
Klebsormidium flaccidum A0A1Y1IGC4 Chloride channel protein
Marchantia polymorpha A0A176VC41 Chloride channel protein
Selaginella moellendorffii D8RUQ1 Chloride channel protein
Selaginella moellendorffii D8QXY8 Chloride channel protein
Physcomitrella patens A9TV34 Chloride channel protein
Physcomitrella patens A9SYH1 Chloride channel protein
Salvinia cucullata v1.1 s0030.g010514
Azolla filiculoides s0019.g015167
Salvinia cucullata v1.1 s0114.g020884
Ginkgo biloba Gb 31543
Amborella trichopoda W1NZQ1 Chloride channel protein
Arabidopsis thaliana P92943 Chloride channel protein CLC-d CLC-D
Oryza sativa indica B8APZ1 Chloride channel protein
Brachypodium distachyon I1GPC7 Chloride channel protein
Brachypodium distachyon I1GPC6 Chloride channel protein
Brachypodium distachyon I1GPC8 Chloride channel protein
Klebsormidium flaccidum A0A1Y1I899 Chloride channel protein
Marchantia polymorpha A0A176WDZ0
Physcomitrella patens A9RQL6 Chloride channel protein
Selaginella moellendorffii D8SGL0 Chloride channel protein
Selaginella moellendorffii D8SCD2 Chloride channel protein
Selaginella moellendorffii D8SQP8 Chloride channel protein
Selaginella moellendorffii D8RTA5 Chloride channel protein
Selaginella moellendorffii D8SZZ9 Chloride channel protein
Selaginella moellendorffii D8RUX8 Chloride channel protein
Azolla filiculoides s0173.g055773
Salvinia cucullata v1.1 s0067.g016309
Azolla filiculoides s0801.g087542
Azolla filiculoides s0045.g030106
Salvinia cucullata v1.1 s0030.g010374
Ginkgo biloba Gb 03967
Ginkgo biloba Gb 03966
Ginkgo biloba Gb 03968
Amborella trichopoda W1NVS8 Chloride channel protein
Brachypodium distachyon I1HTX0 Chloride channel protein
Oryza sativa indica B8A715 Chloride channel protein
Arabidopsis thaliana Q96282 Chloride channel protein CLC-c CLC-C
Oryza sativa indica A2XTU8 Chloride channel protein
Brachypodium distachyon I1IAK3 Chloride channel protein
Brachypodium distachyon I1IY69 Chloride channel protein
Ginkgo biloba Gb 06773
Amborella trichopoda U5CW26 Chloride channel protein
Arabidopsis thaliana P60300 Putative chloride channel-like protein CLC-G
Brachypodium distachyon I1I2P7 Chloride channel protein
Oryza sativa indica A2YTR4 Chloride channel protein
Brachypodium distachyon I1J2I4 Chloride channel protein
Oryza sativa indica A2XY74 Chloride channel protein
Ginkgo biloba Gb 17660
Amborella trichopoda W1PQL9 Chloride channel protein
Brachypodium distachyon I1IIK7 Chloride channel protein
Oryza sativa indica B8BPF4 Chloride channel protein
Arabidopsis thaliana P92942 Chloride channel protein CLC-b CLC-B
Arabidopsis thaliana P92941 Chloride channel protein CLC-a CLC-A
Arabidopsis thaliana Q96325 Chloride channel protein
A

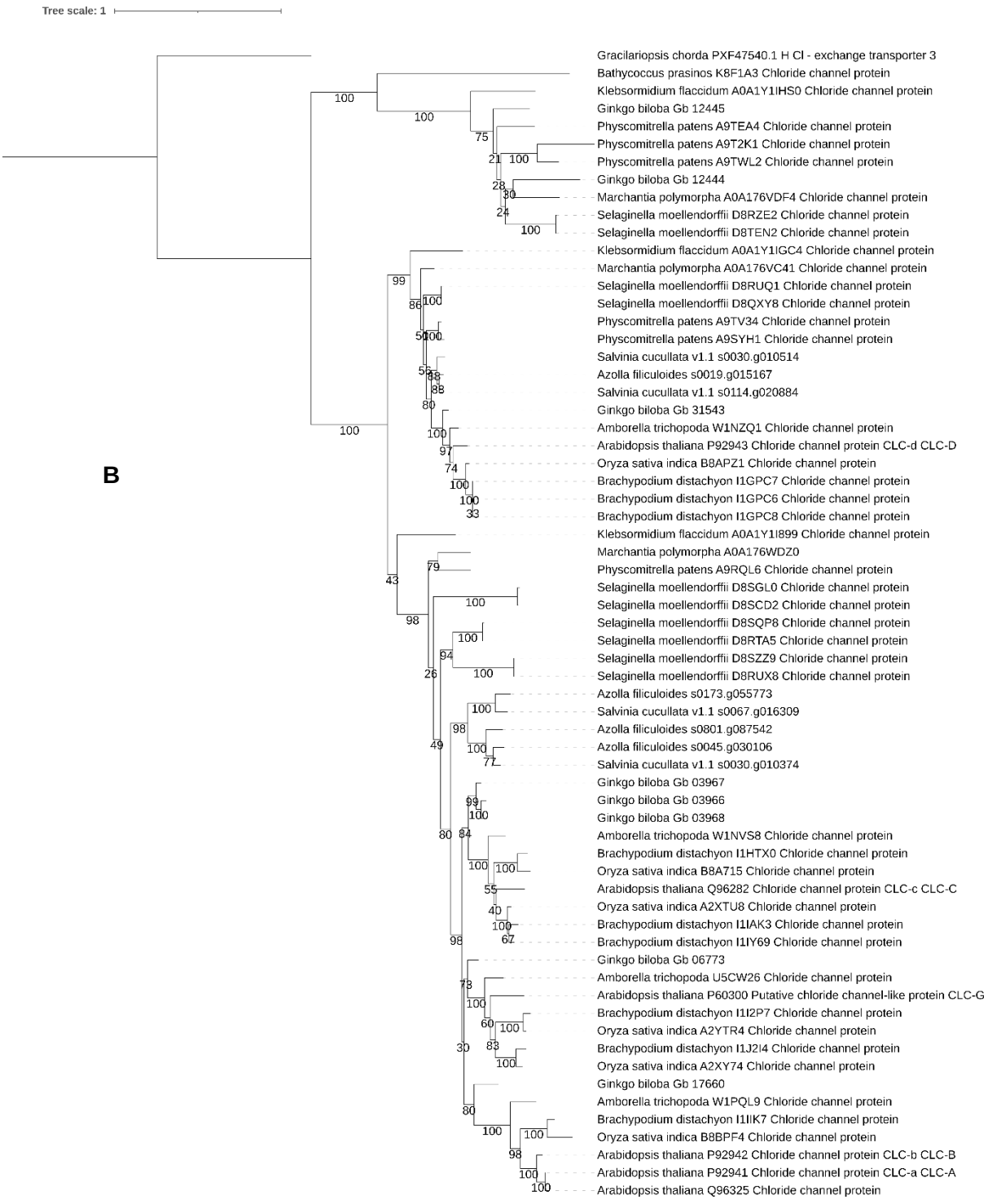


Figure S8.11. Maximum likelihood gene tree for CHLORIDE CHANNEL PROTEIN (CLC-C, Uniprot ID: Q96282) without (A) and with branch lengths (B).

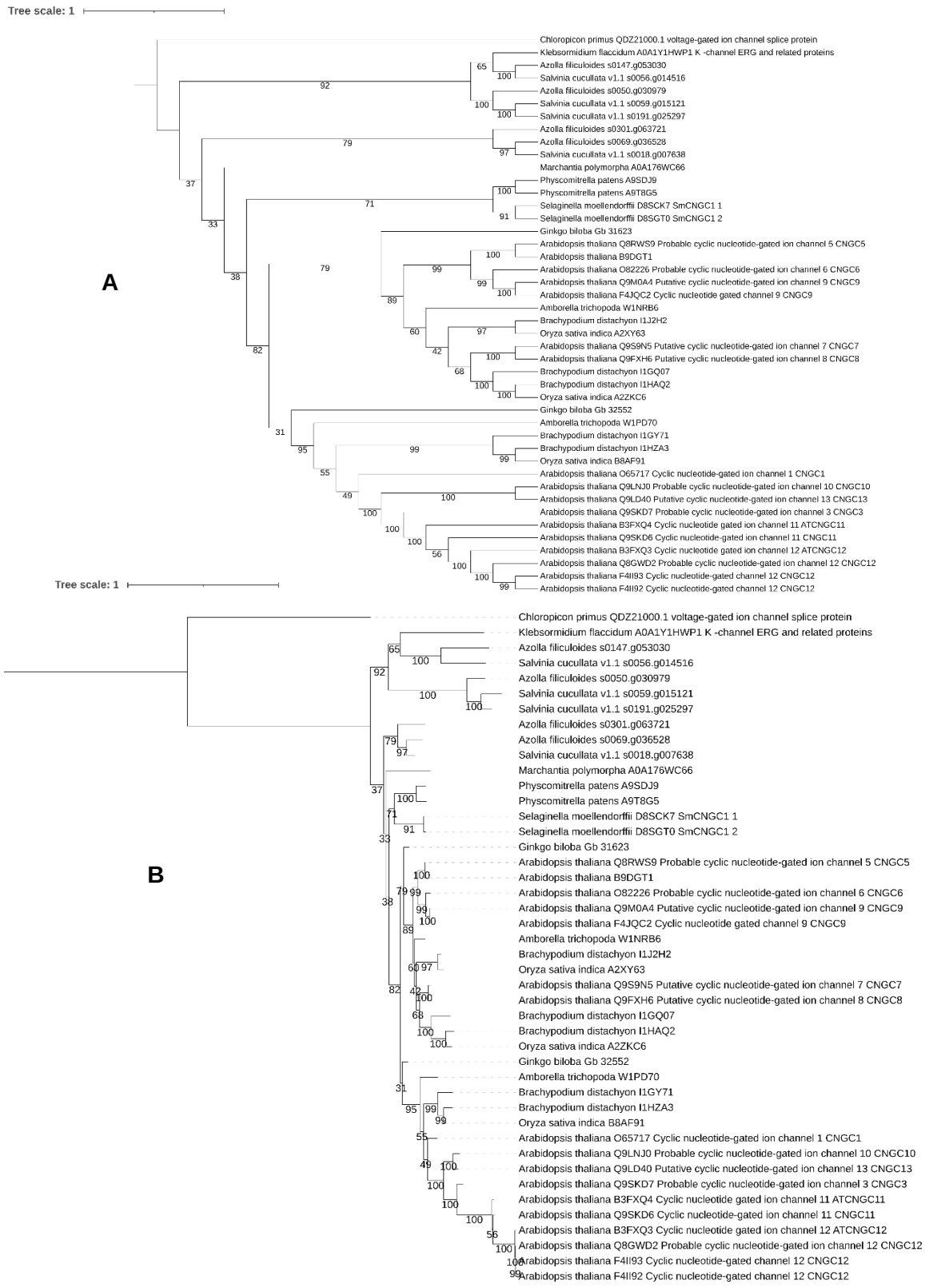


Figure S8.12. Maximum likelihood gene tree for CYCLIC NUCLEOTIDE-GATED ION CHANNEL (CNGC, Uniprot ID: O65717) without (A) and with branch lengths (B).

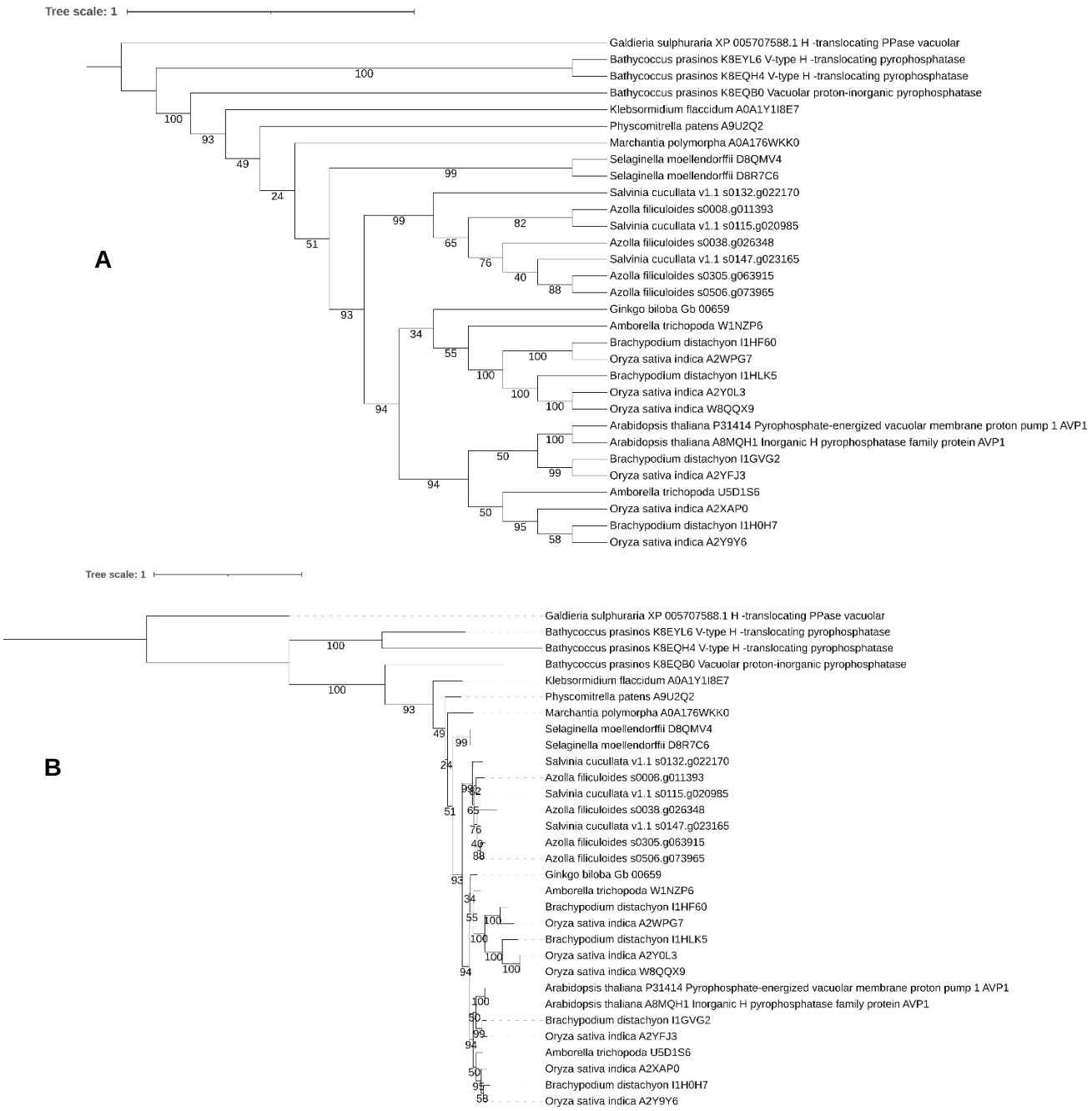


Figure S8.13. Maximum likelihood gene tree for VACUOLAR PROTON PYROPHOSPHATASE (AVP, Uniprot ID: P31414) without (A) and with branch lengths (B).

