Transcriptional regulation of seasonal desiccation tolerance in the fronds and rhizome of the fern *Anemia Caffrorum*.

Farrah Dilshaad Khan

KHNFAR025

Thesis presented for the degree

DOCTOR OF PHILOSOPHY

In the department of Molecular and Cell Biology

February 2023

Supervisor: Professor Jill Farrant, Dr. Suhail Rafudeen

Co-supervisors: Professor Henk Hilhorst

Honorary supervisor: Dr. Rafe Lyall

I, Farrah Khan, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature :

Date : 10/01/2023

ABSTRACT

Agriculture in drought-prone regions of the world is at increasing risk as climate change causes more intense and extended droughts. Desiccation tolerant plants are capable of surviving long periods of severe water loss (> 95% of cellular water) and are therefore prominent models for the production of climate smart crops. The fern *Anemia caffrorum* is a particularly interesting model as it is capable of seasonal desiccation tolerance. In winter, when water availability is high, desiccation sensitive (DS) fronds are produced by the rhizome. In summer, when water availability is low, desiccation tolerant (DT) fronds are produced. Seasonal desiccation tolerance conferred to crops would allow for the maximization of food yields when water is available and prioritization towards survival when water is scarce. In this study transcriptional regulation of *A. caffrorum* desiccation tolerance in the fronds and rhizomes was explored. In particular, rhizomes were of interest as they have been postulated as the regulator of frond phenotype (DS/DT). Fronds and rhizomes were subjected to desiccation and tissues of different water contents were collected. Tissue from the fronds of both phenotypes (DS/DT) at full turgor (approximately 100% RWC), 55% RWC, 30% RWC, 10% RWC and 24-hours post rehydration were chosen. Tissue from the rhizomes producing DS fronds (winter rhizomes) and those producing DT fronds (summer rhizomes) was collected at 100% RWC and 30% RWC. Methods for isolating total RNA from all tissue types, of sufficient purity and yield, for next generation sequencing (NGS) experimentation were established. These samples were used to conduct short- and long-read sequencing for transcriptome assembly and gene expression studies. Transcriptome assembly using long read data outperformed *de novo* short read assembly as it resulted in greater rates of annotation and yielded a less fragmented assembly. The transcriptome was determined to be of a high-quality and was used to perform differential expression (DE) studies. DE studies showed that DS fronds had initiated senescent processes by 55% RWC which was only realised after rehydration. DT fronds were able to survive a desiccation event through several constitutive and inducible effects that were more reminiscent of desiccation tolerance in angiosperms than that of bryophytes. However, a relatively low transcriptional response compared to angiosperm desiccation tolerant plants, was a feature of bryophytic-type desiccation tolerance. Therefore *A. caffrorum*, like other ferns, demonstrated a mixed model of desiccation tolerance. DE studies of the rhizome provided several insights into canonical mechanisms of desiccation tolerance of the underground organ, alternate seasonal strategies, the role of the organ in frond phenotype regulation and cross-organ dynamics. Firstly, at 30% RWC rhizomes were postulated to be fully prepared for desiccation

and achieved this status through many common features observed in angiosperm resurrection plants. The strong presence of chloroplastic protective proteins in this response highlighted the likely transport of such transcripts to newly emerging fronds. In the reverse direction, rhizomes appeared to acquire sucrose and transcripts required for the recruitment of endophytic fungi from the fronds. Finally, severely dehydrated rhizomes showed prominent upregulation of genes that are responsible for leaf senescence, leaf development and initiation of spore production which was highly correlated with ABA-regulation. Also, under the control of ABA, was the production of transcripts that ultimately initiate the production of orange, trichome-like scales which are a key feature of the DT phenotype. Not under the control of ABA regulation, but relevant to the DT phenotype, were genes that promote dwarfism and reduced photosynthetic potential in fronds. Altogether, the study has provided possible candidate genes that may govern the DT phenotype and has underscored the significant interplay between these organs in achieving seasonal tolerance.

ACKNOWLEDGMENTS

Jill, despite your many successes and immeasurable knowledge of plant physiology and desiccation tolerance, you remain humble and kind as a supervisor and scientist. You are an inspiration to me as an aspiring female scientist.

Rafe, simply put, this research would not have been possible without your supervision. Despite your advanced bioinformatic expertise, you were able to train me consistently and patiently as a scientist new to the field.

Suhail Rafudeen, your kindness towards me upon joining the research group was more important to me than you could understand. I appreciate you as a supervisor and confidant.

Henk, your genuine interest, desire to assist throughout my doctorate, and wide range of knowledge on each aspect of this work was highly appreciated. Your continued interest in science, even in retirement, was a great reminder of my passion for science during difficult phases in this research.

Henk, Jill, and Rafe, I am extremely fortunate to have completed this doctorate under your combined expert supervision. I look forward to following the contributions you will all make towards the production of climate smart crops.

To my parents, I continue to thank you for your unwavering support in all my studies to date. I will never forget our time together while I was writing up.

Leigh-anne Derry and Lynn Wambua, my original lab besties. You have brought me immeasurable joy throughout my life.

Stuart Granger, your ongoing patience, and comedic relief were crucial in keeping my sanity intact through the years,

To the Plant Stress Lab - Keren Cooper, thank you for ALWAYS, always, always helping where you can (which was everywhere) and always leaving me chuckling. Robyn Craythorne, I appreciate you taking me under your wing when I joined PSL and always going out of your way to help everyone.

The research presented here was funded by the National Research Foundation (NRF), thank you for your financial support over the years.

LIST OF ABBREVIATIONS

A. caffrorum	*Anemia caffrorum*
AGE	Agarose gel electrophoresis
CCS	Circular consensus sequencing
DS	Desiccation sensitive
DT	Desiccation tolerant
FGS	First generation sequencing
FLNC	Full length non-concatemer
HDT	Homoiochlorophyllous desiccation tolerant
HQ	High quality
Iso-Seq	Isoform sequencing
LRS	Long read sequencing
NGS	Next generation sequencing
PCR	Polymerase chain reaction
PDT	Poikilochlorophyllous desiccation tolerant
PSL	Plant Stress Group
RNA-Seq	RNA sequencing
ROS	Reactive oxygen species
SGS	Second generation sequencing
TGS	Third generation sequencing
UCT	University of Cape Town
VDT	Vegetative desiccation tolerance
RWC	Relative water content
RIN	RNA integrity number
FW	Fresh weight
TW	Turgid weight
DW	Dry weight

TABLE OF CONTENTS

Introduction : *Anemia caffrorum* as a model for climate smart crops

Introduction : *Anemia caffrorum* as a model for climate smart crops

1.1. Desiccation tolerant plants as models to improve drought-tolerant crops

The global population is projected to reach 9.7 billion people by 2050. To sustain our growing population, we must achieve an increase of 60% in agricultural production. Increasing the yields of crop production is essential because 75% of caloric intake is plant-based. This is complicated by the fact that crop production is vulnerable to climate change. Adverse effects of climate change on crop yield has been reported as early as 2014 by the Intergovernmental Panel on Climate Change (IPCC) (Lobell, Schlenker and Costa-Roberts, 2011; Wheeler and Von Braun, 2013). Southern Africa is prone to drought, heat waves and variable rainfall (Nhemachena and Hassan, 2007; Lyon, 2009; Masih *et al.*, 2014). Such events have long thwarted the efforts of small-hold farmers that rely on rain-fed agricultural systems (Runge *et al.*, 2004). As many as 85% of livelihoods in the Southern African region depend on the agricultural activities of small-hold farmers. Because such droughts have devastating socio-economic effects on the region, concerted efforts have been made to mitigate crop losses (Bryceson, 2002; Mangisoni, 2008; Baiphethi and Jacobs, 2009; Silberg *et al.*, 2017). Drought-tolerant cultivars of soybean, groundnut, cowpea and maize have been introduced with moderate success. Such cultivars are bred for shorter growing seasons, greater root length and/or improved resistance to loss of water during dry periods (CIMMYT, 2013; Fisher *et al.*, 2015). Most crops, including drought-tolerant cultivars, can withstand a loss of 10 - 60% of their water before they lose viability (Silva Artur *et al.*, 2019; Farrant and Hilhorst, 2022). As climate change intensifies, droughts are lasting longer and are accompanied by elevated temperatures. Under these conditions even the crops previously developed to survive drought cannot cope. To protect global food security from the effects of climate change we need to develop crops with greater drought resilience (Gaff and Oliver, 2013a; Giarola, Hou and Bartels, 2017). A suitable model for the development of such crops is a group of plants called resurrection plants. ‘Resurrection plants’ is a broad term for any plant capable of desiccation tolerance. Desiccation tolerance is the ability to tolerate dehydration and survive through a prolonged state of desiccation. In the desiccated state plants have lost up to 95% of their cellular water. Furthermore, when water becomes available again, normal metabolic function is reinstated. Theoretically, such plants could survive more intense and extended droughts that are rapidly becoming a reality (Alpert, 2005; Oliver *et al.*, 2020a). Work has been done on the

resurrection fern *Anemia caffrorum* (*A. caffrorum*) as a possible candidate for modelling climate smart crops. However, studying *A. caffrorum* presents a unique set of challenges because it is a fern. Firstly, because ferns hold little economic value, they are understudied and therefore limited fern molecular resources exist. More pertinent to this study, this is limited information on the strategies of desiccation tolerance in ferns. Instead, the few studies of fern desiccation tolerance have been contextualised relative to the highly studied angiosperm and bryophyte resurrection plants. Furthermore, fern desiccation tolerance strategies are typically summarized in the larger Pteridophyte family (Tuba, Protor and Csintalan, 1998; Oliver, Velten and Mishler, 2005; Tuba and Lichtenthaler, 2011; Dinakar and Bartels, 2013; Lyall and Gechev, 2020; Oliver *et al.*, 2020a; Marks *et al.*, 2021).

1.2. Evolution of desiccation tolerance.

Land plants are believed to have transitioned to a terrestrial existence from a freshwater origin. These early basal plants occurred in habitats which exposed them to frequent and rapid cycles of drying. As a result, plants such as bryophytes (liverworts, hornworts, and mosses) evolved a form of desiccation tolerance that allows them to withstand water loss, even if the desiccated state is reach within minutes. In this form of desiccation tolerance, there exist protective processes that are constitutively present and are coupled with an investment in cellular repair mechanisms (Oliver, Tuba and Mishler, 2000; Oliver, Velten and Mishler, 2005; Wood, 2007). (Oliver, Velten and Mishler, 2005) propose that these primitive plants must have been desiccation tolerant (DT) in both their vegetative tissues and reproductive structures to colonize land. As plants developed more sophisticated vascular systems and other morphological adaptations that improve water-holding capacity, there was a decrease in the occurrence of vegetative desiccation tolerance (VDT). Thus, while VDT is common in lower orders such as bryophytes, it is comparatively less common in pteridophytes and angiosperms (Marks *et al.*, 2021). In the latter, desiccation tolerance became confined to the reproductive structures (seeds and pollen) and in most instances, DT in seeds is the default state. DT seeds or 'orthodox seeds' are widespread, while non-DT (known as recalcitrant seeds) are rarer. VDT was likely lost owing to the high metabolic and energy costs associated with constitutive protection (Alpert, 2005; Marks *et al.*, 2021). The low proportion of VDT in contemporary angiosperms is likely the result of more recent, independent evolutionary events in response to the arid environments in which they occur (PROCTOR, 1990; Oliver, Tuba and Mishler, 2000; Proctor and Pence, 2009; Porembski, 2011). Many scientists propose that the re-evolution of VDT is simply the result of a rewiring the pre-existing desiccation tolerance in the seeds of flowering plants.

Theoretically then, crops have all the necessary components to reinstate VDT, but how this process is re-activated is yet to be determined (Illing *et al.*, 2005; Farrant and Moore, 2011; Costa *et al.*, 2017; Han *et al.*, 2017). VDT angiosperms lose water more gradually, typically reaching a state of desiccation over hours to days. This has resulted in the alternative strategy, where many of the mechanisms of tolerance are inducible rather than constitutive (Alpert, 2006; Farrant, 2008). VDT displayed by basal plants and angiosperms are delineated based on the presence of constitutive versus inducible protection, respectively, during the course of drying (Whittaker *et al.*, 2004; Martinelli *et al.*, 2007; Oliver, Guo, *et al.*, 2011; Oliver, Jain, *et al.*, 2011). Of course, the mechanisms are more nuanced, and many exceptions exist, but on a broader scale, these are the defining traits against which fern desiccation tolerance is measured.

1.3. Desiccation tolerance in Pteridophytes

1.3.1. Habitats

Most of the well-known DT plants inhabit rocky outcrops which are characterised by poor soil coverage and volatile water availability. Many of the so called xeric resurrection ferns are found in similar habitats (Porembski and Barthlott, 2000; Poremborski, 2007). The tree canopies of tropical and temperate rainforests also experience rapid fluctuations in water availability but are highly humid (Phillips *et al.*, 2008). In these environments there are a considerable number of DT epiphytic ferns. The mechanisms of desiccation tolerance in epiphytic ferns are considerably different to those studied in their xeric counterparts. (Porembski, 2011).

1.3.2. Anatomical features

DT ferns that have a rudimentary cuticle composed of a single or several cell layers are termed filmy ferns. They occur in regions where drying is rapid, and because of their simplistic cuticles, cannot slow the process of water loss. Because they are subject to cycles of rapid loss, they tend to favour constitutive protective mechanisms much like the DT bryophytes. Ferns with a well-developed cuticle are called non-filmy ferns and can reduce the rate of water loss in their environments. They are akin to DT angiosperms, in that they have developed more inducible desiccation tolerance mechanisms (Hietz, 2010; Fallard *et al.*, 2018; Mkhize, Minibayeva and Beckett, 2020; Alejo-Jacuinde and Herrera-Estrella, 2022).

Bryophytes and filmy ferns can absorb water from dew or rain, whereas most vascular plants can only acquire water via root uptake. Ferns have evolved specific structures that enhance

foliar uptake potential which contribute to the desiccation tolerance response (Alpert, 2000; Hietz, 2010). Trichomes, hairs and peltate scales are structures that have been observed on the frond surfaces and function primarily in facilitating foliar uptake of water. Peltate scales have been shown to play significant roles in managing water loss during dehydration and desiccation. They also mediate the rate of water uptake during rehydration as to avoid cellular damage caused by a rapid influx of water into desiccated structures. Other ferns are incapable of rehydrating purely from foliar rewetting and require hydration of the stele and roots before a reversal of the desiccated state is observed (John and Hasenstein, 2017; Holmlund *et al.*, 2020).

1.3.3. Responses of oxidative stress

As water availability drops, an excess of incoming light results in the production of destructive reactive oxygen species (ROS). This places the plant under significant oxidative stress which is usually mitigated by reducing the amount of incoming light and by the rapid production of antioxidant enzymes (Gaff and Oliver, 2013b; Zhang and Bartels, 2018; Oliver *et al.*, 2020b). To mitigate the production of lethal photosynthetically produced ROS, one common response observed in Pteridophytes is leaf and stem curling (All, 2016). Leaves curl inwardly which shields the adaxial photosynthetic centres from excess light. Plants that opt to maintain the photosynthetic system but reduce exposure to light are termed homoiochlorophyllous desiccation tolerant (HDT). Some examples of stress adaptations that allow for homiochlorophylly is masking of photosynthetic surfaces by leaf curling and/or the accumulation of protective anthocyanins (Farrant, 2000; Cooper and Farrant, 2002). By retaining the photosystems, the plant can quickly reinstate normal photosynthesis when water becomes available, which means they can recover from the air-dry state more rapidly. In contrast, poikilochlorophyllous desiccation tolerant (PDT) plants dismantle their photosynthetic apparatus entirely which offers better protection against destructive ROS. The disadvantage is that they require *de novo* production of the components required to re-establish photosynthesis and normal metabolic function which leads to a slower recovery from the air-dry state (Toldi, Tuba and Scott, 2009). Lower plants more often display HDT, while contemporary angiosperms opt for HDT or PDT. Evidence of stem curling has been observed in the fern *Selaginella lepidophylla (S. lepidophylla*) (Rafsanjani *et al.*, 2015; Brulé *et al.*, 2019). Leaf curling and the production of anthocyanins in orange-coloured scales on the adaxial surfaces of *A. caffrorum* has also been documented (Jill M. Farrant et al., 2009). Several Pteridophytes have also demonstrated the upregulation of antioxidant enzymes that are

commonly observed in the antioxidant response of other resurrection plants. These enzymes include superoxide dismutase, catalase, peroxidase, and glutathione reductase; the specific responses within previously studied species has been reviewed by (Alejo-Jacuinde and Herrera-Estrella, 2022)). Evidence of a primed antioxidant response has been observed in *Selaginella* species (Dinakar and Bartels, 2013). The priming response is not unique to *Selaginella*; it is well documented in other desiccation-tolerant species which have sensitive and tolerant relatives (Oliver, Guo, *et al.*, 2011; Chávez Montes *et al.*, 2022). In addition to the canonical antioxidant enzyme response, Pteridophytes produce other antioxidant molecules that assist in oxidative stress mitigation. Within the *Selaginella* species increases of flavonoids, vanillate, phenylpropanoids and sugar alcohols such as xylitol and sorbitol have been observed. Many of the polyols play a dual role in reducing oxidative stress and reducing the rate of water loss during drying (Shen *et al.*, 1999; Yobi *et al.*, 2012, 2013). A decrease in photosynthesis during dehydration is another conserved response across most DT plants (Challabathula, Puthur and Bartels, 2016). The main distinction between DT and desiccation sensitive (DS) angiosperms is that DT plants actively shut down photosynthesis during the early- to mid-stages of dehydration to reduce excess production of ROS (Dinakar and Bartels, 2013; Challabathula, Puthur and Bartels, 2016). However, in some lycophytes an upregulation of certain photosynthesis- and Calvin cycle-related proteins occurs (Deeba, Pandey and Pandey, 2016). In some filmy ferns, Rubisco remains constant and active through the course of drying and desiccation (Cea *et al.*, 2014; Deeba, Pandey and Pandey, 2016).

1.3.4. Responses to mechanical stress

As water is lost at the cellular level, cell membranes begin to pull away from the rigid cell walls which places cells under mechanical stress. To prevent the eventual rupture of cell membranes resurrection plants, undergo a process of vacuolation (production of many smaller vacuoles) an/or cell wall folding to maintain cell membrane to cell wall contact sites (Rascio and La Rocca, 2005; Fernández-Marín, Holzinger and García-Plazaola, 2016; Georgieva *et al.*, 2017). In terms of mechanical stress adaptions present in Pteridophytes, very few cases have been documented to date. Transcriptomic analyses of *Selaginella* elucidated significant alterations to cell wall structure and composition in the desiccated state. This suggested that cell-wall folding is employed to mitigate mechanical stress (Plancot *et al.*, 2019; Alejo-Jacuinde *et al.*, 2020). An upregulation of arabinose-rich polymers in tolerant fronds of *A. caffrorum* was observed by (Moore *et al.*, 2013), implying increased flexibility of cell walls. Furthermore,

evidence of vacuolation, which enhances stabilization, has been observed in the DT fronds and rhizomes of *A. caffrorum* (Jill M. Farrant et al., 2009).

1.3.5. Vitrification : sugars, osmolytes and proteins

The final stressor is often referred to as macromolecular/or chemical stress. The loss of cellular water effectively reduces the space in which biomolecules occur. This can result in many undesirable reactions which occur due to closer proximity on biomolecules and mass protein aggregations which can lead to a loss of enzyme function. To prevent this outcome, resurrection plants are believed to launch a process called vitrification. Vitrification is the process of massively accumulating osmolytes, sugars, protective proteins and other compatible solutes which results in a biophysically 'glassy state' (du Toit, Bentley and Farrant, 2021). The glassy state maintains molecules in their native conformations and position in the cell in the absence of water. The process of vitrification contributes to the maintenance of cell volume and prevents widespread damage because of ROS; it therefore also reduces oxidative and mechanical stress (Buitink and Leprince, 2008; Fernández-Marín *et al.*, 2013; Leprince and Buitink, 2015; Walters, 2015; Ballesteros, Hill and Walters, 2017). Many of the biomolecules related to the process of vitrification have shown a pronounced increase in certain desiccation tolerant Pteridophytes. A pronounced increase in compatible solute proline is a common feature of many desiccation tolerant plants and has been widely observed in several Pteridophyte species (Alejo-Jacuinde and Herrera-Estrella, 2022). Proline acts as an osmolyte, a ROS scavenger molecule and prevents protein aggregation thereby reducing macromolecular stress imposed by water loss (Le and McQueen-Mason, 2006; Liang *et al.*, 2013). The accumulation of sugars as a replacement for water molecules in the glassy state is a widely conserved feature in all resurrection plants. The sugar composition varies greatly among species, but sucrose is the most observed in most resurrection plants (Illing *et al.*, 2005; Dinakar and Bartels, 2013; Zhang, Song and Bartels, 2016; Costa *et al.*, 2017; du Toit, Bentley and Farrant, 2021). The other sugars that accumulate in response to desiccation in other Pteridophytes are detailed by (Alejo-Jacuinde and Herrera-Estrella, 2022). While many Pteridophytes show an increase in sucrose concentration in response to desiccation, some exceptions exist. *Selaginella bryopteris (S. bryopteris)* decreases the amount of sucrose in the fronds during desiccation and rehydration (Pandey *et al.*, 2010). However, sucrose could be constitutively highly expressed but forfeited during drying as to synthesize other protective sugars. In fact, the constitutively high levels of sucrose have been observed in *S. lepidophylla* (Dinakar and Bartels, 2013). Consistently high concentrations of protective compounds represent one example of bryophyte-type desiccation

tolerance mechanism being present in Pteridophytes (Smirnoff, 1992). Comparisons between DT *Selaginella lepidophylla* and the DS sister species *Selaginella moellendorffii* revealed elevated amounts of sucrose, mono- and polysaccharides, sugar alcohols, aromatic amino acids, osmoprotectant betaine and flavonoids (Yobi *et al*., 2012).

With respect to proteins, the mass accumulation of late embryogenesis abundant (LEA) proteins, early light-inducible proteins (ELIPs) and heat shock proteins (HSPs) has been observed during the dehydration and desiccation stages of most species investigated to date (Oliver *et al.*, 2020b). LEAs are widely accepted as playing prominent roles in the formation of the glassy state, as well as protecting and preserving the structure of nucleic acids during desiccation (Olvera-Carrillo, Reyes and Covarrubias, 2011). ELIPs are present in chloroplasts and are responsible for the protection and repair of photosynthetic structures under desiccation stress (Lee *et al.*, 2020). HSPs maintain proteins in their functional confirmations as plants experience macromolecular stress (Li *et al.*, 2021). The most studied grouping within fern desiccation tolerance are the LEA proteins. The genome of *S. lepidophylla* contains 65 LEA genes. A total of 48 of those genes demonstrate differential expression through the course drying and rehydration (Vanburen *et al.*, 2018). An upregulation of LEA proteins during dehydration was also observed in the desiccation-tolerant *Selaginella tamariscina* (*S. tamariscina*) and the desiccation-sensitive species *Selaginella moellendorffii*. While mass accumulation of LEAs is certainly involved in the desiccation response, the biological role of LEAs is not restricted to resurrection species (Chávez Montes *et al.*, 2022).

1.3.6. Role of phytohormones

Phytohormone regulation of desiccation tolerance within the Pteridophytes is limited to studies focusing on abscisic acid (ABA). One study conducted on *S. bryopteris* demonstrated that there were some similarities in the desiccation responses observed in shoots and roots. The most pronounced similarity was related to the mechanisms of ATP production during drying. Organ-specific responses to dehydration included ABA-induced signal transduction in the shoots, and greater activation of the antioxidant system in the shoots (Deeba, Pandey and Pandey, 2016). An increase in ABA has been associated with responses to dehydration and desiccation stress in sensitive and tolerant plants (Alejo-Jacuinde and Herrera-Estrella, 2022). Dehydration stress has been shown to drastically increase ABA levels in *S. tamariscina* and was required for the expression of certain ELIP and LEA genes (Liu, Chien and Lin, 2008). Currently, ABA has been accepted as a key regulator of the plant desiccation response, but other hormones such as

salicylic acid and/or ROS signalling have been identified as other possible regulators of VDT (Zhang and Bartels, 2018). ABA was also shown to enhance the ability of the fern *Polypodium virginianum* to withstand drying (Reynolds and Bewley, 1993).

1.3.7. Role of the rhizome

Of particular importance to this study are the responses of fern rhizomes to desiccation. Most studies focused on the role of rhizomes in the desiccation response are centred around the effects of hydration dynamics on whole-plant recovery (John and Hasenstein, 2017; Holmlund *et al.*, 2019; Prats and Brodersen, 2021). Research conducted on the land ferns *Pentagramma triangularis* and *Pellaea andromedifolia* suggest that rewetting via the soil is required for whole-plant recovery. The vascular systems are not reinstated via foliar water uptake only. Concurrent foliar uptake does however increase the rate of recovery in fronds. Analysis of the non-structural carbohydrates in both ferns demonstrated that upon rehydration, a decrease in sucrose and increase in starch was observed in the rhizomes (Holmlund *et al.*, 2020). Presumably, a conversion of sucrose to starch is to restore energy reserves for a future desiccation event. This theory is consistent with the primary role of the rhizome in carbohydrate storage (Martínez-Vilalta *et al.*, 2016). To our knowledge, additional studies measuring key biomolecules in the rhizomes of ferns undergoing desiccation do not exist. Proteomic and metabolomic studies of the rhizome of *A. caffrorum* through the course of desiccation have been conducted by the PSL and will be reviewed later. Briefly, proteomic studies suggested a possible role for the rhizome in the seasonal regulation of frond desiccation tolerance. They also highlighted possible stress cross-tolerances that occur in the rhizomes in summer and winter (Shoko *unpublished*, 2007; Wittenberg *unpublished*, 2021).

Presented in this review are the mechanisms of desiccation tolerance that have been elucidated in Pteridophytes to date. Overall, Pteridophytes display a combination of highly conserved constitutive and inducible traits present in basal and flowering resurrection plants (Figure 1). Presumably this is owing to the intermediate evolutionary position of Pteridophytes relative to bryophytes and angiosperms. The review also highlights that although Pteridophytes represent a mixed model of desiccation tolerance responses, they also appear to have highly variable responses within the Pteridophyte grouping. As such, fern desiccation tolerance should be contextualised in terms of constitutive, inducible, and unique processes of desiccation tolerance that may exist within a species.

Constitutive protection	**Mixed model**	**Inducible protection**
Rapid	Rate of drying	Slow
Faster	Rate of recovery	Slower
Simple	Leaf anatomy	Complex
Low	Transcriptional /proteomic response	High
Maintenance of high levels	Metabolomic protective compounds	Primming or accumulated during drying

Figure 1 : Dominant mechanisms of tolerance in resurrection plants over evolutionary time. The key features of bryophytes focus on constitutive protection (left) while angiosperms show greater inducible responses (right). The current framework for Pteridophytes (middle) places them in an intermediate position where each fern species can have any mixture of the strategies observed in basal and flowering plants (adapted from (Alejo-Jacuinde and Herrera-Estrella, 2022).

1.4. Why is Anemia caffrorum a good model for climate smart crops?

A. caffrorum is a resurrection fern that is endemic to South Africa. It grows predominantly in forest margins where it is partially shaded or fully exposed (Roux, 1979). *A. caffrorum* displays a form of desiccation tolerance that is seasonally regulated. In the hot summer months, fronds are DT and, in the cooler, wetter months the fronds are DS (Jill M. Farrant et al., 2009). DT fronds are visually identified by the presence of protective orange scales on the adaxial surface. Varying degrees of coverage of orange hairs suggest possible priming and/or hardening of the DT form through the course of the season. DT fronds also show varying degrees of frond curling as to protect abaxial surfaces from incoming light under conditions of low water availability. DS fronds are characterised by a lack of orange hairs and frond curling capabilities. DS fronds are larger (greater specific leaf area) than DT fronds, presumably owing to more efficient photosynthesis occurring in DS fronds (Jill M Farrant *et al.*, 2009; Nadal *et al.*, 2021) (Figure 2). Experimentally, DS and DT fronds are confirmed by their ability to survive a desiccation event. Excised fronds are allowed to dry overnight, to ensure that they reach the desiccated state. They are subsequently rehydrated by submergence in water. DT fronds are capable of uncurling their pinnae, while DS fronds do not show any visible signs of recovery (Nadal *et al.*, 2021) (Figure 4).

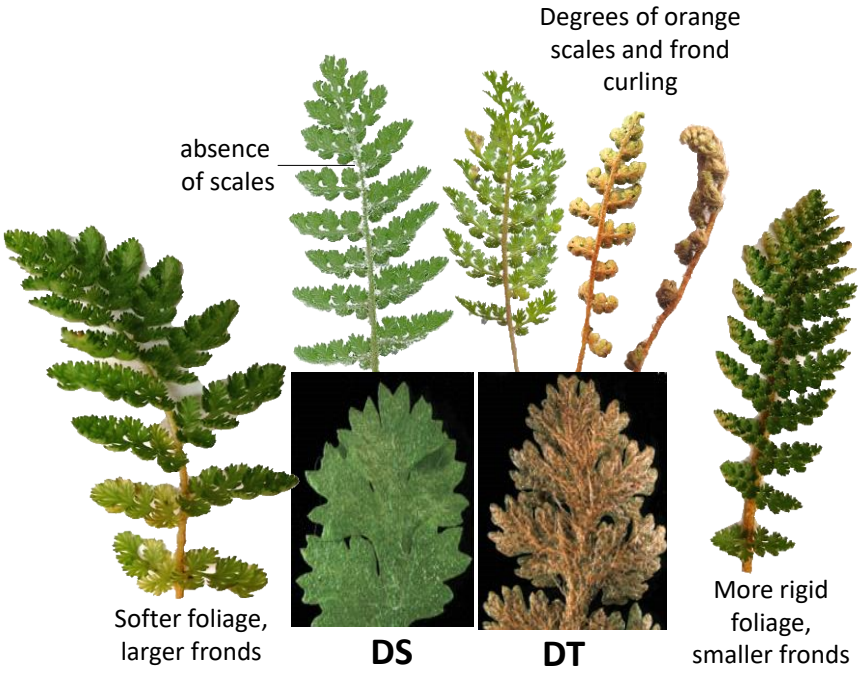

Figure 2: Visual frond phenotypes of *Anemia caffrorum*. DS: desiccation sensitive fronds lack orange abaxial scales, are larger and softer to the touch. DT: desiccation tolerant fronds display degrees of orange abaxial scale coverage and frond curling. They are also more rigid and typically smaller than DS fronds.

In the dry season, DT fronds employ a combination of constitutive and inducible mechanisms of desiccation tolerance which facilitate its survival under drought conditions. The presence of constitutive processes allows for rapid recovery upon rewetting in the rainy season. DS fronds can therefore be produced relatively quickly in the rainy season. DS fronds, which are more photosynthetically active, use excess energy for growth and reproduction (Nadal *et al.*, 2021). Fronds produce sporangia that mature and release DT spores which give rise new DT plants. It has been proposed that during this season, additional carbohydrates are transported to the rhizome for storage. These reserves are utilized during the dry season to produce and maintain new DT fronds (Jill M Farrant *et al.*, 2009; Harvesting, Photosynthetic and Duarte, 2016; Nadal *et al.*, 2021) (Figure 3).

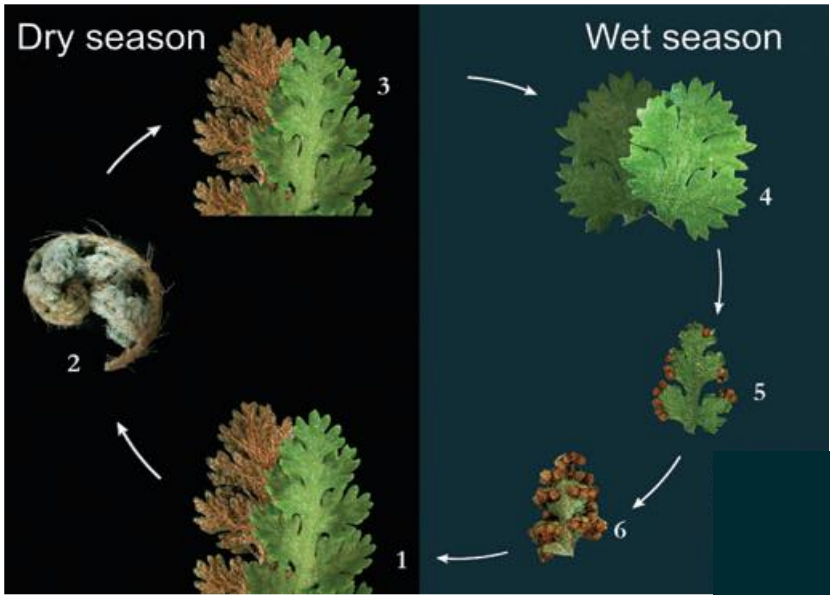

Figure 3 : *Anemia caffrorum* proposed lifecycle. In the dry season, DT fronds focus on drought survival by employing a host of desiccation tolerance mechanisms (2). In the wet season, new DS fronds do not present mechanisms of desiccation tolerance. Comparatively photosynthetic DS fronds produce enough energy to reproduce (4,5). Spores mature and disperse to form new DT plants (6,1) (Jill M. Farrant et al., 2009).

Seasonal regulation of desiccation tolerance provides an opportunity for *A. caffrorum* to escape the energetically taxing processes involved in permanently maintaining the tolerant phenotype. Theoretically, if we could confer seasonal desiccation tolerance to crop plants, we could develop crops that are highly productive during rainy seasons but able to survive unexpected droughts associated with rapid climate change. Finally, *A. caffrorum* is a particularly powerful model for directly comparing DS and DT tissues. Other studies that attempt to do so, have

either compared 1) senescent (DS) and non-senescent regions (DT) of the same species, 2) whole plants (DT) and their detached leaves (DS) or 3) sensitive and tolerant species within the same genus (Willigen et al., 2004; Illing et al., 2005; Balsamo et al., 2006; Le et al., 2007; Jill M. Farrant et al., 2009; Marks et al., 2021). For this reason, the mechanisms of desiccation tolerance of *A. caffrorum* fronds and rhizomes have been explored via proteomic and metabolomic studies. The aim of this thesis is to contribute new information on the desiccation tolerance mechanisms *A. caffrorum* at the transcriptomic level.