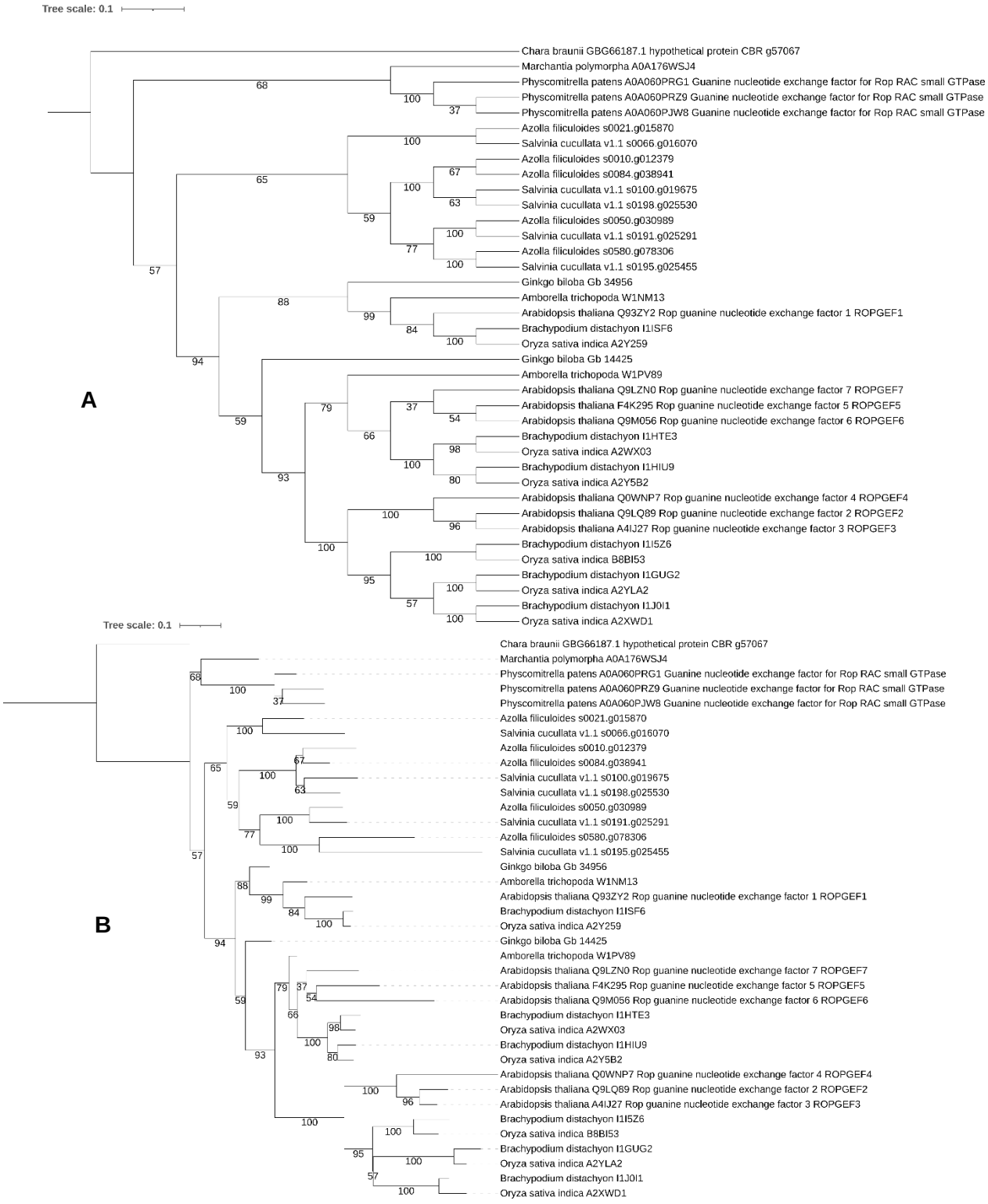
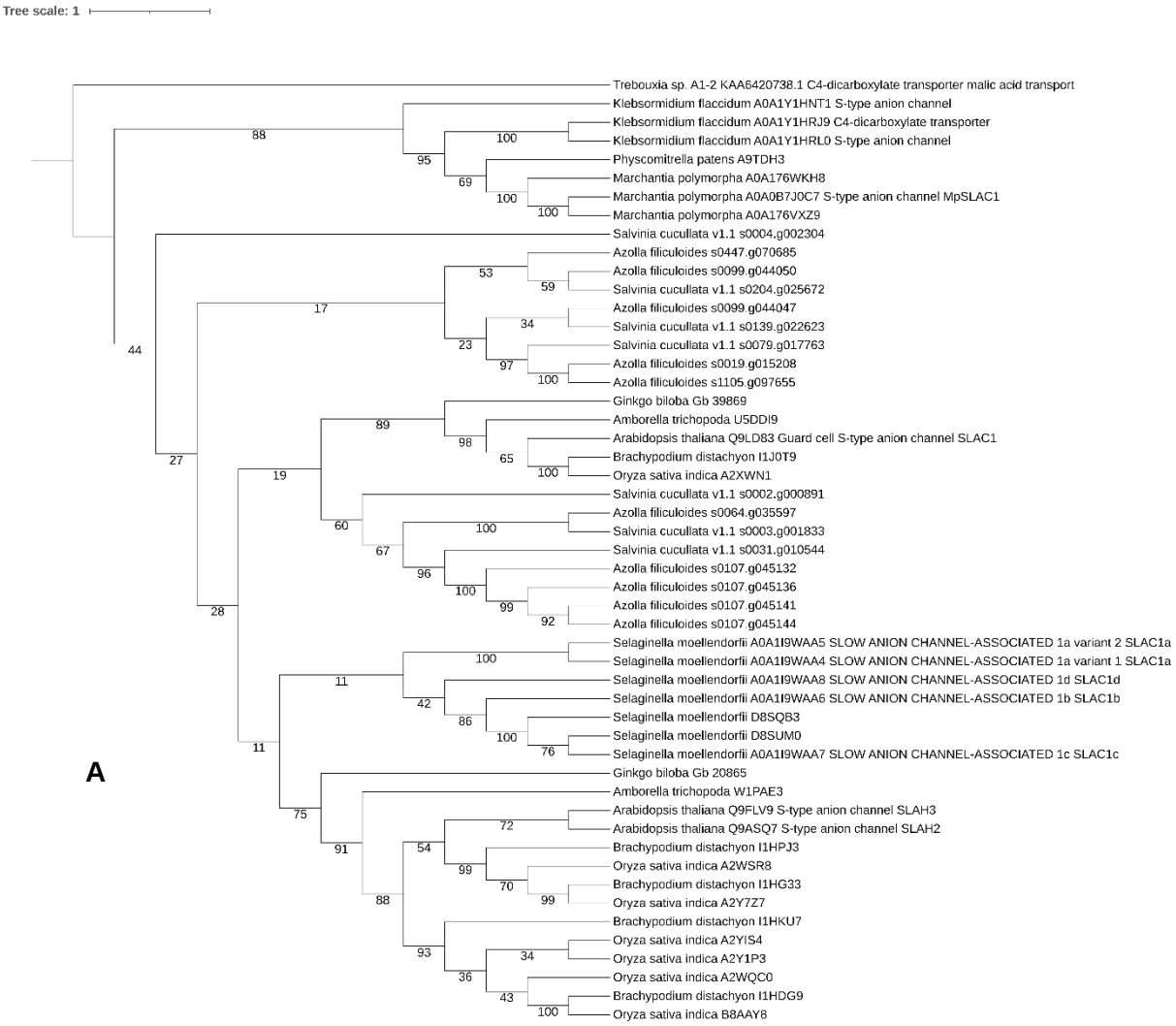


Figure S8.14. Maximum likelihood gene tree for RHO OF PLANTS GUANINE NUCLEOTIDE EXCHANGE FACTOR 1 (GEF1/4, Uniprot ID: Q93ZY2) without (A) and with branch lengths (B).

Appendix 3.1.1.2 Gene duplications in the ancestor of Euphyllophyta



A

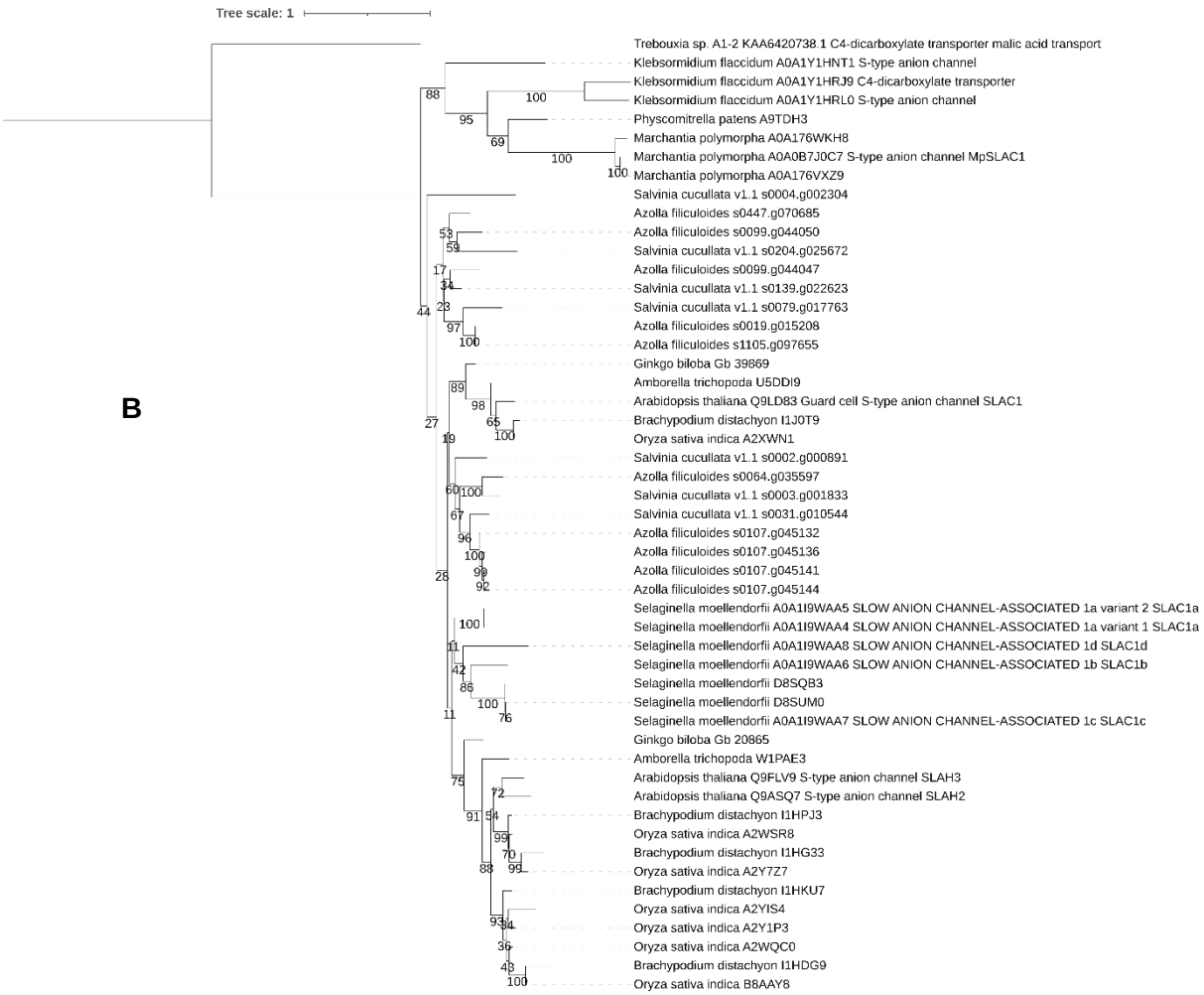


Figure S8.15. Maximum likelihood gene tree for SLOW ANION CHANNEL-ASSOCIATED 1 (SLAC, Uniprot: Q9LD83) without (A) and with branch lengths (B).

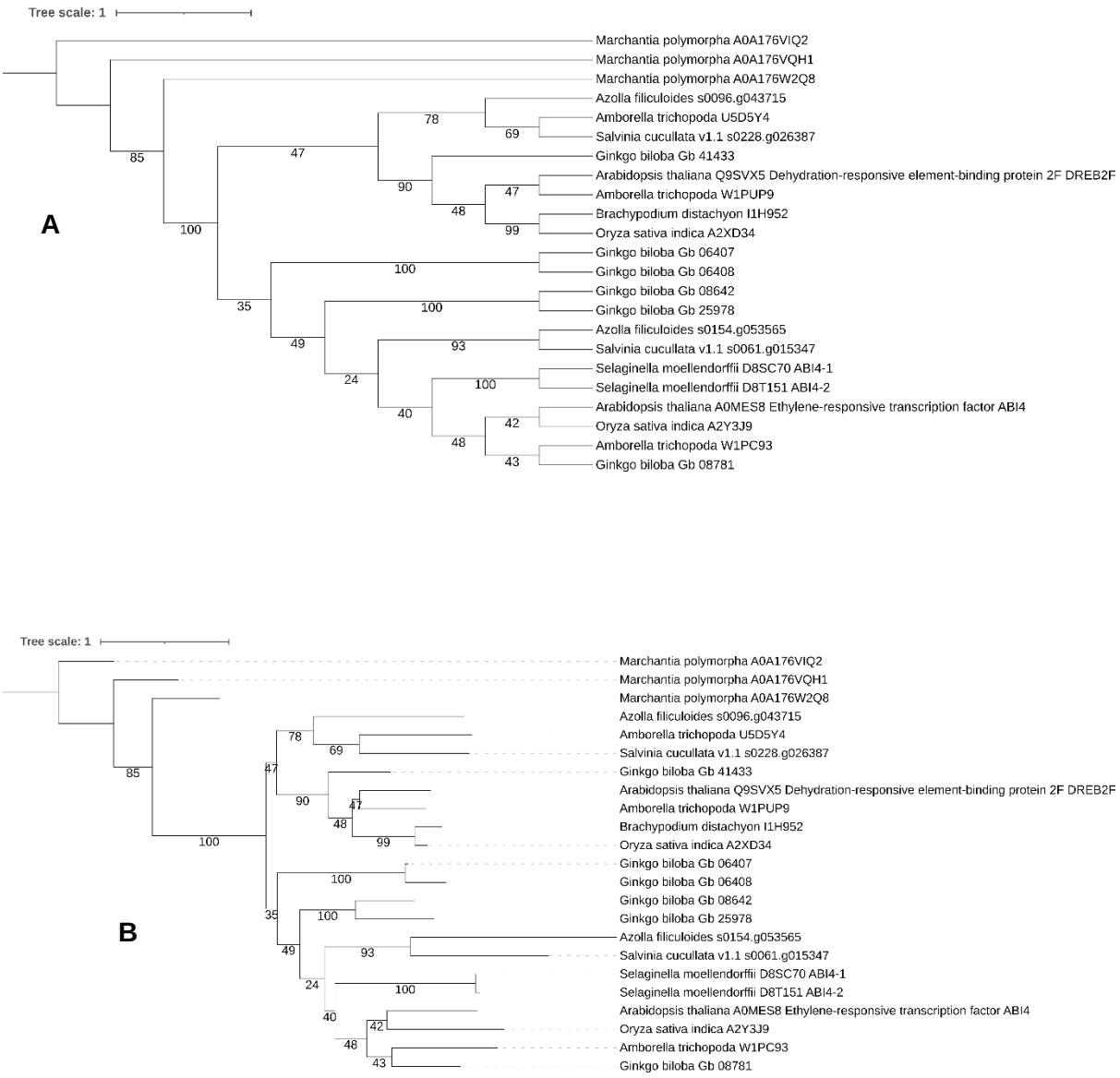
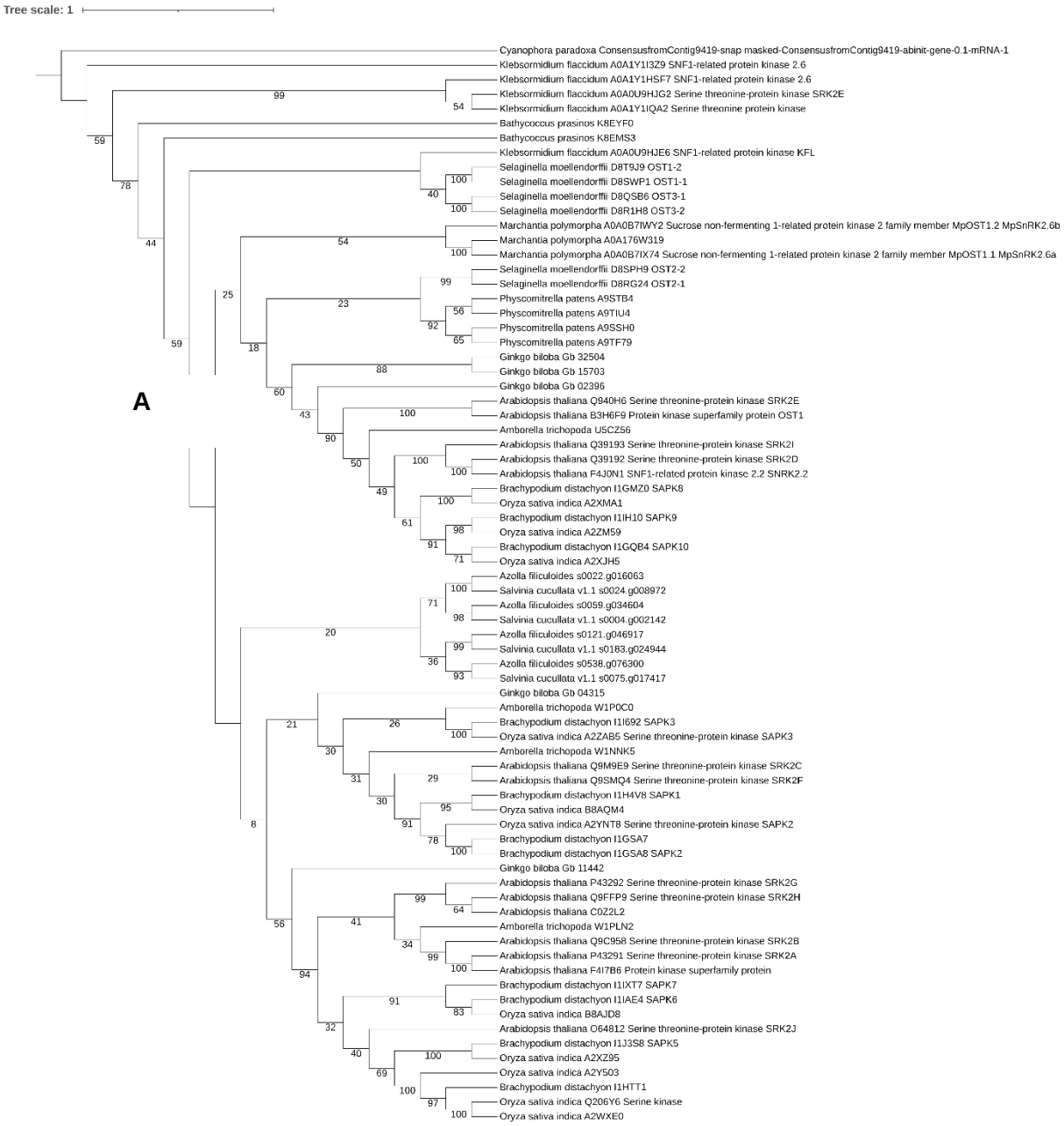


Figure S8.16. Maximum likelihood gene tree for ABSCISIC ACID INSENSITIVE 4 (ABI4, Uniprot: A0MES8) without (A) and with branch lengths (B).