Chapter 1 :

Dehydration experiments and robust sample preparation of *Anemia caffrorum* fronds and rhizomes: a feasibility study

Chapter 1 :

Dehydration experiments and robust sample preparation of *Anemia caffrorum* fronds and rhizomes: a feasibility study

1. Introduction

As reviewed by Marks et al., (2021) several aspects of the methodologies related to desiccation tolerance studies introduce unexpected and unwanted variability. In designing the desiccation experiments the goal is to minimize this variability within the experimental circumstances. This is particularly important in transcriptomic work because next generation sequencing (NGS) experiments are a considerable expense and their success is highly dependent on robust sample preparation (McCombie, McPherson and Mardis, 2019). All previous studies on *A. caffrorum* were conducted using field-collected individuals. This is because a reproducible method to maintain the tolerant phenotype under laboratory conditions has not yet been established. Field studies are fraught with factors of variability that are difficult to minimize and the relative lack of knowledge on how fronds transition between desiccation tolerant (DT) and desiccation sensitive (DS) exacerbates this issue (Marks *et al.*, 2021). Some studies indicate that rearing hardened, field-collected plants under mesic conditions can result in a loss of tolerance (Khandelwal *et al.*, 2010; Stark, 2017). To reduce these effects, field-collected plants were subjected to a shorter acclimation period before tissues samples were harvested. The age of the tissues can also affect the degree of tolerance. Studies have shown that younger plants are more resilient to desiccation, while older plants are more prone to senescence (Blomstedt *et al.*, 2018). Furthermore, the same species occurring in slightly different micro-habitats can display varying degrees of tolerance (Knight *et al.*, 2006; Marks *et al.*, 2021). To minimize the possible effects of plant age, only mature and healthy plants of uniform size were chosen for the study. These plants were assumed to be in similar growth and developmental stages. To mitigate population effects, plants from the same population and in a similar microhabitat were used. In this study, partially shaded plants, as opposed to fully exposed plants, were used. The length of the previous desiccated state, and the condition of the tissue prior to the experimentation can noticeably affect the outcome of a desiccation experiment (Schonbeck and Bewley, 1981; Benko *et al.*, 2002). Unexpected environmental stressors such as an unexpectedly harsh summer period or co-occurring stressors can drastically affect the outcome of desiccation tolerance (Farrant *et al.*, 1999). Therefore, plants collected during different seasons, and even

at different times in the same season may be too dissimilar to act as biological replicates. To mitigate this form of variability, all plants used in the desiccation experiments were collected at the same time and the tissues captured from this experiment were used. Relative water content (RWC) determination is widely adopted as a method to track the stage of water stress for sampling purposes, but been criticised because it does not capture the water potential of tissues (Barrs and Weatherley, 1962; Oliver et al., 2020). Instead, intensive physiological studies were conducted during desiccation of DT and DS fronds to track key responses of the photosystems (Nadal *et al.*, 2021). The use of additional response measures is believed to improve the assessment of the stage of desiccation. Samples used for NGS studies were co-collected during physiological experimentation. Thereafter, the key stages in the desiccation response of angiosperms were used as model for choosing water contents that would be analysed, since such a framework has not been established in ferns. However, Oliver et al., (1998) have proposed that ferns follow similar stages by virtue of the fact that water loss is slow in ferns and angiosperms. Rhizome sample collection could not be tracked via photosynthetic response and was further limited by ethical considerations. Harvest of the rhizome results in the destruction of the whole plant and thus the study was limited to two sample points in each season. As the responses of underground structures such as rhizomes are largely understudied, literature to guide the choice of rhizome water content in ferns was not available. This study will be the first reference for future studies.

All NGS studies are heavily contingent on high quality nucleic acid. Preparation of nucleic acids for sequencing is made up of several steps and quality control measures are usually very strict (McCombie, McPherson and Mardis, 2019). RNA-Seq experiments begin with the isolation of total RNA from the organism of study. Total RNA is typically quantified via more sensitive fluorescent dye-based methods on specialized instruments such as the Bioanalyzer. This is opposed to traditional, less sensitive, UV absorbance methods (Panaro *et al.*, 2000). RNA quality is assessed by a measurement called the RNA integrity number (RIN). The RIN score is determined by first conducting microcapillary electrophoretic RNA separation. Briefly, several key features are considered which essentially determine the integrity of the isolated RNA. RNA is assigned a RIN quality score between 0 and 10 which informs whether RNA quality is sufficient to proceed. It is worth noting that RIN scoring was designed using mammalian tissues. Several studies have indicated that owing to the different rRNA species, RIN scores can often be underestimated for plant tissues. As a result, a lower RIN score for plant NGS experiments is often accepted (Schroeder *et al.*, 2006). Once total RNA is

quantified, and quality assessed it is enriched for the RNA species under study. In this study, total RNA was isolated, and mRNA was enriched for. The preferred method to isolate mRNA is through capture on oligoT magnetic beads (Qing *et al.*, 2013; Zhao *et al.*, 2014). Other popular methods focus on the depletion of ribosomal RNA because 90% of total RNA is comprised of rRNA (O'Neil, Glowatz and Schlumpberge, 2013). Isolated mRNA is converted to cDNA and then fragmented via sonication, nebulization, acoustic shearing, or enzymatic treatment (Hrdlickova, Toloue and Tian, 2017). Fragments are subjected to a process called size selection to capture fragments in the selected size range. End repair is conducted before several adapters are ligated to them. The main functions of the adapters are that they facilitate polymerase chain reaction (PCR) amplification and immobilization on to the solid support structure where sequencing takes place (Van Dijk, Jaszczyszyn and Thermes, 2014). The fragments are then enriched via PCR amplification and the resulting cDNA library is usually quality checked before loading on the sequencing device (Hrdlickova, Toloue and Tian, 2017; Salzberg, 2019; Stark, Grzelak and Hadfield, 2019). To ensure the success of library preparation and ultimate sequencing, prepared RNA samples required considerable optimization to achieve acceptable RIN values. Presented here are the established methods for isolation high-quality RNA from the rhizomes and fronds of *A. caffrorum* for use in future NGS experiments.

2. Aims and objectives

To summarize the aims of this chapter:

1. Assess which relevant water contents were captured based on physiological data and angiosperm desiccation response frameworks.
2. Develop a method to isolate total RNA from frond and rhizome tissue that is of sufficient quantity and quality to be sequenced.

3. Methods

3.1 Anemia caffrorum collection and maintenance

A. caffrorum plants that lacked visible signs of stress and were of similar developmental age were collected from Table Mountain in Cape Town, South Africa. For plants used in frond dehydration studies, similar apical heights and phenotypic presentation were prioritized. For plants used in rhizome dehydration studies, similar rhizome sizes and phenotypic presentation were prioritized. All plants contained between 3-5 health fronds. DS and DT plants were collected in winter and summer, respectively, to capture the true phenotypes found in the field. All plants were potted in soil collected from the field and acclimated in growth chambers (Conviron, Percival, USA) at 28/24^{o}C with a 14-/10-hour photoperiod and 800 µmmol m-2 s-1 of PPFD with optimum irrigation. The plants were acclimated for 10 days before dehydration experiments were started as to avoid the loss of the phenotype observed in the field (Nadal *et al.*, 2021). All plants were kept under the conditions described above through the course of the dehydration experiment (Nadal *et al.*, 2021). Physiological studies from which frond samples were co-collected were conducted by Nadal et al., (2021).

3.2. Anemia caffrorum desiccation experiments and tissue collection

Frond and rhizome dehydration experiments were conducted at separate times from plants within the same population (plants occurring in the same microhabitat on the mountain). A rhizome displaying the DT phenotype in the fronds was considered a DT rhizome. Similarly, rhizomes displaying DS fronds were considered DS rhizomes. Prior to the start of dehydration, the plants were saturated with water. Twenty-four hours later the first set of fronds/rhizomes were collected (full turgor samples). Fronds/rhizomes were collected throughout the 3-week dehydration period until they reached the air-dry state (these are called water stressed (WS) samples). Plants were rehydrated and twenty-four hours later, frond/rhizomes were collected (24-hour post rehydration samples) (Nadal *et al.*, 2021). All samples collected were sectioned; a portion of fresh tissue was used for relative water content (RWC) determination and the remainder was frozen in liquid nitrogen within 1-minute of harvesting. When harvesting the rhizome, the full plant was removed from the pot and the soil was cleared from the organ thoroughly before freezing. These samples were transferred to -80 ^{o}C until further use. To capture the average water content of the frond, four pinnae across the frond's length were used for RWC determination. The remainder of the tissue, including the petiole was frozen and used for RNA isolation. For the rhizomes, the actively growing end of the rhizome was frozen and

used for RNA isolation and the portion immediately adjacent was used for RWC determination. Since part of the rhizomes occurs very close to the soil surface and the remainder fully covered in soil, rhizome tissue was divided into vertical slices to capture the average water content across the organ. In this study, tissues were harvested during the transition from darkness to light in the plant chambers as to capture plants during the time of day that mimics dawn.

3.3. Tissue relative water content (RWC) determination

The fresh weight (FW) was measured on a fine scale. The tissue was soaked in water overnight at 4 ^{o}C, patted dry and the turgid (TW) recorded. Finally, the tissue was incubated at 70^{o}C for 48 hours before the dry weights (DW) were recorded (Barrs & Weatherly, 1962). Four technical replicates were used for each sample and the relative water content was determined using the formula below:

$$\text{RWC} = \frac{\text{FW}-\text{DW}}{\text{TW}-\text{DW}} \text{ x } 100$$

3.4. RNA isolation

Frozen frond and stipe tissue was ground into a fine powder in liquid nitrogen with approximately 2% (w/v) polyvinylpyrrolidone. RNA was isolated from 30 - 50 mg of ground tissue using the Spectrum Plant Total RNA Kit (Sigma-Aldrich) according to the manufacturer's instructions for protocol A in conjunction with the On-Column DNase I Digestion set (Sigma-Aldrich). Where necessary samples were subjected to an additional cleaning step to improve A60/A230 ratios prior to sequencing. To the volume of isolated total RNA, 2.5 volumes of absolute ethanol and 0.026 volumes of 3M sodium acetate (pH 5.0) was added, and the mixture incubated at -20^{o}C for 12 hours. The sample was centrifuged at 12 000 *g* for 20 minutes at 4^{o}C and the pellet washed in 1 mL 75% (v/v) absolute ethanol in DEPC-treated water. Finally, the sample was centrifuged at 12 000 *g* for 10 minutes at 4^{o}C before the supernatant was removed and the pellet air dried. The pellet was resuspended in elution buffer from the Spectrum Plant Total RNA Kit. Rhizome tissue from the actively growing end of the rhizome was ground into a fine powder in liquid nitrogen. To 30 - 50 mg of ground rhizome tissue, 800 uL of extraction buffer (0.2M Na borate decahydrate, 30mM EGTA, 1% (w/v) SDS, 1% (w/v) Na deoxycholate supplemented with 5.7% (w/v) PVP40 and 3mM fresh DTT) pre-heated to 80^{o}C was added and the mixture vortexed briefly. Proteinase K was added to a concentration of 0.1 mg. mL^{-1} and the sample vortexed for 1 minute before incubation at 42^{o}C for 15 minutes with occasional mixing. To this, 64 uL of 2M KCl was added, the samples

incubated on ice for 30 minutes and centrifuged at 12 000 *g* for 20 minutes at 4°C. The supernatant was mixed with 295 uL ice-cold 8M LiCl and incubated overnight at 4°C and on ice. Nucleic acids were pelleted by centrifugation at at 12 000 *g* for 20 minutes at 4°C. The pellets were washed in 2M LiCL, pelleted by centrifugation at 12 000 *g* for 10 minutes at 4°C and DNA digested using the On-Column DNase I Digestion set (methods adpated from Wan & Wilkins, 1994). To each digestion 100 uL of phenol-chloroform (25:24:1 phenol (pH 8.0), chloroform, isoamyl alcohol) was added and the mixture transferred to pre-made phase-lock tubes containing 50 – 100 uL vacuum grease. The samples were centrifuged at 15 000 *g* for 5 minutes at room temperature and the top layer was mixed with binding buffer from the Spectrum Plant Total RNA Kit (Sigma-Aldrich) in a 2:3 ratio. The mixture was loaded onto a binding column and the remainder of the isolation was conducted according to the manufacturer's instructions.

3.5. RNA quality control

The concentration (Abs260) and purity (Abs260/280nm and Abs260/230nm) of purified RNA samples was determined using a NanoDrop 2000 spectrophotometer (ThermoScientific). Total RNA was more accurately quantified using a Qubit RNA BR Assay kit and subsequent analysis on a Qubit 4.0 Fluorometer and (ThermoScientific). To better assess the RNA integrity, the RNA samples were prepared using an Agilent RNA 6000 Nano Kit (5 – 500 ng) and run on an Agilent Bioanalyzer 2100 (Agilent Technologies) according to the manufacturer's instructions. The samples sent for short and long read sequencing, their sample ID and RNA integrity number (RIN) is summarized in Table 1.

Table 1 : Total RNA samples used for short and long read sequencing.

Tissue	Phenotype	Sample ID	RWC	RIN
Frond	DS	FDS_FT_1	91.01	6.30
Frond	DS	FDS_FT_2	90.97	5.90
Frond	DS	FDS_FT_3	89.08	6.60
Frond	DS	FDS_55_1	68.14	6.60
Frond	DS	FDS_55_2	57.25	6.60
Frond	DS	FDS_55_3	55.78	6.00
Frond	DS	FDS_55_4	50.33	6.40
Frond	DS	FDS_30_1	37.91	6.70
Frond	DS	FDS_30_2	23.83	6.80
Frond	DS	FDS_30_3	20.05	6.80
Frond	DS	FDS_10_1	13.28	6.80
Frond	DS	FDS_10_2	12.68	6.20

Frond	DS	FDS_10_3	12.17	6.30
Frond	DS	FDS_10_4	11.05	6.50
Frond	DS	FDS_R24_1	53.72	6.30
Frond	DS	FDS_R24_2	34.55	6.60
Frond	DS	FDS_R24_3	25.97	6.60
Frond	DT	FDT_FT_1	80.42	6.70
Frond	DT	FDT_FT_2	77.31	6.60
Frond	DT	FDT_FT_3	76.35	6.40
Frond	DT	FDT_FT_4	75.71	4.90
Frond	DT	FDT_55_1	61.43	6.90
Frond	DT	FDT_55_2	59.21	7.10
Frond	DT	FDT_55_3	57.68	6.70
Frond	DT	FDT_55_4	52.57	6.70
Frond	DT	FDT_30_1	40.43	6.90
Frond	DT	FDT_30_2	26.61	7.40
Frond	DT	FDT_30_3	24.73	6.00
Frond	DT	FDT_10_1	15.37	6.00
Frond	DT	FDT_10_2	14.88	6.40
Frond	DT	FDT_10_3	11.47	5.80
Frond	DT	FDT_10_4	11.42	5.90
Frond	DT	FDT_R24_1	53.38	5.80
Frond	DT	FDT_R24_2	39.55	6.10
Frond	DT	FDT_R24_3	27.30	6.30
Rhizome	DS	RDS_FT_1	111.99	8.60
Rhizome	DS	RDS_FT_2	101.83	9.00
Rhizome	DS	RDS_FT_3	105.93	7.90
Rhizome	DS	RDS_FT_4	97.11	8.20
Rhizome	DS	RDS_30_1	27.68	7.00
Rhizome	DS	RDS_30_2	34.14	8.40
Rhizome	DS	RDS_30_3	30.88	7.20
Rhizome	DS	RDS_30_4	34.57	7.60
Rhizome	DT	RDT_FT_1	94.19	7.70
Rhizome	DT	RDT_FT_2	93.41	7.00
Rhizome	DT	RDT_FT_3	106.70	6.10
Rhizome	DT	RDT_FT_4	91.13	7.20
Rhizome	DT	RDT_30_1	26.06	7.20
Rhizome	DT	RDT_30_2	28.24	7.70
Rhizome	DT	RDT_30_3	24.78	7.40
Rhizome	DT	RDT_30_4	24.96	7.70
Rhizome	DT	RDT_30_5	28.41	8.20

DS = desiccation sensitive
DT = desiccation tolerant
FT = full turgor
55/30/10 = approximate relative water content for the set of samples
R24 = recovery 24 hours post-rehydration
RWC = relative water content

RIN = RNA integrity number
samples used for SMRT sequencing

4. Results and discussion

4.1. Frond sample selection and preparation for transcriptomics

Figure 1 depicts some anatomical and physiological changes which occur in DT and DS fronds of *A. caffrorum* under dehydrating conditions. Upon full hydration, DT fronds have totally exposed adaxial surfaces but begin to curl, to various degrees. Upon rehydration, the fronds uncurl with an apparent lack of damage to visible surfaces (Figure 1A). In contrast, DS fronds lack organized curling and do not appear to recover following rewetting (Figure 1B). This lack of organized uncurling has also been correlated with a lack of recovery in photosynthetic activity. In contrast, DT fronds did recover some photosynthetic activity (Nadal *et al.*, 2021). This finding supported the visual cues of the desiccation sensitive and tolerant phenotypes. Fully hydrated DS fronds had a relative water content (RWC) of 92.65 ± 0.88% (n = 12) whereas DT fronds had an RWC of 77.45 ± 0.60% (n = 4) (Figure 1C and 1D *green lines*). This discrepancy, in and of itself, speaks to a dichotomy in the plant water relations between sensitive and tolerant phenotypes (Nadal *et al.*, 2021). After approximately 10 days without water, the RWC of DS fronds began to decrease, and by day 15 they approached 10% RWC (Figure 1C). The RWC of DT fronds began to decrease after approximately 8 days and by day 11 they were approaching 10% RWC (Figure 1D). DT fronds began losing water sooner and reached 10% RWC more rapidly. While anecdotal, it is interesting that the period of dehydration (approximately 30 – 80% RWC) appeared more variable (characterised by longer box plots) in DS fronds than DT fronds (Figure 1C and 1D). This could reflect the organized nature of water loss in DT fronds as opposed to drying in DS fronds which appeared uncontrolled (Figure 1A and 1B). Once the frond RWCs approached 10% they were left to dry for 1 more week before they were watered. Twelve-hours after recovery there was no discernible difference in the RWC in both DS and DT fronds. After 24 hours DS fronds reached a RWC of 24.87 ± 4.94% (n=3) and DT fronds had a RWC of 59.73 ± 5.90% (n=7) (Figure 1C and 1D). Some fronds within the DT grouping were able to restore their water contents to that of their fully hydrated state, indicating some variability in timing of rehydration among fronds and plants. After 48-hours there was no discernible difference in the RWC of DS fronds when compared to the 24-hour samples suggesting that there would not be any further recovery (Figure 1C and 1D).

Once all the samples were collected, an interrogation of the captured water contents and their relevance to key points in dehydration, desiccation and rehydration was conducted. Plants under conditions of high-water availability are often referred to as fully hydrated. In this stage the plant is actively growing i.e., undergoing normal metabolism and usually have a RWC of 85 – 100% (Gaff and Oliver, 2013). To interrogate transcriptomic differences between DS and DT fronds under normal metabolism, our first sample grouping was derived from the fully hydrated state. While the RWC of fully hydrated DT and DS samples are significantly different (DS = 92.65 ± 0.88% and DT =77.45 ± 0.60%), the discrepancy can be accounted to the fact that full-hydrated DT fronds having higher leaf mass area and densities than DS fronds, rather than a difference in the osmotic potentials of the fronds (Figure 1E and 3F) (Nadal *et al.*, 2021). Thus, it was deemed that these RWC's represent a fully hydrated state in the respective tissue types and that loss of water below these will depict metabolism associated with dehydration stress. Additionally, these samples served as the control against which tissues experiencing dehydration and desiccation stress were compared. As plants lose internal water and are subjected to dehydration stress, they must alter their metabolism. Omic studies on resurrection plants which comprise extensive sampling points during a dehydration time course suggest a 3 phase response to dehydration, namely an early stage (100-ca 55% RWC), a mid-stage (ca 55-30% RWC) and a late stage (ca 30-5% RWC) (Costa *et al.*, 2017; Farrant and Hilhorst, 2021; Gabier *et al.*, 2021). Unfortunately, many studies, for somewhat valid reasons (e.g. difficulty in catching similar RWC across a sampling population and sampling cost constraints) have been directed at only early and late stages of drying (Ingle *et al.*, 2007; Le *et al.*, 2007; Xu *et al.*, 2021). Nevertheless, collectively, these studies have confirmed that while some mechanisms associated with desiccation tolerance are launched at the start of dehydration, the true determinant of desiccation tolerance are the processes that enable survival below 30 - 40% RWC. Thus, in the present study a second sample grouping to capture the consequences of an early response to dehydration tissues with RWCs of approximately 55% were selected, as was a third grouping which should provide insights into the late response/and or entry into desiccation at approximately 30% (Figure 1E and 3F). Since the period of dehydration between 30 – 80% exhibited considerable biological variability, t-tests were conducted to investigate whether these groupings were statistically distinct. The DS 55% grouping was statistically significant to the DS 30% grouping ($p=0.0048$) which was distinct from the DS 10% group ($p=0.0222$). The DT 55% grouping was statistically significant to the DT 30% grouping ($p=0.0022$) which was distinct from the DT 10% group ($p=0.0104$) (Figure 1E and 1F). The air-dry state ranges from 5 - 13% in the field (Gaff and Loveys, 1984). In the final stages of

drying preceding air-dry, additional protective metabolites accumulate rapidly. The air-dry state can persist for different periods of time depending on the species and is characterised by the total cessation of metabolism and the entry into a quiescent state. The air-dry state was defined at 10% RWC (Figure 1E and 1F) which has been shown to lack any photosynthetic activity in both frond phenotypes (Nadal et al., 2021). When water becomes available again, RWC contents increase, and metabolic function is gradually reinstated. The final grouping was the 24-hour post rehydration state and served to investigate any potential recovery mechanisms (Figure 1E and 3F). (Nadal et al., 2021) highlighted the importance of this stage by demonstrating that only DT fronds of *A. caffrorum* recover photosynthetic activity 24-hours after rehydration which supports the relative lack of RWC recovery in DS fronds (Figure 1B). Of course, the framework of these stages has been defined based on a handful of studies, and inter-species variation has been observed. Furthermore, the process in more basal plants such as bryophytes is considerably different (Whittaker et al., 2004; Martinelli et al., 2007; Oliver et al., 2011b, 2011a). Limited, current evidence indicates that non-filmy ferns behave more similarly to angiosperms than their more basal ancestors, and as such the angiosperm framework was adopted to select samples of interest (Alejo-Jacuinde and Herrera-Estrella, 2022).

After selecting samples for transcriptomic exploration, excess samples were used to optimize the total RNA isolation method that would be suitable for sequencing. This required considerable optimization and ultimately, the Spectrum Plant Total RNA Kit (Sigma-Aldrich) was used for frond samples. To obtain total RNA that was not contaminated with protein (A280/A260 ratio), other contaminants that absorb at 230nm and was of sufficient yield, the total starting tissue was reduced to 30 – 50 mg compared to 100 mg which is routinely used (Wilfinger, Mackey and Chomczynski, 1997). Yields were, on average, 0.11 ± 0.06 mg Total RNA per mg of fresh frond tissue which was sufficient for short read sequencing. RNA integrity number (RIN) values for frond samples were between 4.9 and 7.4. This was not unexpected since plant RIN values are complicated by the presence of additional RNA bands namely cytosolic, chloroplastic and mitochondrial rRNA bands which vary in size from 5S – 25S (Daniell *et al.*, 2005; Schroeder *et al.*, 2006). The developed method was reproducible and can be used in the future to isolate total RNA from both DS and DT *A. caffrorum* fronds at various states of dehydration and desiccation.

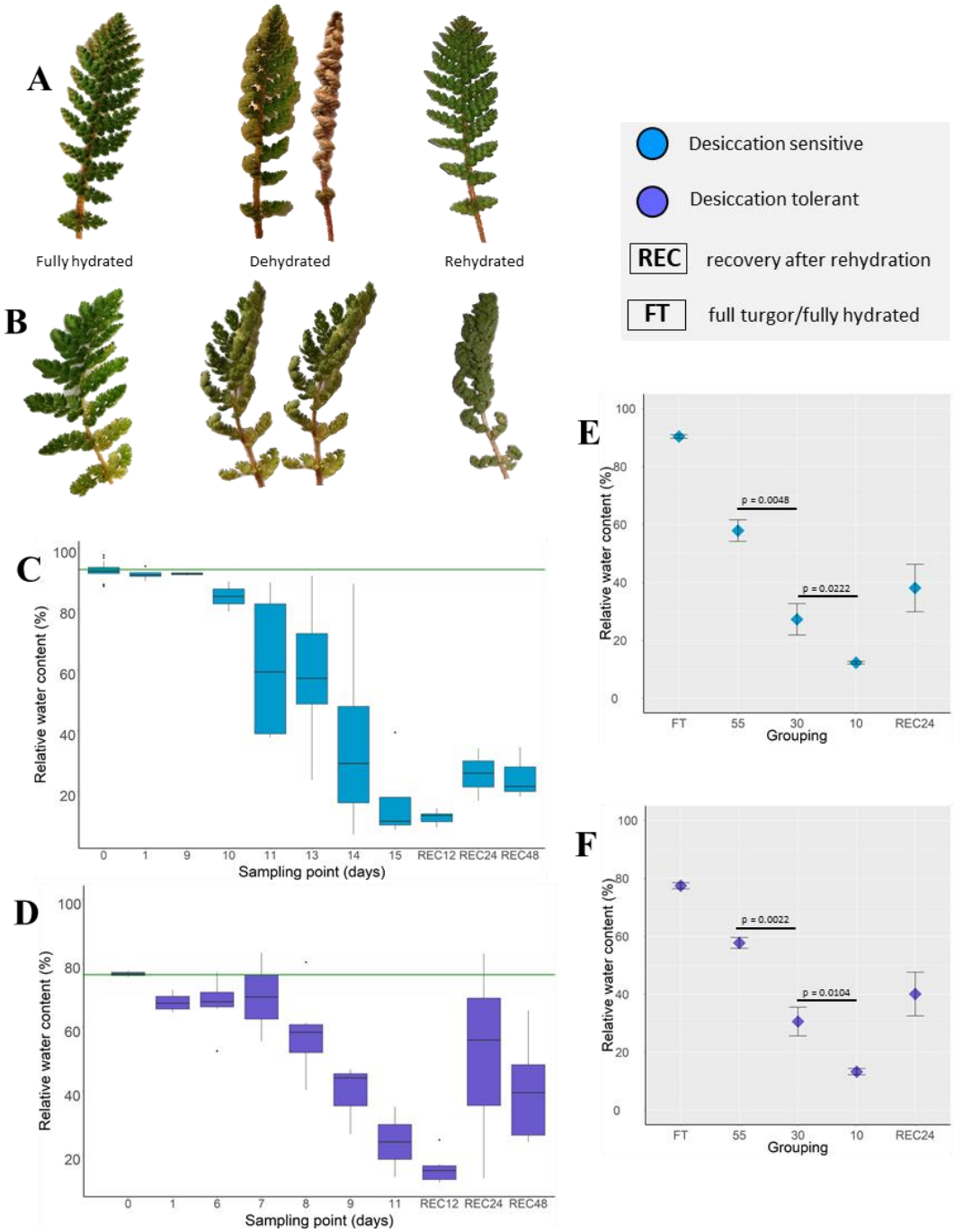

Figure 1: Frond dehydration experiment and samples used for transcriptomics. A: Typical response to dehydration of desiccation tolerant (DT) *Anemia caffrorum* fronds. DT fronds are uncurled during periods of high-water availability but begin organized curling during dehydration. After reaching the air-dry state, once rewatered, DT fronds re-expose adaxial leaf surfaces. B: Typical response to dehydration of desiccation sensitive (DS) fronds. DS fronds show signs of curling but upon rewetting they do not appear to recover from a desiccation event. C: DS fronds had a relative water content (RWC) of approximately 90% in their fully hydrated state (green line) and reached RWCs of 10% after 15 days. After rehydration they reached a RWC of approximately 24.87%. D: DT fronds had a relative water content (RWC) of approximately 80% in their fully hydrated state (green line). DT fronds reached RWCs of 10% after approximately 11 days and could recover to water contents of approximately 60% upon rewetting for 24hr. E: DS frond samples used for transcriptomics were statistically different from one another. F: DT frond samples used for transcriptomics were statistically different from one another.

4.2. Rhizome sample selection and preparation for transcriptomics

A nearby population of plants to those used in the frond dehydration experiments, a second set of dehydration experiments were conducted in the same manner. Rhizome harvesting is a destructive method, meaning that plants cannot be returned to the field-site after experimentation. Owing to these ethical constraints, only two sample points could be investigated in each season. As a result, the groupings chosen for this study were the fully hydrated forms of the DS and DT rhizomes (100% RWC) as well as both phenotypes at 30% RWC (Figure 2A and 2B). Preliminary evidence indicated that the rhizomes took approximately 21 - 25 days to each a relative water content of approximately 30% and persisted in this region for an additional 7 days. Shoko, 2007 (*unpublished*) has however previously shown that the rhizomes can reach 10% RWC, suggesting that *A. caffrorum* rhizomes at 30% RWC are not likely to have reached the air-dry/quiescent state in which little to no metabolic activity would be expected. The 30% RWC therefore may represent some phase between the late dehydration response and entry into desiccation. Interrogation of these samples should contribute to the development of a framework detailing the typical responses through the course of drying and desiccation seen in fern rhizomes. Horizontal slices of the samples collected for RWC determination were briefly observed. Notably, dry (30% RWC) rhizomes appeared 'like glass' in section and the cut surface was highly reflective compared to hydrated rhizomes in both DT and DS phenotypes (Figure 2C and 2D). The glass-like texture and reflective appearance is reminiscent of the state of vitrification that is commonly observed for many resurrection plants in their response to drying. The accumulation of protective sugars and proteins are accredited to the formation of the 'glassy' state (Moore *et al.,* 2007; Peters *et al.,* 2007). The cortex of rhizome, which functions as a storage compartment for carbohydrates and other substances such as resins, latex, oils, and tannins, was also visibly green and reflective at 30% RWC (Figure 2D) suggesting the accumulation of protective sugars and/or proteins and possibly also pigments in this region of the rhizome. Recent metabolomic investigations of this tissue showed that both DT and DS rhizomes at 30% RWC accumulated high levels of sucrose which is one of the primary mechanisms employed by resurrection plants to achieve the glassy state (Wittenberg, 2021, *unpublished*). Taken together, these observations support that possibility that the rhizomes are vitrified at 30% RWC. The samples collected for transcriptomic work, was subjected to the total RNA isolation protocol used for *A. caffrorum* fronds as well as additional other methods used for rhizome organs. None of the methods produced total RNA of sufficient quality and quantity for sequencing (data not shown).

Sufficiently pure RNA of useable yields was ultimately obtained using the hot borate method (Wan and Wilkins, 1994). Yields were, on average, 0.10 ± 0.07mg of total RNA per mg fresh rhizome tissue and the RIN values were between 6.1 and 9.0. The improvement in RIN values, when compared to frond tissue, is likely due to the lack of chloroplasts, and hence absence of chloroplastic RNA, in rhizome tissue (Schroeder *et al.*, 2006). Both sets of RNA were deemed sufficient for short-read and long-read sequencing which was outsourced. The methods developed here were reproducible and can be used in the future for the isolation of total RNA from frond and rhizome tissue of *A. caffrorum*.

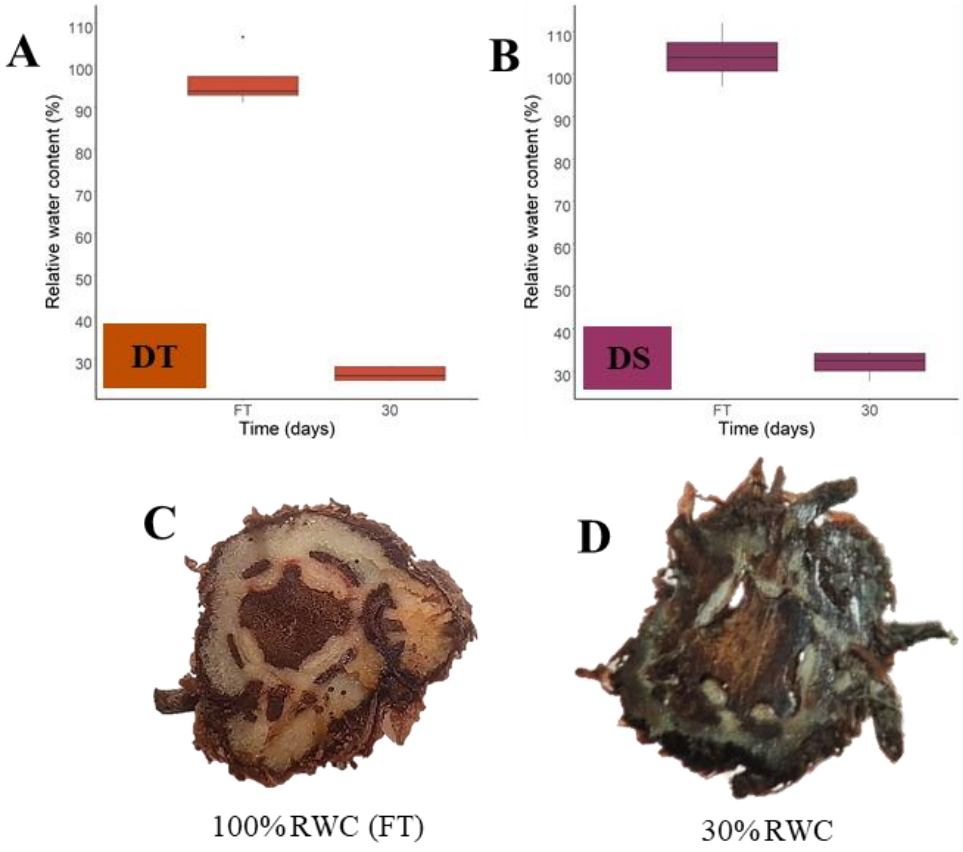

Figure 2: Rhizome dehydration experiment and samples used for transcriptomics. A : DT rhizomes at approximately 100% RWC (fully turgor) and 30% RWC were sampled for transcriptomics. B : DS rhizomes at approximately 100% RWC (fully turgor) and 30% RWC were sampled for transcriptomics. C : Horizontal slices of both DT and DS rhizomes at 100% RWC were visibly similar. D : Horizontal slices of DT and DS rhizomes at 30% RWC had a green cortex with a shiny appearance. In addition to the slices 'cutting like glass', these features suggested possible vitrification of the rhizome at 30% RWC.

5. Conclusions

The required DS and DT frond and rhizome samples of *A. caffrorum* were successfully collected in a single sampling event. Sufficient replicates of fronds in the fully hydrated, mid-phase, late-phase, desiccated and 24-hour post-rehydration states was captured. Rhizomes in the fully hydrated and late-phase/desiccation state were also successfully obtained. RNA isolations were optimized such that the RNA was of sufficient quality was obtained. The samples collected were used short and long read sequencing.

Chapter 2:

Generation of a reference transcriptome

Chapter 2 :

Generation of a reference transcriptome

1. Introduction

1.1. Overview of sequencing technologies used in transcriptome preparation

First generation sequencing (FGS), or Sanger sequencing, was established in the 1970s (Sanger and Nicklen, 1977). Sanger sequencing involved the incorporation of radioactively- or fluorescently- labelled dideoxynucleotides to a template strand. Incorporation of these dideoxynucleotides was chain-terminating and resulted in different size fragments which were analysed by agarose gel electrophoresis (AGE) (Sanger and Nicklen, 1977). While highly accurate, FGS was labour intensive and low throughput (McCombie, McPherson and Mardis, 2019). In the 2005, next generation sequencing (NGS) was introduced. NGS is characterised by massively parallel sequencing which is high throughput and resulted in a dramatic increase in data output (Mardis, 2013). NGS began with second generation or 'short-read' technologies which are led by the Illumina and Ion-torrent platforms. Second generation sequencing (SGS) requires the preparation of DNA and/or RNA through fragmentation, end-repair, and adapter ligation. The resulting fragments are then immobilized on a surface where they are amplified. After amplification they are sequenced, in parallel, on a sequencing machine (Tucker, Marra and Friedman, 2009). Owing to the short read length of SGS this method suffers from the requirement of assembling the fragments *in silico*. Short read sequencing encounters major challenges when presented with low-complexity sequences and structural variants such as transcript isoforms (McCombie, McPherson and Mardis, 2019). Less than a decade later, third generation sequencing (TGS) platforms ushered in the first long read sequencing projects. Long read technologies overcame the challenges of short read sequencing but were initially limited by their read accuracy. Long read sequencing requires minimal library preparation and targets full length DNA and/or RNA which are then sequenced in real-time. Today, some TGS technologies have greater read accuracy than SGS technologies but are not yet widely affordable for the purposes of transcriptome assembly. The leading TGS technologies belong to Pacific Biosciences (PacBio) and Oxford Nanopore technology (ONT) (Logsdon, Vollger and Eichler, 2020).

1.2. De novo transcriptome assembly from Illumina short read data

Currently, *de novo* transcriptome assembly from short read sequencing data is the most prominent method for establishing a transcriptome for non-model organisms that lack a reference genome. This is because RNA sequencing (or RNA-Seq) is less costly and computationally intensive than working with genomic data (Mantione *et al.*, 2014; Stark, Grzelak and Hadfield, 2019). RNA-Seq is also well supported by a vast range of open-source tools which are actively maintained and improved (Han *et al.*, 2015; Raghavan *et al.*, 2022). Raw short read data is collected from the device and assessed using a quality checking tool. The most routinely used tools for this purpose are FastQC and/or MultiQC (Andrews, 2010; Ewels *et al.*, 2016). Even though modern short read sequencers typically achieve 99.9% read accuracy, read quality naturally deteriorates at the end of the read. A trimming tool such as Trimmomatic or TrimGalore/cutadapt is used to trim low accuracy terminal bases. Trimming tools effectively improve the quality of reads according to a user defined quality threshold. They can also be used to remove residual adapters sequences present in the data (Bolger, Lohse and Usadel, 2014; Stoler and Nekrutenko, 2021). The final pre-assembly process that is often applied is kmer correction. Briefly, assembly algorithms use kmers, a sub-sequence of length k derived from the reads, instead of the actual reads in the assembly process. The total set of possible kmers from all the reads are determined and used in the de Bruijn graph reconstruction procedure which results in the reconstructed contigs of the transcriptome. It is therefore essential that a robust set of kmers is available for transcriptome assembly to be most effective (Grabherr *et al.*, 2011; Song and Florea, 2015; Cavallaro *et al.*, 2021). Kmer correction involves the identification of rare/erroneous kmers and their conversion to the most similar, frequently occurring kmer. rCorrector is the main tool used for kmer correction (Song and Florea, 2015).

Transcriptome assembly is the most computationally intensive step and the correction of kmers can drastically reduce graph complexity and therefore the computational processing time (Song and Florea, 2015). The most prominent assembler is Trinity. The main reasons for this are that 1) its performance has been shown as consistently superior to alternative tools, 2) it is highly user friendly and 3) it is well adapted for use with a broad range of downstream tools (Haas *et al.*, 2013; Raghavan *et al.*, 2022). While Trinity has the largest user base, some researchers argue that the best short read assembly tool can only be determined empirically and is unique to each study/set of input data. To simplify this process for the user, there are tools that use multiple assemblers with different parameters to achieve the best possible assembly. High

computational requirements, however, are the barrier to entry to such tools (Haas *et al.*, 2013; Grabherr *et al.*, 2011; Bushmanova *et al.*, 2019; Hölzer and Marz, 2019).

Due to the inherent flaws of each step culminating in an assembly i.e., RNA isolation, library preparation and sequencing, a perfect transcriptome can never be achieved in theory. This is exacerbated by ambiguity in the assembly process that are caused by sequencing artefacts, contamination, repetitive sequences, and chimeras. Since incorrect gene identification can severely affect differential expression studies amongst other downstream analyses, it is vital to assess the quality of an assembled transcriptome (Hansen, Brenner and Dudoit, 2010; Canzar *et al.*, 2016; Mühr *et al.*, 2020; Cavallaro *et al.*, 2021; Freedman, Clamp and Sackton, 2021). There are several approaches used to achieve this. Firstly, a distribution of the transcript model lengths can elucidate the level of fragmentation of an assembly. An assembly with a large proportion of short contigs indicates a high degree of fragmentation, which is usually the result of a poor assembly or poor sequencing (Raghavan *et al.*, 2022). A set of related sequence length statistics such as the N50 are routinely used, but recent updates to this metric have been made and will be discussed in more detail later (Haas *et al.*, 2013; Bryant *et al.*, 2017). The second measurement is the mapping rate. The mapping rate is the percentage of short reads that map back to the assembled contigs. A good quality assembly should have made use of at least 80% of the short reads and a high-quality has a mapping rate of at least 90%. Mapping rates are easily determined using short read mappers such as Salmon and Bowtie2 among other pseudoaligners (Dobin *et al.*, 2016; Patro *et al.*, 2017). Thirdly, the assembly quality can also be measured by the proportion of transcripts that map to 'correct' sequences in an external database. More precisely, the number of transcripts that map to a related target organism as opposed to non-target mappings which could be considered contaminants (Raghavan *et al.*, 2022). In this study for example, *A. caffrorum* mappings to any plant sequence were considered on-target since there is very little representation of fern sequences in external databases (Xu *et al.*, 2018; Marchant *et al.*, 2019). In a similar vein, and the fourth metric commonly used, the benchmarking universal single-copy ortholog (BUSCO) tool can be used to test for the presence of orthologs of genes that are considered universally expressed. The implication is that if a transcriptome has been properly sequenced then orthologs of these curated, universal single-copy genes should be present in the assembly. A good assembly should have at least 80% of the orthologs present, and very few missing or fragmented orthologs. However, it is not uncommon for transcriptome assemblies to have high levels of duplicate orthologs because of the presence of transcript isoforms (Simão *et al.*, 2015; Zdobnov *et al.*, 2021).

De novo transcriptome assembly produces considerably more transcripts than the number of genes present in the genome of an organism (Bryant *et al.*, 2017). This is partly because assemblies can contain sequencing artefacts, contaminant sequences and misassembled transcripts (Freedman, Clamp and Sackton, 2021). Another major contributor to transcript hyperinflation are transcript isoforms (McManus and Graveley, 2011; Zhao, 2019; Zhang *et al.*, 2021). It is not necessary to retain all this additional information when studies such as differential expression analyses are conducted at the gene level. As a result, transcriptome assemblies are routinely subjected to a process of thinning/collapsing. Thinning involves clustering and collapsing similar sequences into a single representative gene for each cluster of sequences likely originating from the same gene. Clustering tools that are routinely used for short read *de novo* assemblies are Corset, Grouper and Compacta (Davidson and Oshlack, 2014; Srivastava *et al.*, 2016; Malik, Almodaresi and Patro, 2018; Razo-Mendivil, Martínez and Hayano-Kanashiro, 2020).

After reducing the transcriptome to the non-redundant set of transcripts/contigs, the final step is transcriptome annotation. Transcriptome annotation is the process of identifying the sequence, sequence features (such as protein domains), gene ontology and the biochemical pathways to which the transcript is connected. This reveals the function, evolutionary properties, and biological relevance of each transcript. Functional annotation commonly leverages sequence translational tools such as Transdecoder to identify protein coding sequences within the transcripts. This distinguishes protein-coding from non-protein coding regions within the transcriptome. Predicted protein sequences can result in greater sensitivity in functional assignments than using only their nucleotide counterparts (Nachtigall, Kashiwabara and Durham, 2021; Raghavan *et al.*, 2022; https://github.com/TransDecoder). There is no standard pipeline for this procedure and the choice of homology search tools and databases are usually determined based on the research question at hand. In this study, the commonly used annotation suite Trinotate was applied with minor adaptation (https://github.com/Trinotate/). Briefly, transcripts and proteins were first identified by conducting homology searches against the highly curated Swissprot database. The Swissport database was prioritized since it contains the highest quality protein sequences of any database (Bateman *et al.*, 2021; Buchfink, Reuter and Drost, 2021). Sequences that were not annotated in the Swissprot database were searched against the UniRef90 database. The UniRef90 database contains protein sequences that have been clustered at 90% sequence identity while hiding redundant sequences. The remaining sequences were searched against NCBI's NR/NT

databases which are lowly curated but is the largest available repository of non-redundant sequence data (Suzek *et al.*, 2007, 2015; Sayers *et al.*, 2021). Sequence features were annotated by performing a sequence profile alignment against the Hidden Markov Model (HMM) database using the tool HMMER3 (Gribskov, McLachlan and Eisenberg, 1987; Eddy, 1998, 2011). Gene ontology annotations were assigned from the top homology search hits (https://github.com/Trinotate/).

Once a transcriptome has been assembled, quality checked and functionally annotated it can be used for downstream analyses. It is worth noting that the methods and tools detailed in this overview represent only a very small set of options available. While there is a very active community creating and maintaining software for these purposes, many researchers believe that long sequencing technology is on the precipice superseding short technology for the purposes transcriptome assembly (Byrne *et al.*, 2019). Long read sequencing (LRS) technology overcomes many of the inherent challenges in working with short read sequencing data and is advancing at a remarkable rate (Byrne *et al.*, 2019). One of the major drawbacks of the short read RNA-Seq approach is that it cannot distinguish transcript isoforms, simply because the reads are too short. Because long read sequencing produces full length transcripts, all splice sites, transcriptional start sites, and polyA sites can be distinguished with relative ease (Kanitz *et al.*, 2015). The principles of the full-length transcriptome generation are the same. The pipeline begins with RNA isolation and library preparation which are assessed very strictly for quality control before sequencing. Post-sequencing quality control is assessed, and the raw data is cleaned so that only high-quality sequences are subsequently used. These sequences are subjected to an 'assembly' process, collapsed and functionally annotated before the non-redundant transcriptome is assessed for quality (Raghavan *et al.*, 2022). Currently, although long read sequencing can overcome some of the major challenges of short read data, it is not without its own limitations.

1.3. The current disadvantages of long read data

The first concern around LRS is RNA integrity. This hurdle is not unique to LRS, but to truly reap the benefits of long read data, mRNA transcripts need to be fully intact. Traditional RNA extraction procedures often isolate RNA that is contaminated with substances such as phenol and guanidium which can compromise RNA integrity. In alternative column-based methods, long mRNA transcripts are usually susceptible to fragmentation. The presence of degraded mRNA reduces sequencing efficacy and results in *in silico* artefacts (Byrne *et al.*, 2017, 2019).

A possible solution for this issue is the coupling of LRS with cap trap technology which reduces 5' degradation. However, cap trap is not yet routinely used in LRS ventures, presumably because it is still too costly. Instead, 5' degraded sequences are more commonly removed during *in silico* processing and the 5' information is lost (Kuo *et al.*, 2019). The second limitation is read length. Longer transcripts occur in comparatively lower abundance within cells. This is exacerbated by the amplification bias and sequencing chemistry that preferentially loads shorter transcripts in the cDNA library to the SMRT cell. Most sequenced transcripts are therefore only approximately 1 - 1.5 kb in length (Picelli *et al.*, 2014; Workman *et al.*, 2019). One method that is routinely used to reduce these effects is a transcript size selection procedure. This process drastically improves the number of sequenced transcripts that are greater than 1.5 kb in size. Size selection is customizable and can therefore cater to the experimental design (Chin *et al.*, 2013). The third consideration is read accuracy. Long reads have gained a reputation for having high error rates. Several tools have been designed to correct inaccurate long reads using highly accurate short reads (Byrne *et al.*, 2019). However, PacBio's circular consensus sequencing (ccs) technology has also been revolutionary in this regard. In the ccs approach, cDNA (derived from extracted RNA) is circularized and sequenced multiple times so that a consensus sequence with 99.0 % accuracy can be inferred. No other competitor has achieved such accuracy in the long-read sequencing realm, but similar technologies are likely to make low accuracy a concern of the past (Tilgner *et al.*, 2014; Gupta *et al.*, 2018). Fourthly, much lower read throughput can limit the ability of long reads to fully capture the true complexity of eukaryotic transcriptomes (Byrne *et al.*, 2019). Sims *et al.*, 2014 argue that a minimum of 30 million long reads are required to perform a surface level analysis of all the isoforms with high and medium expression profiles in human studies. Similar studies have indicated that there are on average 7 isoforms per gene. For some genes there can be as many as 100 isoforms (Pan *et al.*, 2008; Tseng, 2014). Estimates suggest that to truly explore the complexity of the human transcriptome, 100 million long reads would be required. So, LRS would need to produce hundreds of millions of reads at a reasonable cost to supersede short read transcriptome generation that requires greater depth (Byrne *et al.*, 2019). The final and fifth hurdle in the replacement of long read data for transcriptome analysis is data analysis. As full-length transcript sequencing gains traction, more community tools are developed for working with long reads. Currently, this community cannot compete with the range of tools developed for short read *de novo* assembly. Long read tool development is still, comparatively, in its infancy. Nevertheless, interest amongst expert practitioners has increased and so tool development is progressing alongside the technology (Byrne *et al.*, 2019). The tools available

for the development of a reference transcriptome from long reads in a reference-free manner will be discussed here. It is worth noting that tools which can generate a transcriptome with the support of a known genome are also available.

1.4. Transcriptome preparation from PacBio long read data

Library preparation for long reads is based on a similar set of principles as those used for RNA-Seq. Briefly, RNA preparation does not require fragmentation since the focus is to produce full-length transcripts. In addition, size selection and adapter ligation procedures will be specific to the sequencing chemistry and experimental design that is ultimately applied (McCombie, McPherson and Mardis, 2019). Similarly, the sequencing chemistry and related platform will depend on the end-goal of the user.

LRS data has only very recently become highly accurate. Quality control tools therefore focused heavily on correcting the high rate of sequencing errors in long reads. Error-prone long reads were previously error-corrected against highly accurate short read data using existing tools such as Proovread, LSC and LoRDEC. However, if highly accurate long read sequencing is available, then short read quality error correction is no longer required (Hackl *et al.*, 2014; Salmela and Rivals, 2014; Song, Florea and Langmead, 2014). It is worth mention that, in fact, many tools employed to work with long reads still account for the fact that long reads were typically relatively inaccurate. After error correction, the IsoSeq3 package is the most prominent tool used to 'clean' raw sequencing data. IsoSeq3 isolates high-quality, full-length transcripts from the raw circular consensus data. These transcript isoforms are subjected to a hierarchical clustering approach where identical transcript isoforms are collapsed (https://github.com/PacificBiosciences/IsoSeq). Some studies have suggested that IsoSeq3 tends to over-report potential isoforms and therefore this set requires more stringent clustering to collapse the isoforms and remove 5' degradation products that are misinterpreted as unique isoforms (Tardaguila *et al.*, 2018). Alignment of full-length transcripts to one another for such purposes is usually conducted using the tools GMAP and/or BLAT (Kent, 2002; Wu and Watanabe, 2005). The most recently released aligner, *minimap2*, which has comparable performance to GMAP but higher speed, has recently become more widely adopted for the purposes of long-read alignment (Li, 2018).

In the absence of a genome there are two options for collapsing redundant isoforms. The first option is to use CD-HIT to further group and collapse similar isoforms. CD-HIT provides a quick, easy solution to cluster sequences by sequence similarity. Because CD-HIT clusters

based only on percentage of sequence similarity it is not able to discern transcript isoforms that belong to the same gene (Fu *et al.*, 2012; Chen, Wan, Lei, Zobel and Verspoor, 2017). The second approach requires the use of a tool called Cogent which has a slower turnaround due to its more complex nature. Cogent first groups similar sequences into bins using *minimap2* and then applies a more stringent grouping of sequences into unique gene families using the program Mash (Ondov *et al.*, 2016; Chen, Wan, Lei, Zobel, Verspoor, *et al.*, 2017). Using the set of transcripts belong to a gene family, Cogent reconstructs artificial contigs representing all exonic information available for each gene. The collection of these reconstructed contigs is referred to as the 'fake genome'. The transcript isoforms are then collapsed in a mapping-based manner against the 'fake genome' (Tseng, 2014) (Figure 2). Unlike CD-HIT, Cogent provides exon-level information which allows isoforms belonging to each gene to be determined. Once the collapsed transcriptome is available it is subjected to an annotation pipeline desired by the user. In theory, annotation using long read data should have superior performance due to the longer sequence lengths (Raghavan *et al.*, 2022).

As long read sequencing continues to take a foothold, improvements in read accuracy and sequencing depth can be expected. Furthermore, researchers will likely conduct systematic studies to determine appropriate RNA extraction methods which maintain RNA integrity and bioinformatic tool development will likely improve in a similar capacity. Arguably the major hurdle for long read technology is that of read throughput at accessible cost. As the scale of operation of long read sequencing technology increases, cost advantages should allow for a decrease in the cost to provide high read throughput technologies.

1.5. Rationale for the sequencing strategy applied in this study

In this study, Illumina's short read sequencing was chosen for RNA-Seq as it is the most accurate short read technology on the market with a read accuracy of 99.9% (McCombie, McPherson and Mardis, 2019). High accuracy should result in superior performance in the *de novo* short read assembly of a transcriptome and subsequent differential expression studies. The only weakness of Illumina platforms compared to their competing chemistry, Ion-torrent, is the relatively long run time which was not considered detrimental to the study (McCombie, McPherson and Mardis, 2019). The second technology chosen was PacBio's Single Molecule, Real-Time (SMRT) sequencing using the circular sequencing approach which results in 99.8% accurate long reads (Wenger *et al.*, 2019). Although ONT sequencing can produce longer reads, their read accuracy is only between 92 - 93% and was therefore not considered appropriate for

the aims of this study. Furthermore, full-length transcripts sequenced by ccs technology can delineate individual transcript isoforms. The technology has been aptly named Iso-sequencing for this reason (McCombie, McPherson and Mardis, 2019). PacBio offers an affordable compromise of the Iso-sequencing method to assist in the generation of transcriptomes belonging to non-model organisms. In this approach select RNA samples are pooled and sequenced on a single SMRT cell. This method has the potential to result in the loss of important lowly expressed transcripts owing to dilution effect and to further reduce depth of sequencing (Jain *et al.*, 2016). The overall aim of this chapter was to determine whether this approach was suitable for the generation of a high-quality transcriptome of *A. caffrorum*.

2. Aims and Objectives

The overall aim of this chapter was to explore the best possible method to generate a high-quality reference transcriptome for *A. caffrorum*. To this end there were three main experiments that were performed with the following questions in mind :

1. ***De novo* assembly using Illumina short reads**. Currently, this is the most prominent method used to assemble the transcriptome of a non-model organism such as *A. caffrorum*. How does the assembled transcriptome of *A. caffrorum* compare to the transcriptome of other ferns generated by similar methods?
2. **Generation of a reference transcriptome from long reads**. Can a high-quality and complete transcriptome be generated by pooling RNA samples and sequencing on a single SMRT cell? How does the *A. caffrorum* transcriptome compare to other full-length fern transcriptomes?
3. **Hybrid assembly combining short and long read data**. The first major drawback of the Iso-Seq method is low depth of sequencing. Low depth of sequencing can result in the loss of biologically relevant transcripts that are lowly expressed. Furthermore, in this study, long read sequencing was conducted on a single SMRT cell due to budget constraints. This required the pooling of RNA isolated from different tissues (frond and rhizome) under different stressors (dehydration, desiccation, and rehydration) which can further contribute to the loss of lowly expressed transcripts due to dilution effect. A second drawback is the higher rate of 5' degradation that occurs when long read sequencing is not conducted in conjunction with cap trap technology as was done in this study. To explore whether short and long read data can be combined to reduce the effects of low depth and 5' degradation a hybrid assembly method was attempted. In the hybrid assembly method, the raw long read data was combined with the *de novo* assembled transcript models and was used as input to the Cogent/cDNA_Cupcake pipeline. Furthermore, to explore the identity of the possible additional genes that may exist in the short read data, the complement of the long-read transcriptome in the short read data set was investigated. The two major questions explored here were: Can the higher depth of sequencing from short read sequencing be leveraged to capture low abundant transcripts that may be lost due to lower depth of long read sequencing? Can the short read data also be used to capture information lost to the effects of 5' degradation?
4. **Comparison of the assemblies**. Which assembly had the highest quality?

3. Methods

3.1. Library preparation and Single Molecule Real Time (SMRT) sequencing

A subset of the plant material was used for Single Molecule, Real-Time (SMRT) sequencing. The 8 samples highlighted in pink in Table 1 (See Chapter 1 methods) were used for SMRT sequencing. These samples represent those with the highest RIN values. Thereafter the most diverse set of tissue (frond/rhizome) and conditions (RWCs) was chosen. Quality control, library preparation and sequencing were conducted by Inqaba Biotec (Pretoria, South Africa). The concentration and purity (Abs260/280nm and Abs260/230nm) of the total RNA samples was assessed using a NanoDrop 2000 spectrophotometer (ThermoScientific). The quality i.e., RIN values, were determined using a TapeStation System. Total RNA from all the samples was pooled before cDNA conversion and amplification was conducted using the NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Module according to the manufacturer's instructions. cDNA was size selected for 2kb length transcripts and PacBio libraries were prepared with the SMRTbell Template Prep kit 2.0. Qubit RNA BR Assay kit and subsequent analysis on a Qubit 4.0 Fluorometer and (ThermoScientific) was used to quantify the library and the quality was assessed using a TapeStation System. Sequencing primer v4. Binding kit 2.1/control 1.0 was used to prepare the library for sequencing. Polymerase bound complex was loaded onto a SMRTcell 8M and sequenced using Sequencing plate 2.0 on the Sequel IIe system.

3.2. Pre-processing the long-read data

The IsoSeq3 pipeline was used to extract the high-quality (HQ) isoforms from the full set of circular consensus sequences (ccs) generated by the Sequel IIe instrument (Figure 1A) (https://github.com/PacificBiosciences/IsoSeq). Extracthifi (version 1.1.0) was used to filter HiFi reads from the ccs reads (https://github.com/PacificBiosciences/extracthifi). HiFi reads have quality score of Q>20 and are considered 99.9% accurate. Reads containing the 5' and 3' cDNA primers as well as a polyA tail of at least 20bp was considered full-length (FL) and were isolated from the HiFi set using Lima (version 2.4.0-1) (https://lima.how/). After identification of FL reads, Lima removed primers were and reorientated the sequences to the 5' to 3' direction. IsoSeq3 version 3.4.0 was used to remove polyA tails and concatemers before clustering the isoforms to generate the set of full-length non-concatemer (FLNC) high-quality isoforms (HQ isoforms) (Figure 1A).

3.3. Generation of the transcriptome from long reads

The FLNC HQ isoforms were used in the Cogent assembly pipeline to generate a 'fake genome' (Figure 2B). *Minimap2* (2.23-r1117-dirty) was used to roughly group FLNC HQ isoforms into clusters based on sequence similarity (Li, 2018). Sequences that were not assigned to a cluster were either considered 'orphans' (those that did not share sequence similarity with any other sequence in the set) or 'tucked' (sequences which are sub-sequences contained within another transcript in the set). Mash (version 2.3) was used to identify likely gene families within each cluster (Ondov *et al.*, 2016). Sequences in each cluster that were not assigned to an identified gene family was considered 'unassigned' (Figure 1B). Coding genome reconstruction tool (Cogent) version 8.0.0 was used to reconstruct contigs from sequences belonging to each gene family. Reconstructed contigs are artificial sequences composed of all possible exons observed in the set of sequences (Figure 2). The reconstructed contigs, 'unassigned' sequences and 'orphan' transcripts were combined to serve as the 'fake genome' (Tseng, 2014). The FLNC HQ isoforms were mapped against the 'fake genome' using *minimap2*, and the isoforms that mapped to genome were collapsed using cDNA_Cupcake 28.0.0 (https://github.com/Magdoll/cDNA_Cupcake). These isoforms were further reduced by removing sequences that were supported by less than two FL reads and by the removal of isoforms representing 5' degradation products of other isoforms. The collapsed, non-redundant set of isoforms was used for annotation (Figure 3).

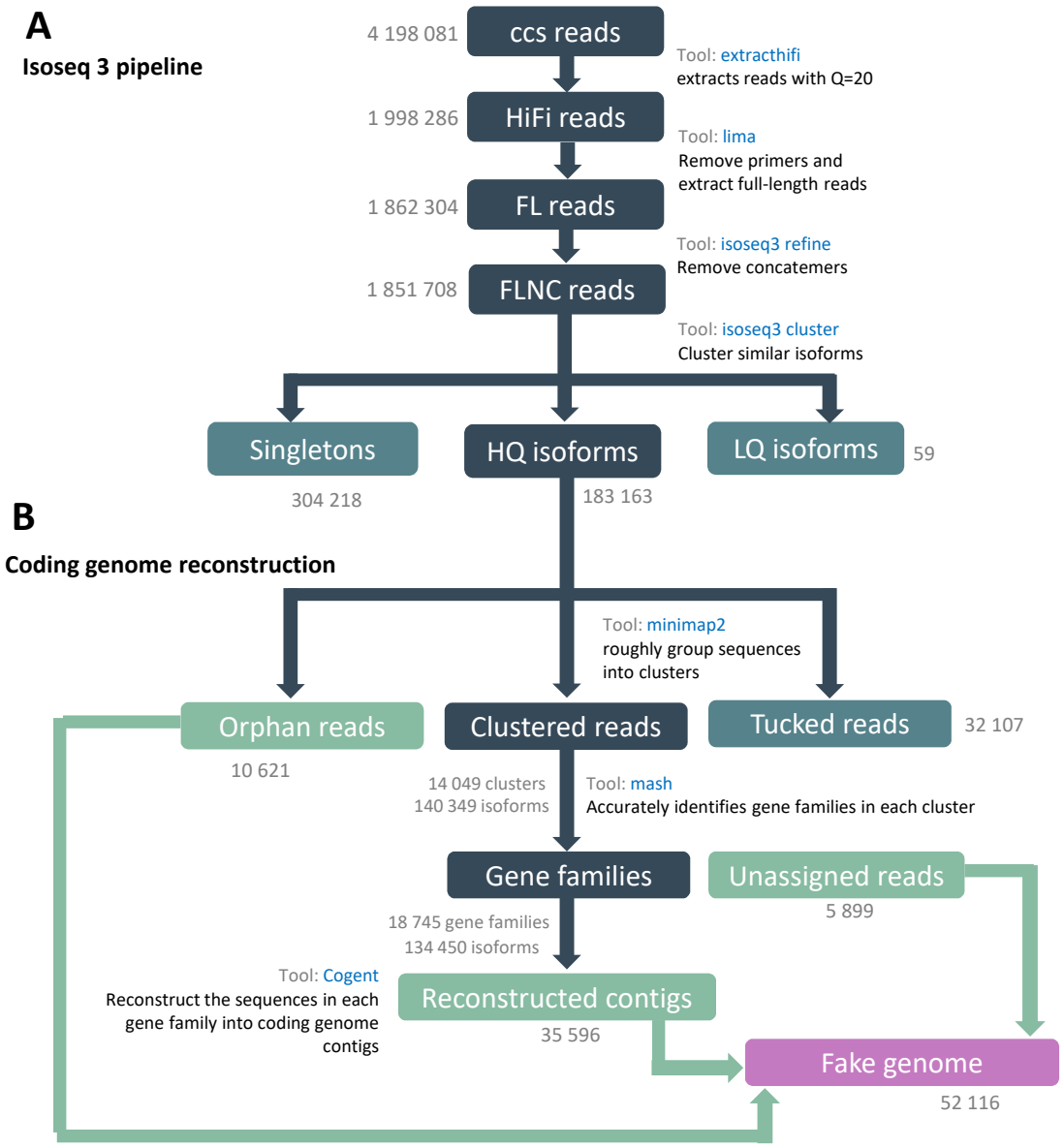

Figure 1: Generation of the fake genome. **A**: The IsoSeq3 pipeline starts by extracting HiFi reads from the circular consensus sequences (ccs). Transcripts that have a 5' primer, 3' primer and a polyA tail are extracted as full-length (FL) reads before their primers are removed. Concatemer reads are removed before the remaining full-length non-concatemer (FLNC) are clustered using the IsoSeq3 cluster algorithm. FLNC reads are categorized as high, or low-quality (H/LQ) isoforms and singletons are noted. **B**: Coding genome reconstruction begins by roughly grouping the HQ isoforms into clusters based on sequenced similarity. Reads that are sub-sequences (tucked) or do not share sequence similarity (orphans) to any other sequence in the set are removed. Clustered reads are more finely clustered into possible gene families. Sequences that do not fit in a gene family are termed unassigned. Isoforms that are part of a gene family are used to reconstruct the minimal number of contigs required to represent the gene family. The unassigned isoforms and the reconstructed contigs are combined to serve as the 'fake genome'.

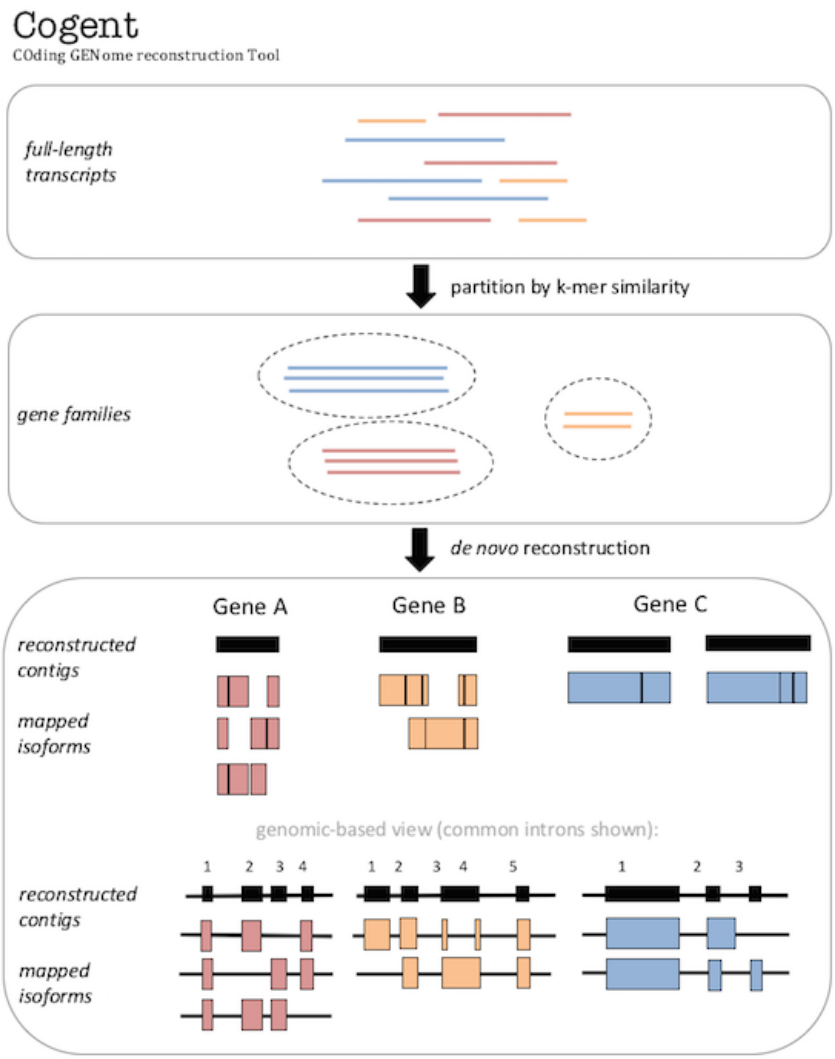

Figure 2: Coding Genome reconstruction tool (Cogent) principle. Cogent used a set of full-length transcripts, such as those generated after the IsoSeq3 pipeline, as input and partitions them into gene families based on kmer similarity. Cogent then uses the De Bruijn graph approach to reconstruct the transcribable regions of each gene into contigs. It attempts to construct 1 contig per gene family (orange and red) but in cases where there is insufficient information to resolve a single path, Cogent will output the minima number of contigs required to represent all possible exons in the gene family (blue; an example where Cogent produce two separate contigs for one gene). Image reproduced from (Tseng, 2014).

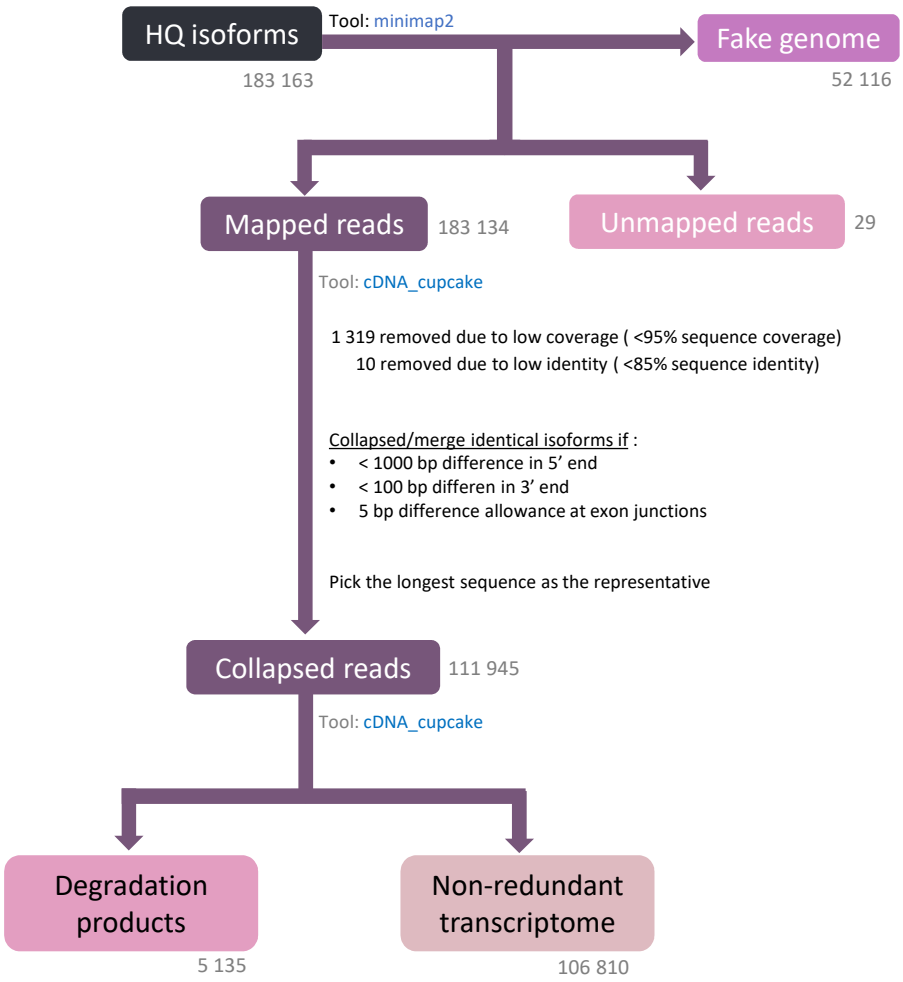

Figure 3: Generation of the non-redundant transcriptome from high-quality isoforms. The total set of high-quality isoforms from the IsoSeq3 pipeline were mapped against the fake genome using the long-read aligner *minimap2*. Unmapped reads and reads that mapped with sequence coverage < 95% and sequence identity < 85% were discarded. The remaining mapped reads were collapsed based on aligned exonic structure. More specifically, isoforms were considered identical if they had 1) < 1000 bp difference in their 5' end, 2) less than < 100bp difference in their 3' end and no more than a 5bp difference at exonic junctions. The isoforms chosen to represent the set of isoforms was the one with the longest sequence. The collapsed set was further interrogated to identify isoforms that represented 5' degradation products of one another as opposed to unique isoforms. The degradation products were removed to generate the final non-redundant long-read transcriptome.