Tree scale: 1
A
Cyanophora paradoxa ConsensusfromContig9419-snap masked-ConsensusfromContig9419-abinit-gene-0.1-mRNA-1
Klebsormidium flaccidum A0A1Y1I3Z9 SNF1-related protein kinase 2.6
Klebsormidium flaccidum A0A1Y1HSF7 SNF1-related protein kinase 2.6
Klebsormidium flaccidum A0A0U9HJG2 Serine threonine-protein kinase SRK2E
Klebsormidium flaccidum A0A1Y1IQA2 Serine threonine protein kinase
Bathycoccus prasinos K8EYF0
Bathycoccus prasinos K8EMS3
Klebsormidium flaccidum A0A0U9HJE6 SNF1-related protein kinase KFL
Selaginella moellendorffii D8T9J9 OST1-2
Selaginella moellendorffii D8SWP1 OST1-1
Selaginella moellendorffii D8QSB6 OST3-1
Selaginella moellendorffii D8R1H8 OST3-2
Marchantia polymorpha A0A0B7IWY2 Sucrose non-fermenting 1-related protein kinase 2 family member MpOST1.2 MpSnRK2.6b
Marchantia polymorpha A0A176W319
Marchantia polymorpha A0A0B7IX74 Sucrose non-fermenting 1-related protein kinase 2 family member MpOST1.1 MpSnRK2.6a
Selaginella moellendorffii D8SPH9 OST2-2
Selaginella moellendorffii D8RG24 OST2-1
Physcomitrella patens A9STB4
Physcomitrella patens A9TIU4
Physcomitrella patens A9SSH0
Physcomitrella patens A9TF79
Ginkgo biloba Gb 32504
Ginkgo biloba Gb 15703
Ginkgo biloba Gb 02396
Arabidopsis thaliana Q940H6 Serine threonine-protein kinase SRK2E
Arabidopsis thaliana B3H6F9 Protein kinase superfamily protein OST1
Amborella trichopoda U5CZS6
Arabidopsis thaliana Q39193 Serine threonine-protein kinase SRK2I
Arabidopsis thaliana Q39192 Serine threonine-protein kinase SRK2D
Arabidopsis thaliana F4J0N1 SNF1-related protein kinase 2.2 SNRK2.2
Brachypodium distachyon I1GMZ0 SAPK8
Oryza sativa indica A2XMA1
Brachypodium distachyon I1IH10 SAPK9
Oryza sativa indica A2ZM59
Brachypodium distachyon I1GQB4 SAPK10
Oryza sativa indica A2XJH5
Azolla filiculoides s0022.g016063
Salvinia cucullata v1.1 s0024.g008972
Azolla filiculoides s0059.g034604
Salvinia cucullata v1.1 s0004.g002142
Azolla filiculoides s0121.g046917
Salvinia cucullata v1.1 s0183.g024944
Azolla filiculoides s0538.g076300
Salvinia cucullata v1.1 s0075.g017417
Ginkgo biloba Gb 04315
Amborella trichopoda W1P0C0
Brachypodium distachyon I1I692 SAPK3
Oryza sativa indica A2ZAB5 Serine threonine-protein kinase SAPK3
Amborella trichopoda W1NNK5
Arabidopsis thaliana Q9M9E9 Serine threonine-protein kinase SRK2C
Arabidopsis thaliana Q9SMQ4 Serine threonine-protein kinase SRK2F
Brachypodium distachyon I1H4V8 SAPK1
Oryza sativa indica B8AQM4
Oryza sativa indica A2YNT8 Serine threonine-protein kinase SAPK2
Brachypodium distachyon I1GSA7
Brachypodium distachyon I1GSA8 SAPK2
Ginkgo biloba Gb 11442
Arabidopsis thaliana P43292 Serine threonine-protein kinase SRK2G
Arabidopsis thaliana Q9FFP9 Serine threonine-protein kinase SRK2H
Arabidopsis thaliana C0Z2L2
Amborella trichopoda W1PLN2
Arabidopsis thaliana Q9C958 Serine threonine-protein kinase SRK2B
Arabidopsis thaliana P43291 Serine threonine-protein kinase SRK2A
Arabidopsis thaliana F4I7B6 Protein kinase superfamily protein
Brachypodium distachyon I1IXT7 SAPK7
Brachypodium distachyon I1IAE4 SAPK6
Oryza sativa indica B8AJD8
Arabidopsis thaliana O64812 Serine threonine-protein kinase SRK2J
Brachypodium distachyon I1J3S8 SAPK5
Oryza sativa indica A2XZ95
Oryza sativa indica A2Y503
Brachypodium distachyon I1HTT1
Oryza sativa indica Q206Y6 Serine kinase
Oryza sativa indica A2WXE0

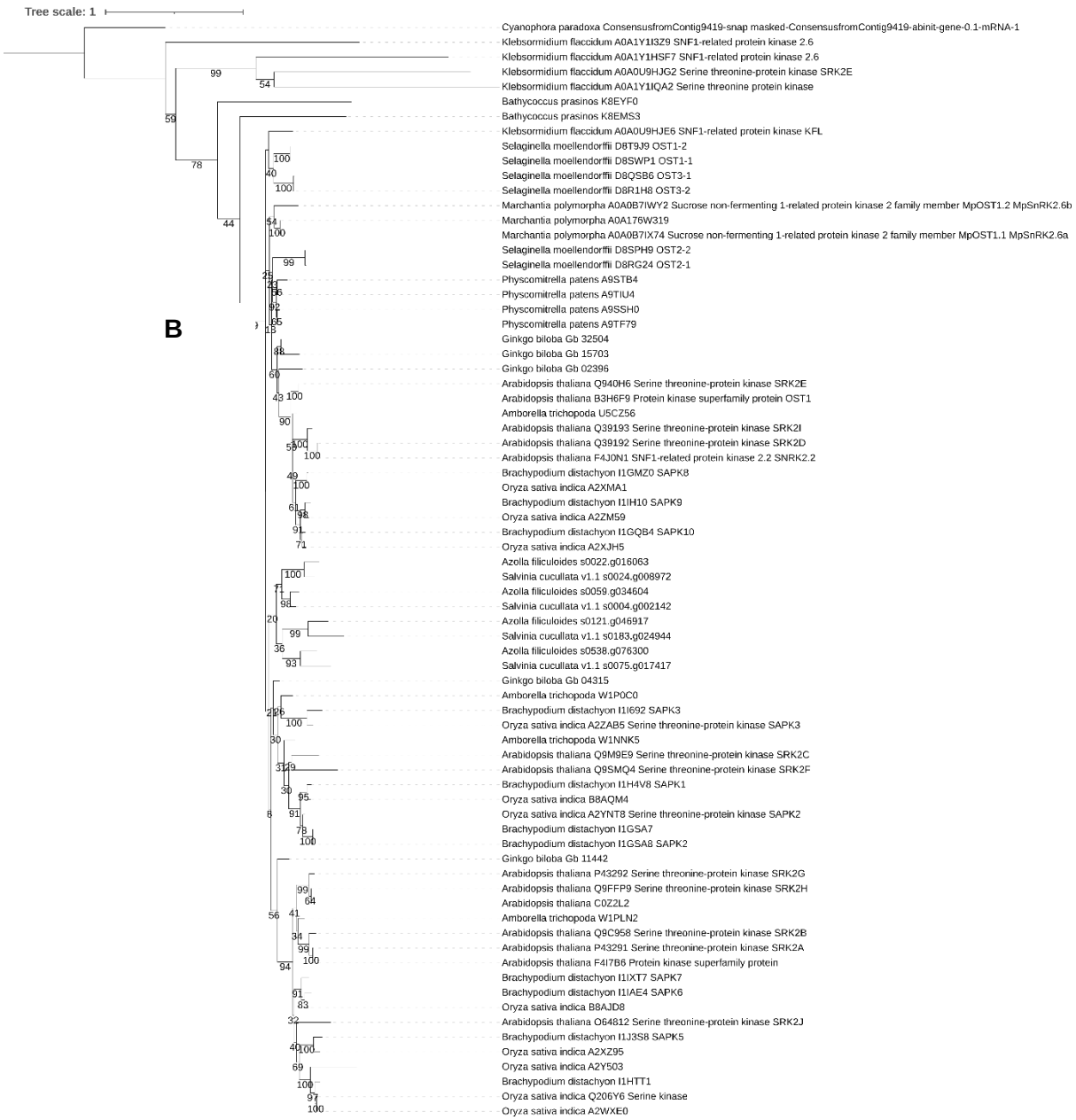


Figure S8.17. Maximum likelihood gene tree for SNF1-RELATED KINASE (SnRK2, Uniprot: P43291) without (A) and with branch lengths (B).

Appendix 3.1.1.3 Gene duplications in both the ancestors of Euphyllophyta and Spermatophyta

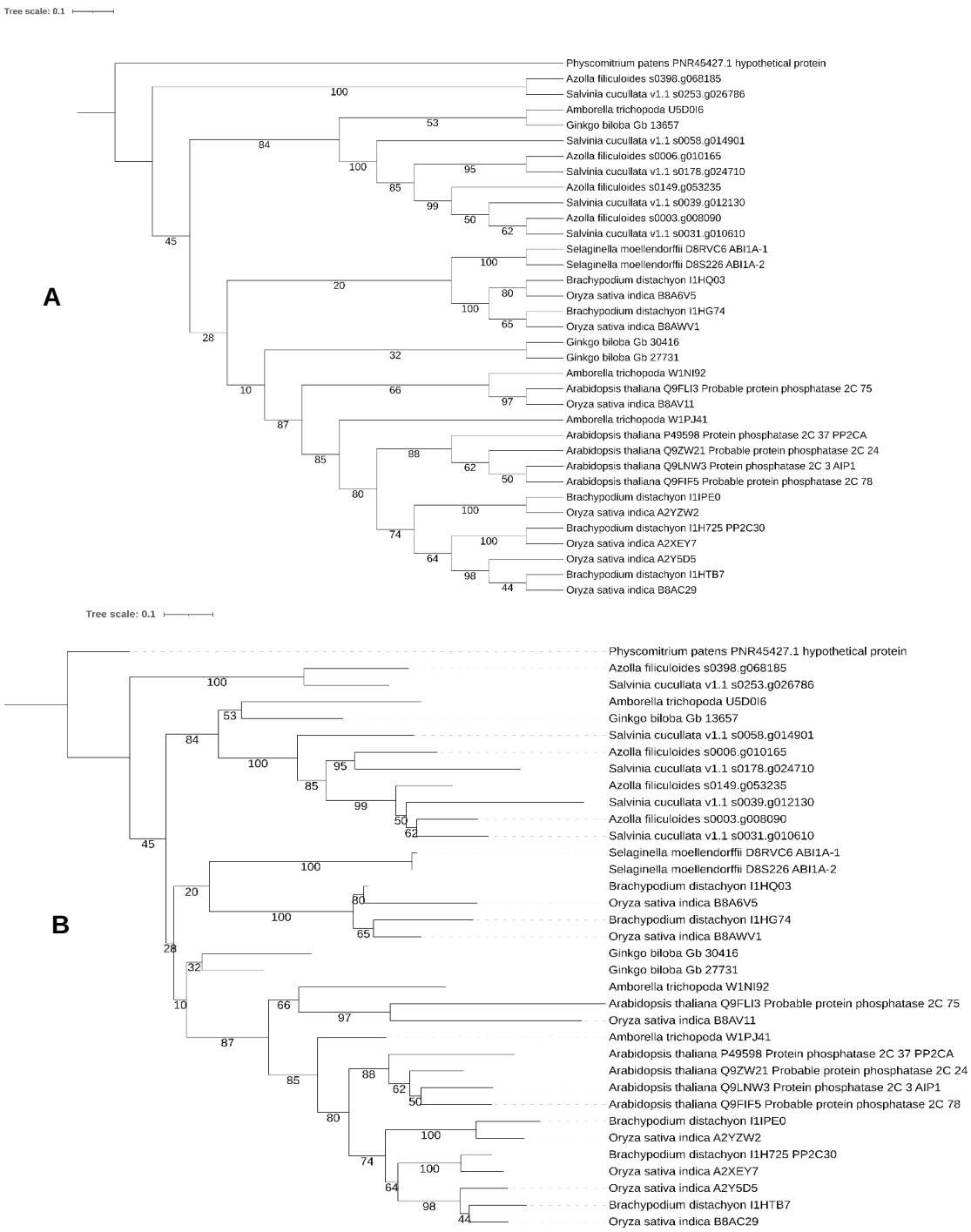
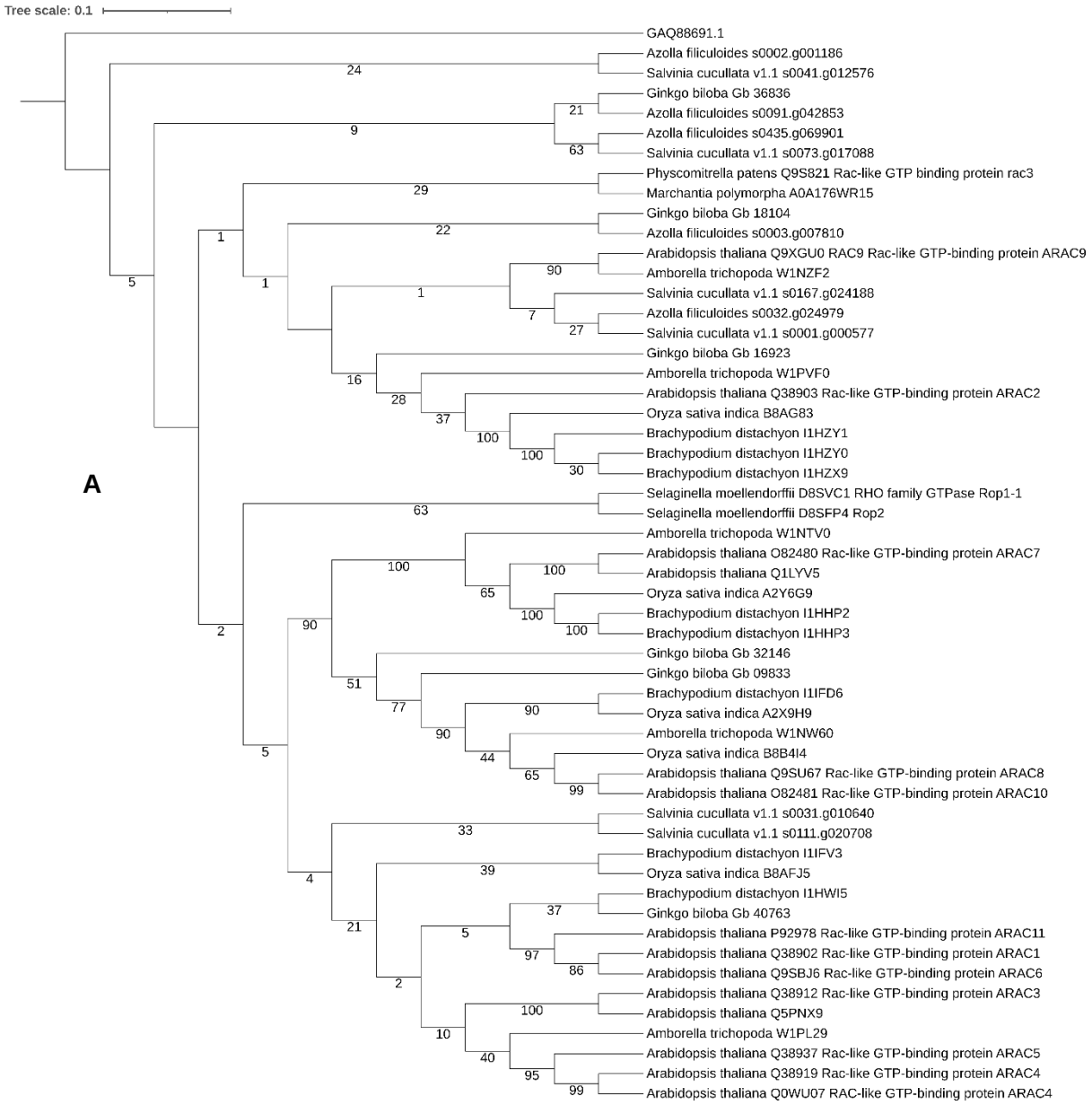


Figure S8.18. Maximum likelihood gene tree for SLOW ANION CHANNEL-ASSOCIATED 1 (PP2C, Uniprot ID: P49598) without (A) and with branch lengths (B).