3.4. Library preparation and short read sequencing

Messenger RNA (mRNA) was isolated from total RNA using the NEXTflex™ Poly(A) Beads and to prepare the directional, strand specific RNA libraries the NEXTFLEX® Rapid Directional RNA-Seq Kit was used according to the manufacturer's instructions. The libraries were measured using the Qubit DNA HS Assay and fragment size distribution of each library was determined by Bioanalyzer High Sensitivity DNA Chip Kit (Agilent). The concentration and fragment size distribution of the pooled libraries were confirmed using the qPCR-based Roche Library Quantification Kit and the Bioanalyzer High Sensitivity DNA Chip Kit. A PhiX positive control (Illumina) was spiked into the final pooled library at 1% to evaluate the efficiency of the sequencing reaction. Sequencing was performed on the Illumina NextSeq 550 instrument using two NextSeq 500/550 High Output Reagent Kits v2.5 chemistry (150 cycle). Up to approximately 21 million 2X 75 bp paired-end sequencing reads per sample were generated.

3.5. Processing the short read data and transcriptome assembly

The quality metrics of the raw short read data was assessed using FastQC v0.11.08 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) (Figure 4A). Rare kmers in the reads were converted to more frequently occurring kmers using rCorrector v1.0.4 (Song & Florea, 2015). Erroneous kmers that could not be fixed were removed using a python script in the TranscriptomeAssemblyTools collection. TrimGalore v0.6.7 was used to remove the adapters: AGATCGGAAGAGCACACGTCTGAACTCCAGTCA and AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT, and low-quality base pairs using the following parameters: --paired --retain_unpaired --phred33 --length 36 -q 5 --stringency 1 -e 0.1 (Martin, 2011; Krueger, 2012). Trimmed reads were mapped against the rRNA SILVA database (downloaded on the 12th January 2022) to remove contaminant rRNA sequences. To achieve this, the SSUParc and LSUParc fasta files were concatenated, converted to DNA and indexed using Bowtie2 v2.4.5. The following parameters were used in Bowtie2 to remove reads that mapped to the rRNA database: --quiet --very-sensitive-local --phred33 –nofw (Langmead & Salzberg, 2012; Langmead et al., 2019). The final, processed sequences were assessed using FastQC v0.11.08 before they were used to assemble the transcriptome (Figure 4B). *In-silico* normalization was run using Trinity version 2.10.0 with the following arguments: --max_cov 30, --pairs_together and --SS_lib_type RF and thereafter, assembly was conducted (Haas *et al.*, 2013; Grabherr *et al.*, 2011). Salmon version 1.6.0 was used to determine the mapping rate of

the short reads to the trinity assembly (Roberts et al., 2011; Roberts & Pachter, 2012; Love et al., 2015; Chen et al., 2017; Patro et al., 2017; Suzuki & Kasahara, 2018). Salmon quantification data was used to generate gene expression matrices for each sample. The trimmed mean of m-values (TMM) value was averaged for each condition (to obtain the TMM on a per biological replicate basis) and the full trinity transcript set was filtered using a TMM value of 1. If the greatest TMM value for all conditions was less than 1 the transcript was removed; otherwise, the transcript was retained (Figure 4A). The short reads were mapped against the trinity (TMM > 1) set of transcripts and the quantification data was used to cluster the transcripts using Corset version 1.06 using the –hardFilter argument. Corset was run with the -D parameter set to 999999 to ensure that transcripts within a gene with differential transcript usage were not split into different clusters (Davidson and Oshlack, 2014) (Figure 4B).

3.6. Generation of the hybrid assembly

The set of trinity-generated transcript models (TMM > 1) was combined with the HQ isoforms and used as input to the Cogent/cDNA_Cupcake pipeline to generate the hybrid assembly (Figure 1 and 3).

3.7. Complement of the long-read transcriptome from short read data

The short read data was mapped against the trinity contigs with a TMM > 1. The generated mapping data was used to cluster the transcripts using Corset version 1.06 as described above (Davidson and Oshlack, 2014). Concurrently, the TMM > 1 transcript set was mapped against the HQ isoform set that was obtained through the IsoSeq3 pipeline using blastn (Figure 1A). Corset clusters (all transcripts within the cluster) were removed if at least one transcript mapped to the HQ isoform set. The final set of transcripts was considered complement of the long-read transcriptome generated from the short read data (Figure 4B).

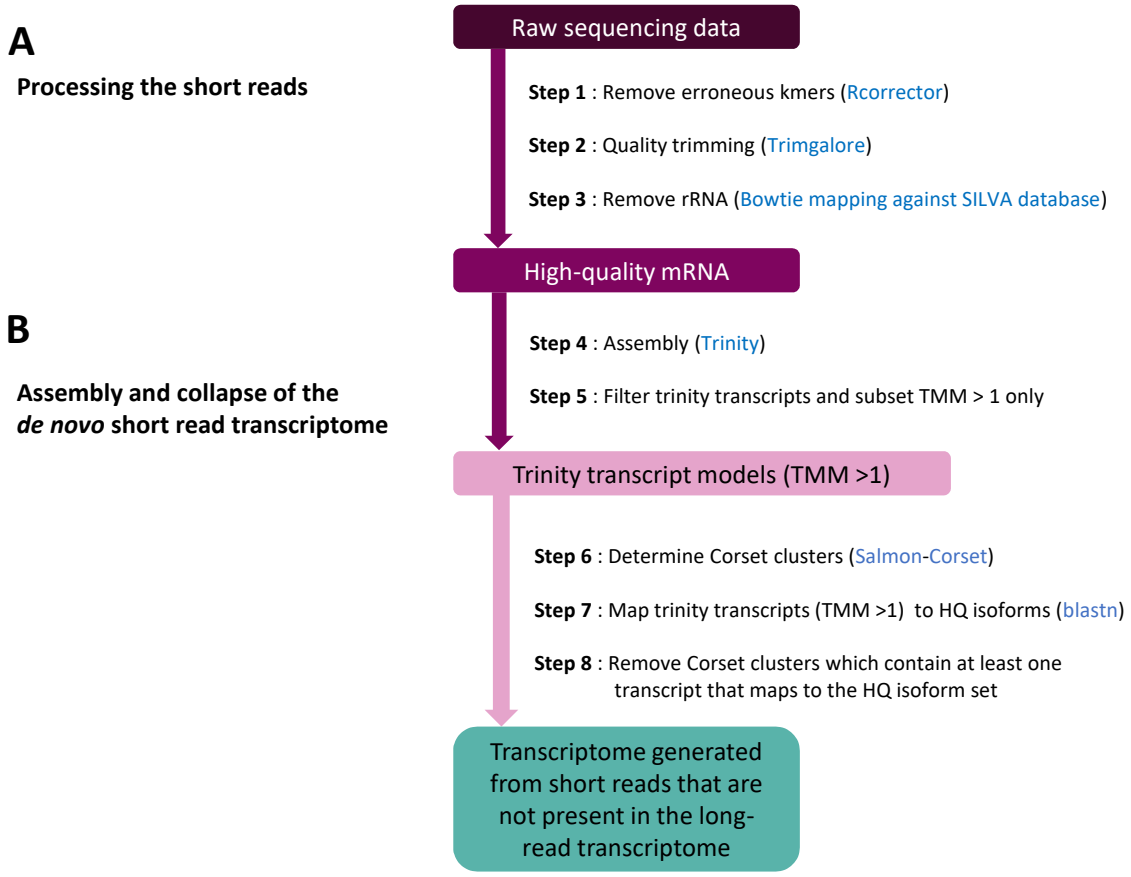

Figure 4: Generation of the complement of the long-read transcriptome from short read data. **A**: High-quality mRNA sequences were obtained by processing the raw sequencing data. Erroneous kmers were removed from the raw reads using Rcorrector before they were quality trimmed using TrimGalore. Transcripts representing rRNA were identified using Bowtie2 to map the short reads against the LSU and SSU SILVA database. rRNA was removed from the raw reads for transcriptome assembly. **B**: High-quality mRNA was assembled using Trinity, which was guided by the high-quality, full-length non-concatemer isoforms (HQ isoforms) obtained from the IsoSeq3 pipeline. The average TMM for each biological condition was calculated for each trinity-assembled transcript and the full set of transcripts were subjected to a TMM filter. Transcripts with a TMM < 1 over each condition were removed. **C**: To further reduce the short read transcriptome several additional filtering steps were conducted. The short reads were mapped to the TMM > 1 transcript set using Salmon and gene clusters were identified using Corset. Concurrently, the TMM > 1 transcript set was mapped to the HQ isoforms using blastn. Corset clusters that contained at least one mapped transcript were removed.

3.8. Annotation of the transcriptomes

Likely coding regions within the set of collapsed, non-redundant transcripts were identified using Transdecoder v5.5.0 (https://github.com/TransDecoder/TransDecoder0) (Figure 5). The single best opening reading frame (ORF) was determined using the --single_best_only argument. The set of transcripts and ORFs were subjected to a series of sequence homology searches. In the first phase, all transcript/protein sequences were aligned by blastx/blastp respectively to the Swiss-Prot database (downloaded on the 28th of April 2022). Sequences that were not annotated in this database were then aligned to the Uniref90 database (downloaded on the 28th of April 2022), and the remaining set were finally aligned to the nr/nt database (downloaded on the 28th of April 2022). Hmmer version 3.3.2 was used to identify conserved domains by running the predicted protein sequences against the Pfam 35.0 database (downloaded on the 28th of April 2022). Trinotate version 3.2.2 was used to generate an annotation report from all the sequence homology search results (https://github.com/Trinotate) (Figure 5).

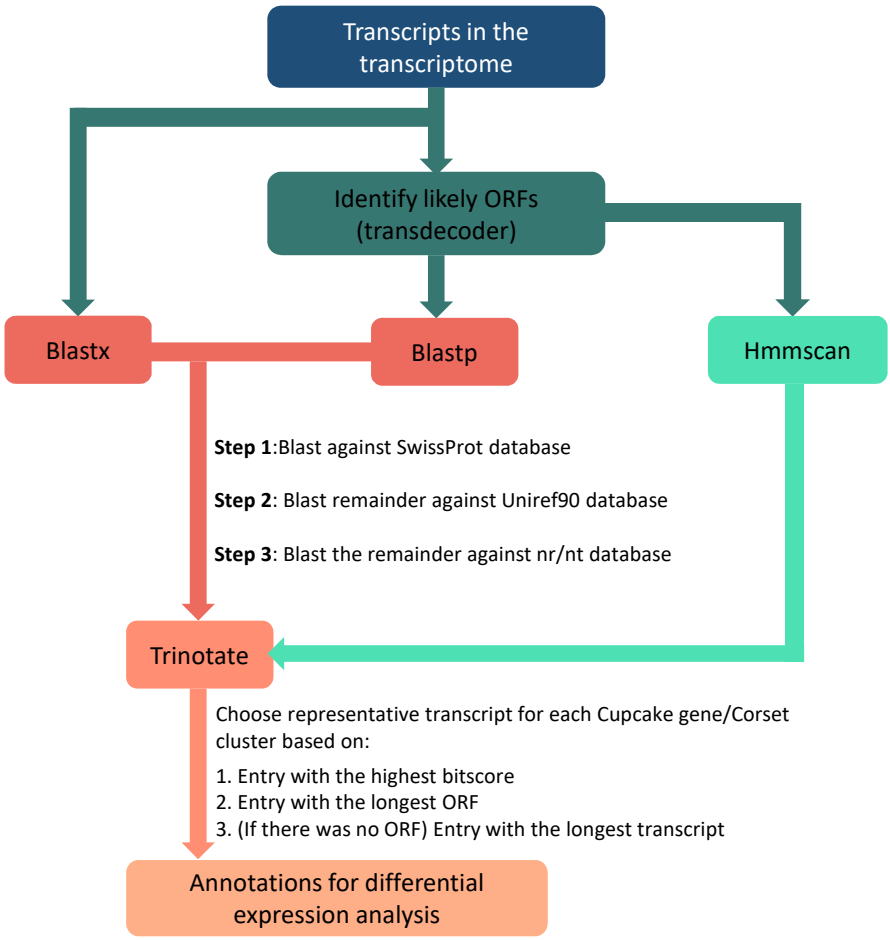

Figure 5: Annotation of the transcriptomes. The final transcriptome annotated through a sequence of homology searches. First, likely opening read frames were identified using transdecoder. Predicted ORFs were searched against the Pfam database using hmmscan in the hmmer suite. The predicted ORFs and the transcripts were then subjected through a sequential set of homology searches using blastp and blastx respectively. First all the transcripts and predicted protein sequences were searched against the Swissprot database. The sequences that were not annotated were then searched against the Uniref90 database and the remaining sequences were searched against the nr/nt databases. All the homology search results were compiled into a report using trinotate and the representative transcript for each gene cluster was chosen based on the highest bit score, then ORF length and transcript length where there was no ORF.

4. Results and Discussion

4.1. Assembling the long-read transcriptome

The IsoSeq3 pipeline was used to process 5.96 GB of raw sequencing data into a set of FLNC HQ isoforms. A total of 4 198 081 raw circular consensus sequencing (ccs) reads were obtained from the Sequel II instrument (Figure 1A). Approximately half (1 998 286) of these were HiFi reads (reads with a quality score of Q20 ≥) and a total of 1 862 304 were considered full length (contained the 5' primer, 3' primer and a polyA tail of at least 20 bp). Owing to amplification bias during long read sequencing, most sequenced transcripts are approximately 1 - 1.5 kb in length (Picelli *et al.*, 2014; Workman *et al.*, 2019). Size selection can be used to reduce amplification bias. The sequence length distribution of the HiFi reads demonstrates the successful use of 2 kb size selection procedure (Figure 6). From this set, 10 596 concatemer sequences were removed before the remaining reads were categorized into 183 158 high-quality (HQ) isoforms, 304 218 singletons and 59 low-quality (LQ) isoforms (Figure 1A).

Previously, LQ isoforms were error corrected against highly accurate short reads but because the proportion of LQ isoforms to HQ isoforms was so low (approximately 0.03% of the total) they were discarded (Hackl *et al.*, 2014; Salmela and Rivals, 2014). The negligible amount of LQ isoforms produced by the sequencing is a testament to the success of the circular ccs approach which produces highly accurate long read data (Tilgner *et al.*, 2014; Gupta *et al.*, 2018). Singletons are a set of full-length transcripts derived from a template that was sequenced only once in the sample. Singletons are therefore discarded owing to insufficient read support. A total of 89.9% singletons mapped to the fake genome suggesting that most of the sequencing information in this set was captured in the long-read transcriptome. Mapped singletons could represent orthologs or degradation products that were slightly too divergent to be categorized as an isoform. The 10.1% unmapped singletons were subjected to homology search against the Uniref90 database where 82.2% had a blastx hit. Of these, 20.8% matched plant proteins and the remaining 79.2% were non-plant. At least 63.0% of the non-plant species were fungal, 11.5% were metazoan and 0.9% were bacterial. This suggests that the unmapped singletons were likely contaminant sequences derived from other organisms.

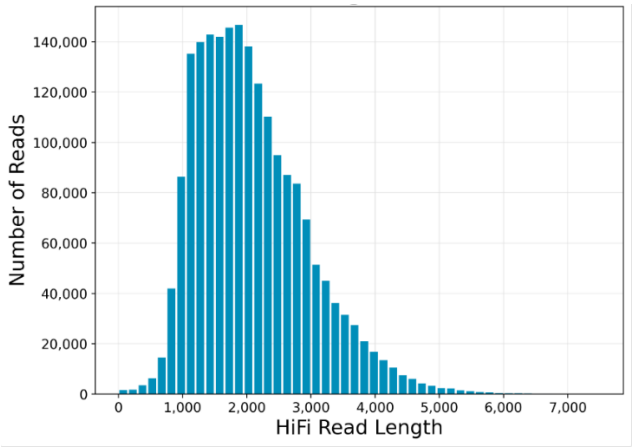

Figure 6: Most transcripts from the 2kb size selected library sequenced on the Sequel IIe sequencing platform were greater than 1.5 kb in size.

The remaining HQ isoforms are full-length transcripts that have sufficient read support to be used for reference transcriptome assembly. Because IsoSeq3 tends to over-report potential isoforms (Tardaguila *et al.*, 2018), the HQ isoforms were subjected to clustering using the two available tools – CD-HIT and the Cogent/cDNA_Cupcake pipeline. The set of HQ isoforms were collapsed by each method and the resulting sets were compared. To generate the fake genome required by the Cogent/cDNA_Cupcake pipeline, the full set of 183 158 HQ isoforms were aligned to one another and roughly grouped into clusters based on sequence similarity (Li, 2018) (Figure 1B). A total of 14 049 clusters encompassing 140 349 of the HQ isoforms were identified. Isoforms that did not share sequence similarity to any other sequence in the set were assigned to the 'orphan' category. Orphans are therefore transcripts which were sequenced at least twice, but do not share sequence similarity with any other transcript in the set. Orphans have a high likelihood of representing rare artifacts or low-level contaminants. Theoretically, they could also represent paralogs that are simply too divergent to be identified as similar, but this is less likely. A total of 10 621 orphans were identified (approximately 5.7% of the total HQ isoforms). Approximately 90% of orphans did not map to fake genome. Of the unmapped orphans only 10% were identified as plant proteins in the Uniref90 database. Approximately 80% were identified as having fungal origins. Despite orphans representing a low confidence set of *A. caffrorum* transcripts they are still retained in the fake genome according to the software's instructions. Many researchers caution against the removal of such low confidence sets on the basis of high levels of off-target taxa since there may still be interesting and biologically important transcripts within the set (Wang and Huang, 2020). The

final set of transcripts were termed ‘tucked’ reads. They represented isoforms that were sub-sequences of at least one other isoform in the set. There were 32 107 (approximately 17.5% of the total HQ isoforms) tucked reads and since they do not contribute any new information to the transcriptome they were removed (Figure 1B). Within each of the 14 049 clusters, the programme Mash was used to robustly identify separate gene families based on kmer similarity (Ondov *et al.*, 2016). A total of 18 745 gene families were identified which encompassed 134 450 HQ isoforms. Sequences in a cluster that were not identified as belonging to a defined gene family were categorized as ‘unassigned’. Unassigned sequences may represent paralogs that share sequence similarity to other sequences in the cluster but are slightly too divergent to belong to a clearly defined gene family. Because these sequences were not similar enough to contribute to the reconstruction of contigs, they were not included in the assembly process (Figure 1B). The transcripts assigned to each gene family were then used in a de Bruijn graph reduction procedure that preserves the exon order information from all the transcript isoforms and outputs the minimal amount of reconstructed contigs required to represent the gene family (Figure 2). The 18 745 gene families produced 35 569 reconstructed contigs, which is an average of 2 contigs per gene. The fake genome was defined as the combination of the 35 569 reconstructed contigs, 10 621 orphan reads and the 5 899 unassigned sequences (Tseng, 2014) (Figure 1B).

To generate the non-redundant reference transcriptome, the 183 158 HQ isoforms were aligned to the fake genome using *minimap2* (Figure 3). Identical isoforms identified based on aligned exonic structure were collapsed using the cDNA_Cupcake algorithm (https://github.com/Magdoll/cDNA_Cupcake) (Figure 3). A total of 29 isoforms did not map to the fake genome, therefore there was a mapping rate of 99.9% which indicates that the reconstructed fake genome is a high-quality, accurate representation of the total HQ isoforms. Of the 183 134 mapped isoforms, 1 319 were ignored due to low coverage (< 95% sequence coverage) and 10 were ignored due to low sequence identity (< 85%) (Figure 3). Isoforms ignored due to low sequence coverage or identity represented less than 1% of the total set. The remaining 181 805 sequences were collapsed to 111 945 transcript isoforms. A total of 5 135 degradation products (approximately 4.3% of the total HQ isoforms) were identified in the collapsed set and removed. After removal of 5’ degraded isoforms, the final non-redundant set contained 106 810 transcript isoforms which represented 51 483 gene clusters (Figure 3). The reduction of 183 158 HQ isoforms to 51 483 genes was a collapse rate of 3.6X (Table 1). To

compare the efficacy of the Cogent/cDNA_Cupcake pipeline in clustering and collapsing the long-read data, CD-HIT was employed as an alternative.

Table 1 : Comparison of the performance of the short read, long read and hybrid assemblies

	Short read assembly			**Long read assembly**			**Hybrid assembly**	
	Transcripts	Genes (Trinity)	Genes (Corset)	HQ isoforms	Genes (CD-HIT)	Genes (Cogent-Cupcake)	Transcripts	Genes (Cogent-Cupcake)
Total transcripts/genes	502 119	336 307	336 331	183 158	67 537	51 483	334 897	250 967
Collapse rate (transcripts/gene)	1.0	1.5	1.5	1.0	2.7	3.6	1.0	1.3
Total ORFs	185 700	99 728	126 078	174 589	62 939	47 630	185 720	112 271
Percentage ORFs (%)	37	30	37	95	93	93	55	45
Number annotations in Uniref90 database	133 298	71 908	89 081	133 298	53 735	40 287	153 461	88 685
Annotation rate (%)	72	72	71	76	85	85	83	53
Plant (%)	43	24	30	78	67	65	63	49
Fern (%)	6	5	5	12	6	10	10	8
Fungal (%)	42	60	54	6	21	19	22	35
Other (%)	9	11	11	4	6	6	6	8
Mapping rate	95	95	95	92	92	92	96	96
Minmum mapping rate	90	90	90	84	84	84	90	90
N50 (bp)	1 406		1 501	2 415	2 573	2 490	2 252	
E90N50 (bp)	2 167		2 172	1 858	2 183	1 944	2 210	
Number of genes corresponding to the E90 peak	21 266		19 427	46 855	17 313	27 398	79 337	

As discussed in the introduction, CD-HIT is the quick and easy solution to collapsing transcripts based on sequence similarity (Fu *et al.*, 2012; Chen, Wan, Lei, Zobel and Verspoor, 2017). However, one must be cognizant of the difference between genome-based and sequence-based collapse methods (such as those used by CD-HIT). While sequence-based methods have their utility, they also reduce the amount of important information that can be used to collapse a set of sequences. Firstly, since no genome mapping is available, they cannot discern exon-level differences/similarities between transcript isoforms. The second set of information that is lost in sequence-based approaches is the ability to quality filter for alignment coverage and identity (Tseng, 2014). While it is possible to customize CD-HIT parameters to achieve near or similar performance to tools such as Cogent/cDNA_Cupcake, the analytical time and cost of such an approach heavily outweigh the benefits. It is also worth highlighting that some studies still use CD-HIT exclusively to reduce transcriptomes; presumably owing to its comparative ease of use. As an example, the non-redundant full-length transcriptome of the fern *Adiantum flabellulatum* (*A. flabellulatum*) was generated using only CD-HIT (Cai *et al.*, 2022). In a similar study the full-length transcriptome of the fern *Alsophila spinulosa* (*A. spinulosa*) was first collapsed using CD-HIT. Thereafter, Cogent was used to generate a fake genome and GMAP (an earlier iteration if the long-read aligner *minimap2*) was used to further thin the transcriptome (Hong *et al.*, 2022). Interestingly, both approaches performed similarly in terms of their collapse rate. In this study, CD-HIT was simply used as a benchmark against which to compare the Cogent/cDNA_Cupcake pipeline that was chosen owing to its greater depth of exon-level information. Firstly, CD-HIT was used to collapse the 183 158 HQ isoforms at several sequence identities (Figure 7). At the highest sequence identity of 99%, the set of transcripts were reduced 2.1X to a set of 89 498 sequences. At 80% sequence identity the set of transcripts were reduced to 44 230 sequences (or 4.1X) (Figure 7). Interestingly, CD-HIT clustering at 85% resulted in a 3.6X collapse rate (Figure 7) which was like the 3.6X collapse rate observed in the Cogent/cDNA_Cupcake pipeline which also uses a sequence identity of 85% (Table 1). But because the Cogent/cDNA_Cupcake pipeline employs several more advanced filtering and clustering steps this is not directly comparable. Instead, a more conservative sequence identity value was chosen to compare to the Cogent/cDNA_Cupcake generated transcriptome. At 95% sequence identity, 67 537 transcripts remained which was a collapse rate of 2.7X (Table 1). In both transcript sets, 93% of the sequences contained likely protein coding sequences and 85% of the ORFs were annotated in the Uniref90 database. The annotated set of transcripts in the CD-HIT collapsed transcripts contained similar amounts of ‘plant’ and ‘fungal’ annotations. However, the CD-

HIT collapsed set had a lower number of 'fern' annotated transcripts (Table 1). This could be an indication of the oversimplification of the CD-HIT algorithm. Both transcriptomes had similar mapping rates which indicated that they were both a highly quality representation of the underlying data (Table 1). The N50 and Ex90N50 metrics were designed to assess transcriptomes generated by short read *de novo* assembly and will be discussed in more detail later. For now, only the Ex90N50 and the corresponding E90 peak value is of interest. Briefly, the Ex90N50 is the average transcript length for all the transcripts that contribute to 90% of the expression data. The E90 peak value is the actual number of transcripts that account for 90% of the expression data. The Ex90N50 for each non-redundant long read transcriptome was similar but the Ex90 peak value distinguishes the CD-HIT sequence-based collapse from the Cogent/cDNA_Cupcake genome-based collapse approach (Table 1). The CD-HIT collapsed transcriptome accredited only 17 313 genes to 90% of the expression data, while the Cogent/cDNA_Cupcake collapsed transcriptome ascribed 27 398 genes (Table 1). The E90 value of the CD-HIT collapsed long read transcriptome is the lowest of any transcriptome generated in this study (Table 1). It is also considerably lower than the E90 peak value for the short read transcriptomes of *A. caffrorum* and the fern *Vandenboschia speciosa* (*V. speciosa*) (Kunkel, 2021) (Table 2). These comparisons may indicate that the CD-HIT collapse has been too simplistic and as a result may have collapsed transcripts belonging to separate, but highly similar gene families into a single cluster. While this has not been definitively proven, it supports a major and common criticism of CD-HIT – that is that its algorithm is too simplistic to accurately reduce complex transcriptomes into a redundant set.

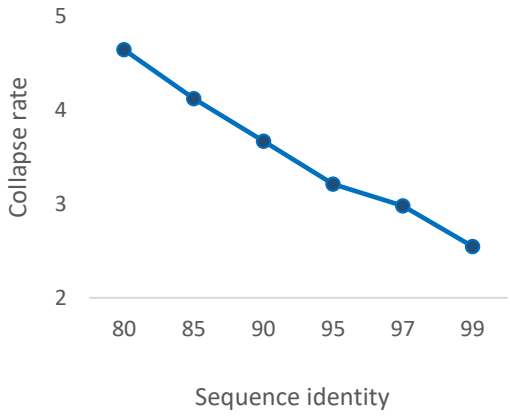

Figure 7: The collapse rate of the HQ isoforms by CD-HIT under different sequence identity scores. The collapse is the total number of HQ isoforms divided by the number of transcript isoforms remaining after being clustered by CD-HIT.

The Cogent/cDNA_Cupcake reduced transcriptome was therefore chosen moving forward. To support this decision, the long read transcriptome was subjected to a more in-depth analysis and compared to two other fern transcriptomes generated by similar methods (Cai *et al.*, 2022; Hong *et al.*, 2022). Briefly, the full-length transcriptome of *A. spinulosa* was generated by sampling root, rachis and pinna tissue and sequencing each tissue type on its own SMRT cell (Hong *et al.*, 2022). It is therefore reasonable to expect greater depth from this study. The depth of sequencing of the *A. flabellulatum* transcriptome was not clearly indicated but compromised only of gametophytic tissue cultured under various lighting conditions (Cai *et al.*, 2022). As previously mentioned, the *A. spinulosa* transcriptome was also built by first running all the FLNC reads through CD-HIT. Subsequently, Cogent was used to reconstruct contigs and GMAP was used to collapse redundant transcripts. Whereas the transcriptome of *A. flabellulatum* was collapsed only using CD-HIT. From a total of 183 158 HQ isoforms, 51 438 gene families were identified in the transcriptome of *A. caffrorum* (Table 1). The *A. spinulosa* transcriptome contained 278 357 HQ isoforms which were reduced to 38 470 gene families. The collapse rate of the *A. caffrorum* transcriptome was 3.6X which was lower than the *A. spinulosa* and *A. flabellulatum* transcriptomes which were both approximately 7.2X. So, although CD-HIT seemed to over-collapse the *A. caffrorum* transcriptome, Cogent/cDNA_Cupcake may have not fully reduced the transcriptome either (Table 1). However, over-filtering can result in the removal of important, biologically relevant genes which can be more detrimental than a somewhat inflated transcriptome. A total of 31 375 (61%) of gene families contained only 1 isoform per gene, 16 905 (33%) contained 2 - 5 isoforms per gene and 3 179 (6%) contained between 10 - 50 isoforms per gene. A similar proportion of gene families containing only one isoform was observed in the root, rachis, and pinna tissues of *A. spinulosa* (Hong *et al.*, 2022). The largest gene families containing more than 40 isoforms were annotated as ‘transcribed RNA sequence’ and ‘pentatricopeptide repeat-containing protein’. The pentatricopeptide repeat proteins (PPR) contain tandem repeat modules, are the largest family in the plant genome and have been implicated in playing a role in environmental adaptation in plants (Aubourg *et al.*, 2000; Nakamura *et al.*, 2003; Manna, 2015). It is therefore not surprising that this gene family was so large. It either contained many isoforms or the graph reconstruction procedure was too complicated (due to the highly repetitive nature of the gene) to reduce the family into fewer contigs. Of the 51 438 gene families, 93% contained likely open reading frames (ORFs) (Table 1). Of the 47 630 predicted protein sequences, 70.3% were annotated in the Swissprot database, 17.7% in the Uniref90 database and only 1.6% in the NR/NT databases (Figure 8A). A total of 41 802 gene families

(or 90%) were therefore annotated by the sequential annotation pipeline (Figure 5). Only 10% of the transcripts/ORFs were not annotated in any of these external databases (Figure 8A).

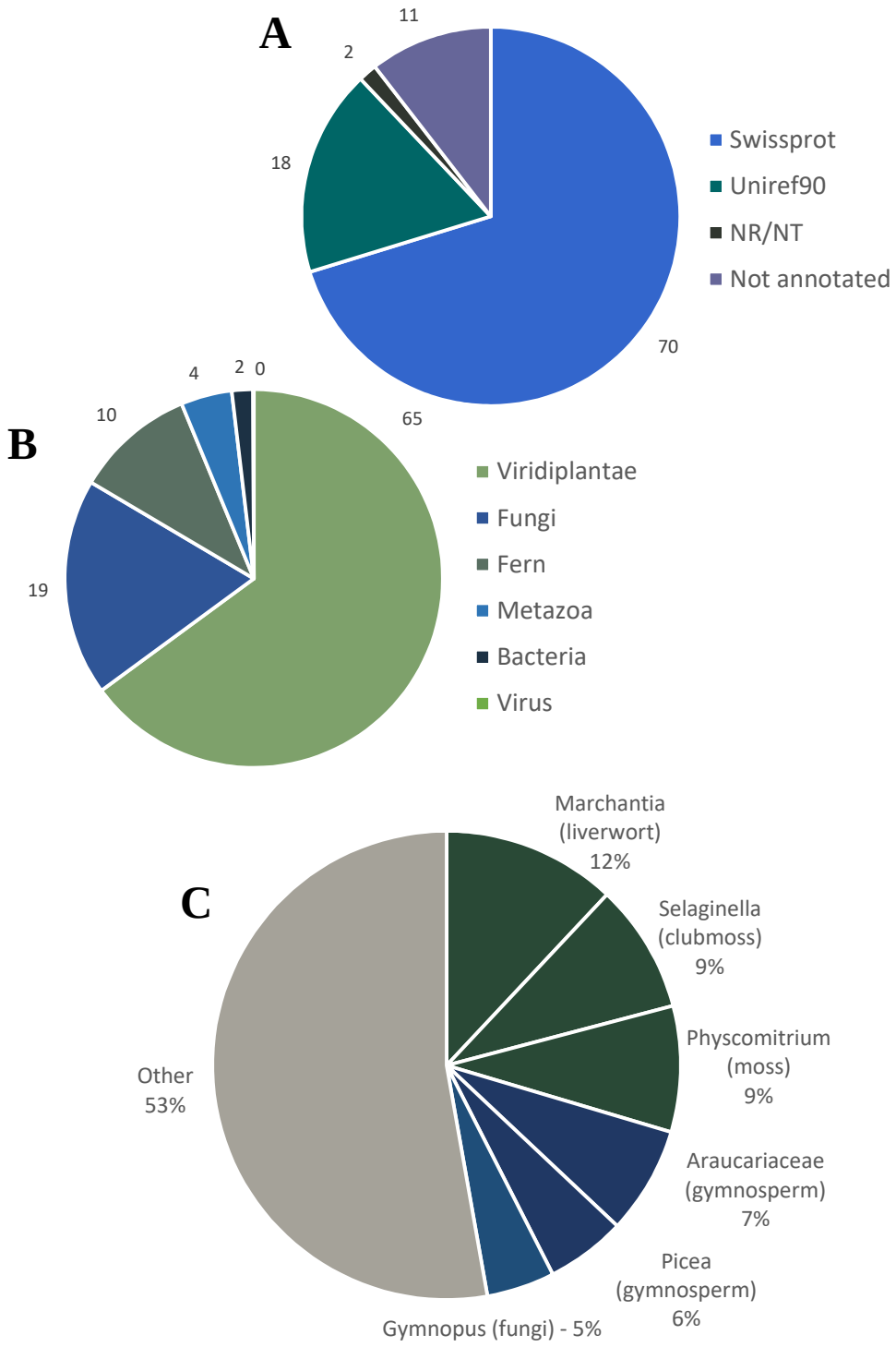

Figure 8: Annotation statistics of the long-read transcriptome. **A**: The proportion of genes that were annotated in each external base used. **B**: The proportion of genes belonging to target and non-target taxonomic groupings assigned through the Uniref90 database. **C**: The six most dominant genera present in the gene annotations.

All the transcript isoforms were separately searched against the Uniref90 database to investigate the taxonomic distribution of the genes in the long-read transcriptome. The Uniref90 database was chosen to interrogate taxonomic trends since the highly curated Swissprot database would inflate the counts of well-studied organisms such as *Arabidopsis thaliana* – which would not truly distinguish how many genes aligned to basal plants when compared to higher order plants (Suzek *et al.*, 2015; Sayers *et al.*, 2021). Approximately 64.9% of the annotated genes belonged to 'Viridiplantae' which was considered the target organism (Figure 8B). A total of 10.3% of genes aligned to fern proteins which demonstrated the relative lack of such resources in current databases and was similarly observed by Hong *et al.*, 2022. The second most dominant taxonomic grouping was 'Fungi' with a 18.6% occurrence in the gene set (Figure 8B). 'Metazoa', 'Bacteria' and 'Viruses' were represented at 4.4%, 1.8% and 0.1% respectively and their contributions were considered negligible (Figure 8B). The high proportion of plant annotations was an indicator of good transcriptome quality (Raghavan *et al.*, 2022).

Two possible reasons for the presence of fungal genes in the transcriptome are 1) they represent contaminant sequences or 2) they may be products of horizontal gene transfer (HGT) which has been documented in ferns (Li *et al.*, 2014; Wisecaver *et al.*, 2016; Gonc *et al.*, 2018). Currently there is no method available to confirm whether the genes belonging to non-target taxon are truly a result of HGT (Zhou, 2021). It is also worth highlighting that *A. caffrorum* tissues were not sterilized as to avoid possible confounding effects of a sudden influx of water into dehydrated/desiccated structures. Furthermore, one can reasonably expect soil on rhizomes to be a considerable contributor of contaminant sequences. While it is not possible to truly determine whether the presence of fungal genes is the result of horizontal gene transfer (HGT) or contaminants in field samples, it is entirely possible that the transcriptome of *A. caffrorum* may contain more contaminant sequences because of the sampling methods used in the study. To explore the possible effects of contaminant sequences on downstream differential expression studies, the relative contribution of non-plant transcripts in the counts matrix was determined. Briefly, the sum of the raw count data belonging to transcripts which were assigned as non-plant was determined for each library (Figure 9). On average 'Viridiplantae' transcripts accounted for 69.83 ± 3.01% of the expression data and 'Fern' transcripts encompassed 9.58 ± 0.67% (Figure 9B). 'Fungi' only contributed 0.36 ± 0.33% of the count data and the largest

non-plant contributor, 'Metazoa', was 1.93 ± 0.20% (Figure 9B). These results indicated that the high number of fungal transcripts in the long-read transcriptome may simply be lowly expressed contaminant sequences and that they should not negatively impact differential expression studies.

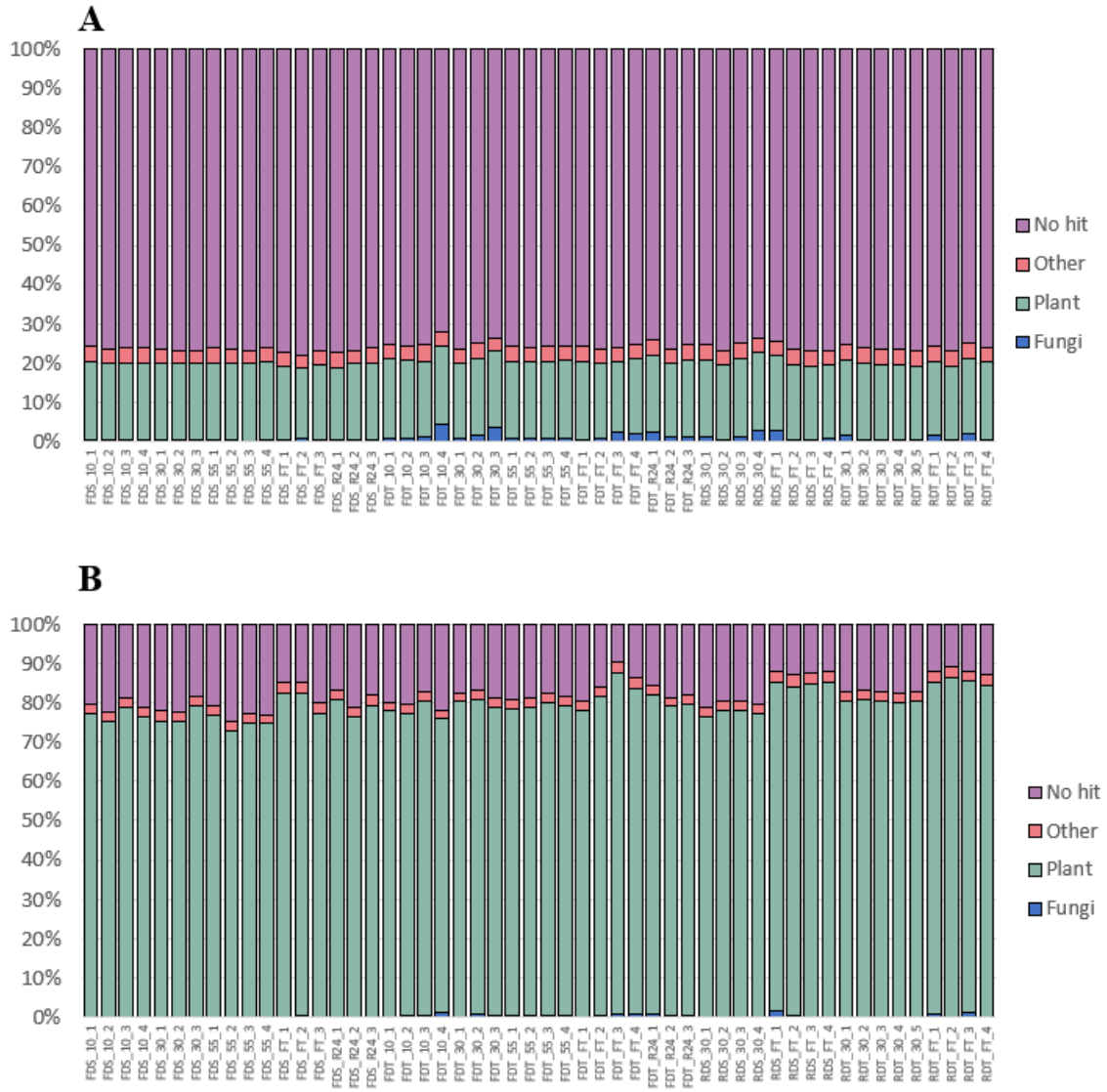

Figure 9: Taxonomic representation of count data. Raw count data for each taxonomic grouping was summed based on transcripts that were annotated in the Uniref90 database. The plant grouping contains 'Viridiplantae' and 'Fern' transcripts. The 'Other' grouping contains 'Metazoa', 'Bacteria' and 'Virus' transcripts. No hit transcripts are those that were not annotated in the Uniref90 database. **A**: Taxonomic distribution of the count data belonging to the short read *de novo* assembly. **B**: Taxonomic distribution of the count data belong to the long-read generated transcriptome.

The organisms which represented the highest proportions of homology hits belonged to lower order plants. Approximately 12.0%, 8.9% and 8.7% of genes belonged to *Marchantia,*

Selaginella and *Physcomitrium* genera (Figure 8C). Taken together approximately 30% of the most abundant genera belonged to lower order plants. Thereafter, approximately 7.4% and 5.5% of the most dominant annotations belonged to the gymnosperms *Picea* and *Araucariaceae*. The long-read transcriptome *A. spinulosa* contained high dominance of *Marchantia, Physcomitrium* and *Picea*, while the transcriptome of *A. flabellulatum* contained these same species in addition to a high proportion of *Selaginella* identified transcripts (Table 2). The close evolutionary relationship of ferns with bryophytes and gymnosperms would explain why many of the transcripts are annotated by these genera (Wickett *et al.*, 2014). The sixth most dominant genus was *Gymnopus*, a fungus. *Gymnopus* was not significantly present in the full-length transcriptomes of *A. spinulosa* and *A. flabellulatum*, although the authors did also indicate high levels of fungal transcripts (Table 2). The fruiting bodies of *Gymnopus* are often found in leaf and woody litter similar to the site from which *A. caffrorum* plants were collected (Petersen, 2016). *Gymnopus* annotated genes could be contaminants from the study site that propagated owing to the chosen sample preparation methods or could still be a product of HGT.

Table 2: The most dominant genera in fern long read generated transcriptomes

Genus	**Classification**	*Anemia Caffrorum*	*Alsophila spinulosa*	*Adiantum flabellulatum*
Marchantia	Liverwort	12.0	15.0	59.3
Selaginella	Spike/clubmoss	8.9	N/A	11.0
Physcomitrium	Moss	8.7	10.3	9.5
Araucariaceae	Gymnosperm	7.4	N/A	N/A
Picea	Gymnosperm	5.5	7.8	8.7
Gymnopus	Fungi	4.7	N/A	N/A

Lastly, the completeness of the gene-level, long-read transcriptome was assessed using BUSCO (Simão *et al.*, 2015). BUSCO assesses the completeness of a transcriptome based on the presence and composition of genes that are expected to be evolutionary conserved amongst plants. The full transcriptome was assessed against the viridiplantae_odb10 database. There was a completeness rate of 92.3% with 61.2% complete and single copy orthologs, 31.1% complete and duplicated orthologs, 3.5% fragmented orthologs and 4.2% missing orthologs. This suggests that the *A. caffrorum* transcriptome is intact and suitable for downstream differential expression studies. The inflated rate of complete and duplicated orthologs was not a cause for concern since this phenomenon is fairly common when working with transcriptomes due to the presence of isoforms which cannot always be successfully delineated (Raghavan *et*

al., 2022). A similar completeness score and rate of duplicated orthologs was observed in the full-length generated *A. spinulosa* transcriptome (Hong *et al.*, 2022).

4.2. Assembling the short-read transcriptome

Pre-processing of the raw short read data was conducted prior to assembly. An average of 5.45 ± 1.61 % of the raw reads were removed from each library because they represented erroneous kmers, while less than 0.2% of all reads required quality trimming (Figure 10A). The average percentage of rRNA was 15.01 ± 10.42 % and several samples demonstrated an rRNA content exceeding 30% (Figure 10A). An rRNA content exceeding 30% of the total mRNA library, despite the use of the widely popular polyA capture method, has been observed in other NGS experiments (Lahens *et al.*, 2014). Closer inspection of the rRNA percentage within each sample suggested that total rRNA content could also be related to a mechanism of desiccation tolerance. The average proportion of rRNA in the desiccation sensitive fronds (FDS) did not exceed 2% (Figure 10B). A similar trend was seen in desiccation tolerant fronds (FDT) at relative water contents (RWCs) of approximately 100% and 55%. Interestingly, as the RWC reached 30% in FDT fronds, a marked increase in rRNA was observed (Figure 10B). To test for statistical significance, t-tests were conducted against DT fronds at 55%. DT Fronds at 30% RWC, 10% RWC and 24-hours post-rehydration contained significantly more rRNA ($p=0.0201$, 0.0138 and 0.0018 respectively) than DT fronds with a RWC of 55% (Figure 10B). The proportion of rRNA between rhizome samples was not statistically different from one another but the average amount of rRNA in rhizome samples exceeded 5% which was noticeably higher than that of hydrated DT fronds (100 and 55% RWC) and all DS fronds (Figure 10B). The higher proportion of rRNA in the rhizome tissues relative to the frond tissue may be an artefact of sampling below ground tissues which will likely contain more microbial contamination owing to the presence of soil on the organ during processing. An increase in rRNA in DT fronds where no apparent change occurs in DS fronds could indicate a possible mechanism of desiccation tolerance as opposed to random events due to library preparation. It is, however, not impossible that the rRNA content may still be a result of better poly-A enrichment in some samples and may even be an artefact of batch processing. Information about how samples were processed was not available so this could not be tested. This finding is the first result that indicates that a possible delineation between DS and DT fronds has at least partially been experimentally captured. Interestingly, the shift is only apparent after the first response (approximately 55% RWC) and potentially during the second response (approximately 30% RWC) to drying observed in many angiosperm resurrection plants which

marks the entry into desiccation (Costa *et al.*, 2017; Farrant and Hilhorst, 2021; Gabier *et al.*, 2021). While interesting, this finding may only be better contextualized alongside the results of the differential expression studies.

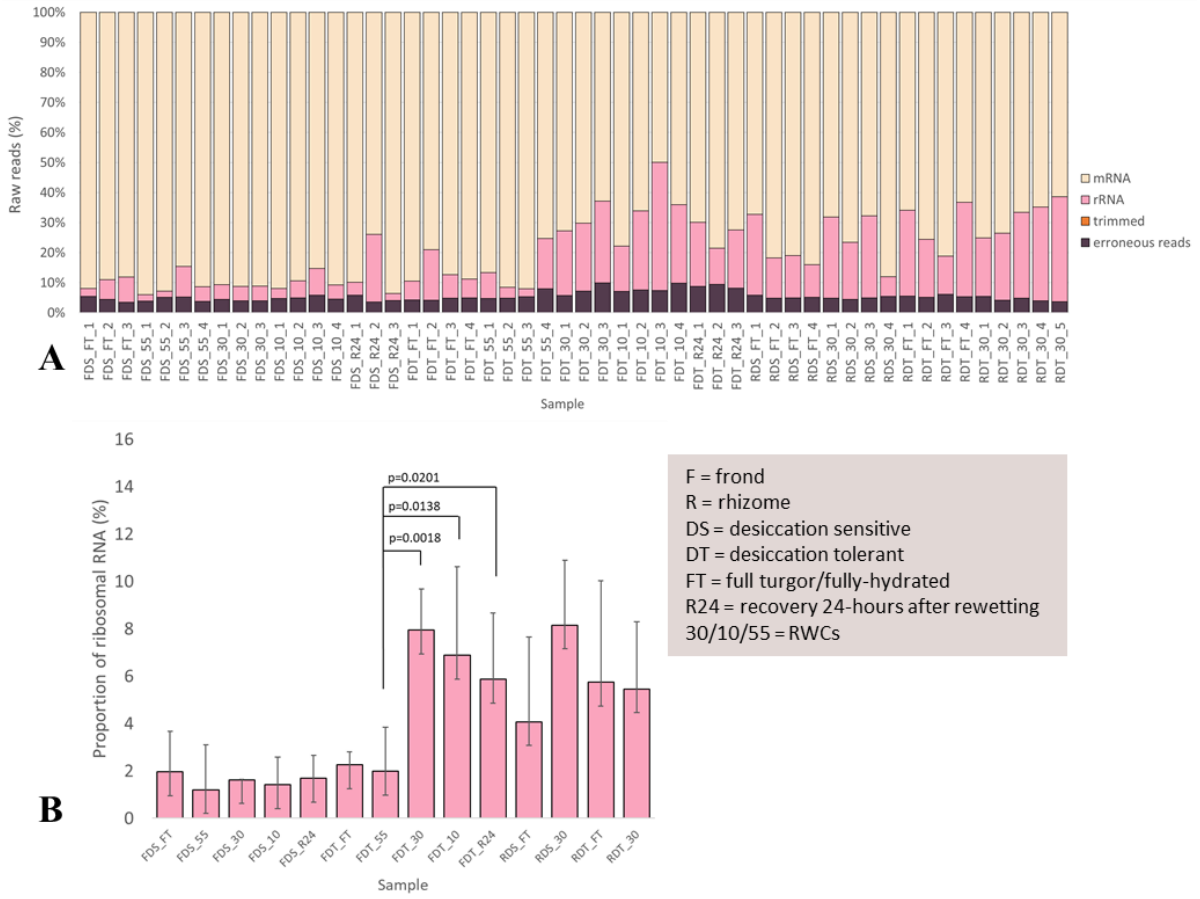

Figure 10: Processing raw read data to mRNA sequences used for *de novo* assembly. A: Raw reads obtained after short-read sequencing contained low levels of erroneous kmers (*purple*) and low-quality bases (*contribution too small to visualize*). Approximately 2-10% of all the reads represented rRNA (*pink*) and the remainder was presumed to be reads originating from mRNA (*light pink*). B: Average amounts of rRNA within each sample grouping. DS fronds contained approximately less than 5% rRNA in all groupings. rRNA levels in DT fronds at RWC 100% and 55% was also relatively low but a marked increase was observed at 30% RWC, 10% RWC and rehydration after 24-hours. Statistical t-tests comparing DT fronds at 55% RWC to DT fronds at 30% RWC (p=0.0018), 10% RWC (p=0.0138) and 24-hours post rehydration (p=0.0201) indicate that the changes were statistically significant. Changes in rRNA levels between rhizome samples was not statistically significant but still noticeably higher than that of DS fronds. Error bars indicate standard deviation between samples.

The 850 322 958 (850 million) mRNA sequences remaining after pre-processing were assembled using Trinity (Figure 4B) (Haas *et al.*, 2013; Davidson and Oshlack, 2014; Signal

and Kahlke, 2021). A total of 502 119 contigs representing 336 307 trinity ‘genes’ were assembled. The 502 119 trinity-generated transcripts were clustered into gene families using Corset (Davidson and Oshlack, 2014). Corset clustering resulted in the removal of 71 184 (14%) of the trinity contigs and clustered the remaining 431 010 transcripts into 336 307 gene families (Table 1). The average mapping rate of the Corset-collapsed transcriptome was approximately 95% and the lowest mapping library had a rate of 90% (Table 1). This implied that the transcriptome was a high-quality representation of the underlying data. A total of 126 078 (37%) Corset gene families contained likely open reading frames (ORFs) and 89 081 were annotated via blastp against the Uniref90 database (Table 1). Approximately 30%, 5% and 54% of the gene families were identified as belonging to ‘Viridiplantae’, ‘Fern’ and ‘Fungi’ respectively (Table 1). By comparison, the trinity ‘genes’ only had a 30% ORF prediction rate and markedly lower plant annotated genes. Trinity gene annotations were 24% ‘Viridiplantae’, 5% ‘Fern’ and 60% ‘Fungi’ (Table 1). Although Trinity and Corset performed similarly in reducing the transcriptome, Corset produced gene families with higher protein prediction rates and more taxonomically accurate annotations. Furthermore, even though Corset appeared to have conservatively reduced the transcriptome, after the removal of low-level contaminants and transcripts that were not annotated in an external database, a total of 26 460 ‘Viridiplantae’ genes and 4 590 ‘Fern’ genes were identified. Therefore, a total of 31 050 plant genes were identified in the Corset gene set (Table 3). The total number of plant-assigned genes was similar to that of the *Athyrium yokoscense* and *Vandenboschia speciosa* fern transcriptome assemblies (Table 3) (Ukai *et al.*, 2020; Kunkel, 2021). Approximately double the number of genes were identified in the transcriptome assemblies of *Microsorum punctatum*, *Microsorum scolopendria,* and *Ceratopteris richardii* (Table 3) (Sripinyowanich *et al.*, 2021; Zhou, 2021). While it is difficult to predict why more genes were identified between fern species, two possibilities are the different amounts of input data used for the assembly process and differences in the complexity of the tissue sets used for sequencing (Table 3). Overall, the total number of genes identified in the transcriptome of *A. caffrorum* was similar to that observed in similar studies. This implied that *de novo* short read assembly of the *A. caffrorum* transcriptome using Trinity was successful at least in terms of the number of identified plant genes.

Table 3: Comparison of short read *de novo* assemblies of fern transcriptomes using Illumina reads assembled by Trinity

Species name	*Anemia Cafforurm*	*Athyrium yokoscense*	*Botrychium lunaria*	*Ceratopteris richardii*	*Microsorum punctatum*	*Microsorum scolopendria*	*Vandenboschia speciosa*
Year of publication	2022	2020	2020	2021	2021	2021	2021
Total raw reads	850 million paired-end	661 million paired-end	49.5 million paired-end	2.6 billion paired-end	71 million paired-end	70 million paired-end	66 million paired-end
Read length (bp)	75	100	150	150	100	100	100
Total genes identified	31 050	35 681	25 677	64 974	57 252	54 618	36 430
% GC	47		44	43	46	46	
N50 (bp)	1 501	2 668	1 995		1 934	2 148	2 085
Ex90N50 (bp)	2 172						2 299
Number of transcripts at the Ex90 peak	19 427						21 543
BUSCO completeness score (%)	96	71	97	71	71	74	94

Both short read *de novo* transcriptomes displayed considerable proportions of fungal contigs (Table 1). To explore whether the fungal genes represented lowly expressed contaminant sequences, a counts matrix was generated and the sum of the counts from each taxon was calculated for each library (Figure 9A). On average, fungal identified genes only represented 0.79 ± 0.62 % of the count data while plant derived genes accounted for 19.36 ± 0.49 % (Figure 9A). This supports the hypothesis that many of the fungal assigned, trinity-generated contigs were simply the product of low-level contaminant sequences. Therefore, the large proportion of off-target, fungal genes in the transcriptome was not a concern for subsequent differential expression studies since their contribution to overall expression was negligible.

Common transcriptome assessment metrics of the non-redundant transcriptome was compared to the *de novo* assemblies of six other fern species (Mossion *et al.*, 2020b; Ukai *et al.*, 2020; Kunkel, 2021; Sripinyowanich *et al.*, 2021; Zhou, 2021). Each transcriptome was generated using Illumina reads and Trinity was the chosen assembler. Notably, the total number of short reads and the read lengths varied between each assembly (Table 3). The GC content of the *A. caffrorum* assembly was 47% which was similar to the GC percentage of the other fern transcriptome assemblies (Table 3). The N50 value of the contigs was 1 501 bp (Table 3). The contig N50 is a weighted statistic such that 50% of the entire assembly is contained within contigs equal to or larger than this value i.e., half of the assembled bases is contained in contigs of at least 1 501 bp in length (Haas et al., 2013). The contig N50 for the *A. caffrorum* transcriptome was considerably lower than that of the other fern transcriptome assemblies and is likely a product to the shorter read lengths used (Table 3). While the N50 metric was relatively lower than the other fern assemblies, it was not considered an unacceptable value. It is also worth noting that the contig N50 has often been criticized for its oversimplification. The N50 value may be inflated due to trinity assembling too many isoforms per transcript, with inherent biases towards those that are longer. For this reason, it is recommended to calculate the N50 of the longest isoforms per 'gene'. The N50 based only on the longest isoform per trinity 'gene' was 805 bp with the average contig length of 434 bp. As expected, the N50 based on the longest isoform per gene is much lower than that of all assembled contigs. Relatively low N50 values based on the longest isoforms are often because of the presence of rarer transcripts in the assembly. Lowly expressed transcripts tend to produce shorter contigs owing to the effects of low coverage on their assembly (Haas *et al.*, 2013). The large proportion of lowly expressed fungal transcripts is an example of the transcripts that can affect metrics such as the N50 (Figure 9A). While this metric is recommended as a superior alternative, it was not

provided in the other fern transcriptome reports and therefore omitted in Table 3. The most accurate measure for contig length distribution is the Ex90N50 metric which measures the N50 as described above but is limited to the most highly expressed transcripts which represent 90% of the total normalized expression data. The Ex90N50 circumvents the confounding effects of rarer, lowly-expressed transcripts in the assembly and is considered the most useful indicator of transcriptome assembly quality. This statistic is measured at the level of the gene i.e., the value for a gene's expression is the sum of the expression of all isoforms (Grabherr et al., 2011; Haas et al., 2013). The Ex90N50 value was 2 172 bp for the 19 427 genes that it represented (Table 3). The fact that the Ex90N50 was higher than the N50 demonstrated that the contigs which are responsible for 90% of the expression data are larger. This would suggest that the short read length used for sequencing did not appear to hinder the assembly of the most highly expressed genes (Table 3). Presumably, the removal of the lowly expressed fungal transcripts has also resulted in the improved N50 length. The Ex90N50 of the *A. caffrorum* assembly was comparable to that found for the *de novo* assembly of the fern *Vandenboschia speciosa* (Table 2) (Kunkel, 2021). Surprisingly, despite the Ex90N50 being the improved measure for assembly assessment this was not assessed in the other studies under review.

Finally, the completeness of the short read assembly was assessed using BUSCO (Simão *et al.*, 2015). BUSCO assesses the completeness of a transcriptome based on the presence and composition of genes that are expected to be evolutionary conserved amongst plants. The full transcriptome was assessed against the viridiplantae_odb10 database. There was a completeness rate of 96.0% with 35.3% complete and single copy orthologs, 60.7% complete and duplicated orthologs, 2.8% fragmented orthologs and 1.2% missing orthologs (Table 3). The inflated rate of complete and duplicated orthologs was not a cause for concern since this phenomenon is fairly common when working with transcriptomes due to the presence of isoforms which cannot always be successfully delineated (Raghavan *et al.*, 2022). The *A. caffrorum* transcriptome had a similarly high completeness scores to the *Botrychium lunaria* and *Vandenboschia speciosa* fern assemblies and was considered more complete than the other assemblies under review (Table 1) (Mossion *et al.*, 2020a; Kunkel, 2021).

4.3. Exploring a hybrid transcriptome

While one can reasonably expect a reference transcriptome generated by long read data to be superior to those generated via *de novo* short read assembly, the concern in this study was whether sufficient sequencing depth was achieved for this to be realized (Byrne *et al.*, 2019). Pooling of samples and sequencing on a single SMRT cell is advertised as being both cost-effective (and therefore accessible) and a plausible means of capturing a full plant transcriptome. However, pooling too many samples and/or using an insufficient range of biological treatments can result in the loss of total genes or important tissue-specific genes. To this end, the main goal was to explore whether there were any 'missing' genes that were not captured by the Iso-Seq method. Long read sequencing also suffers from 5' degradation when it is not coupled with cap trap technology (Kuo *et al.*, 2019). Since cap trap technology was not applied in this study, the secondary goal was to attempt to fix gene models that have degraded 5' ends using short read data. To explore both goals the long and short reads were combined in various ways.

4.3.1. Combining the short and long read data in the short-read assembly pipeline

In the first approach, short read data was used for assembly via Trinity but was guided by the HQ isoforms. Table 4 displays the performance metrics of the Trinity assembly against the Trinity assembly guided by HQ isoforms. The long-read guided approach was virtually identical to the base assembly in every regard. This is because when Trinity incorporates long read data into the assembly process it does not use the full-length sequences in the actual graph reconstruction steps (https://github.com/trinityrnaseq/trinityrnaseq/issues/210). The reason for this is that at the time of development long read data still had higher error rates that would propagate through the assembly process and negatively affect the quality of the generated contigs. The long-read guided functionality of Trinity has not been updated since long read sequencing has become more accurate than even some forms of short read sequencing. While this has no real effect on the study at hand, it is interesting to note that very little development has been made to incorporate the use of highly accurate long reads into short read *de novo* assembly software. Since this approach did not yield any fruitful results, two alternative approaches were attempted.

Table 4: Trinity assembly with and without long read guidance

	Trinity base assembly	Trinity assembly guided by HQ isoforms
Total transcripts	502 119	502 194
Total genes	336 415	336 307
Collapse rate (transcripts/gene)	1.5	1.5
Total ORFs	185 700	184 706
Percentage ORFs (%)	37	37
Number annotations in Uniref90 database	133 298	132 330
Annotation rate (%)	72	72
Plant (%)	43	43
Fern (%)	6	6
Fungal (%)	42	42
Other (%)	9	9
Mapping rate	95	85
Minmum mapping rate	90	90
N50 (bp)	1 406	1 399
E90N50 (bp)	2 500	2 400
Number of genes corresponding to the E90 peak	21 266	21 832

4.3.2. Combining the short and long read data in the long-read assembly pipeline

In the second approach the HQ isoforms were combined with the short read generated contigs and processed through the Cogent/cDNA_Cupcake pipeline (Figure 1). Because Trinity notoriously generates many more transcripts that one would expect to find in an organisms full transcriptome, several filtering options are often applied to the data depending on the downstream application (Haas *et al.*, 2013). To mimic the effects of singleton removal from the HQ isoform set, the trinity base assembly was subjected to an expression filter. A second motivation for the expression filter was to remove as many of the lowly expressed fungal transcripts that were observed in the short read *de novo* assembly. These were expected to complicate graph reconstruction procedures.

Briefly, the short reads were mapped to all trinity generated transcripts to generate a TMM expression matrix. Transcript models that had a TMM < 1 over every biological condition were removed. A total of 258 884 transcripts had a TMM < 1 and 243 310 had a TMM > 1 for at

least one biological condition. Considering the high rates of fungal contigs observed in the short read *de novo* assembly, the TMM > 1 and TMM < 1 transcript sets were screened for the relative proportions of fungal contamination. In the TMM > 1 set, approximately 41% of annotated transcripts had a plant assignment while 48% were fungal in nature. The TMM < 1 set contained 54% fungal transcripts and 28% plant transcripts. This indicated that the TMM filter resulted in a greater selection for plant transcripts and slightly reduced fungal contaminants in the set. Since the aim of this approach was to capture any 'missing' genes in the short read data, the TMM filter was not repeated under more stringent conditions since this may result in the loss of biologically relevant genes.

The 243 310 contigs with a TMM > 1 and were combined with the 183 163 HQ isoforms, so a total of 426 468 sequences were used as input to the Cogent pipeline (Figure 11). A total of 118 543 of all the trinity contigs (approximately 50%) were orphaned. Recall that orphans are transcripts that do not appear to share sequence similarity with any other transcript in the set and therefore have a high likelihood of representing rare artifacts or low-level contaminants. It is possible that the *minimap2* algorithm did not classify many trinity contigs as similar sequences since they may be too dissimilar to full-length sequences to be identified as possible transcript isoforms. A similar finding can be seen in the number of tucked sequences which represented sub-sequences of other sequences in the set. A relatively small number of sub-sequences were derived from the Trinity contigs despite the high degrees of redundancy in Trinity assemblies. This trend extends to the number of clustered sequences and identified gene families. In both cases, the number of clusters and gene families which contain only long read sequences was similar when running the Cogent pipeline with and without Trinity contigs (Figure 11). Most of the additional clusters and gene families contained Trinity derived transcripts. These mixed clusters/gene families do however indicate that, in some cases, Trinity transcripts were being combined with full-length sequences and may in turn improve the characteristic 5' degraded ends. These mixed groupings only made up a very small proportion of the clusters/gene families. On the other hand, clusters/gene families that only contain Trinity contigs show that *minimap2*/mash algorithms cannot effectively identify them as isoforms of the long-read transcripts. While the short read assembly had produced longer transcripts, in the same length range as the long reads, shorter transcripts and transcripts that do not contain clear splice junctions may be resulting in spurious alignments (Ondov *et al.*, 2016; Li, 2018; Raghavan *et al.*, 2022). This may be due to the presence of SNPs/indels in the trinity generated contigs or simply heterozygous sites that were only captured by the higher depth of Illumina

sequencing. Alternatively, they could represent close orthologs and/or paralogs that are in fact separate genes. An additional 22 165 clusters were identified when the trinity transcript models were incorporated to the pipeline and 28 344 additional gene families were identified; this was approximately double and triple the values, respectively, when using only the HQ isoforms (Figure 11). Because of the large proportion of orphaned reads and the inflated number of gene families, the fake genome contained 253 426 sequences, more than quadruple the previously generated fake genome (Figure 11).

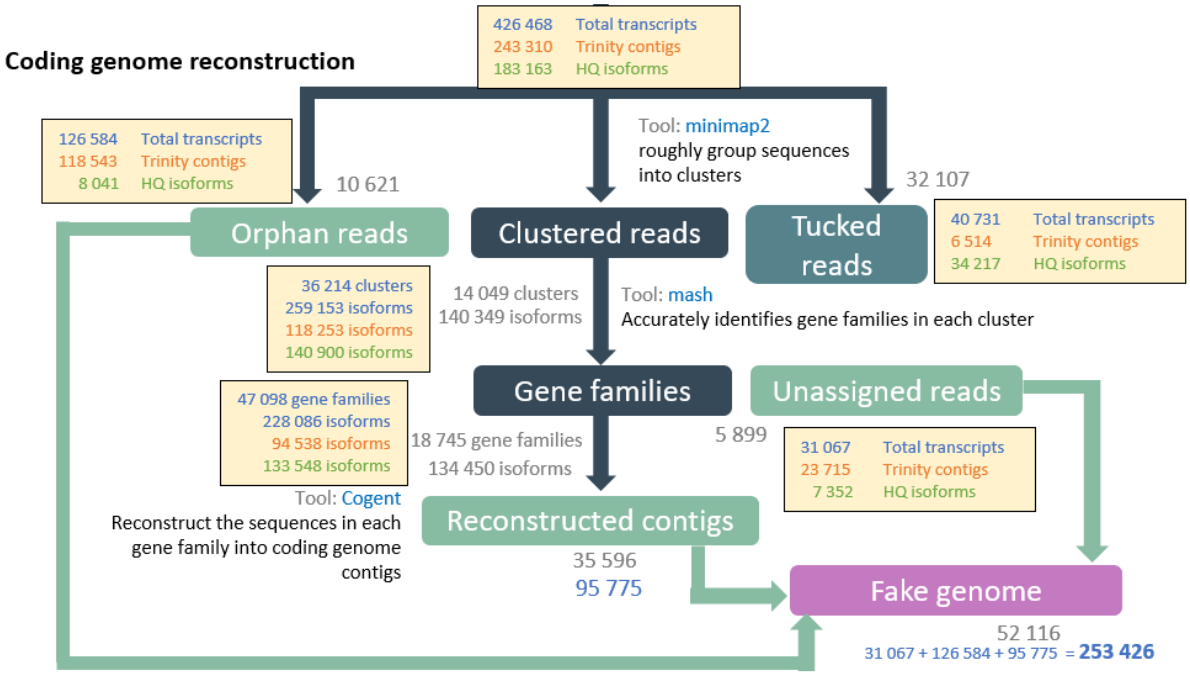

Figure 11: Generation of the fake genome using short and long read data. Short-read generated trinity contigs and HQ isoforms were combined and used as input to the Cogent pipeline. Orange text indicates trinity generated contigs. In the case of clusters/or gene families which contain multiple sequences, orange text indicates that the group of sequences contains at least one trinity contig. Green text represents transcripts derived from the IsoSeq3 pipeline in a similar manner. Blue text represents the total trinity contigs and long read sequences. Grey text are the results of the Cogent pipeline when using the HQ isoforms as the only output and are provided to simplify comparisons. Approximately 50% of trinity contigs were orphaned by *minimap2*. The number of sequence clusters and gene families generated by the original pipeline was almost identical to the number of sequence clusters and gene families which contained only long read sequences generated in the hybrid process. The total number of clusters, gene families, reconstructed contigs and transcripts in the fake genome were highly inflated.