Tree scale: 0.1
A
GAQ88691.1
Azolla filiculoides s0002.g001186
Salvinia cucullata v1.1 s0041.g012576
Ginkgo biloba Gb 36836
Azolla filiculoides s0091.g042853
Azolla filiculoides s0435.g069901
Salvinia cucullata v1.1 s0073.g017088
Physcomitrella patens Q9S821 Rac-like GTP binding protein rac3
Marchantia polymorpha A0A176WR15
Ginkgo biloba Gb 18104
Azolla filiculoides s0003.g007810
Arabidopsis thaliana Q9XGU0 RAC9 Rac-like GTP-binding protein ARAC9
Amborella trichopoda W1NZF2
Salvinia cucullata v1.1 s0167.g024188
Azolla filiculoides s0032.g024979
Salvinia cucullata v1.1 s0001.g000577
Ginkgo biloba Gb 16923
Amborella trichopoda W1PVF0
Arabidopsis thaliana Q38903 Rac-like GTP-binding protein ARAC2
Oryza sativa indica B8AG83
Brachypodium distachyon I1HZY1
Brachypodium distachyon I1HZY0
Brachypodium distachyon I1HZX9
Selaginella moellendorffii D8SVC1 RHO family GTPase Rop1-1
Selaginella moellendorffii D8SFP4 Rop2
Amborella trichopoda W1NTV0
Arabidopsis thaliana O82480 Rac-like GTP-binding protein ARAC7
Arabidopsis thaliana Q1LYV5
Oryza sativa indica A2Y6G9
Brachypodium distachyon I1HHP2
Brachypodium distachyon I1HHP3
Ginkgo biloba Gb 32146
Ginkgo biloba Gb 09833
Brachypodium distachyon I1IFD6
Oryza sativa indica A2X9H9
Amborella trichopoda W1NW60
Oryza sativa indica B8B4I4
Arabidopsis thaliana Q9SU67 Rac-like GTP-binding protein ARAC8
Arabidopsis thaliana O82481 Rac-like GTP-binding protein ARAC10
Salvinia cucullata v1.1 s0031.g010640
Salvinia cucullata v1.1 s0111.g020708
Brachypodium distachyon I1IFV3
Oryza sativa indica B8AFJ5
Brachypodium distachyon I1HWI5
Ginkgo biloba Gb 40763
Arabidopsis thaliana P92978 Rac-like GTP-binding protein ARAC11
Arabidopsis thaliana Q38902 Rac-like GTP-binding protein ARAC1
Arabidopsis thaliana Q9SBJ6 Rac-like GTP-binding protein ARAC6
Arabidopsis thaliana Q38912 Rac-like GTP-binding protein ARAC3
Arabidopsis thaliana Q5PNX9
Amborella trichopoda W1PL29
Arabidopsis thaliana Q38937 Rac-like GTP-binding protein ARAC5
Arabidopsis thaliana Q38919 Rac-like GTP-binding protein ARAC4
Arabidopsis thaliana Q0WU07 RAC-like GTP-binding protein ARAC4

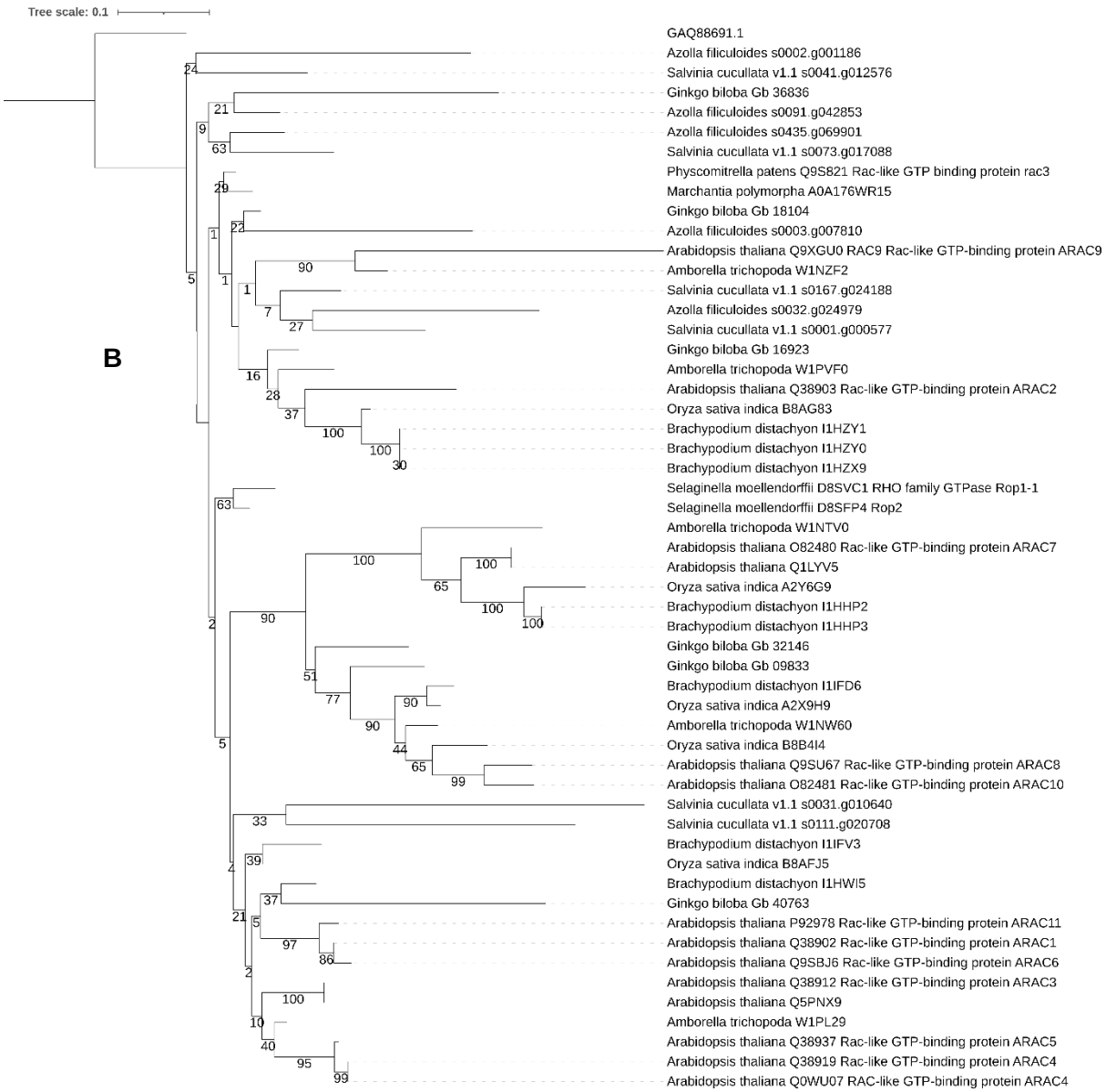


Figure S8.19. Maximum likelihood gene tree for RHO-RELATED PROTEIN FROM PLANTS 11 (ROP11, Uniprot: O82481) without (A) and with branch lengths (B).

Appendix 3.5.2. Phylogenetic trees for vascular tissue development genes discussed in Chapter 3.

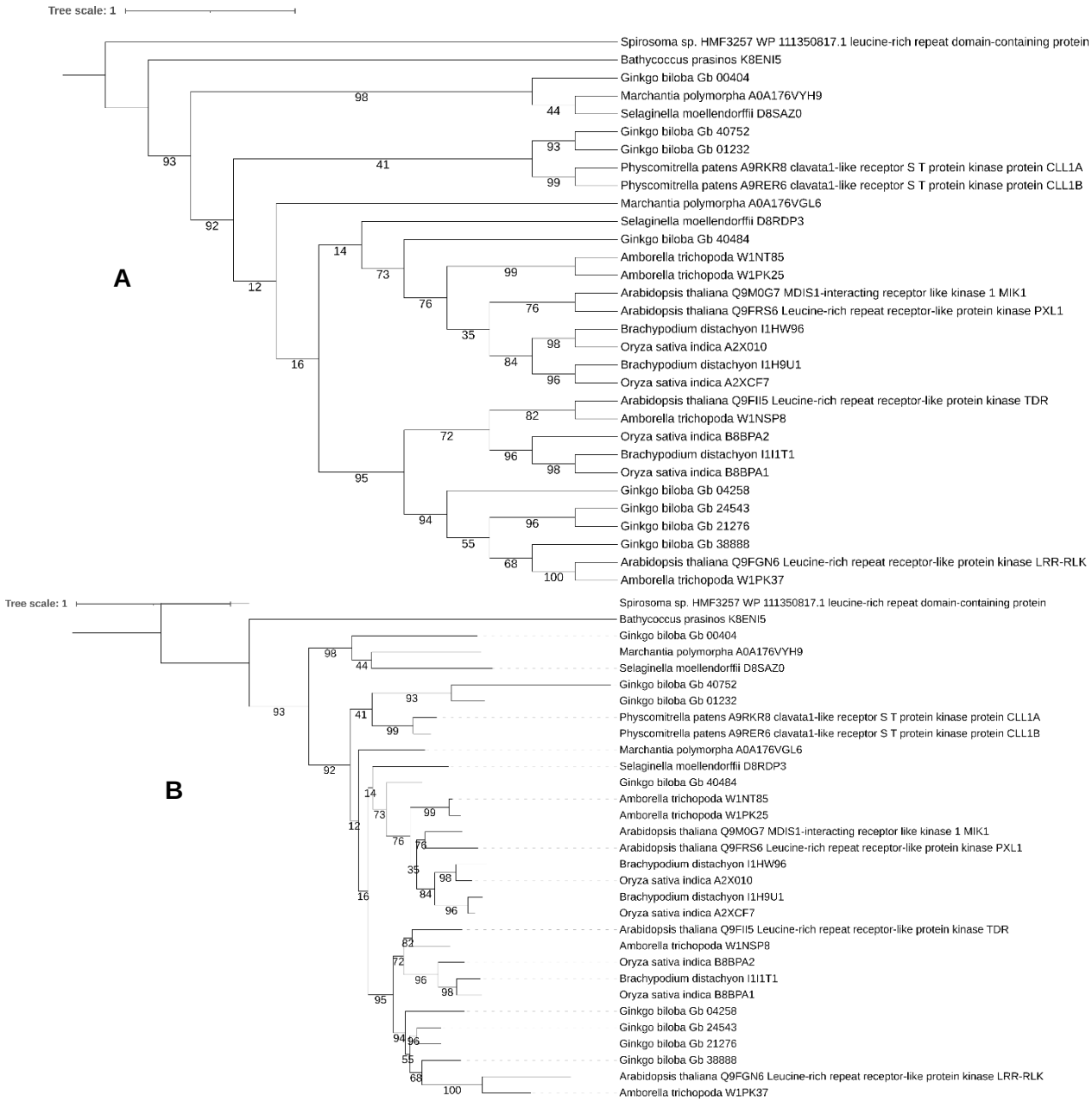


Figure S8.20. Maximum likelihood gene tree for PHLOEM INTERCALATED WITH XYLEM (PXY, Uniprot ID: Q9FII5) without (A) and with branch lengths (B).

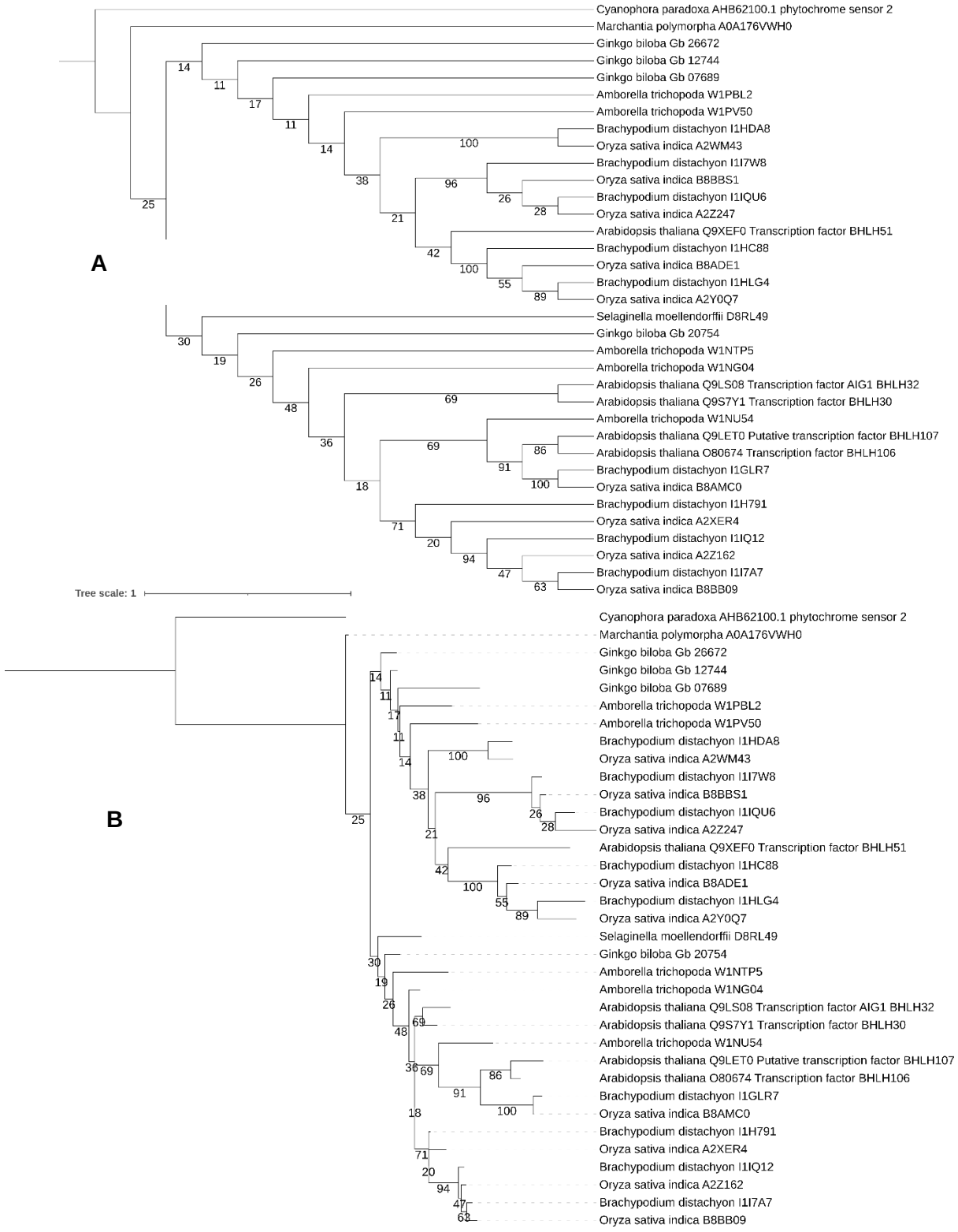


Figure S8.21. Maximum likelihood gene tree for TARGET OF MOOPTEROS 5 (TMO5, Uniprot ID: Q9FLS08) without (A) and with branch lengths (B).

Appendix 3.5.2. Phylogenetic trees for root hydrotropism genes discussed in Chapter 3.

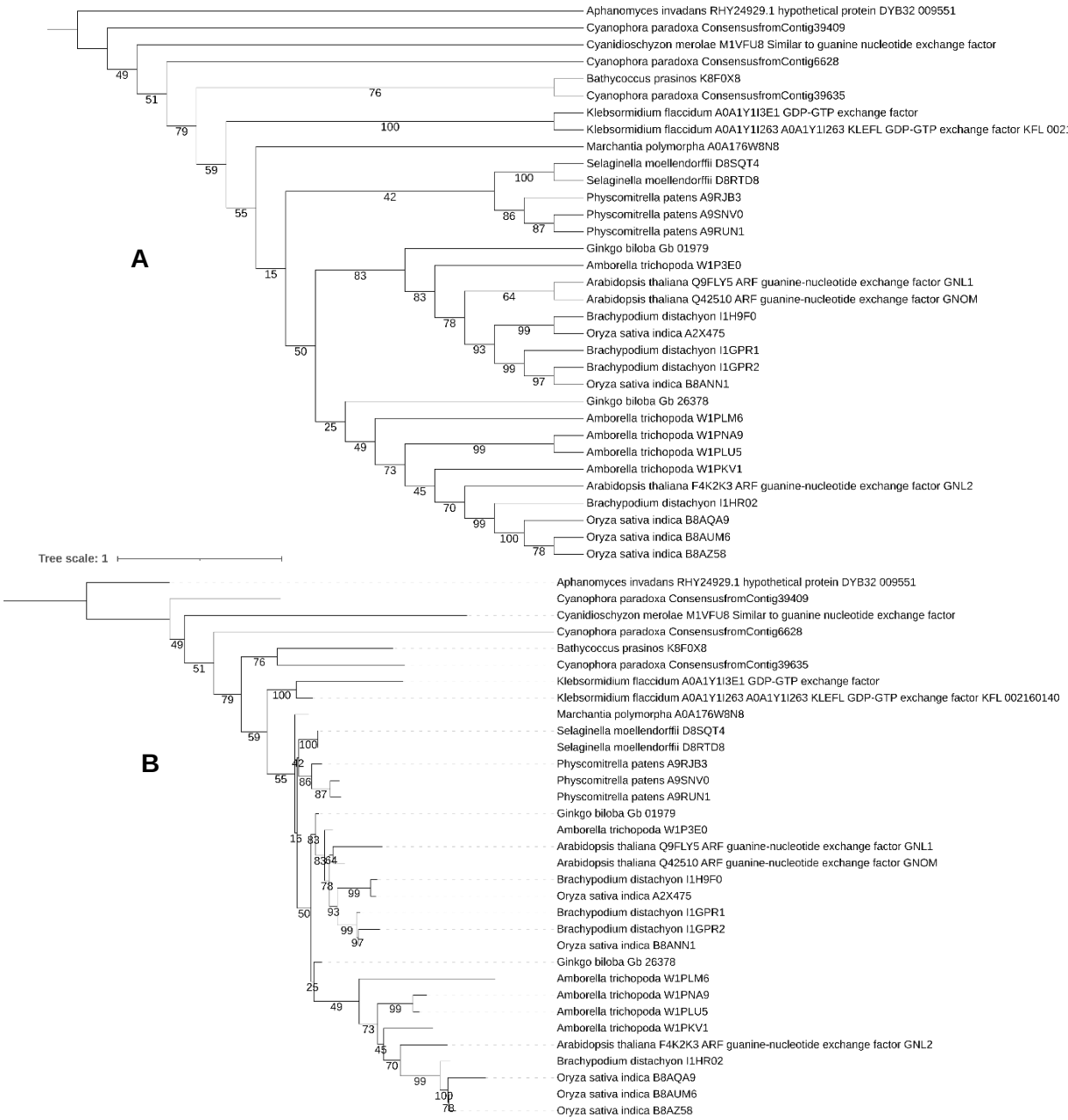
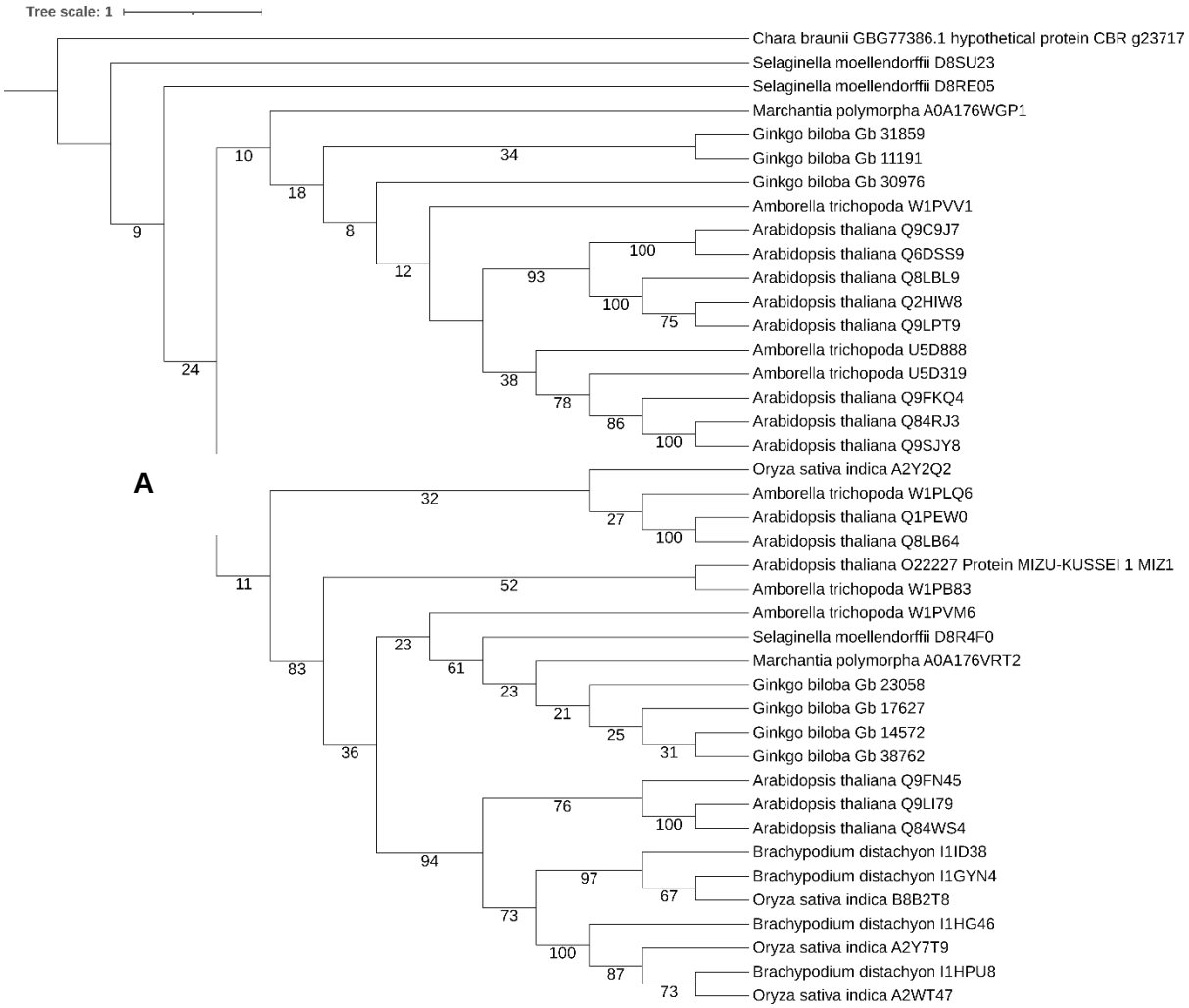


Figure S8.22. Maximum likelihood gene tree for MIZU-KUSSEI 2 (MIZ2, Uniprot ID: Q42510) without (A) and with branch lengths (B).


Tree scale: 1
Chara braunii GBG77386.1 hypothetical protein CBR g23717
Selaginella moellendorffii D8SU23
Selaginella moellendorffii D8RE05
Marchantia polymorpha A0A176WGP1
Ginkgo biloba Gb 31859
Ginkgo biloba Gb 11191
Ginkgo biloba Gb 30976
Amborella trichopoda W1PVV1
Arabidopsis thaliana Q9C9J7
Arabidopsis thaliana Q6DSS9
Arabidopsis thaliana Q8LBL9
Arabidopsis thaliana Q2HIW8
Arabidopsis thaliana Q9LPT9
Amborella trichopoda U5D888
Amborella trichopoda U5D319
Arabidopsis thaliana Q9FKQ4
Arabidopsis thaliana Q84RJ3
Arabidopsis thaliana Q9SJY8
A
Oryza sativa indica A2Y2Q2
Amborella trichopoda W1PLQ6
Arabidopsis thaliana Q1PEW0
Arabidopsis thaliana Q8LB64
Arabidopsis thaliana O22227 Protein MIZU-KUSSEI 1 MIZ1
Amborella trichopoda W1PB83
Amborella trichopoda W1PVM6
Selaginella moellendorffii D8R4F0
Marchantia polymorpha A0A176VRT2
Ginkgo biloba Gb 23058
Ginkgo biloba Gb 17627
Ginkgo biloba Gb 14572
Ginkgo biloba Gb 38762
Arabidopsis thaliana Q9FN45
Arabidopsis thaliana Q9LI79
Arabidopsis thaliana Q84WS4
Brachypodium distachyon I1ID38
Brachypodium distachyon I1GYN4
Oryza sativa indica B8B2T8
Brachypodium distachyon I1HG46
Oryza sativa indica A2Y7T9
Brachypodium distachyon I1HPU8
Oryza sativa indica A2WT47

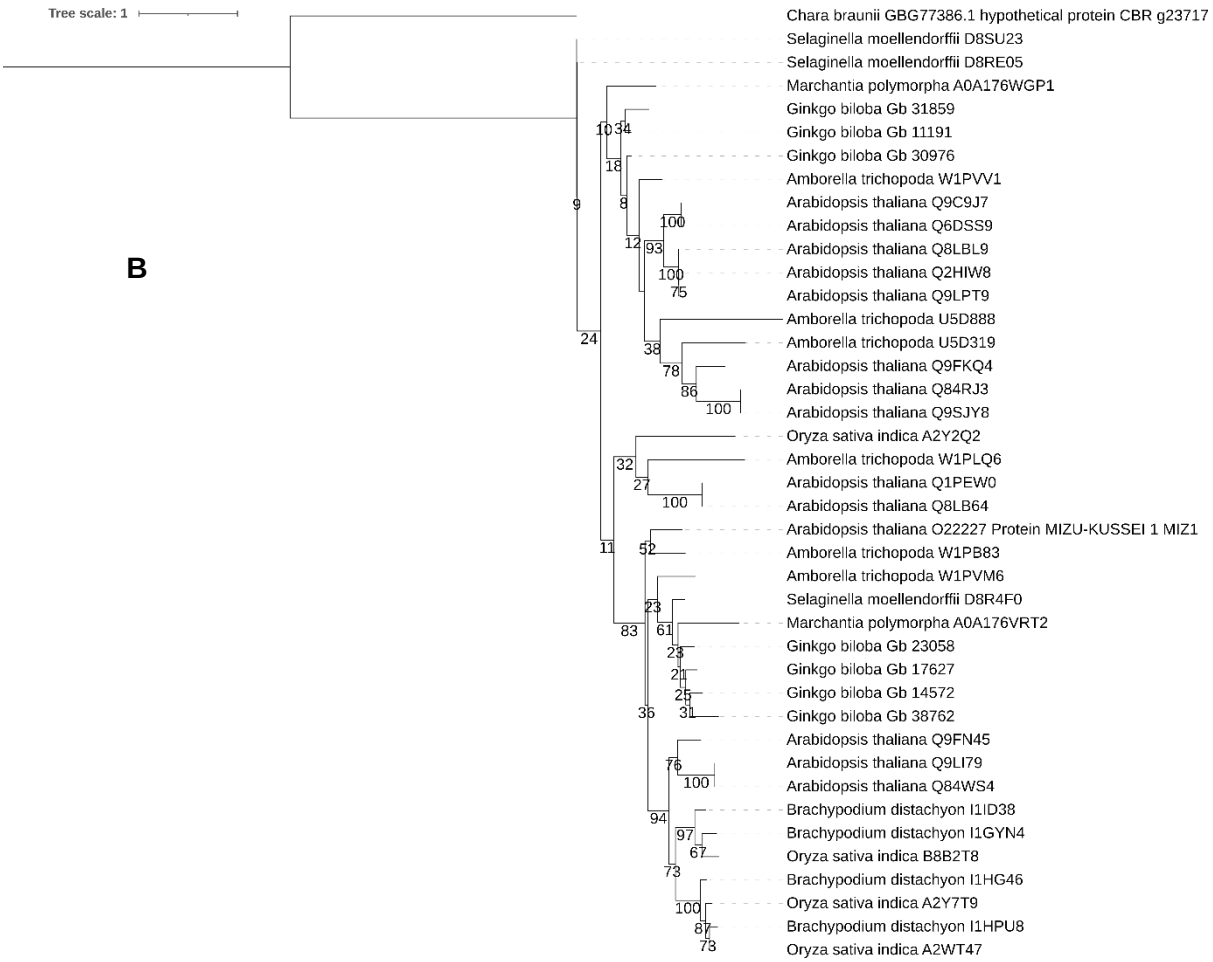


Figure S8.23. Maximum likelihood gene tree for MIZU-KUSSEI 1 (MIZ1, Uniprot ID: O22227) without (A) and with branch lengths (B).

Appendix 4: related to work in Chapter 4

Table S8.6. A list of the Archaeplastida species in the genomic dataset, a four letter species code, a drought adaptation status, the literature for any drought status and the cultivation status. For the drought adaptation status, plants were characterised as drought adapted, drought sensitive or for plants, with no clear response drought response uncertain, undefined. Briefly this characterisation was completed with a literature search to create a collective definition of a drought adapted and drought sensitive plant. The literature search queried the literature with the name of each species and a relevant drought search term (as detailed in Table 4.1). These databases included PubMed, Web of Science and Google Scholar. An example for a search identifying a drought adapted species would be the query for *Kalanchoë fedtschenkoi* with the resulting text reporting 'Our findings (from the genome of *Kalanchoë fedtschenkoi*) hold tremendous potential to accelerate the genetic improvement of crops for enhanced drought avoidance and sustainable production of food and bioenergy on marginal lands' (Yang *et al.*, 2017c). An example of a search identifying a drought sensitive species would be the query for *Gnetum montanum* with the resulting text reporting 'Extant species of Gnetum are unusual among gymnosperms in being restricted to warm, mesic habitats' (Wan et al., 2018a). Species were said to be undefined if a drought term could not be clearly associated with a particular species in a paper in the literature. Additionally species were said to be undefined if there had been no attempts to characterise their response to drought. For the cultivation status, plant were defined as wild, cultured or cultivated. These were defined based on evidence from the genome papers of these species. These genome papers are listed in Appendix 2.1.

Species	**Species Code**	**Drought status**	**Source**	**Cultivation**
Galdieria sulphuraria	gsul	Drought response uncertain	NA	Cultivated
Galdieria phlegrea	gphl	Drought response uncertain	NA	Wild
Cyanidioschyzon merolae	cmer	Drought response uncertain	NA	Cultured
Porphyridium purpureum	ppur	Drought response uncertain	NA	Cultured
Cyanophora paradoxa	cpar	Drought response uncertain	NA	Cultured

Bathycoccus prasinos	bpra	Drought response uncertain	NA	Wild
Ostreococcus lucimarinus	oluc	Drought response uncertain	NA	Cultured
Ostreococcus tauri	otau	Drought response uncertain	NA	Cultured
Micromonas commoda	mcom	Drought response uncertain	NA	Cultured
Micromonas pusilla	mpus	Drought response uncertain	NA	Cultured
Auxenochlorella protothecoides	apro	Drought response uncertain	NA	Cultured
Chlorella variabilis	cvar	Drought response uncertain	NA	Cultured
Picochlorum sp.	psp.	Drought response uncertain	NA	Cultivated
Coccomyxa subellipsoidea	csub	Drought response uncertain	NA	Cultured
Chromochloris zofingiensis	czof	Drought response uncertain	NA	Cultured
Gonium pectorale	gpec	Drought response uncertain	NA	Cultured
Volvox carteri	vcar	Drought adaptation	(Jaenicke *et al.*, 1982)	Cultured
Chlamydomonas reinhardtii	crei	Drought response uncertain	NA	Cultured
Chlamydomonas eustigma	ceus	Drought response uncertain	NA	Wild
Klebsormidium flaccidum	kfla	Drought adaptation	(Holzinger *et al.*, 2015a)	Cultured
Marchantia polymorpha	mpol	Drought adaptation	(Bowman *et al.*, 2017)	Wild
Physcomitrella patens	ppat	Drought adaptation	(Li *et al.*, 2017b)	Wild
Selaginella moellendorffii	some	Drought sensitive	(Dinakar *et al.*, 2013)	Wild