A total of 334 897 transcripts mapped to the fake genome and cDNA_Cupcake collapsed these into 250 967 gene loci (Figure 2). The hybrid assembly had a collapse rate of 1.3X while the long-read assembly had a collapse rate of 3.6. The collapse rate for the hybrid assembly was even lower than that of the short read assembly suggesting that cDNA_Cupcake could not accurately cluster and collapse trinity contigs and full-length long reads effectively (Table 1). In terms of ORF predictive capabilities and annotated genes, the hybrid assembly seemingly performed better than the long-read assembly. In the long-read assembly 75% of 40 287 hits were assigned plant annotations, which is 30 215 plant genes. In the hybrid assembly, 57% of 88 685 hits were assigned plant annotations which is 50 550 plant genes (Table 1). At the gene-level, only 53% of the genes mapped to a sequence in the Uniref90 database, and equal amounts were identified as plant and fungal proteins (Table 1). The number of genes corresponding to 90% of the expression data was 79 337 (Table 1). This highly inflated value is a clear indicator that the extra gene families that were identified were most likely the result of fragmented gene families that were not successful reduced. Furthermore, a lot of non-useful off-target transcripts were propagated through in the hybrid assembly because of ineffective incorporation of trinity contigs to the assembly at large. Figure 12 also demonstrates how the hybrid assembly was more fragmented than the long-read assembly, suggesting that many of the shorter transcripts were not successfully incorporated into larger contigs in the hybrid assembly process. In terms of improving missing 5' information, manual inspection of the data showed that in some cases the short read data did in fact improve some of the missing 5' regions. However, the improvement of the 5' regions in some instances (*data not shown*) was not considered sufficient justification to adopt the hybrid transcriptome for differential expression studies. It was concluded that the HQ isoforms and trinity transcripts models could not be reliably integrated by the Cogent/cDNA_Cupcake pipeline. It is possible that with more in-depth fine-tuning that these programmes could perform more efficiently but it was not within the scope of this study. The analytical burden of the removing the additional non-useful information far outweighed the information that might be gained. Furthermore, at the time of this study, there were no alternative tools available that specifically combined short and long read data suggesting that the combination of such data sets has yet to be explored for good reason.

4.3.3. Complement of the long read transcriptome in the short read data

To circumvent the difficulties around integrating short and long read data, a third approach was attempted. The short read generated contigs were mapped to the long-read transcripts and categorized as mapped and unmapped contigs (Figure 4B). Unmapped contigs were considered

possible ‘missing’ genes that were in the short read data but lost due to low coverage in the Iso-Seq approach. More specifically, short reads were first mapped to the TMM-filtered transcript set (TMM > 1) and Corset was run on the TMM > 1 set to identify gene clusters. Concurrently, the trinity generated contigs were subjected to a nucleotide blast against the HQ isoforms to determine which assembled transcripts were already present in the long-read data. If a transcript mapped to the HQ isoforms, it was considered redundant. A redundant transcript and the set of transcripts in the Corset gene family that it belonged to were all removed. A total of 64 213 trinity transcripts representing 55 736 Corset genes remained. Only 18 324 (or 32.9%) Corset genes contained likely ORFs and 15 659 were functionally annotated by homology search in the Uniref90 database. Approximately 87% of the annotated genes represented fungal transcripts. Taken all together, the findings seem to suggest that virtually each Trinity-generated transcript was contained in the Iso-Seq data and only lowly expressed fungal transcripts were additionally assembled by Trinity owing to the higher coverage of short read sequencing.

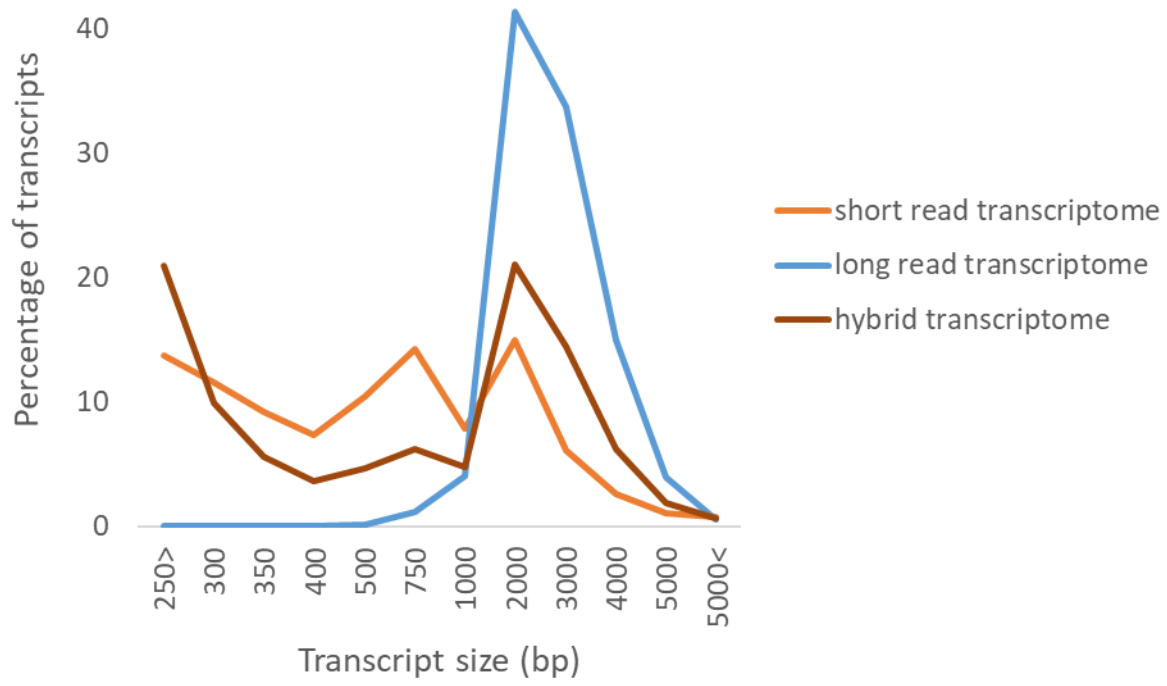

Figure 12: Comparison of the degrees of fragmentation of the short, long and hybrid assemblies based on sequence length distribution.

5. Conclusion

The short read transcriptome was characterised by low annotation rates and high numbers of fungal assigned genes. Furthermore, the plant annotated genes accounted for a very low proportion of the expression data which remained largely un-annotated. One likely reason for this is the short read length applied in the short read sequencing. The short read length may have resulted in the highly fragmented transcriptome which in turn resulted in low annotation rates. The *de novo* short read transcriptome did however have a high mapping rate and BUSCO completeness score. Despite its performance being poorer than the long-read transcriptome, it was similar to what has been observed in other *de novo* generated fern transcriptomes. Owing to the longer read length (and therefore minimal fragmentation) the long-read transcriptome had very high plant annotation rates which accounted for most of the expression data. Furthermore, the most dominant organisms present in the data agreed with what was observed for other full-length fern transcriptomes. The *A. caffrorum* transcriptome was not collapsed as effectively as the other transcriptomes in literature but the slight gene inflation was not a major cause for concern. The transcriptome was also characterised by a good BUSCO completeness score and was a high-quality representation of the short read data. Attempts to generate a hybrid assembly approach were not successful. Mixed model software is not yet available, and integration of short and long read data using the current tools proved ineffective. The hybrid assembly generated using the Cogent/cDNA_Cupcake pipeline resulted in a highly inflated gene counts because of this inefficacy. Exploration of the complement of the long-read transcriptome existing within the short read transcriptome was extremely insightful in terms of the overall success of the long-read sequencing approach applied in this study. The complement demonstrated that all additional contigs captured by high-depth short read sequencing were simply lowly expressed fungal contigs. This implied that the *A. caffrorum* full-length transcriptome generated on a single SMRT cell from a pooled sample captured all the information present in the *de novo* short read assembly. Furthermore, it captured the information more effectively owing to the longer read lengths. Altogether the findings suggested that the goals of the study were achieved – the more financially accessible option of pooling a mixture of samples, and sequencing on a single SMRT, was sufficient to capture the transcriptome of a non-model plant for differential expression studies. These findings support the notion that long read technology may very soon supersede *de novo* short read transcriptome assembly.

Chapter 3 :

Frond transcriptional responses to desiccation

Chapter 3 :

Frond transcriptional responses to desiccation

1. Introduction

In hot summer months, the fronds of *A. caffrorum* are desiccation tolerant (DT), while in the wet, winter months they are desiccation sensitive (DS). DS fronds lose membrane integrity once relative water (RWC) drops below 60%. Several typical desiccation tolerant features allow DT fronds to maintain membrane integrity at RWCs as low as 8% (Farrant *et al.*, 2009). DT fronds curl in such a way that they mask the abaxial surfaces from incoming light. Excess light under low water conditions results in the production of destructive reactive oxygen species (ROS), and so, frond curling is an adaption that mitigates oxidative stress. Furthermore, DT fronds display a prominent upregulation in lipophilic antioxidants. DS fronds have been shown to be more photosynthetically efficient than DT fronds (Nadal *et al.*, 2021). Similar ultra-structures are observed in the mesophyll cells of fully hydrated DS and DT fronds. The vacuoles in both phenotypes contain spherical bodies which house osmophilic polyphenols. Upon drying below 60% RWC, clear rupture of the plasmalemma and tonoplast membranes was observed in DS mesophyll cells. In DT cells, an increase in phenolic inclusions and vacuolation was observed lining the periphery of the vacuole (Farrant *et al.*, 2009). In other angiosperm resurrection plants these structures have been proposed to play a role in stabilization of cell structure during mechanical stress (Moore *et al.*, 2005). Similarly, the chloroplasts of DS fronds lost their organization during drying, whereas those in DT fronds maintained chloroplast structure, lacked starch and displayed an accumulation of plastoglobuli (Farrant *et al.*, 2009). Both phenotypes contained chlorophyll throughout the course of desiccation. These are all typical responses of homoiochlorophyllous resurrection angiosperms (Sherwin and Farrant, 1998; Farrant *et al.*, 2003). In DT fronds, the efficiency of photosystem II starts to decline at 70% RWC and seemingly loses all activity by 30% RWC. Photosystem II is fully reinstated 6 hours following rehydration from the desiccated state. DS fronds show rapid photosynthetic decline at 40% RWC and no signs of recovery thereafter (Farrant *et al.*, 2009). The antioxidant enzymes, catalase (CAT), glutathione reductase (GR) and superoxide dismutase (SOD), displayed low activity in hydrated DS fronds. They were highly active in hydrated DT fronds, and less active in dry DT fronds, suggesting a possible constitutive

protection mechanism (Farrant *et al.*, 2009). In response to drying, DT fronds produce several heat-stable LEA/sHSP like proteins as detected by SDS-PAGE analysis. A putative chaperonin was also identified. Chaperones are known to participate in the stabilization of proteins under heat and desiccation stress (Waters and Vierling, 2020). Induction of these proteins in response to drying suggests at least that there are also inducible mechanisms of tolerance occurring in *A. caffrorum* DT fronds (Farrant *et al.*, 2009). The drying of DT fronds showed marked increases in sugars known to play a role in protection – sucrose, raffinose and cyclitols. There was also a decline in reducing sugars, which could have damaging effects – fructose and glucose (Farrant *et al.*, 2009). The induction of heat-stable protein and the accumulation of sugars are features of the vitrification process.

2. Aims and objectives

To supplement the current knowledge of the responses of *A. caffrorum* fronds to desiccation, ongoing work to prepare the proteome and metabolome are currently underway. As such, the study of transcriptional mechanisms of desiccation tolerance in the fronds of *A. caffrorum* is highly desirable. Because ferns can display both constitutive and inducible responses to desiccation, the objectives of this chapter were to determine the

1. Inducible processes that confer desiccation tolerance
2. Constitutive protection of the system to desiccation

3. Methods

3.1. Quality control and outlier detection

Frond material used in transcriptome analysis was prepared and analysed as described in Chapter 1. Transcript quantification was conducted by mapping the short reads to the long read transcriptome using Salmon version 1.6.0 (Patro *et al.,* 2017). To assess intragroup variability the biological replicates were compared to one another. The raw counts belonging to each sample were subjected to a CPM transformation and then were log2 transformed. The correlation between biological replicates was measured using Pearson's coefficient. Intragroup variability assessments were facilitated by the 'PtR' script belonging to the Trinity toolkit (Bryant *et al.*, 2017). To further explore inter- and intra-group variability principal component analysis (PCA) was used. Examination of cross-organ relationships was attempted using the Weighted Gene Correlation Network Analysis (WGCNA) tool but were not fruitful (Langfelder and Horvath, 2008). Therefore, all analyses were conducted by separating the rhizome and frond data. Salmon counts were imported to DESeq2 version 1.36.0 using the tximport package and the dds object was generated using the design formula : design =~ condition (Love, Huber and Anders, 2014). Each biological grouping was considered as a condition as described below in Table 1. The dds object for each tissue was filtered by removing any genes that had less than 1 read in each sample. The filtered dds objects were subjected to variance stabilizing transformation (vst) and then used for PCA analysis. Sample FDT_FT_1 (fronds, desiccation tolerant, replicate 1) was the only sample that was considered an outlier and was removed from all further analyses.

Table 1: Grouping of biological data for differential expression analyses

Dataset	**Conditions**
Fronds	FDT_FT(desiccation tolerant, full turgor) (n = 4) FDT_55 (desiccation tolerant, 55% RWC) (n = 4) FDT_30 (desiccation tolerant, 30% RWC) (n = 3) FDT_10 (desiccation tolerant, 10% RWC) (n = 4) FDT_R24 (desiccation tolerant, 24-hours post-rehydration) (n = 3) FDS_FT (desiccation sensitive, full turgor) (n = 3) FDS_55 (desiccation sensitive, 55% RWC) (n = 4) FDS_30 (desiccation sensitive, 30% RWC) (n = 3) FDS_10 (desiccation sensitive, 10% RWC) (n = 4) FDS_R24 (desiccation sensitive, 24-hours post-rehydration) (n = 3)

3.2.Differential expression testing using DESeq2

3.2.1. Pairwise comparisons

The most differentially expressed genes were extracted using a p-value threshold for the false discovery rate (FDR) of 0.01 and an LFC of 1 (i.e., a 4-fold change in differential expression). LFC shrinkage using the lfcshrink() was conducted using the apeglm method (version 1.18.0) (Love, Huber and Anders, 2014; Zhu, Ibrahim and Love, 2019). For rhizome tissues, each condition was compared to one another (4 comparisons as described in the results section). For frond tissues, each phenotype was compared to one another at each water content i.e., the following pairwise comparisons were conducted : DT_FT vs DS_FT, DT_55 vs DS_55, DT_30 vs DS_30, DT_10 vs DS_10 and DT_R24 vs DS_R24. These were termed 'cross' comparisons.

3.2.2. Likelihood ratio test

To identify genes that showed a change in expression across different time points, a likelihood ratio test (LRT) was performed. The frond data was further split between phenotypes (DS/DT) and a dds object was generated for each tissue as described above. A likelihood ratio test using a reduced model was performed on each dataset using DESeq2 v 1.36.0 (Love, Huber and Anders, 2014).

3.3. Functional enrichment analysis

Statistically overrepresented gene ontology (GO) terms in gene sets were determined using TopGO version 2.48.0. A weighted fisher statistic of < 0.001 was considered a significantly enriched term (Rahnenfuhrer, 2022).

3.4. Heatmap preparation and kmeans clustering

Unless stated otherwise, variance-stabilized normalized count data was used as input for heatmaps. Before plotting the data was centred by subtracting each value in the matrix by the row mean. Heatmaps were visualized using ComplexHeatmap version 2.12.1 (Gu, Eils and Schlesner, 2016; Gu, 2022). Kmeans clustering was either conducted using ComplexHeatmap or via the kmeans function in R and the resulting clusters plotted using ggplot2 version 3.3.6 (Wickman, 2016).

3.5. Constitutive expression testing

To determine genes that were constitutively expressed in the frond tissue over time, the list of differentially expressed genes (DEGs) from the LRT analysis was overlapped with the cross

pairwise comparisons described in section 3.2.1. DEGs in the cross comparisons represent genes that are significantly different between phenotypes but not necessarily through the course of drying. This set of DEGs therefore represent constitutive and inducible genes related to desiccation. The genes in the LRT analyses represent genes that are differentially expressed (DE) during at least one time point and are therefore likely inducible by nature. First, the unique set of genes in the union of the cross comparisons was isolated as candidate genes for constitutive expression (Figure 1A). The candidate genes were overlapped with the genes that are DE in DS and the genes that are DE in DT determined by LRT analysis. From here, genes that are DE in DS, DE in DT, and DE in neither were separated (Figure 1B). DE in DS contains genes that could be constitutively expressed in DT. Similarly, DE in DT contains genes that could be constitutively expressed in DS. DE in neither could contain genes that are constitutively expressed in either phenotype. These groups of genes were extracted from the dds matrix (i.e. normalized read counts that were not transformed) and subjected to kmeans clustering using ComplexHeatmap version 2.12.1 to determine clusters with interesting constitutive expression profiles (Figure 1C). An unscaled vst matrix was used to generate Figure 7.

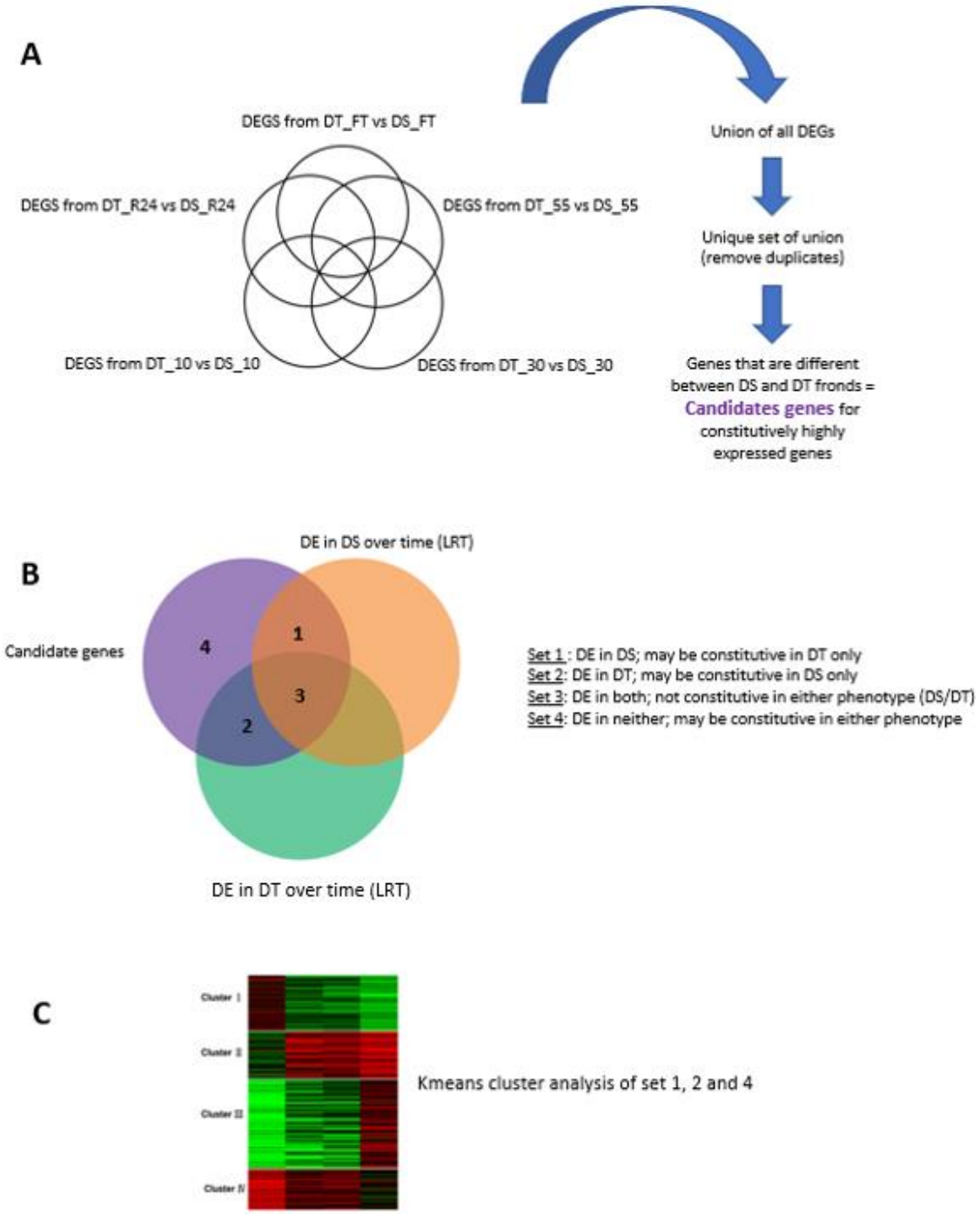

Figure 1: Exploring constitutively expressed genes in each phenotype. A: The differentially expressed genes (DEGs) in each 'cross' pairwise comparison are, by definition, genes that are different in each phenotype at least at one water content. The unique union of these represent candidate genes for different constitutive expression patterns across phenotype. B: The candidate genes are separate into four sets. 1 – Candidate genes that overlap with genes that are differentially expressed (DE) in desiccation sensitive (DS) fronds are possibly constitutively expressed in DT only. 2 – Candidate genes that overlap with genes that are DE in DT are possible constitutively expressed in DS only. 3 – Candidate genes that are DE in both DS and DT are therefore not constitutive in either phenotype. 4 – Candidate genes that are not DE in DS or DT, may be constitutive in both phenotypes. C: Sets 1,2 and 4 were subjected to kmeans clustering to determine clusters of constitutively expressed genes with interesting expression patterns.

4. Results and preliminary discussion

4.1. Sample quality control and outlier detection

To assess intragroup variability the biological replicates were compared to one another. The raw counts belonging to frond samples were subjected to a CPM transformation and then were log2 transformed. The correlation between biological replicates was measured using Pearson's coefficient. The intragroup Pearson correlation coefficients (r) for DS frond samples was >= 0.79. This implied that the DS frond replicates were well-correlated within each biological grouping. The DT frond samples had an r >= 0.76, except for one correlation which was 0.69. The lowest correlation belonged to the comparison of the replicates FDT_FT_3 and FDT_FT_1. This suggested that FDT_FT_1 and/or FDT_FT_3 were possible outliers. To further explore this and assess whether each biological grouping was distinct from one another, a PCA plot was generated. PCA was conducted using a vst transformation while varying the total number of top variable genes in the set. After including the top 30 000 variable genes in the PCA, clear separation of clusters relating to each biological condition was observed (Figure 2). PC1 accounted for 11% of the total variance and represented the distinction between frond phenotypes (DT versus DS). PC2, which accounted for a further 8% of the variance in the data, delineated the fronds according to RWC. More specifically, the more hydrated DT tissues (RWC >= 55%) were generally separated from the drier tissues (RWC <= 30%) (Figure 2). Generally, DT frond groupings were more clearly separated than DS fronds. Each DT grouping appeared to form their own clusters aside from the DT_30 and DT_R24 groups. This may be accredited to the fact that average RWC for the DT_R24 grouping was 30.05%. One might expect there to be unique signatures in the DT_R24 condition that are related to post-rehydration recovery processes. This will be explored later. The clear separation of DT samples indicated that despite the lower correlation (r) between frond samples, intergroup variability was still sufficient to capture unique gene profiles related to the stages of water loss. Conversely, the DS fronds showed little to no separation aside from the DS_FT grouping. One possible interpretation is that once sensitive fronds experience at least 55% water loss, senescent processes are activated and are consistent until loss of overall viability. This is further supported by the fact that the DS_R24 grouping does not form its own cluster. This hypothesis will also be explored in more detail later. The PCA plot corroborated the FDT_FT_1 replicate as a clear outlier. The FDT_FT_1 sample was clustering on the side of PC1 which represented sensitive tissues and on the side of PC2 which represented lower water contents (Figure 2). For

this reason. It was removed from further analyses. The FDT_FT grouping still contained three biological replicates so no changes to the analysis pipeline was required.

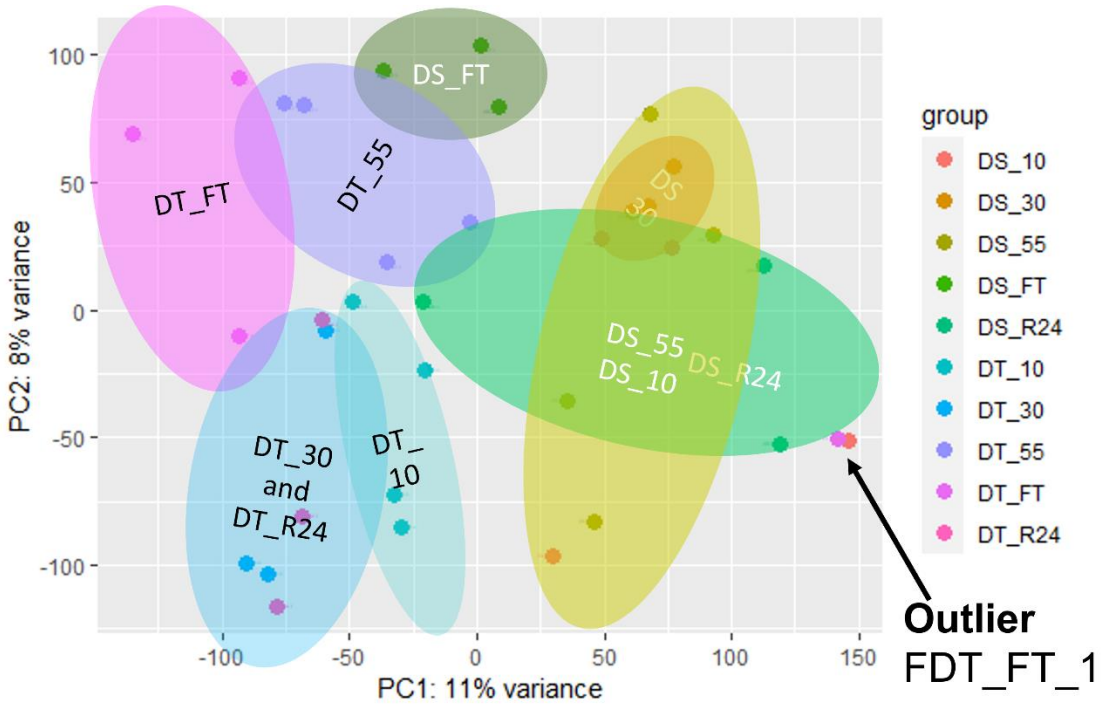

Figure 2 : PCA plot of frond tissues representing the 30 000 most variable genes. PCA analysis indicated that each biological grouping of DT fronds was distinct from one another and that the replicates within each grouping were well-correlated. The exception was the DT fronds at RWC = 30% (DT_30) and those sampled 24-hours post-rehydration (DT_R24). The DS fronds only showed a clear distinction between fully hydrated fronds (DS_FT) and the fronds experiencing any form of water stress did not appear to have clear gene expression profiles. Approximately 11% of the variance in the data (PC1) was accredited to phenotypic differences and a further 8% (PC2) was a result of differences in water stress. FDT_FT_1 was placed in a quadrant that did not match the phenotype or water stress level observed for the other replicates in the FDT_FT group. It was therefore considered an outlier in the frond data set. DS = desiccation sensitive; DT = desiccation tolerant; 30/55/10 = % RWC; R24 = 24-hours post-rehydration and FT = full turgor (or 100% RWC).

4.2. Inducible responses to desiccation over time (time-series analysis)

After the removal of outlier samples, the frond sequencing data was used to conduct differential expression analysis using the likelihood ratio test (LRT). LRT analysis tests for any gene that is differentially expressed across several factors and therefore suitable for time course experiments (Love, Huber and Anders, 2014). LRT analysis of the DS samples across time yielded 3374 DEGs, 2064 and 1310 were significantly ($p > 0.05$) up- and down-regulated respectively (Figure 3A). DT samples subjected to the LRT method contained 1961 and 863 significantly ($p > 0.05$) up- and down-regulated genes, respectively. A total of 2824 DEGs were identified over the time course of DT drying, desiccation and rehydration (Figure 3A).

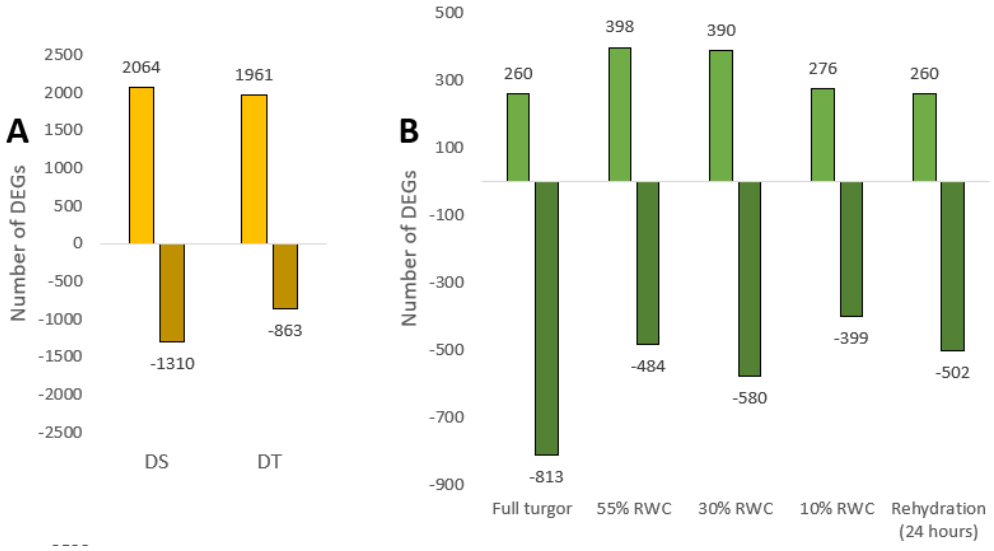

Figure 3: Number of differentially expressed genes (DEGs) observed in fronds. **A**: Number of DEGs in fronds tested using the likelihood ratio test (LRT). B: Number of DEGs in fronds compared in a pairwise fashion.

4.2.1. Tolerant fronds

Each set of DEGs was subjected to cluster analysis using the k-means method. A range of k-values (6 - 12) were tested for clustering. At k = 8, clusters that strongly resembled common inducible patterns that are characteristic of angiosperm resurrection plants were observed. Omics studies on resurrection plants have highlighted a 3 phase response to dehydration, namely an early stage (100-ca 55% RWC), a mid-stage (ca 55- 30% RWC) and a late stage (ca 30-5% RWC) (Costa *et al.*, 2017; Farrant and Hilhorst, 2021; Gabier *et al.*, 2021). These clusters were also subjected to functional enrichment analysis to determine whether specific

biological processes were enriched in each cluster. Clusters 1, 3, 5 and 6 were explored as they had the most interesting gene expression profiles.

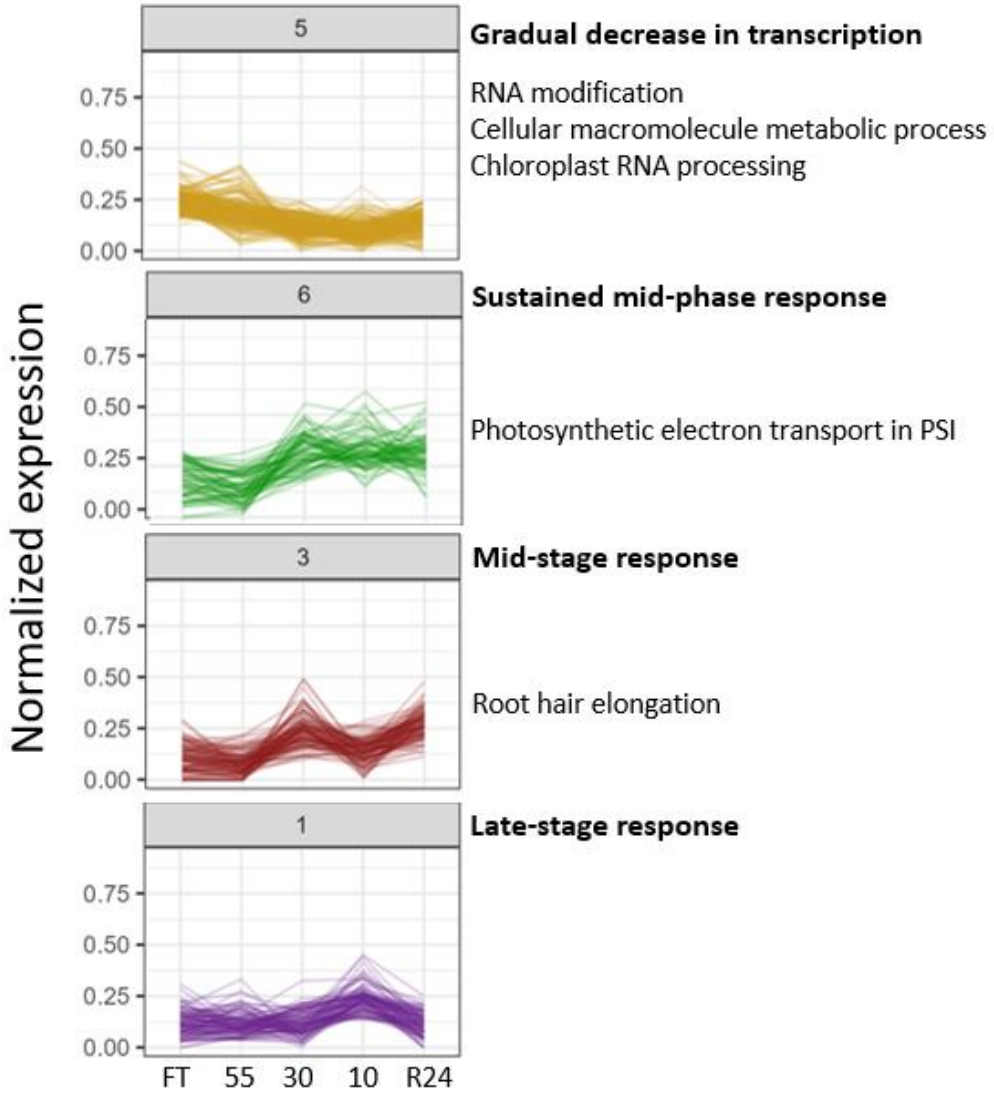

Figure 4: Gene expression clusters that demonstrate evidence of inducible gene expression in DT fronds across drying, desiccation and rehydration. Enriched GO terms are indicated right of each graph. Cluster 5 demonstrates a gradual decline in transcription-related processes through the course of drying. Cluster 6 and 3 represent mid-stage responses that are induced between 55 and 30% relative water content (RWC). Cluster 6 responses are sustained through the late stage and rehydration. Cluster 3 response drops between 30% and 10% RWC. Cluster 1 represent genes that are late-stage responses. Significantly enriched biological processes in each cluster are presented. FT = full turgor, 55, 30 and 10 are relative water content values, R24 = 24-hours post-rehydration.

Transcription

Cluster 5 genes (Figure 4) slowly decrease over the course of drying and desiccation but increase after rehydration. The enriched GO processes were highly linked to RNA processing which demonstrated that transcription gradually slows down during drying. The increase in transcription after rehydration implies that DT fronds were able resume metabolic function after drying to 10% RWC.