Ginkgo biloba	gbil	Drought adaptation	(Shan-An *et al.*, 1997)	Wild
Picea abies	pabi	Drought adaptation	(Kohler *et al.*, 2010)	Cultivated
Gnetum montanum	gmon	Drought sensitive	(Wan *et al.*, 2018a)	Wild
Amborella trichopoda	atri	Drought response uncertain	NA	Wild
Spirodela polyrhiza	spol	Drought sensitive	(Cheng, 2011)	Wild
Zostera marina	zmar	Drought sensitive	(Leuschner *et al.*, 2017)	Wild
Zostera muelleri	zmue	Drought sensitive	(Lee *et al.*, 2016a)	Wild
Dioscorea rotundata	drot	Drought adaptation	(Loko *et al.*, 2015)	Cultivated
Apostasia shenzhenica	ashe	Drought adaptation	(Zhang *et al.*, 2017a)	Wild
Dendrobium catenatum	dcat	Drought adaptation	(Wan *et al.*, 2018b)	Wild
Phalaenopsis equestris	pequ	Drought adaptation	(Wan *et al.*, 2018b)	Cultivated
Asparagus officinalis	aoff	Drought adaptation	(Whitmore, 2000)	Cultivated
Phoenix dactylifera	pdac	Drought adaptation	(Safronov *et al.*, 2017)	Cultivated
Elaeis guineensis	egui	Drought adaptation	(Murugesan *et al.*, 2017)	Cultivated
Cocos nucifera	cnus	Drought adaptation	(Gomes *et al.*, 2007)	Cultivated
Ananas comosus	acom	Drought adaptation	(Ming *et al.*, 2015)	Cultivated
Oryza brachyantha	obra	Drought response uncertain	NA	Wild
Oryza punctata	opun	Drought response uncertain	NA	Wild

Oryza glumipatula	oglu	Drought response uncertain	NA	Wild
Oryza rufipogon	oruf	Drought adaptation	(Biaolin *et al.*, 2010)	Wild
Oryza meridionalis	omer	Drought adaptation	(Vaughan *et al.*, 2003)	Wild
Oryza barthii	obar	Drought response uncertain	NA	Wild
Oryza glaberrima	ogla	Drought adaptation	(Bimpong *et al.*, 2011)	Cultivated
Oryza nivara	oniv	Drought response uncertain	NA	Wild
Oryza sativa Indica	osai	Drought sensitive	(Wei *et al.*, 2016)	Cultivated
Oryza sativa Japonica	osaj	Drought sensitive	(Wei *et al.*, 2016)	Cultivated
Leersia perrieri	lper	Drought response uncertain	NA	Wild
Phyllostachys edulis	pedu	Drought adaptation	(Wu *et al.*, 2018b)	Wild
Brachypodium distachyon	bdis	Drought adaptation	(Bertolini *et al.*, 2013)	Cultivated
Hordeum vulgare	hvul	Drought response uncertain	NA	Cultivated
Aegilops tauschii	atau	Drought response uncertain	NA	Cultivated
Triticum urartu	tura	Drought response uncertain	NA	Cultivated
Triticum aestivum	taes	Drought sensitive	(He *et al.*, 2009)	Cultivated
Triticum turgidum	ttur	Drought response uncertain	NA	Wild
Lolium perenne	lpee	Drought adaptation	(Cheplick *et al.*, 2000)	Cultivated
Echinochloa crus-galli	ecru	Drought response uncertain	NA	Wild

Setaria italica	sita	Drought adaptation	(Li *et al.*, 2014a)	Cultivated
Zea mays	zmay	Drought sensitive	(Agrama *et al.*, 1996)	Cultivated
Sorghum bicolor	sbic	Drought adaptation	(Abdel-Ghany *et al.*, 2020)	Cultivated
Eragrostis tef	etef	Drought adaptation	(Degu *et al.*, 2008)	Cultivated
Zoysia japonica	zjap	Drought adaptation	(Patton *et al.*, 2017)	Cultivated
Zoysia matrella	zmat	Drought adaptation	(Ntoulas *et al.*, 2012)	Cultivated
Zoysia pacifica	zpac	Drought adaptation	(Patton *et al.*, 2017)	Cultivated
Oropetium thomaeum	otho	Drought adaptation	(Vanburen *et al.*, 2015)	Wild
Musa itinerans	miti	Drought adaptation	(Wu *et al.*, 2018a; Kew Science, 2020b)	Wild
Musa balbisiana	mbal	Drought adaptation	(Nansamba *et al.*, 2020)	Wild
Musa acuminata	macu	Drought response uncertain	NA	Wild
Eschscholzia californica	ecal	Drought adaptation	(Wilts *et al.*, 2018)	Cultivated
Macleaya cordata	mcor	Drought response uncertain	NA	Wild
Nelumbo nucifera	nnuc	Drought response uncertain	NA	Cultivated
Kalanchoe fedtschenkoi	kfed	Drought adaptation	(Yang *et al.*, 2017c)	Cultivated
Rhodiola crenulata	rcre	Drought adaptation	(Zhang *et al.*, 2019b)	Wild
Vitis vinifera	vvin	Drought adaptation	(Gambetta *et al.*, 2020)	Cultivated
Lupinus angustifolius	lang	Drought adaptation	(Jensen *et al.*, 1990; Kalandyk *et al.*, 2017)	Cultivated

Arachis duranensis	adur	Drought adaptation	(Guimarães *et al.*, 2012)	Wild
Arachis ipaensis	aipa	Drought adaptation	(Azevedo Neto *et al.*, 2010)	Wild
Cajanus cajan	ccaj	Drought adaptation	(Varshney *et al.*, 2012)	Cultivated
Phaseolus angularis	pang	Drought adaptation	(Cortés *et al.*, 2013)	Cultivated
Phaseolus vulgaris	pvul	Drought response uncertain	NA	Cultivated
Vigna radiata	vrad	Drought adaptation	(Iseki *et al.*, 2018)	Cultivated
Glycine max	gmax	Drought sensitive	(Wang *et al.*, 2017a)	Cultivated
Glycine soja	gsoj	Drought adaptation	(Ji *et al.*, 2010)	Cultivated
Cicer arietinum	cari	Drought adaptation	(Varshney *et al.*, 2014)	Cultivated
Medicago truncatula	mtru	Drought response uncertain	NA	Cultivated
Trifolium pratense	tpra	Drought response uncertain	NA	Cultivated
Fragaria vesca	fves	Drought response uncertain	NA	Cultivated
Prunus avium	pavi	Drought response uncertain	NA	Cultivated
Prunus mume	pmum	Drought response uncertain	NA	Cultivated
Prunus persica	pper	Drought sensitive	(Eldem *et al.*, 2012)	Cultivated
Pyrus bretschneideri	pbre	Drought adaptation	(Cao *et al.*, 2018)	Cultivated
Pyrus communis	pcom	Drought adaptation	(Paudel *et al.*, 2019)	Cultivated
Malus domestica	mdom	Drought response uncertain	NA	Cultivated
Ziziphus jujuba	zjuj	Drought adaptation	(Cruz *et al.*, 2012)	Cultivated

Morus notabilis	mnot	Drought response uncertain	NA	Wild
Cucurbita maxima	cmax	Drought adaptation	(Yasar *et al.*, 2014)	Cultivated
Cucurbita moschata	cmos	Drought adaptation	(Cao *et al.*, 2017)	Cultivated
Citrullus lanatus	clan	Drought sensitive	(Zhang *et al.*, 2011)	Cultivated
Lagenaria siceraria	lsic	Drought adaptation	(Mashilo *et al.*, 2017)	Cultivated
Cucumis melo	cmel	Drought adaptation	(Kusvuran, 2012)	Cultivated
Cucumis sativus	csat	Drought sensitive	(Wang *et al.*, 2012a)	Cultivated
Populus trichocarpa	ptri	Drought adaptation	(Tang *et al.*, 2015)	Cultivated
Populus pruinosa	ppru	Drought adaptation	(Yang *et al.*, 2017b)	Wild
Linum usitatissimum	lusi	Drought sensitive	(Dash *et al.*, 2014)	Cultivated
Jatropha curcas	jcur	Drought adaptation	(Sapeta *et al.*, 2016)	Cultivated
Manihot esculenta	mesc	Drought adaptation	(Okogbenin *et al.*, 2013)	Cultivated
Hevea brasiliensis	hbra	Drought adaptation	(Kew Science, 2020a; Plants for a Future, 2020)	Cultivated
Ricinus communis	rcom	Drought response uncertain	NA	Cultivated
Cephalotus follicularis	cfol	Drought response uncertain	NA	Wild
Punica granatum	pgra	Drought adaptation	(Catola *et al.*, 2016)	Cultivated
Eucalyptus grandis	egra	Drought response uncertain	NA	Wild
Dimocarpus longan	dlon	Drought adaptation	(Wiriya-Alongkorn *et al.*, 2013)	Cultivated

Atalantia buxifolia	abux	Drought adaptation	(Newton *et al.*, 1989)	Wild
Citrus clementina	ccle	Drought response uncertain	NA	Cultivated
Citrus sinensis	csin	Drought response uncertain	NA	Cultivated
Citrus ichangensis	cich	Drought response uncertain	NA	Wild
Citrus medica	cmed	Drought response uncertain	NA	Cultivated
Carica papaya	cpap	Drought response uncertain	NA	Cultivated
Tarenaya hassleriana	thas	Drought adaptation	(Kocacinar, 2015)	Cultivated
Eutrema salsugineum	esal	Drought adaptation	(Yang *et al.*, 2013)	Wild
Thellungiella parvula	tpar	Drought adaptation	(Griffith *et al.*, 2007)	Wild
Brassica napus	bnap	Drought response uncertain	NA	Cultivated
Brassica oleracea	bole	Drought response uncertain	NA	Cultivated
Brassica rapa	brap	Drought response uncertain	NA	Cultivated
Sisymbrium irio	siri	Drought response uncertain	NA	Wild
Barbarea vulgaris	bvug	Drought response uncertain	NA	Wild
Capsella rubella	crub	Drought response uncertain	NA	Cultivated
Capsella grandiflora	cgra	Drought response uncertain	NA	Wild
Arabidopsis thaliana	atha	Drought sensitive	(Marín-de la Rosa *et al.*, 2019)	Cultivated
Arabidopsis lyrata	alyr	Drought response uncertain	NA	Wild

Theobroma cacao	tcac	Drought sensitive	(Bae *et al.*, 2008)	Cultivated
Corchorus capsularis	ccap	Drought sensitive	(Yang *et al.*, 2017d)	Cultivated
Corchorus olitorius	coli	Drought adaptation	(Yang *et al.*, 2017d)	Cultivated
Durio zibethinus	dzib	Drought sensitive	(Wan Nazri *et al.*, 2014)	Cultivated
Gossypium arboreum	garb	Drought adaptation	(Maqbool *et al.*, 2009)	Cultivated
Gossypium hirsutum	ghir	Drought sensitive	(Li *et al.*, 2017a)	Cultivated
Gossypium raimondii	grai	Drought sensitive	(Chen *et al.*, 2013b)	Cultivated
Fagopyrum esculentum	fesc	Drought sensitive	(Jamwal *et al.*, 2015)	Cultivated
Dianthus caryophyllus	dcay	Drought adaptation	(Wan *et al.*, 2015)	Cultivated
Beta vulgaris	bvul	Drought sensitive	(Pidgeon *et al.*, 2006)	Cultivated
Spinacia oleracea	sole	Drought sensitive	(Schwab *et al.*, 1984)	Cultivated
Chenopodium quinoa	cqui	Drought adaptation	(Al-Naggar *et al.*, 2017)	Cultivated
Amaranthus hypochondriacus	ahyp	Drought adaptation	(Sunil *et al.*, 2014)	Cultivated
Camptotheca acuminata	cacu	Drought response uncertain	NA	Wild
Camellia sinensis	csie	Drought sensitive	(Liu *et al.*, 2016)	Cultivated
Actinidia chinensis	achi	Drought sensitive	(Mills *et al.*, 2009)	Cultivated
Lactuca sativa	lsat	Drought sensitive	(Kizil *et al.*, 2012)	Cultivated
Erigeron breviscapus	ebre	Drought response uncertain	NA	Cultivated
Helianthus annuus	hann	Drought adaptation	(Badouin *et al.*, 2017)	Cultivated

Panax ginseng	pgin	Drought response uncertain	NA	Cultivated
Daucus carota	dcar	Drought response uncertain	NA	Cultivated
Coffea canephora	ccan	Drought sensitive	(Cheserek *et al.*, 2012)	Cultivated
Calotropis gigantea	cgig	Drought adaptation	(Tezara *et al.*, 2011; Mutwakil *et al.*, 2017)	Wild
Fraxinus excelsior	fexc	Drought adaptation	(Dobrowolska *et al.*, 2011)	Wild
Olea europaea	oeur	Drought adaptation	(Sofo, 2011)	Cultivated
Boea hygrometrica	bhyg	Drought adaptation	(Xiao *et al.*, 2015)	Wild
Mimulus guttatus	mgut	Drought sensitive	(Hughes *et al.*, 2001)	Cultivated
Sesamum indicum	sind	Drought adaptation	(Golestani *et al.*, 2015; Dossa *et al.*, 2017)	Cultivated
Handroanthus impetiginosus	himp	Drought adaptation	(Dombroski *et al.*, 2014)	Wild
Genlisea aurea	gaur	Drought response uncertain	NA	Wild
Utricularia gibba	ugib	Drought response uncertain	NA	Wild
Ipomoea nil	inil	Drought response uncertain	NA	Cultivated
Ipomoea trifida	itri	Drought response uncertain	NA	Wild
Petunia inflata	pinf	Drought response uncertain	NA	Cultivated
Petunia axillaris	paxi	Drought response uncertain	NA	Cultivated
Nicotiana obtusifolia	nobt	Drought adaptation	(Su *et al.*, 2017)	Wild
Nicotiana sylvestris	nsyl	Drought response uncertain	NA	Cultivated

Nicotiana tabacum	ntab	Drought response uncertain	NA	Cultivated
Nicotiana tomentosiformis	ntom	Drought response uncertain	NA	Wild
Solanum lycopersicum	slyc	Drought sensitive	(Mishra *et al.*, 2016)	Cultivated
Solanum pennellii	spen	Drought adaptation	(Egea *et al.*, 2018)	Wild
Solanum pimpinellifolium	spim	Drought response uncertain	NA	Wild
Solanum tuberosum	stub	Drought sensitive	(Boguszewska-Mańkowska *et al.*, 2018)	Cultivated
Capsicum annuum	cann	Drought adaptation	(Sahitya *et al.*, 2019)	Cultivated
Capiscum baccatum	cbac	Drought response uncertain	NA	Cultivated
Capiscum chinense	cchi	Drought response uncertain	NA	Cultivated

Appendix 5: related to work in Chapter 5

Appendix 5.1: Schematic representation of T-DNA insertions for all loss of function mutants. Black bars represents exons whilst the black line represent introns. The white boxes represent the translational start and stop regions. The white triangle represents the T-DNA insert and the arrow highlights the direction of the insert. Information about gene models and tDNA inserts sourced from The Arabidopsis Information Resource (Berardini *et al.*, 2015). For each *A. thaliana* gene, the HG it corresponds with is in brackets and two SALK lines are listed for each gene.

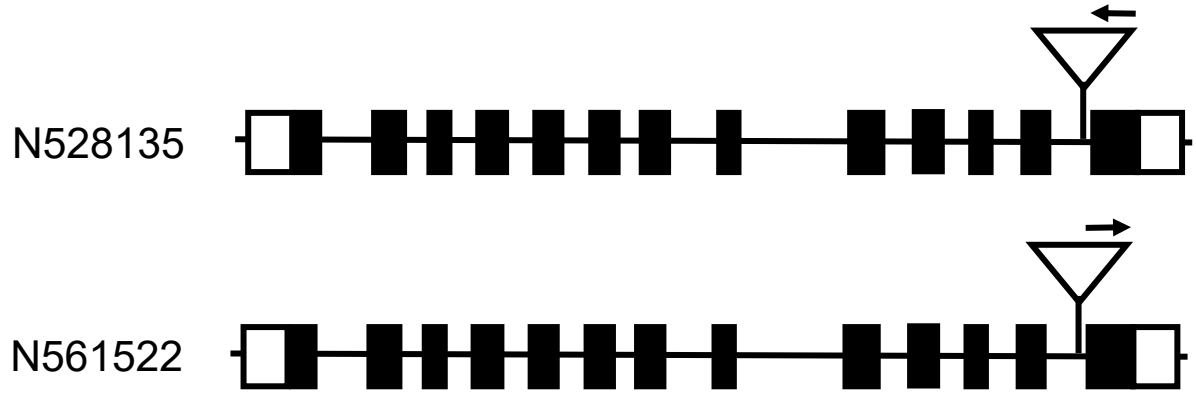


Figure S8.24. Schematic diagram demonstrating the location of the T-DNA insertion for two mutant lines of At2g26300 (HG_72).

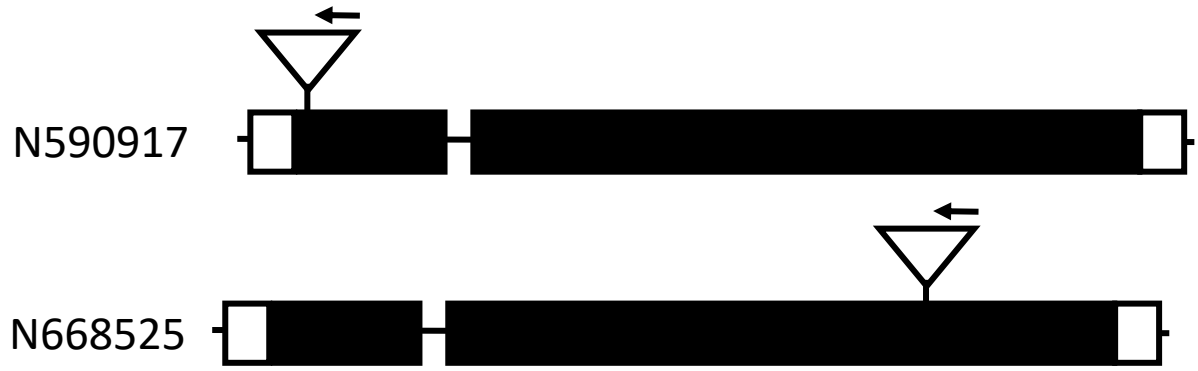


Figure S8.25. Schematic diagram demonstrating the location of the T-DNA insertion for two mutant lines of At5g37830 (HG_72).

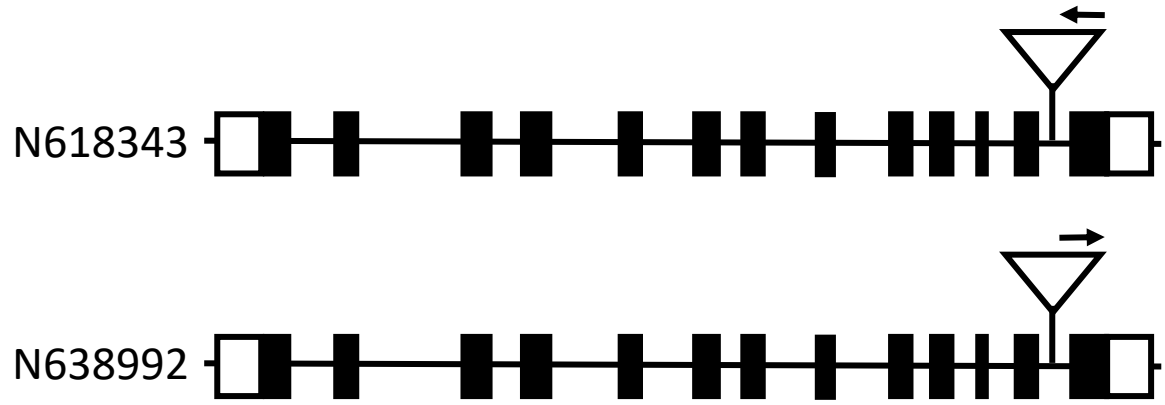


Figure S8.26. Schematic diagram demonstrating the location of the T-DNA insertion for two mutant lines of At5g37850 (HG_72).

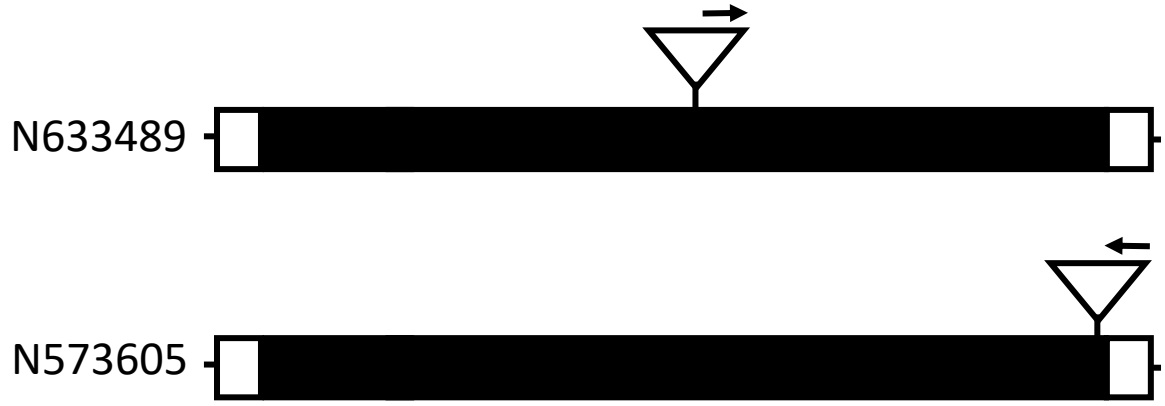


Figure S8.27. Schematic diagram demonstrating the location of the T-DNA insertion for two mutant lines of At4g16515 (HG_10098).

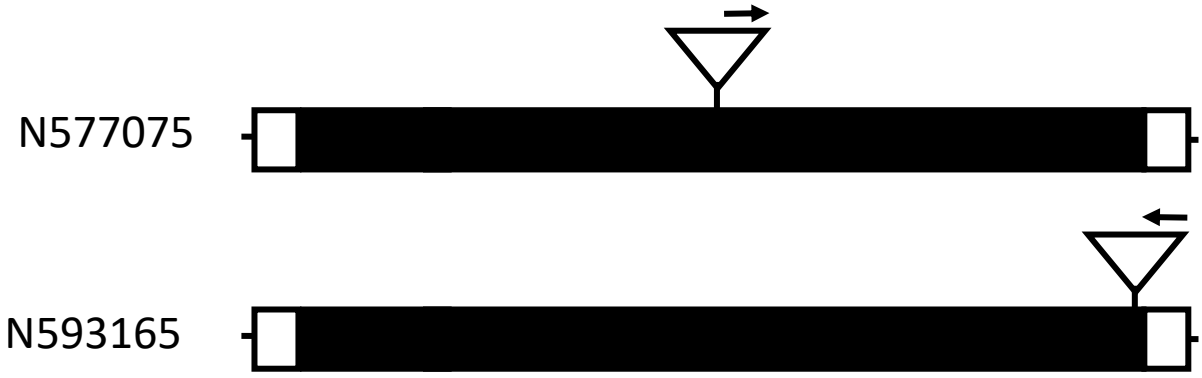


Figure S8.28. Schematic diagram demonstrating the location of the T-DNA insertion for two mutant lines of At4g16530 (HG_10098).

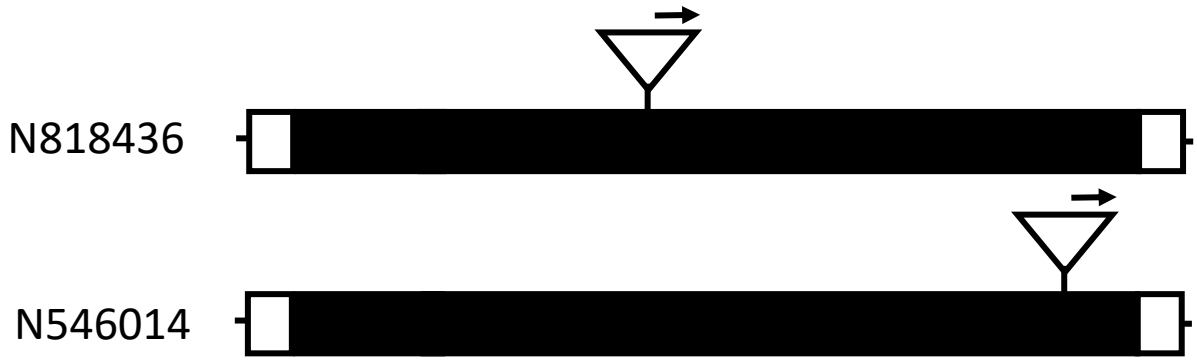


Figure S8.29. Schematic diagram demonstrating the location of the T-DNA insertion for two mutant lines of At5g48890 (HG_2909).

Figure S8.30. Schematic diagram demonstrating the location of the T-DNA insertion for two mutant lines of At1g10460 (HG_5775).

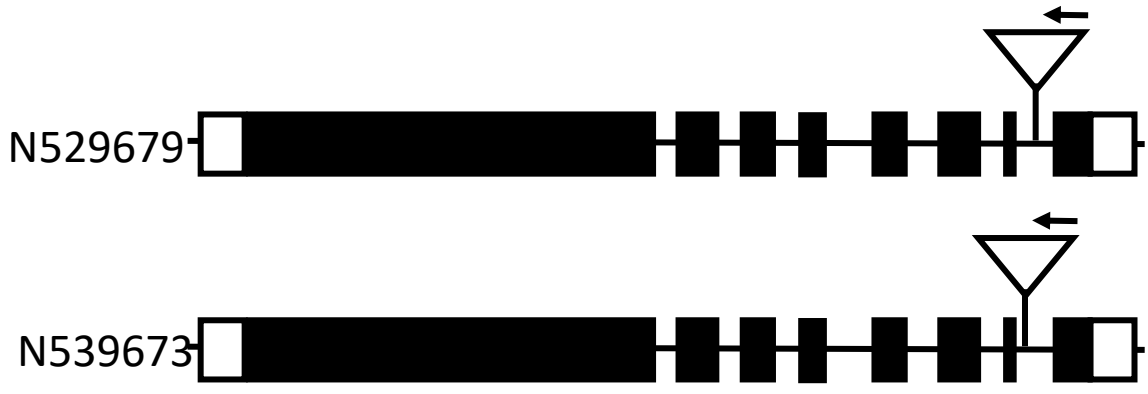


Figure S8.31. Schematic diagram demonstrating the location of the T-DNA insertion for two mutant lines of At5g25100 (HG_7522).

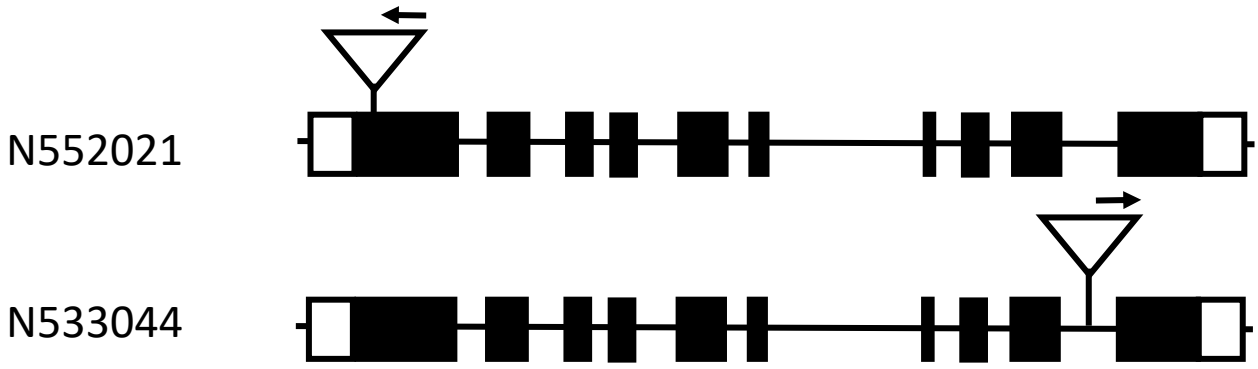


Figure S8.32. Schematic diagram demonstrating the location of the T-DNA insertion for two mutant lines of At4g09340 (HG_9215).

Appendix 5.2: Gel confirmation of tDNA mutant lines for all mutants used in the preliminary drought experiment described in 5.47 (N529679 (At5g25100), N546014 (At5g48890), N572453 (At1g10460), N593165 (At4g16530), N633489 (At4g16515)). For each tDNA mutant line, multiple plants were grown and confirmed. This is denoted by the letter after the line ID (e.g. N633489A). For each mutant lines, three primer combinations were run. 1: Left border and gene specific forward primer, 2: Left border and gene specific reverse primer and 3: Gene specific forward and reverse primers. For gene specific primers, see Table 5.1 and for left border primers, see Table 5.3 in the main text.

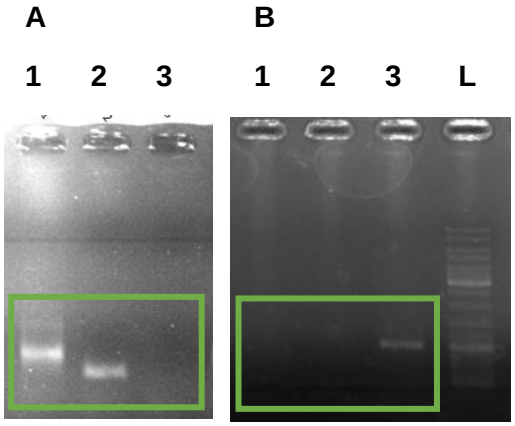


Figure S8.33. Example PCR of successful confirmation of a mutant line (A) and wild type *A. thaliana*.

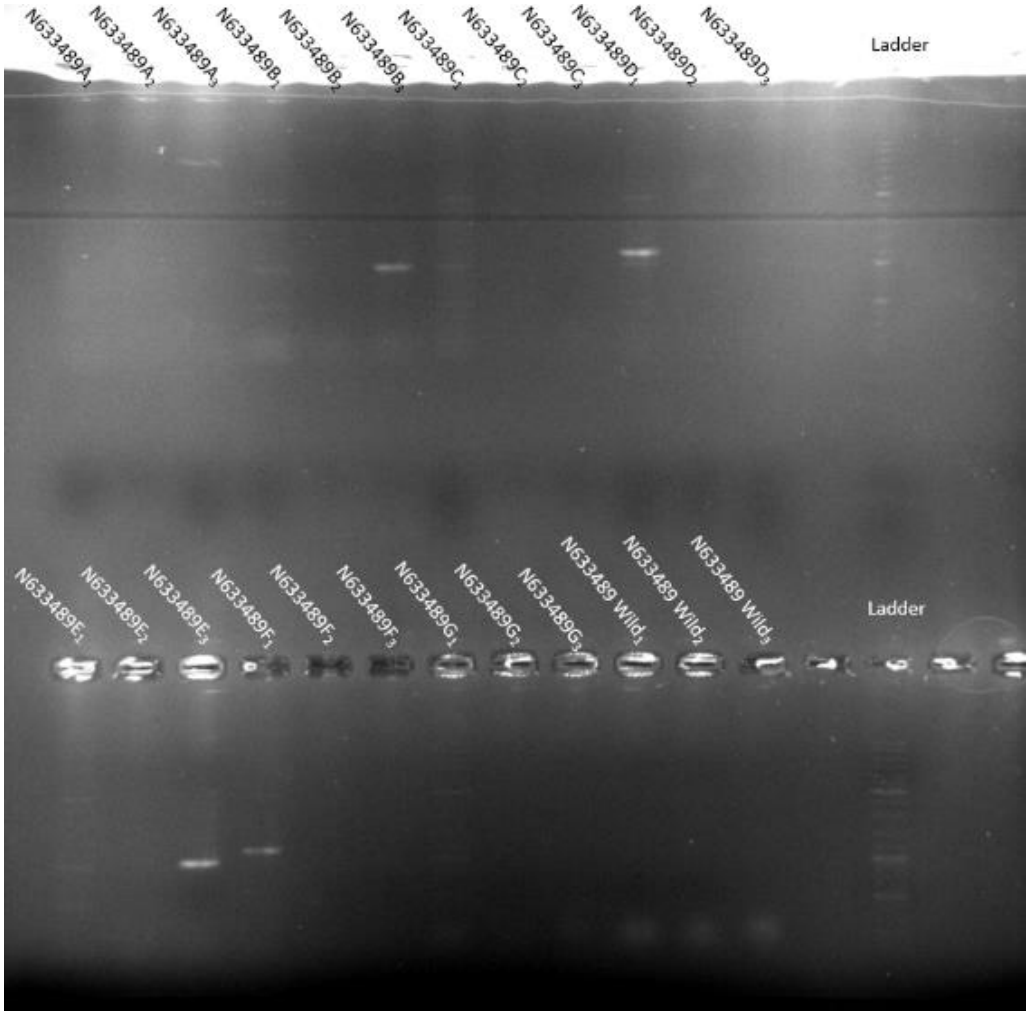


Figure S8.34. PCR for mutant line N633489 associated with At4g16515. Plants C, D and F were identified as homozygous mutants and chosen for analysis.