Protection of photosystem I

Photosynthetic electron transport in PSI occurred in cluster 6 (Figure 4) and was related to the upregulation of the protein 'PHOTOSYNTHETIC NDH SUBUNIT OF SUBCOMPLEX B3' (NDH) in the chloroplasts ($p = 3.98 \times 10^{-4}$) (Table S3). NDH is involved in cyclic electron flow which allows for the continued generation of ATP without accumulation of NADPH and elimination of ROS formed in PSI. A shift from linear to cyclic electron transport occurs in the angiosperm *Craterostigma pumilum* as plants dehydrate below 40% RWC (Zia et al., 2016). The balancing of ATP/NADPH is proposed to protect both photosystems from damage and a similar protective role for NDH has also been observed in *Craterostigma plantagineum* (*C. plantagineum*) (Nikkanen *et al.*, 2018; Xu *et al.*, 2021). The spores of the ferns *Osmunda japonica* (*O. japonica*) and *Adiantum capillus-veneris* (*A. capillus-veneris*) accumulate a fern-specific protein called 22 KDA PROTEIN OF CHLOROPLASTS IN THYLAKOID MEMBRANES prior to germination (Minamikawa, Koshiba and Wada, 1984). The role of 22 KDA PROTEIN OF CHLOROPLASTS has not been elucidated but the degradation of this protein during germination suggests a role for this protein in chloroplast maintenance in the desiccated state (Haupt, 1990; Inoue et al., 2000; Minamikawa et al., 1984). The presence of 22 KDA PROTEIN OF CHLOROPLASTS in a sustained mid-phase response (cluster 6) ($p = 5.19 \times 10^{-4}$) (Table S3) suggests that this protein participates in the protection of frond chloroplasts during desiccation and rehydration. In the late stage (cluster 1) chloroplastic protein FERRITIN-3 is found ($p = 5.19 \times 10^{-6}$) (Table S3). This protein is responsible for the accumulation of a soluble, non-toxic form of iron (Kim et al., 2021; Vigani et al., 2013). This could be to protect the cell from ferroptosis (the process of cell death that is a result of iron accumulation and lipid peroxidation) (Doll *et al.*, 2019; Bersuker *et al.*, 2020; Li *et al.*, 2020). This theory is supported by the upregulation and sustained presence of FERROPTOSIS SUPPRESSOR PROTEIN 1 (FSP1) which appears in cluster 6 ($p = 3.58 \times 10^{-4}$) (Table S3). FSP1 is an oxidoreductase enzyme that prevents cell death via ferroptosis. These iron reserves

could also facilitate the rapid reinstatement of photosynthesis upon rehydration as was previously observed in *A. caffrorum* (Farrant *et al.*, 2009).

Antioxidant responses

Catalase-3, cytochrome P450 709B2 and 734A1, and NADPH--cytochrome P450 reductase 1 are known antioxidant enzymes, the transcripts of which are significantly increased in the mid-stage and maintained through to rehydration (cluster 6) (Table S3). A probable *PHOSPHOLIPID HYDROPEROXIDE GLUTATHIONE PEROXIDASE 6* (PHGPx) displayed the same expression pattern ($p = 6.09 \times 10^{-8}$) (Table S3). PHGPx reduces hydrogen peroxide, lipid peroxides and organic hydroperoxide, by glutathione and is specifically localised in the mitochondria (Yang, Dong and Liu, 2006). Late-stage antioxidant responses include the upregulation of aldehyde dehydrogenase (Yang et al., 2012) and cytochrome P450 93A3, for which the exact function has not been determined (Table S3). The antioxidant producing enzyme 7-deoxyloganetin glucosyltransferase (7-DG) in the late stage was also observed ($p = 1.88 \times 10^{-3}$) (Table S3). 7-DG has been implicated in the production of secologanin, a monoterpene indole alkaloid (MIA) with antioxidant activity (Asada et al., 2013; Courdavault and Besseau, 2022; Matsuura et al., 2014).

Cellular glass formation

HEAT SHOCK 70 KDA PROTEIN 1 (*HSP70-1*) ($p = 3.03 \times 10^{-4}$) (Table S3) transcripts accumulated in the mid-stage and remained at high levels through desiccation and rehydration (cluster 6). HSP70-1likely stabilizes and refolds proteins through these phases. Upregulation of a second HSP70 was also observed in the late stage, alongside *D-XYLOSE-PROTON SYMPORTER-LIKE 3* (D-XPS) (Table S3). D-XPS has been implicated in the transport of sucrose reserves from the chloroplasts to the cytoplasm (Patzke *et al.*, 2019). Sucrose and HSP70 are the likely main drivers of cellular glass formation in tolerant fronds.

Lipid metabolism

Choline-phosphate cytidylyltransferase 1 is responsible for the regulation of phosphatidylcholine (PC) in membranes. It is upregulated in the mid-stage and sustained through to rehydration (cluster 6) ($p = 9.74 \times 10^{-5}$) (Table S3). It phosphorylates diacylglycerol (DAG) to phosphatidic acid (PA). PC and PA lipids have both been shown to confer tolerance to water loss (Foka *et al.*, 2020; Sun *et al.*, 2020). Also in the mid-stage response (cluster 3),

synaptotagmin-2 prevents the accumulation of toxic DAG at the plasma membrane ($p = 3.45 \times 10^{-4}$) (Table S3) (Lara *et al.*, 2018; Ruiz-Lopez *et al.*, 2021). The enzyme 3-ketoacyl-CoA synthase (KCS) 4 is required to produce very long-chain fatty acids (VLCFAs). The presence of VLFCAs during desiccation and recovery of *Xerophyta humilis* (*X. humilis*) has been observed, suggesting a possible role for KCS-4 as a late-stage (cluster 1) lipid response ($p = 1.94 \times 10^{-7}$) (Table S3) (Kim et al., 2021; Markham et al., 2011; Tshabuse et al., 2018). During the late stage, acyl-CoA--sterol O-acyltransferase 1 (ACAT) esterifies toxic cholesterol to cholesteryl esters ($p = 1.15 \times 10^{-3}$) (Table S3). In several cell-types, this prevents the accumulation of cholesterol which is toxic to the cellular environment. In the way of membrane integrity maintenance, phosphoinositide phosphatase SAC6 (SAC6) localizes to the tonoplast where it protects vacuolar shape (Nováková *et al.*, 2014). This occurs in the late stage ($p = 2.80 \times 10^{-7}$) (Table S3).

Responses in the roots

Cluster 3 represented genes that are consistent with a mid-stage response i.e., genes that are upregulated at some point between 55 and 30% RWC (Figure 4). These transcripts decline by 10% RWC, suggesting a transient role during dehydration/desiccation. Enriched terms included 'root hair elongation'. This process was related to the increased abundance of *XYLOGLUCAN-SPECIFIC GLUCURONOSYLTRANSFERASE 1* transcripts which are required for the development of root hairs ($p = 2.10 \times 10^{-4}$) (Table S3). Furthermore, a putative glucuronosyltransferase PGSIP8 which has xylan glucuronosyltransferase activity is also present ($p = 1.56 \times 10^{-3}$) (Table S3) (Peña *et al.*, 2012; Rennie *et al.*, 2012). Xyloglucans are also associated with cell wall folding. For example, Vicré et al (2004) showed that cell wall folding occurs simultaneously with a modification in the organization of xyloglucan-associated epitopes within the wall. Such modification in the cell wall architecture appears at the onset of dehydration and subsequently increases upon severe water loss to remain constant to air dryness (Moore et al., 2022). These enzymes show evidence for an increase in root hair length which could be in search of soil water and/or may reflect an organization of the cell wall. The movement of molecules between source (frond) and sink (rhizome/roots) tissues is implied by the presence of the cutin-monomer producing protein glycerol-3-phosphate acyltransferase (GPAT) RAM2 ($p = 6.92 \times 10^{-5}$) (Table S3). GPAT RAM2 is involved in the production of cutin monomers which are essential in preparing roots for colonization by arbuscular mycorrhizal fungi. Cutin also sends signals to fungal endophytes to promote their entry (Wang

et al., 2012). Presumably, cutin signalling required to recruit these fungi to root structures must be present in roots, therefore a source-sink dynamic is implied through the course of the mid-stage, late-stage, and rehydration (cluster 6).

Phytohormone regulation

Finally, several elements that belong to ABA-signalling in drought resistance are present in cluster 6. Pyrabactin resistance (PYR)/pyrabactin resistance-like (PYL)/regulatory component of ABA receptors (RCAR), protein phosphatase 2C (PP2C) and Sucrose non-fermenting (SNF) 1-related protein kinase 2 (SnRK2) are the three major components of ABA signalling. PP2Cs act as negative regulators and SnRK2s act as positive regulators of PYR/PYL/RCAR, respectively (references). The presence of several SnRK2s (*LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.2* (*LRL*), *RR RECEPTOR-LIKE SERINE/THREONINE-PROTEIN KINASE GSO1* (GSO1), transcription factor TGA2.2) and the relative absence of PP2Cs suggest that ABA-induced regulation has occurred at some point between 55% and 30% RWC, and these elements are seemingly not degraded 24-hours following rehydration (Table S3). At which time point ABA levels rise is yet to be measured, but this finding suggests that it must occur before 30% RWC (i.e., before entry into desiccation). Other than its role as a positive regulator of ABA signalling, the precise mechanism of LRL has yet to be determined. GSO2 has been shown to play a role in modulating the root endodermis. Effectively GSO2 increases waterproofing capacity of roots which protects roots from volatile influxes of water and ions during drying. It therefore plays an important role in ion homeostasis in roots (Tsuwamoto, Fukuoka and Takahata, 2008; Pfister *et al.*, 2014; Racolta, Bryan and Tax, 2014; Nakayama *et al.*, 2017). Finally, TGA2.2. has an implied function in salicylic acid- (SA) and auxin-inducible gene expression (Niggeweg *et al.*, 2000), therefore extending the possible phytohormone regulation in the fronds of *A. caffrorum* to ABA, SA and auxin. Mitogen-activated protein kinase kinase kinase 17 (MAPKKK17) is an ABA-induced signal transduction element (Danquah *et al.*, 2015), that is present in the mid-stage response and absent in the late stage (cluster 1) ($p = 4.48 \times 10^{-4}$) (Table S3). Other regulatory elements in this cluster include *HEAT STRESS TRANSCRIPTION FACTOR A-1* (HSF A1) and *PROTEIN TIFY 10A* and *10B* (Table S3). HSFs are known to regulate the expression of stress-related genes such as the HSPs (Guo *et al.*, 2016). Protein TIFY 10a and 10b have been determined as a repressor of jasmonic acid signalling, suggesting a negative

regulatory role of JA during desiccation (Chini *et al.*, 2007; Thines *et al.*, 2007; Hoo and Howe, 2009).

4.2.2. Sensitive fronds

While DT fronds displayed several of the canonical responses to drying over time, such patterns were not observed in DS fronds. Instead, cluster analysis produced eight gene expression clusters that were highly similar to one another. Cluster 2, 8, 3 and 6 (Figure 5) all show a decline in expression between full turgor and 55% RWC, and an increase in expression between 10% RWC and the 24-hour post-rehydration state. The sharpness of the incline and decline in expression varied between clusters. The remaining clusters were noisy (i.e., the expression profiles of all genes were too variable to be considered a robust cluster) and were therefore not chosen for this study (*data not shown*). Clusters 2, 6 and 3 were highly connected to processes relating to RNA processing (Figure 5). This highlights a considerable downregulation in transcription, and presumably, translation very early during drying. The downregulation of 'response to light stimulus' (Figure 8, *cluster 2*) suggests that photosynthesis must be declining during drying. The reduction in cluster 6 (Figure 5) which is highly associated with photosynthesis, carbon fixation and photorespiration does further assert this notion. The pentose phosphate pathway, which is believed to play a fundamental role in redox homeostasis in cells (Ge *et al.*, 2020), is also downregulated by at least 55% RWC. This finding suggested that limited action against increasing oxidative stress occurs in DS fronds and may imply that DS fronds were experiencing senescence below 55% RWC. The set of genes that were considered differentially expressed in DS fronds did contain senescence-associated elements such as '*AEROLYSIN-LIKE PROTEIN*' and '*BAG FAMILY MOLECULAR CHAPERONE REGULATOR 6* (BAG)' (Table S3). Aerolysin-like proteins form pores in membranes causing loss of membrane integrity (Podobnik, Kisovec and Anderluh, 2017). BAG family chaperones are regulators of several processes, including programmed cell death (PCD) (Kang *et al.*, 2006). Cluster 6 (Figure 5) shows seeming evidence of recovery in the photosystems, carbon fixation, and photorespiration. However, Nadal et al., (2021) showed that DS fronds did not ultimately remain photosynthetically viable upon rehydration, verifying the suggestion that DS fronds likely induce senescence upon drying below 55% RWC. Recent studies on angiosperm resurrection plants in which distinction is made between non-senescent tissues and those destined to senesce during desiccation have shown that senescence processes are induced in mid to late stages during dehydration, but senescence per se is only realised on rehydration (Radermacher et al., 2019; 2021; du Toit et al., 2022). Therefore, a seeming recovery of carbon

metabolism could be temporary, if the same pattern of senescence that occurs in angiosperms appears here (du Toit et al., 2022). Alternatively, this may simply be due to an increase in the relative abundance of these transcripts after rehydration. Metabolism associated with nutrient mobilization may have been initiated on rehydration, possibly enabling the whole plant to re-absorb nutrients from senescent fronds into the rhizome or other tissues. As discussed in the next chapter, evidence for the control of leaf senescence is in fact present in the rhizomes, suggesting considerable interplay between these two organs. Anabolism and subsequent export of macromolecules from senescent fronds to the rhizome, could include the degradation of cytosolic transcripts. A decrease in other transcripts could then inflate the relative abundance of transcripts related to photosynthesis, carbon fixation, and photorespiration (Figure 5, *cluster 6*).

Other proteins that are highly associated with desiccation tolerance were observed in the DS clusters (i.e., a general decrease in expression after 55% RWC) but not differentially expressed in DT plants. These proteins include *EARLY LIGHT INDUCIBLE PROTEIN* (ELIP), a LEA_2 *DOMAIN-CONTAINING PROTEIN* (Table S3). Although such proteins are not exclusive to the tolerant phenotype and have been observed in DS plants (Chávez Montes *et al.*, 2022), one might expect an inducible increase in transcripts related to these key proteins in tolerant fronds. As previously alluded to, fronds may contain a mixture of inducible and constitutive desiccation responses and it is possible that these proteins are constitutively expressed in DT fronds. In fact, the relatively low number of DEGs in the time-series studies (Figure 3A) also speaks to a muted transcriptional response in fronds. A similar 'low level' response was observed in the proteomic studies and is characteristic of desiccation tolerance observed in bryophytes (Alejo-Jacuinde and Herrera-Estrella, 2022) (Figure 1 in the introduction). To further explore the underwhelming response in the transcript expression patterns, owing to a higher level of constitutive gene expression, direct comparison of sensitive and tolerant tissues was subsequently conducted.

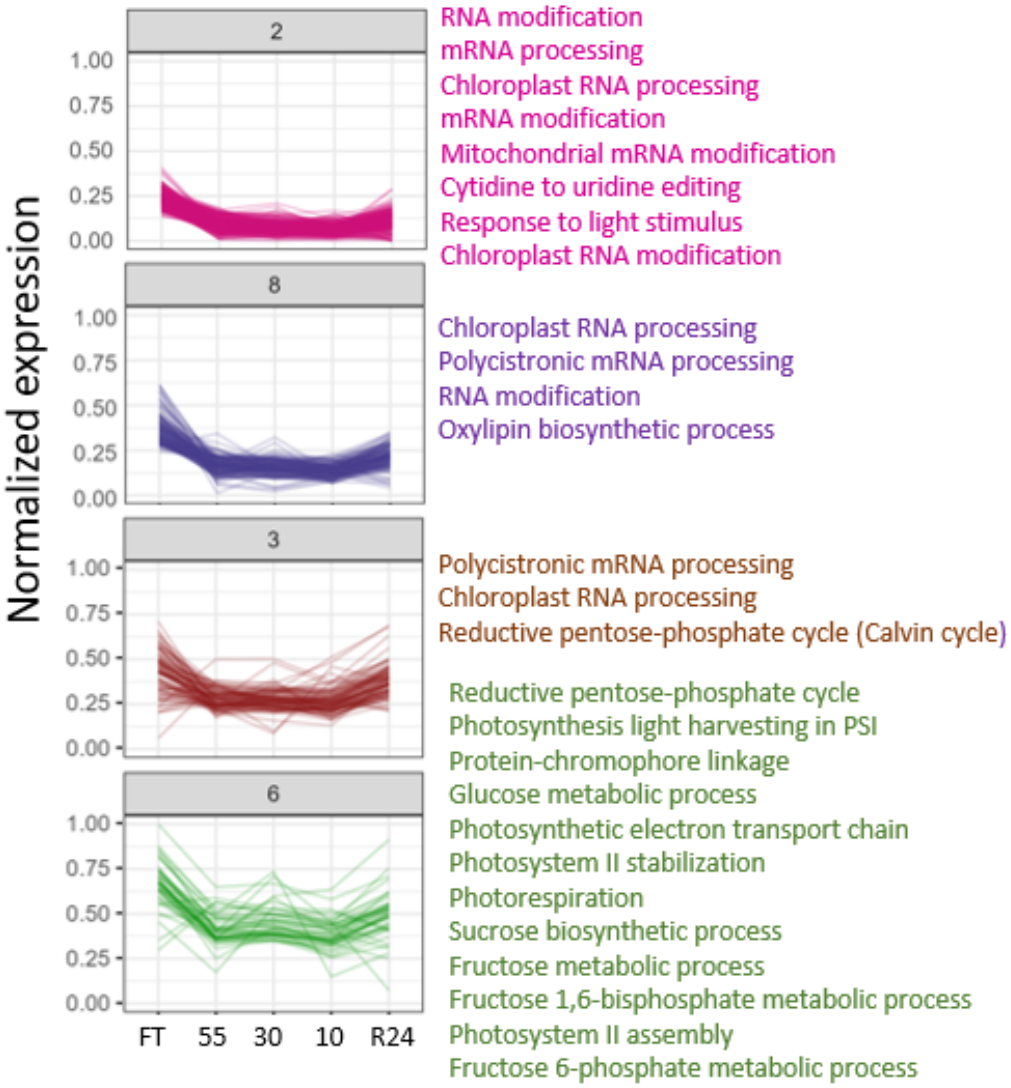

Figure 5: Dominant expression patterns in DS fronds during drying, desiccation and rehydration. All clusters show a decrease in expression between full turgor and 55% RWC, maintenance at a low state, and then an increase in expression between 10% RWC and 24-hours post rehydration. Significantly enriched biological processes in each cluster are presented. FT = full turgor, 55, 30 and 10 are relative water content values, R24 = 24-hours post-rehydration.

4.3. Direct comparison of DS/DT tissues and constitutive protection mechanisms

Arguably, one of the most important comparisons in this study, is the comparison of DT fronds to DS fronds. No other study, to date, has been able to directly compare DS and DT tissues of the same organism. Traditional approaches usually compare senescent vs non-senescent regions of the same leaf, whole DT plants with their detached (senescent) leaves and the DS/DT leaves of a sister species in the same genus (Willigen *et al.*, 2004; Illing *et al.*, 2005; Balsamo *et al.*, 2006; Le *et al.*, 2007; Farrant *et al.*, 2009; Yobi *et al.*, 2017; Marks *et al.*, 2021; Chávez Montes *et al.*, 2022). Pairwise analysis was used to compare DT tissues to DS tissues of the same water content. The resulting set of DEGs were subjected to GO enrichment analysis and the gene sets were studied. A total of 1073, 882, 970, 675 and 762 DEGs were observed in the full turgor, 55% RWC, 30% RWC, 10% RWC and 24-hour post rehydration stages respectively (Figure 3B). Figure 6 displays the enriched processes across each comparison.

Broadly speaking, at full turgor, DT fronds appear to prioritize processes related to regulating light dynamics and downregulate process related to structural changes in cell walls (Figure 6, *right column*). As DT plants experience dehydration, desiccation and presumably vitrification, they prioritize regulation of cellular respiration and focus on increasing the proportion of unsaturated fatty acids in cellular membranes. In addition, at 10% RWC a response to heat stress was enriched and 24-hours post-rehydration, a possible removal of the mass accumulated unsaturated fatty acids was observed (Figure 6, *right column*). Despite these independent comparisons being highly useful in demonstrating how DT fronds cope with drying and desiccation relative to DS fronds, the comparisons do not speak to whether involved genes contribute to desiccation tolerance via inducible or constitutive processes. To ascertain which genes are contributing to the DT phenotype via constitutive mechanisms, the genes within these pairwise comparisons were overlayed with the genes identified as DE in the time-series analysis (Figure 1). Genes that were not differentially expressed in DT, DS or both DS/DT (DE in neither) were subjected to kmeans clustering and visualized using a heatmap (Figure 7). Genes that were DE in DS contained two clusters (Figure 7A) and the second cluster (cluster 2) showed evidence of higher constitutive expression in DT fronds. Genes that were DE in DS had two clusters, cluster 3 and 4, which had noticeably higher constitutive expression patterns in DS (Figure 7A). Genes that were not DE in either DT or DS fronds each contained three clusters of genes that were more highly constitutively expressed in DS fronds (cluster 2,3 and 8) or more highly constitutively expressed in DT fronds (cluster 4, 5 and 7) (Figure 7B). Broadly speaking, several genes related to water stress were constitutively expressed in DS

fronds but typically more highly regulated in DT fronds as will be discussed below. These gene sets were used to determine if the identified key genes in desiccation tolerance were constitutively active.

UPREGULATED IN DS/ DOWNREGULATED IN DT	UPREGULATED IN DT / DOWNREGULATED IN DS
FULL TURGOR	
cellulose biosynthetic process	acropetal auxin transport
cell wall organization	basipetal auxin transport
plant-type primary cell wall biogenesis	stamen development
plant-type secondary cell wall biogenesis	positive gravitropism
xylem development	response to blue light
regulation of stomatal complex development	response to far red light
cutin biosynthetic process	anthocyanin accumulation in tissues in response to UV light
suberin biosynthetic process	photomorphogenesis
defense response to fungus	formation of plant organ boundary
	defense response
55% RWC	
lipoprotein metabolic process	regulation of cellular respiration
lipid transport	unsaturated fatty acid biosynthetic process
protein kinase C-activating G protein-coupled receptor signaling	
30% RWC	
	regulation of cellular respiration
	unsaturated fatty acid biosynthetic process
	energy coupled proton transport down electrochemical gradient
10% RWC	
positive regulation of erythrocyte aggregation	regulation of cellular respiration
	unsaturated fatty acid biosynthetic process
	cellular response to heat
Post-rehydration 24 hours	
flavonoid biosynthetic process	fatty acid omega-oxidation

Figure 6: Enriched biological processes in each phenotype at each time point measured in this study. At each time point a pairwise comparison between desiccation sensitive (DS) and desiccation tolerant (DT) tissue was conducted. The resulting up- and down-regulated genes were subjected to functional enrichment.

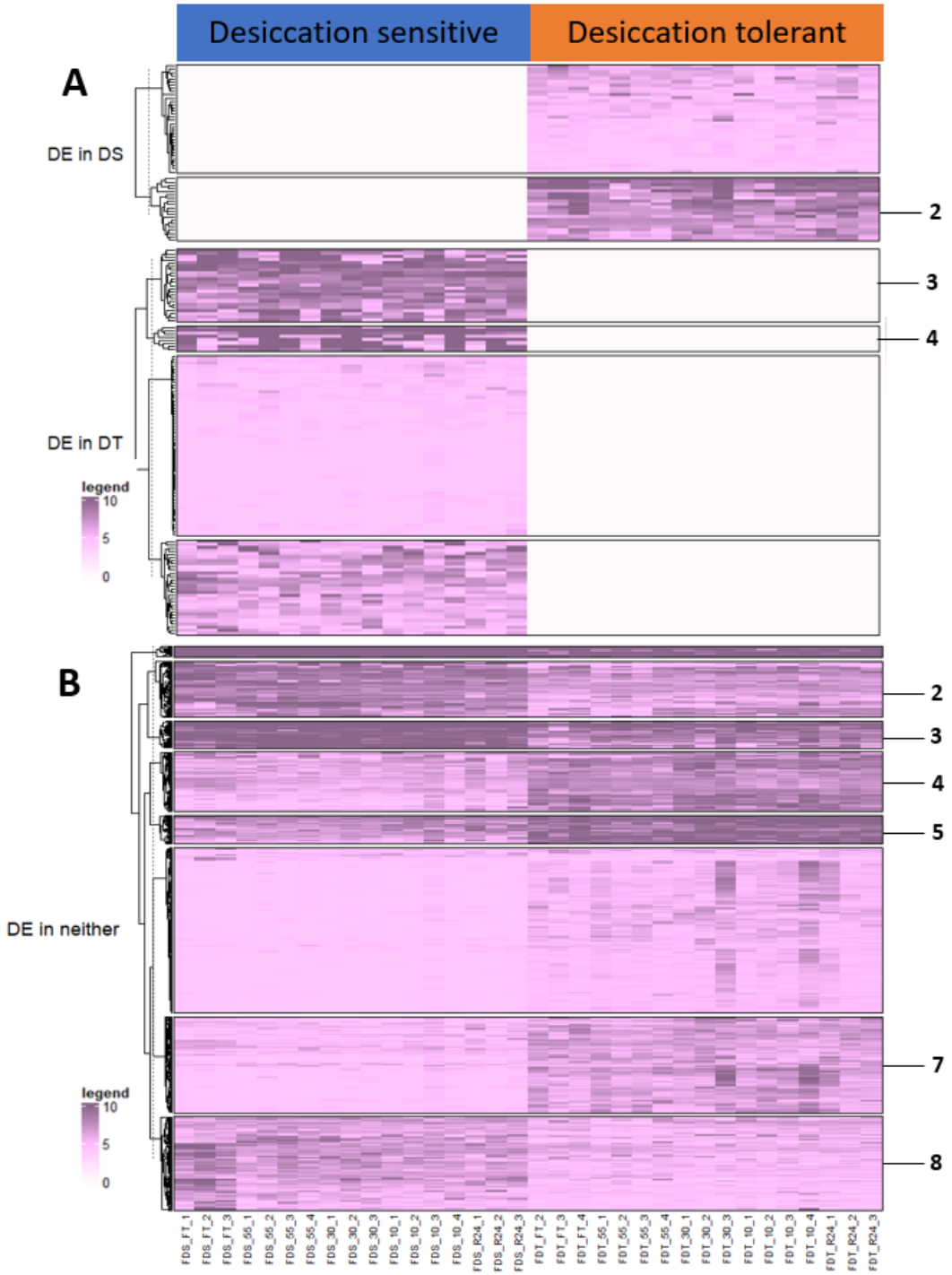

Figure 7: Relative patterns of constitutive expression in desiccation sensitive (DS) and desiccation tolerant (DT) fronds. A: Genes that were differentially expressed (DE) in either DS or DT. Cluster 2 were more constitutively active in DT fronds, while cluster 3 and 4 were more constitutively active in DS fronds. B: Genes that were not DE in DS or DT fronds. Clusters 2, 3 and 8 are more highly, constitutively expressed in DS fronds than DT fronds. Clusters 4, 5 and 7 are more highly constitutively expressed in DT fronds than DS fronds.

Maintenance of anthocyanins in DT fronds in the full turgor state

The enriched processes in DT fronds at full turgor (Figure 6, *right column*, *tan*) are mostly related to light dynamics. *ABC TRANSPORTER B FAMILY MEMBER 19* (ABC-B19) protein (log2foldchange (LFC) = 4.93, $p = 6.0 \times 10^{-5}$) was upregulated in DT fronds relative to DS fronds under conditions of high-water availability. ABC-B19 is an auxin efflux transporter that is negatively regulated by light. It controls the accumulation of anthocyanins and chlorophyll in response to light stimulus (Noh, Murphy and Spalding, 2001; Scherl *et al.*, 2002; Lin and Wang, 2005). However, it was also observed in cluster 2 (Figure 7B), suggesting that it has higher levels of constitutive expression in DS fronds. Tolerant fronds are subjected to higher irradiance during summer and produce peltate scales and orange hairs on the adaxial surfaces on the frond in response (Farrant *et al.*, 2009). On the other hand, sensitive fronds have greater chlorophyll contents compared to DT fronds (Farrant *et al.*, 2009; Nadal *et al.*, 2021). Constitutive expression of ABC-B19 suggests that it plays a role in chlorophyll maintenance in DS fronds. Seemingly, DS fronds do not alter chlorophyll levels during desiccation which implies a lack of an organized response to desiccation in DS fronds. In DT fronds, ABC-B19 likely influences the balance between chlorophyll and anthocyanins, particularly those that likely accumulate in the protective orange hairs. Four of the 20 most highly expressed genes in DT fronds at full turgor were UDP-glycosyltransferases (Table S1). UDP-glycosylation is the transfer of a glycosyl group to an acceptor molecule (aglycone). Glycosylation by UDP-glycosyltransferases has been shown to play a significant role in the stabilization, availability, and biological activity of anthocyanins (Le Roy *et al.*, 2016; Wang *et al.*, 2019; Dong *et al.*, 2020). It is possible then that the accumulated UDP-glycosyltransferases 86A1, 85A2, 85A3 and 85A5 play a role in maintaining high concentrations of anthocyanins under high light conditions. Furthermore, UDP-glycosyltransferases 85A2 and 85A3 showed higher levels of constitutive expression in DT fronds (Figure 7B, cluster 4). Noticeably, probable WRKY transcription factors 48, 12, 6 and 24 were also significantly upregulated in DT fronds, with at least 9-fold expression differences (Table S1). WRKY 42 is constitutive and constitutively more highly expressed in DT fronds (Figure 7B, *cluster 4*). WRKY24 has been shown to act as a negative regulator of ABA and gibberellic acid (GA) signalling (Zhang et al., 2015). WRKY12 confers drought and salt tolerance to transgenic soybean plants via an increase in proline content and decrease in malondialdehyde during drought (Shi *et al.*, 2018). WRKY6 controls processes related to leaf senescence, pathogen defence and ABA signalling during seed germination and development (Huang et al., 2016; Robatzek and Somssich, 2002).

WRKY48 is induced under stress and represses plant defence (Xing *et al.*, 2008). While many of the determined roles for the WRKYs in this dataset have relevance to the desiccation response, the regulatory mechanisms of desiccation tolerance are highly complex, and these signalling pathways have not yet been fully elucidated (Zhang et al., 2015). It is simply worth noting that multiple WRKY transcription factors were highly expressed in DT fronds (relative to DS fronds) in the full turgor state. The processes that are significantly downregulated in DT fronds relative to DS fronds are all related to processes that involve cell wall synthesis (Figure 6, *left column, dark and light green*). The most frequently observed enzyme in each GO term was cellular synthase (CSC). CSC is involved in cellulose synthesis which is a major component of plant cell walls. Suberin and cutin, which were also present, are glycerolipids that are also highly associated with the cell-wall (Huang and Cheng, 2019; Philippe *et al.*, 2020). Cell-wall modifying proteins were not concurrently observed at full turgor and therefore the downregulation in these processes likely represent a trade-off in growth to maintain the tolerant phenotype that must respond to high irradiance.

DT fronds are primed for desiccation stress by hydrogen peroxide at higher water contents

DT fronds at full turgor and 55% RWC showed very high fold changes in *ALCOHOL OXIDASE* transcripts (LFC = 9.74 and 11.85 respectively) (Table S1). Alcohol oxidase oxidizes methanol to produce formaldehyde and hydrogen peroxide (H_2O_2). Alcohol oxidase has been observed in the desiccated leaves of the resurrection angiosperm *Harbelea rhodopensis* (*H. rhodopensis*) (Dinakar and Bartels, 2013). Aldehyde oxidase (GLOX), another H_2O_2-producing enzyme (Whittaker *et al.*, 1996), showed higher levels of constitutive expression in DT fronds (Figure 7B, cluster 5). H_2O_2 is a very soluble ROS, which first acts as a signal to turn on drought associated mechanisms in all plants, including resurrection plants (Hossain *et al.*, 2015). Elevated levels of H_2O_2-producing enzymes in DT fronds suggest an ongoing signal for drought tolerance that is not present to the same degree in DS fronds.

DT fronds display regulated cellular respiration and increased unsaturated fatty acid content

Figure 6 clearly demonstrates two major responses of DT fronds to dehydration and desiccation stress that was not enriched in DS fronds. In the first instance, there is a strong prioritization of converting fatty acids to the unsaturated forms which has been observed in several other resurrection plants as it increases membrane fluidity during water loss (Figure 6, *right column, yellow*) (Platt, Oliver and Thomson, 1994; Navari-Izzo *et al.*, 1995; Quartacci *et al.*, 1997; Gasulla *et al.*, 2013; Neeragunda Shivaraj *et al.*, 2018; Tshabuse *et al.*, 2018; Mihailova *et al.*,

2022). The enzymes connected to this response are acyl-CoA desaturase and delta(12) fatty acid desaturase which both had an LFC of at least 9 at 55% RWC. Delta(12) fatty acid desaturase is required to desaturate fatty acids of the mitochondrial membrane (Caiveau *et al.*, 2001; Matos *et al.*, 2007), and in seeds has been shown to increase linoleic acid content in several membranous structures (Dehghan Nayeri and Yarizade, 2014). Acyl-CoA desaturases have a similar role and are usually associated with the membranes of endoplasmic reticulum (ER) (Wang *et al.*, 2014; Bai *et al.*, 2015; Berestovoy, Pavlenko and Goldenkova-Pavlova, 2020). In fact, *ACYL-COA DESATURASE*, and *DELTA(12) FATTY ACID DESATURASE* were more highly constitutively expressed in DT fronds (Figure 7B, cluster 7). Together, these enzymes may have a constitutive role in protecting the membranes of the mitochondria and ER, which implies maintenance of redox balance and protection of protein synthesis. Secondly, regulation of cellular respiration (Figure 6, *pink*) further demonstrated a prioritization towards the maintenance of redox homeostasis in the mitochondria. The proteins associated with the regulation of cellular respiration were protein TAR1, a late embryogenesis (LEA) protein, PPR (pentatricopeptide repeat) protein PDM1/SEL1 and apolipoprotein A1/A4/E domain. Protein TAR1 has been implicated in the regulation of mitochondrial gene expression and stability of mtDNA species (Coelho *et al.*, 2002). Mitochondrial LEA proteins have been shown to stabilize mitochondrial membranes in the dry state, and therefore playing a protective role, rather than a regulatory one (Tolleter, Hincha and Macherel, 2010). PPR protein PMD1/SEL1 is important in the post-transcriptional regulation of chloroplast transcripts (Zhang et al., 2015). Owing to the related roles of the chloroplast and mitochondria in ultimately converting light energy to ATP, it is not surprising then that the regulation of one organelle is connected to regulation of the other. The apolipoprotein A1/A4/E domain has not been studied in plants, but in mammals it has been implicated in the storage and trafficking of lipids within lipoproteins (Ding *et al.*, 2012). The persistence of these genes from 55% RWC through to 10% RWC suggest that they are likely required to prevent the uncontrolled production of ROS in the cell during dehydration, desiccation and in the vitrified state. Their presence in the vitrified state also suggests a possible requirement of these genes upon rehydration, and hence may be important in the maintenance of these structures during rehydration.