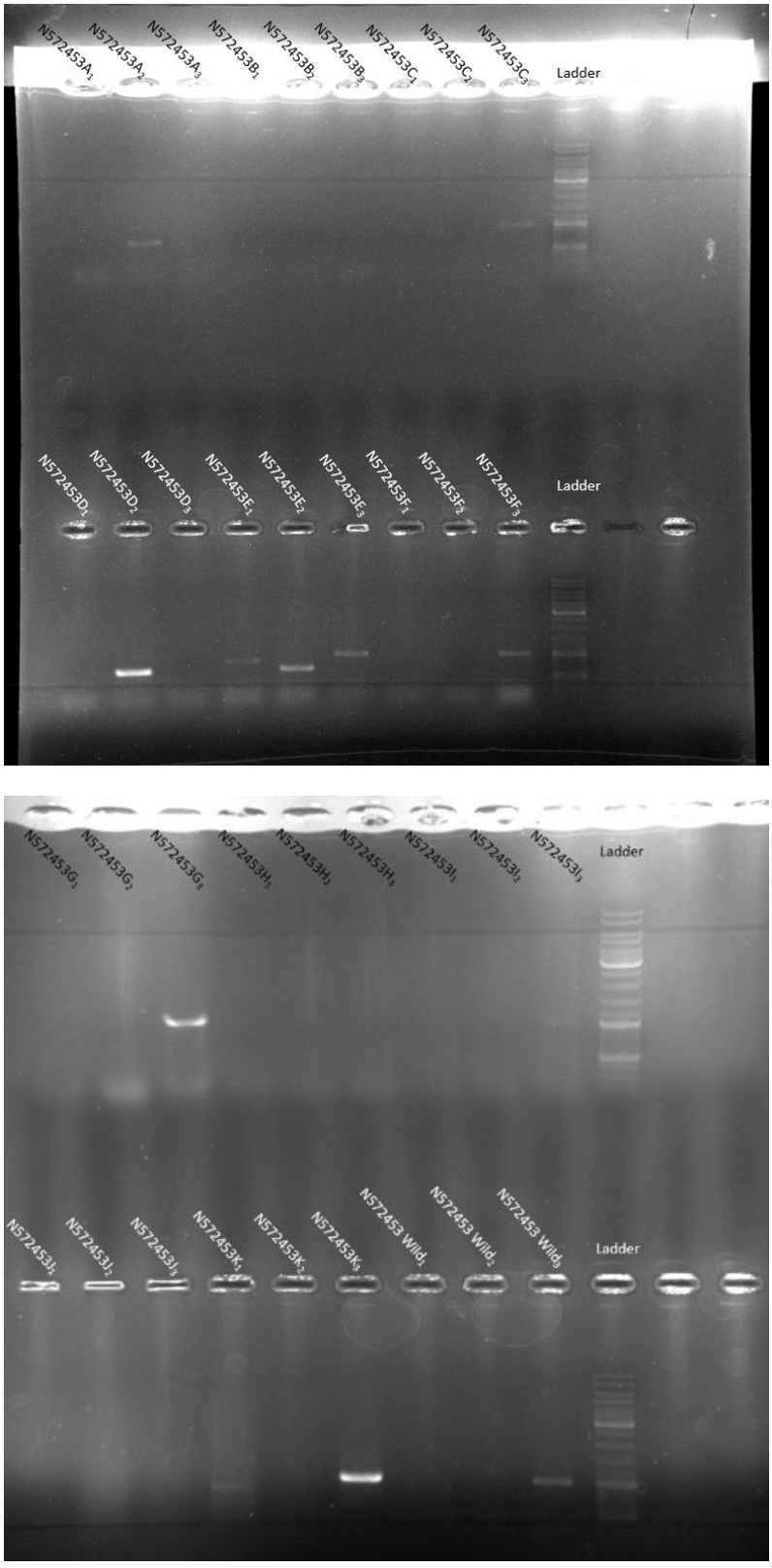


Figure S8.35. PCR for mutant line N572453 associated with At1g10460 for plants A-F (A) and G-K (B). Plants A and D were identified as homozygous mutants and chosen for analysis.

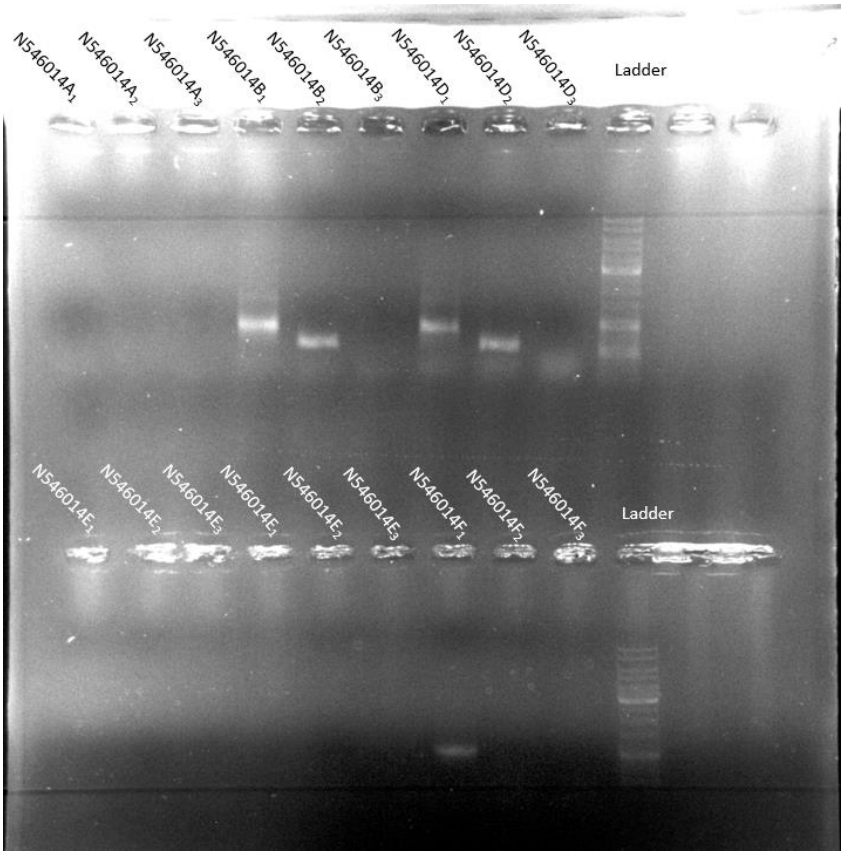


Figure S8.36. PCR for mutant line N546014 associated with At5g48890 for plants A-F. Plants A and D were identified as homozygous mutants and chosen for analysis.

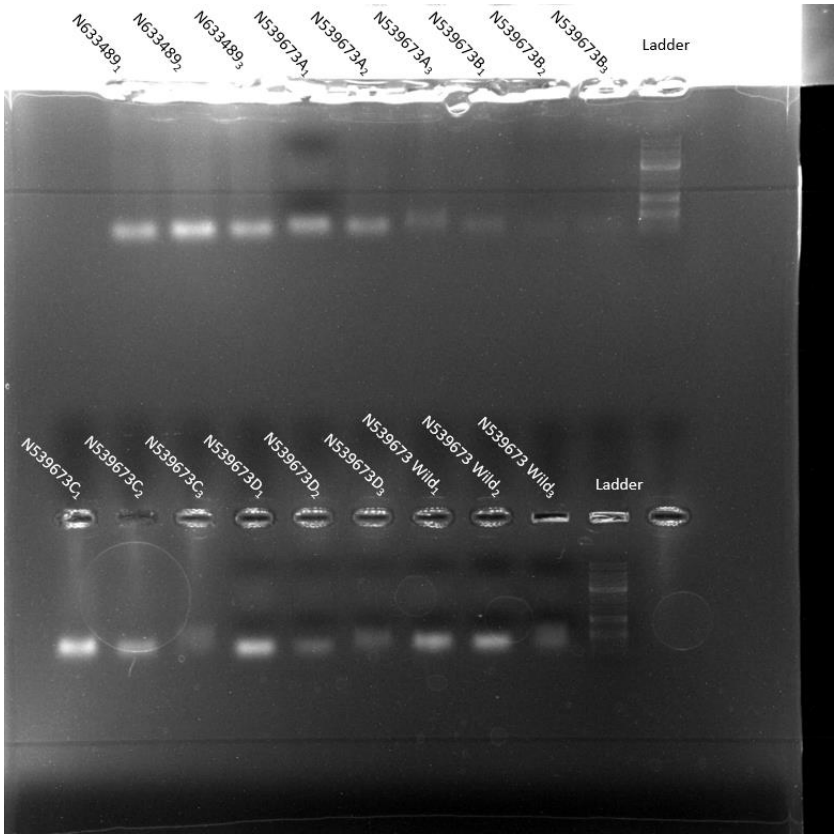


Figure S8.37. PCR for mutant line N633489 associated with At1g146515 for plants A and mutant line N539673 associated with At5g25100. Plants N539673 C and D were identified as homozygous mutants and chosen for analysis.

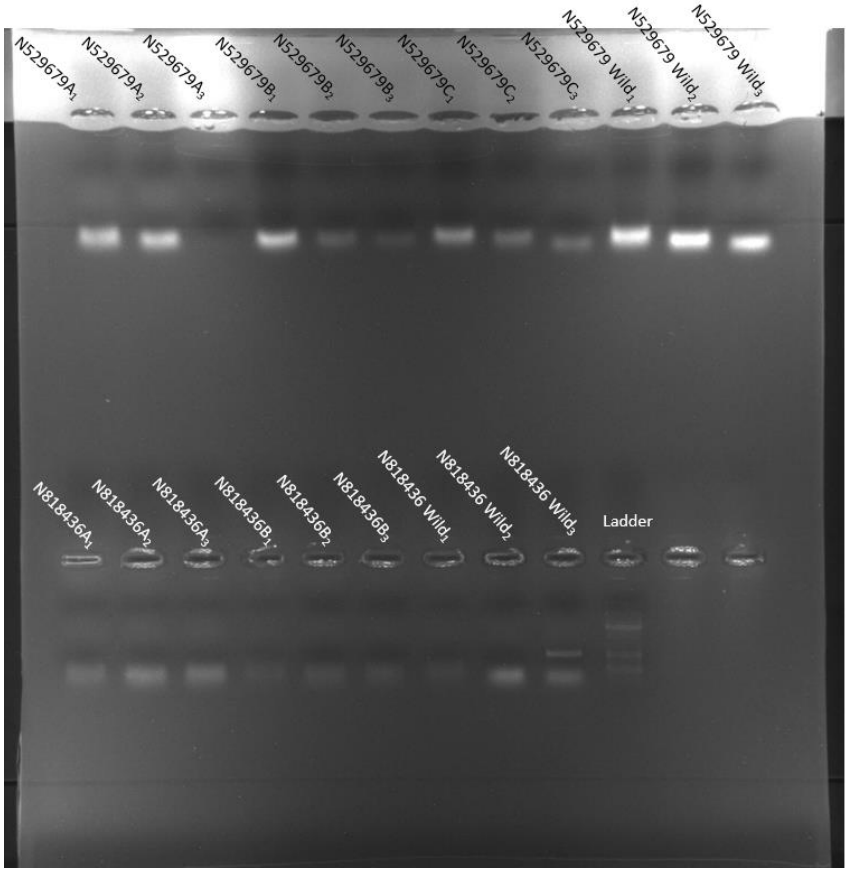


Figure S8.38. PCR for mutant line N529679 associated with AT5G25100 for plants A and mutant line N818436 associated with AT5G48890. Plants N529679 A and N818436 B were identified as homozygous mutants and chosen for analysis.

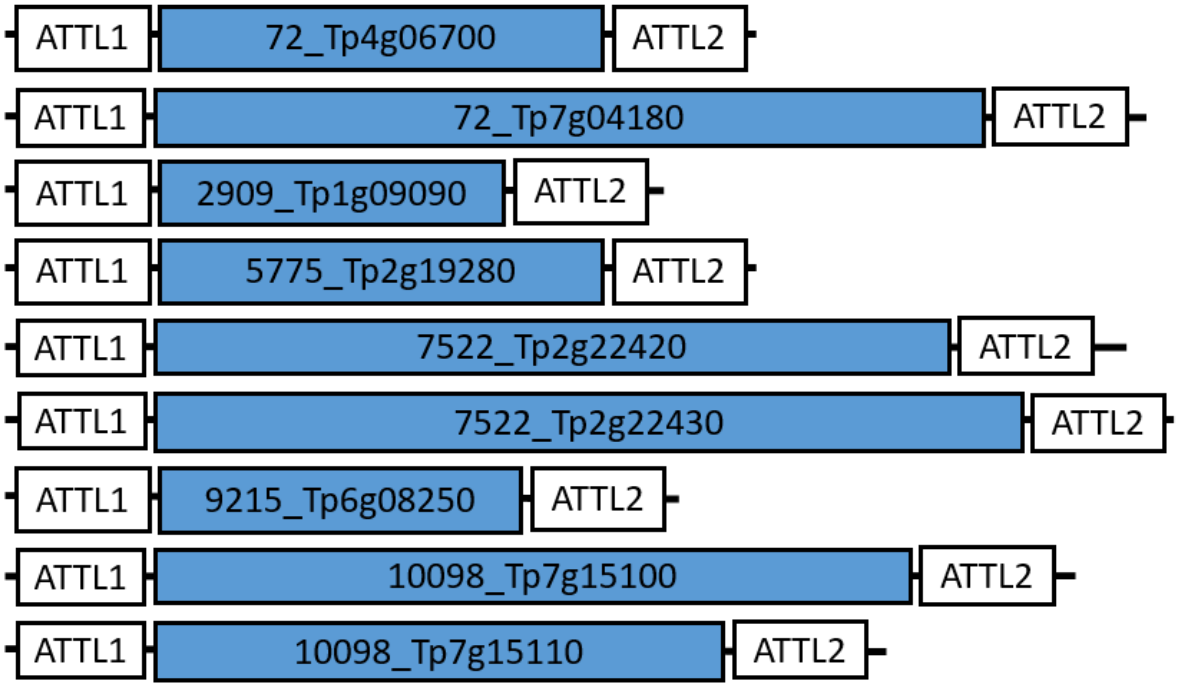


Figure S8.39. Schematic representation of the gene constructs for expression in A. thaliana. ATTL1 and ATTL2 sites were inserted for recombination during gateway cloning. The candidate drought gene is in between these two sites. The HG that each candidate gene relates to is placed in the ID (e.g. 72_Tp4g06700 is found in HG_72). Constructs were used to transform *A. thaliana*.

Appendix 6: Supplementary data

Appendix 6.1 Supplementary Data for Chapter 2

Supplementary Data 2.1. Genomes used in this study, species code, BUSCO scores and phylogenetic relationships of species. Related to Methods: Compiling genomic dataset.

Supplementary Data 2.2. Number of each evolutionary distinct class of Homology Groups. Related to Methods: Phylogenetically Aware Parsing Script.

Supplementary Data 2.3. Assessing impact of altering granularity score. Related to Methods: Homology assignment. A) Comparison with Catarino et al. Green bars indicate the origin of transcription factors identified in Catarino et al. For each granularity score (1.2, 2, 4, 6), an output was created. The values in brackets indicates the number of HGs for each transcription factor. B) Comparison with Banks et al. Analysis conducted using criteria specified in Banks et al for each granularity score.

Supplementary Data 2.4. GO analysis of Homology Groups. Related to Methods: Functional annotation A) Analysis for Novel Core HGs B) Analysis of Novel HGs C) Analysis of Ancestral HGs D) Analysis of Ancestral Core HGs E) Analysis of Lost HGs GO analysis. For GO analysis, *Arabidopsis thaliana* was used as an extant representative to assign biological functions. For Monocot, *Brachypodium distachyon* (core novelties) and *Oryza sativa* (novelties) genes were used to assign functions.

Supplementary Data 2.5. BLASTP of Novel Core HGs to validate their identification. Related to Methods: Novel Core HG validation.

Supplementary Data 2.6. GO analysis of HGT HGs. Related to Methods: Inferring Horizontal Gene Transfer.

Supplementary Data 2.7. Function of all Novel Core HGs with sources. Related to Methods: Functional annotation.

Supplementary Data 2.8. BLASTP of protein coding genes of the genome of the charophyte, *Chara braunii* and first two fern genomes, *Azolla filiculoides* and *Salvinia cucullata.* Related to Methods: Novel Core HG validation.

Supplementary Data 2.9. Processed data for 208 eukaryote genomes from this thesis including BLASTP output, MCL analysis and all Homology Groups. Additionally included in this folder are the scripts used to reproduce the results presented in this chapter.

Appendix 6.2 Supplementary Data for Chapter 3

Supplementary Data 3.1. Occupancy of genes linked to roots, vascular tissue and stomata across all species within the genomic dataset.

Supplementary Data 3.2. Protein sequences, trimmed alignments and treefiles for all genes in the study.

Supplementary Data 3.3. Outputs from blast queries against two fern genomes to confirm presence or absence in the LCA of Euphyllophyta.

Appendix 6.3 Supplementary Data for Chapter 4

Supplementary Data 4.1. Sequence data used to build species tree.

Supplementary Data 4.2. Newick file of the species tree used for Bayesian approach to ancestral state reconstruction.

Appendix 6.4. Supplementary Data for Chapter 5

Supplementary Data 5.1. Pipeline script used to query the genomic dataset in relation to the occupancy of drought adaptations.

Supplementary Data 5.2. Gene occupancy for each queries of the genomic dataset, revealing 238 HGs differentially retained between drought tolerant and sensitive species.

Supplementary Data 5.3. Gene occupancy for each queries of the genomic dataset, after specifying presence in the drought tolerant *Thellungiella parvula* and absence in the drought sensitive *Arabidopsis thaliana* (50 HGs).

Supplementary Data 5.4. Protein domain analysis for 50 HGs.

Supplementary Data 5.5. 42 genes of HG 72 and syntenic blocks of genes in *A. thaliana*.