Transcription and translation below 30% RWC in DT fronds

Evidence for an increase in transcription and translation was observed at 30% RWC. Firstly, *HEAT STRESS TRANSCRIPTION FACTOR A-1* (*HSFA1*) was noticeably upregulated during dehydration and through to desiccation. HSFA1 has previously been shown to be a master

regulator of the heat stress response. Its upregulation at 55% RWC (LFC = 9.30) and 30% RWC (LFC = 8.69) suggests that at least some transcription is continued during drying and dehydration (Liu, Liao and Charng, 2011). *WHITE COLLAR 1 PROTEIN* (LFC = 8.95) which is responsible for DNA repair further suggests active transcription at 30% RWC (Gudjonsson *et al.*, 2012). Presumably, continued transcription is primarily focused on the production of essential genes in the response to desiccation. For example, at 30% RWC, transcripts encoding for the protective proteins *HSP70-2* and *HSP90-3* have accumulated (LFC = 9.76 and 8.97 respectively) and are likely required for cellular glass formation. *HSP70, HSP70-2, HSP-90* and heat shock protein *HSS1* all showed higher levels of constitutive expression in DT fronds (Figure 7B, cluster 7). In addition, *ELONGATION FACTOR 1-ALPHA* and 3 which are upregulated at 55% and 30% RWC (LFC = 10.91 and LFC = 10.49 respectively) are responsible for the recruitment of tRNAs to ribosomes during protein synthesis (Table S1) (Andersen, Nissen and Nyborg, 2003). In fact, *ELONGATION FACTOR 1-ALPHA*, *ELONGATION FACTOR-2* and *ELONGATION FACTOR-3* were constitutively highly expressed in DT fronds (Figure 7B, cluster 7). This further implied that the functional products of mRNA are likely still being synthesized at 30% RWC. Together, upregulation of these proteins in DT fronds indicated efforts to continue transcription and translation as key elements in surviving desiccation. Ongoing transcription at such low water contents have been reported for many angiosperm resurrection species (Dace et al., 1998; Radermacher et al., 2021; Xu et al., 2021; Farrant and Hilhorst, 2022) and it has been proposed that this is due to presence of, inter alia, NaDES forming metabolites that facilitate this (du Toit et al., 2021).

Antioxidant responses in DT fronds

Highly abundant transcripts encoding antioxidant enzymes at 55% RWC included *ALDEHYDE DEHYDROGENASE* (LFC = 11.85) (Table S1). *ALDEHYDE DEHYDROGENASE* is well known as an antioxidant enzyme which detoxifies malondialdehyde, acetaldehyde, propionaldehyde during water loss. Furthermore, an *ALDEHYDE DEHYDROGENASE* was found to be upregulated in the resurrection angiosperm *Xerophyta schlecteri* during drying (Kirch, Nair and Bartels, 2001; Mundree *et al.*, 2002; Moore *et al.*, 2009; Rodriguez *et al.*, 2010). *ALDEHYDE DEHYDROGENASE* was still highly upregulated at 30% RWC (LFC = 9.40), at which point catalase T had also accumulated (LFC = 11.39). Microsomal *GLUTATHIONE S-TRANSFERASE 3* and chloroplastic *THIOREDOXIN F-TYPE ENZYME* which have also been implicated in the antioxidant response were constitutively expressed in DT fronds (Nikkanen and Rintamäki, 2019) (Figure 7B, cluster 5). Transcripts encoding *7-*

DEOXYLOGANETIN GLUCOSYLTRANSFERASE, a key enzyme in the production of the antioxidant secologanin, were also constitutively expressed in DT fronds (Figure 7B, cluster 4 and 5). In DS fronds, *GLUTATHIONE S-TRANSFERASE F10* was more highly constitutive than in DT fronds (Figure 7B, cluster 8) which demonstrates that antioxidant processes occur in both fronds, but the ones described here may be specific to the DT phenotype.

Carbon metabolism in DT fronds

A mitochondrial iron-sulphur (Fe-S) cluster biogenesis chaperone was significantly upregulated at 30% RWC (LFC = 10.34). In fact, this chaperone was constitutively more active in DT fronds (Figure 7B, cluster 7). Fe-S clusters are protein-clusters built on an Fe-S scaffold that are often components of electron transfer proteins. They are found in chloroplasts and mitochondria and are important in energy harvesting reactions in photosynthetic organisms. Some Fe-S clusters are inactive in electron transfer processes, and instead play a structural role in the cell. Fe-S cluster biogenesis chaperone participates in the correct assembly of Fe-S clusters (Braymer and Lill, 2017; Puglisi and Pastore, 2018; Shi *et al.*, 2021; Marengo *et al.*, 2022). Increases in Fe-S cluster biogenesis could indicate continued energy production to fuel essential processes such as mass accumulation of the components required for vitrification. It could also indicate a role in sequestering iron into non-toxic forms to prevent ferroptosis as previously discussed. At 30% RWC the upregulation of transcripts coding for other enzymes involved in energy metabolism in the mitochondria (*PYRUVATE CARBOXYLASE*; LFC = 9.69 and *FRUCTOSE-BISPHOSPHATE ALDOLASE* 1; LFC = 10.00) was also observed. Pyruvate carboxylase converts pyruvate to oxaloacetate which fuels the TCA cycle (Valle, 2017). Fructose-bisphosphate aldolase 1 is an intermediate glycolytic enzyme which eventually reduces glucose to pyruvate which is another requirement for the TCA cycle (Ziveri *et al.*, 2017). Xu et al. (2021) observed the accumulation of several TCA intermediates during dehydration in the resurrection angiosperm *C. plantagineum* and concluded that cellular respiration is likely maintained during dehydration. The same authors observed increases in succinate dehydrogenase during dehydration, further supporting the notion that the TCA cycle is maintained during water stress. In fact, in a survey of resurrection plants, it has been pointed out that respiration is the last metabolism to be shut down during desiccation, and the first to be started up upon rehydration (Farrant and Hilhorst, 2022). In the DT fronds this enzyme is upregulated at 10% RWC (LFC = 7.21), suggesting that the DT fronds have not yet reached a state of quiescence at 10% RWC. It is possible that their air-dry state occurs at some time point below 10%, and that carbon metabolism may be occurring at water contents as low as 10%

RWC. Alternatively, late transcription could be taking place for translation on rehydration. In fact, *FRUCTOSE-BISPHOSPHATE ALDOLASE* and *SUCCINATE DEHYDROGENASE* were both constitutively expressed in DT fronds (Figure 7B, cluster 7). As mentioned above, this is likely owing to the formation of NaDES metabolites at lower water contents (du Toit et al., 2021). *SUCCINATE-SEMIALDEHYDE DEHYDROGENASE* was also upregulated at 10% RWC (LFC = 7.18) and has been implicated in the 4-aminobutyric acid (GABA) shunt. Glutamate decarboxylase produces GABA from glutamate (Zhang *et al.*, 2017) and was constitutively expressed in DT fronds (Figure 7B, cluster 5). The GABA shunt, and possible continuous preparation of GABA, has been proposed as a mechanism to replenish succinate in the TCA cycle which maintains energy production during water stress (Fait *et al.*, 2008). Finally, *RIBOSE-5-PHOSPHATE ISOMERASE A* (RPIA) had accumulated at 10% RWC (LFC = 6.32). RPIA functions in the pentose phosphate pathway where, ultimately, it produces NADPH and protects against the formation of destructive ROS species. RPIA also plays a role in the Calvin cycle in the ultimate conversion of glucose into storage forms such as starch (Chen *et al.*, 2020). An accumulation of starch was not observed in the mesophyll cells of dry DT fronds (Figure 1). This suggests that if starch is being accumulated for storage, it is likely transported to the rhizome. Alternatively, the upregulation of RPIA at 10% RWC is likely in the production of energy for biosynthetic processes.

Trehalose and heat shock proteins are components of the glassy state

In the desiccated state, *HSP70-4* was upregulated (LFC = 7.06) (Table S1). As previously discussed, *HSP70, HSP70-2, HSP-90* and heat shock protein HSS1 all showed higher levels of constitutive expression in DT fronds (Figure 7B, cluster 7). *XYLOSE ISOMERASE*, which has been implicated in the production of trehalose, was also upregulated at 10% RWC (LFC = 8.21) (Reina-Bueno *et al.*, 2012). Together, these findings suggest that trehalose and HSP accumulation are key drivers in the vitrification process in DT fronds.

Possible recovery processes

The accumulation of transcripts related to cellular respiration as detailed under the '*Carbon metabolism in DT fronds*' subheading could also be interpreted as a mechanism to facilitate rapid metabolic recovery. DT fronds of *A. caffrorum* have been shown to resume metabolic activity as early as 4-hours post rehydration (Farrant *et al.*, 2009). Accumulation of these transcripts before quiescence could be for the rapid reinstatement of energy production when water becomes available again. *HIGH-AFFINITY NITRATE TRANSPORTER 2.1* and *ACID*

PHOSPHATASE were also significantly upregulated at 10% RWC (LFC = 6.98 and 8.15 respectively) which could similarly facilitate the rapid return of phosphate and nitrogen metabolism. Elongation factors that were upregulated and those that were constitutively expressed in DT fronds likely mediate the resumption of translation following rehydration. Cell wall protein PIR5 and protein *HOTHEAD* were the only genes that were significantly upregulated in DT fronds compared to DS fronds at the 24-hour rehydration stage (LFC = 7.26 and 10.01 respectively). Protein *HOTHEAD* was associated with fatty acid omega-oxidation that was enriched in DT fronds during the recovery process (Figure 6, *left column*). Omega-oxidation of fatty acids is an alternative to β-oxidation which occurs during periods of high energy demand in the human body. Omega-oxidation is believed to remove xenobiotics and lipid structures such as sphingolipids and other constituents of the plasma membrane. This process likely reverses the mass preparation of unsaturated fatty acids and may recycle these components for further use in the TCA cycle (Xu et al., 2017; Talley and Mohiuddin, 2022).

5. Supplementary material

Table S1: Most highly upregulated genes in DT fronds relative to DS fronds			
Full turgor			
gene id	gene name	log2FoldChange	padj
PB_29785	Alcohol oxidase 1	-9.741050864	0.000558
PB_30184	UDP-glycosyltransferase 86A1	-8.707206846	5.37E-06
PB_38364	Cell wall protein PIR5	-7.770236786	0.003982
PB_30189	UDP-glycosyltransferase 85A2	-7.061090337	0.000756
PB_28733	Linamarin synthase 2	-6.730656342	1.25E-07
PB_4892	Germin-like protein subfamily 1 member 16	-6.017043008	0.001045
PB_1078	ADP,ATP carrier protein 1, mitochondrial	-5.544355751	5.17E-06
PB_28742	UDP-glycosyltransferase 85A3	-5.264162426	0.004192
PB_21869	Berberine bridge enzyme-like 13	-5.186206462	0.012325
PB_30957	Cytochrome P450 716B1	-5.175023777	3.16E-09
PB_34399	UDP-glycosyltransferase 85A5	-5.172906848	0.012406
PB_27488	Mutanase Pc12g07500	-5.046730568	0.000647
PB_5860	Probable WRKY transcription factor 48	-4.978581788	0.007874
PB_28291	Jasmonate-induced oxygenase 4	-4.949457265	0.007206
PB_12955	ABC transporter B family member 19	-4.933551565	0.001166
PB_11614	U-box domain-containing protein 13	-4.733521124	0.034915
PB_16558	ABC transporter B family member 19	-4.659343944	6.79E-05
PB_85	Probable WRKY transcription factor 12	-4.508729323	0.000299
PB_31067	Polygalacturonase QRT3	-4.493712057	0.007652
PB_3725	pathogenesis-related protein 1-like	-4.429170673	0.010011
55% RWC			
gene id	gene name	log2FoldChange	padj
PB_29785	Alcohol oxidase 1	-11.85385743	1.98E-06
PB_39355	Tubulin alpha chain	-10.8970462	1.78E-05
PB_37059	BTB domain-containing protein	-10.43487114	3.27E-05
PB_39204	Tubulin beta chain	-10.41957185	0.000161
PB_16020	Uncharacterized protein	-10.35425588	4.78E-05
PB_36638	Acyl-CoA desaturase	-10.1804145	0.000206
PB_35699	Acyl-CoA desaturase	-10.15178131	1.44E-05
PB_37189	Aldehyde dehydrogenase	-10.00936926	0.000492
PB_38692	Probable pathogenesis-related protein ARB_02861	-9.853206552	0.000103

PB_36550	Major facilitator-type transporter ecdD	-9.83742705	0.000364
PB_553	Aldehyde dehydrogenase	-9.536621196	0.000368
PB_39111	Uncharacterized protein	-9.432300785	0.000781
PB_50681	Repressible high-affinity phosphate permease	-9.407291631	0.000278
PB_39028	Heat shock cognate 70 kDa protein	-9.338722383	0.001139
PB_48776	Heat stress transcription factor A-1	-9.304560346	5.19E-05
PB_40303	Acid phosphatase	-9.289746708	0.001637
PB_40473	Elongation factor 3	-9.262553222	0.001183
PB_38756	Acid phosphatase	-9.20626936	0.003754
PB_37078	Uncharacterized protein	-9.203124101	0.006079
PB_37931	Delta(12) fatty acid desaturase	-9.164076908	0.002517
30% RWC			
gene id	gene name	log2FoldChange	padj
PB_36479	Catalase T	-11.39129887	5.03E-05
PB_33048	Elongation factor 1-alpha	-10.91310083	0.001281
PB_5224	Allergen Asp f 7 homolog	-10.84084627	1.29E-05
PB_41492	Elongation factor 3	-10.48996966	0.000288
PB_17209	Iron-sulfur cluster biogenesis chaperone, mitochondrial	-10.34360695	0.000449
PB_35982	Putative inorganic phosphate transporter C8E4.01c	-10.21121691	0.000183
PB_40962	Fructose-bisphosphate aldolase 1	-9.99171904	0.000611
PB_49678	Heat shock 70 kDa protein 2	-9.762908974	3E-05
PB_50685	Pyruvate carboxylase	-9.685445517	0.002666
PB_33055	Elongation factor 1-alpha	-9.665558471	0.001566
PB_36888	Aldehyde dehydrogenase	-9.400345032	0.00055
PB_39272	GTP-binding protein EsdC	-9.270386362	0.00188
PB_38926	Uncharacterized protein	-9.181162843	0.008447
PB_38714	Actin, gamma	-9.042770455	0.003274
PB_49132	Heat shock protein 90-3	-8.974184722	5.35E-05
PB_39008	White collar 1 protein	-8.95638443	0.003224
PB_48404	E3 ubiquitin-protein ligase TRIP12	-8.896763013	0.002053
PB_50186	Heat shock 70 kDa protein 2	-8.884246978	0.009516
PB_37757	Aldehyde dehydrogenase	-8.839865166	0.000286
PB_48494	Heat stress transcription factor A-1	-8.692593766	0.004078
10% RWC			
gene id	gene name	log2FoldChange	padj
PB_51016	Xylose isomerase	-8.212328277	0.002093
PB_39355	Putative pentatricopeptide repeat-containing protein At1g68930	-8.186791116	0.000242
PB_8325	Acid phosphatase	-8.149339718	0.01229
PB_5224	Uncharacterized protein	-8.130717508	3.33E-05

PB_39111	Probable receptor-like protein kinase At1g30570	-7.82850619	0.000918
PB_29785	uncharacterized protein LOC110114147	-7.620571655	0.000237
PB_28462	Jasmonate-induced oxygenase 4	-7.509093895	0.000149
PB_38683	Uncharacterized protein	-7.445972306	0.002517
PB_18585	RRM domain-containing protein	-7.307960701	7.74E-06
PB_42316	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	-7.215002736	0.000806
PB_35346	Succinate-semialdehyde dehydrogenase, mitochondrial	-7.184463101	0.020363
PB_16008	Heat shock 70 kDa protein 4	-7.064100701	1.15E-05
PB_16007	High-affinity nitrate transporter 2.1	-6.975683439	7.64E-07
PB_37922	Phragmoplastin DRP1E	-6.942841201	0.018967
PB_44676	Chitin synthase D	-6.941517765	0.034316
PB_37285	pathogenesis-related protein 1-like	-6.69147906	0.000734
PB_44481	Primary amine oxidase	-6.628254098	0.02324
PB_36888	Elongation factor 1-alpha	-6.364671413	0.005657
PB_3671	1,3-beta-glucanosyltransferase gel1	-6.324865949	1.26E-06
PB_38756	Ribose-5-phosphate isomerase A	-6.319060203	0.035348
24 hours post rehydration			
gene id	gene name	log2FoldChange	padj
PB_12788	Protein HOTHEAD	-10.01206179	0.001599
PB_38364	Cell wall protein PIR5	-7.260097645	0.017947

Table S2: Most highly upregulated genes in DS fronds relative to DT fronds			
Full turgor			
gene id	gene name	log2FoldChange	padj
PB_603	Uncharacterized protein	11.09609	3.39E-04
PB_13582	Ricin B-like lectin R40C1	10.26723	1.24E-04
PB_34439	Microtubule-associated protein 70-1	9.986803	6.04E-06
PB_42741	glycine-rich protein DOT1-like isoform X1	9.586121	2.30E-02
PB_13435	Cation/H(+) antiporter 19	8.843937	1.06E-02
PB_1506	Aldo-keto reductase family 4 member C10	8.705107	3.43E-05
PB_30109	Aquaporin NIP3-1	8.413776	2.19E-05
PB_31682	Probable serine/threonine-protein kinase PBL18	8.082725	3.11E-05
PB_29507	Aldehyde oxidase GLOX	7.943922	5.75E-08
PB_13859	IPD083Cf	7.880906	1.73E-02
PB_50748	Transcription factor MYB61	7.841758	1.63E-04
PB_26390	Glucomannan 4-beta-mannosyltransferase 9	7.675999	1.81E-06
PB_7759	Aspartic proteinase nepenthesin-2	7.339524	1.77E-05
PB_42218	Glucan endo-1,3-beta-glucosidase	7.304844	8.14E-03
PB_50749	Glycerol-3-phosphate acyltransferase RAM2	7.162345	1.05E-02
PB_31586	Aldehyde oxidase GLOX	7.046892	2.79E-04
PB_38481	Omega-hydroxypalmitate O-feruloyl transferase	6.893884	2.94E-02
PB_34575	Peroxidase 53	6.827776	1.43E-03
PB_600	Ornithine decarboxylase	6.755722	1.29E-03
PB_9564	GDSL esterase/lipase At4g28780	6.6371	2.05E-03
55% RWC			
gene id	gene name	log2FoldChange	padj
PB_24690	GEM-like protein 1	12.8281	1.24E-05
PB_23784	Phenylcoumaran benzylic ether reductase IRL1	10.24999	3.59E-06
PB_29507	Aldehyde oxidase GLOX	9.097545	3.18E-09
PB_30109	Aquaporin NIP3-1	9.052646	3.29E-04
PB_1553	Isoprenyl transferase 1	8.320173	2.85E-02
PB_4275	Soluble inorganic pyrophosphatase 4	7.914917	6.53E-04
PB_7823	uncharacterized protein LOC9635598	7.653943	3.56E-03
PB_31682	Probable serine/threonine-protein kinase PBL18	7.252982	1.17E-04
PB_28309	Probable starch synthase 4, chloroplastic/amyloplastic	7.167547	2.76E-04
PB_24962	Sodium/hydrogen exchanger 3	6.782993	3.54E-02
PB_31586	Aldehyde oxidase GLOX	6.726899	1.04E-02

PB_5556	Protein DOG1-like 3	6.475737	4.35E-04
PB_10051	Membrane protein PM19L	6.386205	1.49E-03
PB_1332	Altered inheritance of mitochondria protein 32	6.308045	8.23E-03
PB_28879	Signal recognition particle subunit SRP68	6.246283	9.45E-03
PB_13280	Glutamine--tRNA ligase	5.882085	2.08E-03
PB_28661	Allantoate deiminase 1	5.840188	7.57E-03
PB_7010	Peptidase A1 domain-containing protein	5.796915	3.17E-03
PB_11190	Peptidyl-prolyl cis-trans isomerase CYP21-1	5.685605	3.20E-03
PB_34708	AAI domain-containing protein	5.59821	6.22E-03
30% RWC			
gene id	gene name	log2FoldChange	padj
PB_603	Uncharacterized protein	10.26238	3.22E-03
PB_33741	Lactamase_B domain-containing protein	9.157319	3.55E-05
PB_33499	22-kDa protein of chloroplasts in green spores	8.282573	1.10E-03
PB_17621	Putative membrane protein	7.903489	3.59E-03
PB_29507	Aldehyde oxidase GLOX	7.672153	3.01E-05
PB_48640	Polyadenylation and cleavage factor homolog 4	7.305895	4.40E-02
PB_43060	Alpha-amylase	7.25259	6.69E-03
PB_30109	Aquaporin NIP3-1	7.167456	1.12E-02
PB_13068	uncharacterized protein LOC112342872	6.609645	9.99E-03
PB_3603	Chalcone synthase	6.522946	1.31E-03
PB_29090	Heparanase-like protein 3	6.47753	6.44E-04
PB_5	Ferritin-3, chloroplastic	6.375285	9.11E-03
PB_34708	AAI domain-containing protein	6.234333	5.63E-03
PB_15559	7-deoxyloganetin glucosyltransferase	6.213493	8.84E-03
PB_8177	Uncharacterized protein	6.179007	4.41E-03
PB_31584	Aldehyde oxidase GLOX	6.171041	9.08E-04
PB_14948	Uncharacterized protein	6.149859	1.05E-03
PB_22861	Web family protein	6.072832	7.13E-03
PB_31599	Uncharacterized protein	5.73636	1.44E-02
PB_43100	Protein SPIRAL1-like 2	5.534468	1.92E-02
10% RWC			
gene id	gene name	log2FoldChange	padj
PB_7687	ATP-dependent protease ATPase subunit HslU	9.202836	9.98E-07
PB_33182	Cyanohydrin beta-glucosyltransferase	8.527164	7.82E-04
PB_29507	Aldehyde oxidase GLOX	7.036898	2.45E-05
PB_9489	Protein BOBBER 1	6.726742	3.58E-02

PB_34433	Peptidyl-prolyl cis-trans isomerase FKBP16-4, chloroplastic	6.456052	7.08E-03
PB_3603	Chalcone synthase	6.373267	3.66E-04
PB_23485	Tropinone reductase homolog At1g07440	5.827395	1.39E-03
PB_15559	7-deoxyloganetin glucosyltransferase	5.783927	7.83E-03
PB_30948	Tubulin beta-1 chain	5.302451	2.52E-03
PB_11069	CVNH domain-containing protein	5.211902	6.04E-03
PB_16824	Phenylalanine ammonia-lyase	5.125456	1.07E-03
PB_8614	Chitin-binding protein CbpD	5.048187	1.38E-03
PB_29820	Shikimate O-hydroxycinnamoyltransferase	4.641493	3.21E-05
PB_5042	Xylose isomerase	4.55164	2.57E-03
PB_8804	Protein COBRA	4.527699	4.46E-03
PB_16894	Tetratricopeptide repeat protein 38	4.183004	4.22E-03
PB_17792	Linoleate 9S-lipoxygenase 2	4.142726	3.05E-03
PB_31584	Aldehyde oxidase GLOX	4.088457	4.29E-02
PB_16065	Hypothetical protein (Fragment)	3.848855	1.61E-02
PB_11201	Glutathione S-transferase F10	3.840799	1.56E-02
24 hours post rehydration			
gene id	gene name	log2FoldChange	padj
PB_13582	Ricin B-like lectin R40C1	7.989058	1.21E-02
PB_17506	Chalcone synthase	7.1514	1.80E-02
PB_43060	Alpha-amylase	6.851867	9.53E-03
PB_29507	Aldehyde oxidase GLOX	5.692972	7.38E-03
PB_11201	Glutathione S-transferase F10	5.204632	9.23E-04
PB_3603	Chalcone synthase	5.034426	3.46E-02
PB_32385	Flavonoid 3',5'-hydroxylase 2	4.250595	5.90E-03
PB_12980	Coatomer subunit alpha-1	4.022107	4.52E-03
PB_51178	Calcineurin B-like protein 1	3.870053	3.51E-02

Table S3: Significant genes in LRT analysis	
Gene name	**p-value**
Photosynthetic NDH subunit of subcomplex B 3, chloroplastic	3.99E-04
22-kDa protein of chloroplasts in green spores	5.35E-04
Ferritin-3, chloroplastic	5.19E-06
Ferroptosis suppressor protein 1	3.59E-04
Catalase-3	6.42E-05
Cytochrome P450 709B2	5.70E-04
Cytochrome P450 734A1	1.03E-05
NADPH--cytochrome P450 reductase 1	1.36E-03

Probable phospholipid hydroperoxide glutathione peroxidase 6, mitochondrial	6.09E-08
Aldehyde dehydrogenase	4.26E-04
Cytochrome P450 93A3	6.38E-04
7-deoxyloganetin glucosyltransferase	1.89E-03
Heat shock 70 kDa protein 1	3.28E-04
Heat shock 70 kDa protein	9.29E-04
D-xylose-proton symporter-like 3, chloroplastic	8.32E-06
Choline-phosphate cytidylyltransferase 1	9.74E-05
Synaptotagmin-2	3.45E-04
3-ketoacyl-CoA synthase 4	1.94E-07
Acyl-CoA--sterol O-acyltransferase 1	1.15E-03
Phosphoinositide phosphatase SAC6	2.80E-07
Xyloglucan-specific galacturonosyltransferase 1	2.10E-04
Putative glucuronosyltransferase PGSIP8	1.56E-03
Glycerol-3-phosphate acyltransferase RAM2	6.92E-05
LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.2	7.77E-04
Transcription factor TGA2.2	4.69E-04
Mitogen-activated protein kinase kinase kinase 17	4.48E-04
Heat stress transcription factor A-1	1.04E-03
Protein TIFY 10a	1.94E-04
Protein TIFY 10b	1.99E-03
Aerolysin-like protein	4.29E-02
BAG family molecular chaperone regulator 6	3.02E-03
Early light-induced protein, chloroplastic	2.38E-03
LEA_2 domain-containing protein	4.41E-03

Chapter 4 :

Rhizome transcriptional responses to desiccation

Chapter 4 :

Rhizome transcriptional responses to desiccation

6. Introduction

Rhizomes are specialized underground stems with meristematic tissue within nodes that occur along the organ. Meristematic regions can give rise to new shoots (fronds) and roots. The primary function of the rhizome is the storage of carbohydrates (Holmlund *et al.*, 2020). However, an unpublished proteomic investigation of the rhizomes of *A. caffrorum*, provided evidence in support of a possible regulatory role of the rhizomes in frond phenotype (Shoko, 2007). Electrolyte leakage studies of DT (rhizomes collected in summer and producing tolerant fronds) and DS (rhizomes collected in winter and producing sensitive fronds) rhizomes illustrated that both phenotypes of the organ maintain membrane integrity through dehydration, and desiccation up to 15% RWC (Shoko, 2007). These studies were supplemented with transmission electron microscopy (TEM) investigations which further demonstrated a lack of membrane damage in the cortical cells of rhizomes. In the hydrated state, cells of both phenotypes did not show any evidence of stress. In desiccated cells, not only were cell walls and membranes preserved, but plasmodesmata contact sites were more prominent, suggesting an increase in cell-to-cell communication. Decline in starch and lipid bodies occurred during drying in both phenotypes, the breakdown products of which have been proposed to be used in subcellular protection and ultimate vitrification of the cytoplasm. A point of difference is that large numbers of polyphenol bodies were observed in DT rhizomes compared to DS rhizomes, suggesting a greater need for antioxidant protection in summer. Proteomic studies showed a clear upregulation of heat shock proteins and other antioxidant enzymes in response to desiccation. In addition, cross-tolerances against pathogen stress and heat stress were observed in desiccated winter and summer rhizomes respectively. A relatively low number of proteins were identified during proteomic study, which is likely indicative of the limited fern rhizome molecular resources that are available for annotation (Shoko, 2007). To supplement the proteomic work, metabolomic studies were conducted in 2021 (Wittenberg, 2021; Honours dissertation). The same samples collected in the current study were used for metabolomic work as illustrated in Figure 1. The metabolomic profile indicated that sucrose is the main sugar employed for cellular glass formation.

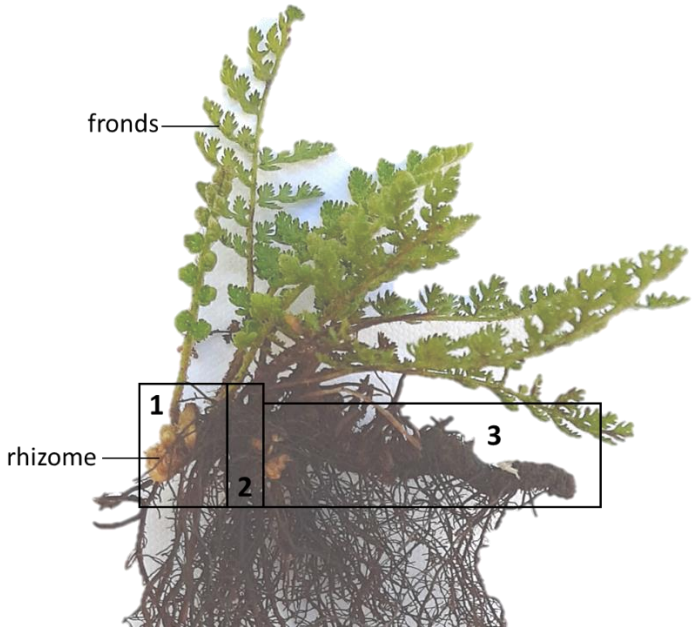

Figure 1: Portioning of the rhizome for transcriptomics, relative water content determination (RWC) and metabolomics. The meristem was used for transcriptomic analysis (section 1), the area directly adjacent to this was used for RWC (section 2). The remainder of the rhizome (section 3) was used for metabolomic work.

7. Aims and objectives

Previous proteomic and metabolomic studies demonstrated the responses of the rhizome to desiccation at the protein and metabolite level. Proteomic data further positioned the rhizome as the major regulator of desiccation in the whole organism. As such this study aims to:

1. Characterise the response to desiccation at the level of transcription
2. Explore transcriptional regulatory elements in the rhizome that may pertain to frond phenotype.

8. Methods

Rhizome material was prepared as described in Chapter 2 and data analysis was conducted using the same pipeline as detailed in the Chapter 3. Rhizome biological replicates were grouped as described in Table 1 below:

Table 1: Grouping of biological data for differential expression analyses

Dataset	Conditions
Rhizomes	RDT_FT (desiccation tolerant, full turgor) RDT_30 (desiccation tolerant, 30% RWC) RDS_FT (desiccation sensitive, full turgor) RDS_30 (desiccation sensitive, 30% RWC)

The following pairwise comparisons were conducted:

1. 30% RWC DT rhizomes against full turgor DT rhizomes (summer comparison)
2. 30% RWC DS rhizomes against full turgor DS rhizomes (winter comparison)
3. 30% RWC DS rhizomes against 30% RWC DT rhizomes (desiccated comparison)
4. Full turgor DS rhizomes against full turgor DT rhizomes (hydrated comparison)

9. Results and preliminary discussion

9.1.Quality control and outlier detection

To assess intragroup variability the biological replicates were compared to one another. The raw counts belonging to rhizome samples were subjected to a CPM transformation and then were log2 transformed. The correlation between biological replicates was measured using Pearson's coefficient. The Pearson correlation coefficient (r) was >= 0.80 for each intragroup comparison. This indicated that each biological grouping had expression profiles that were similar enough for the samples to be considered consistent between replicates. PCA was conducted to further explore inter- and intragroup variability using the vst-transformed dataset. Approximately 56% of the variance in the data was accounted for by the first two principal components. PC1, which accounted for 44% of the variance appeared to delineate rhizomes based on RWC (Figure 2). PC2, which accounted for an additional 12% of the variance seemed to separate the tissues based on phenotype (DS versus DT rhizomes). No clear outliers were observed in the PCA plot. Taken together, the PCA and Pearson correlation analyses suggested that there was high correlation between biological replicates and that each biological grouping was distinct.

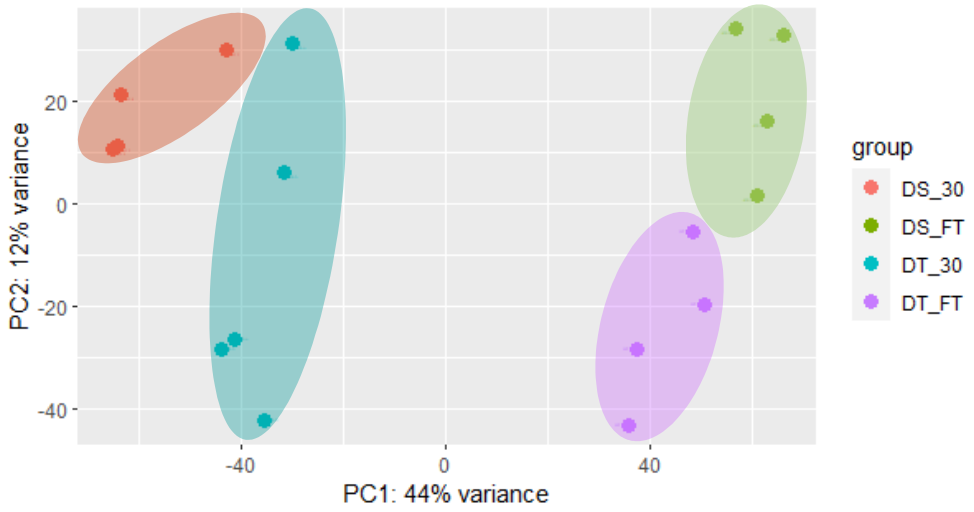

Figure 2: PCA plot of rhizome tissues representing the 500 most variable genes. PCA analysis indicated that each biological grouping of rhizomes was distinct from one another and that the replicates within each grouping were well-correlated. Approximately 44% of the variance in the data (PC1) was accredited to RWC differences and a further 12% (PC2) was a result of phenotypic differences. No clear outliers were detected by PCA analysis. DS = desiccation sensitive; DT = desiccation tolerant; 30 = 30% RWC and FT = full turgor (or 100% RWC).

9.2. Up- and down-regulated processes in desiccated rhizomes

Of the 51 459 genes in the transcriptome of *A. caffrorum*, 5 107 genes had a count of < 1 across all rhizome samples and were therefore filtered out. The remaining 46 352 (or 90.1%) genes were subjected to differential expression analysis via pairwise comparisons detailed in the methods section. A total of 1929 and 2102 genes were up- and down-regulated in the summer comparison, while 3111 and 3112 were up- and downregulated in the winter comparison respectively (Figure 3). The hydrated comparison identified 426 up- and 209 down-regulated genes. The dehydrated comparison yielded 235 upregulated and 125 downregulated genes (Figure 3). The much lower number of differentially expressed genes (DEGs) observed when comparing samples within the same phenotype supports the notion that the rhizomes are constitutively tolerant (Shoko, 2007).

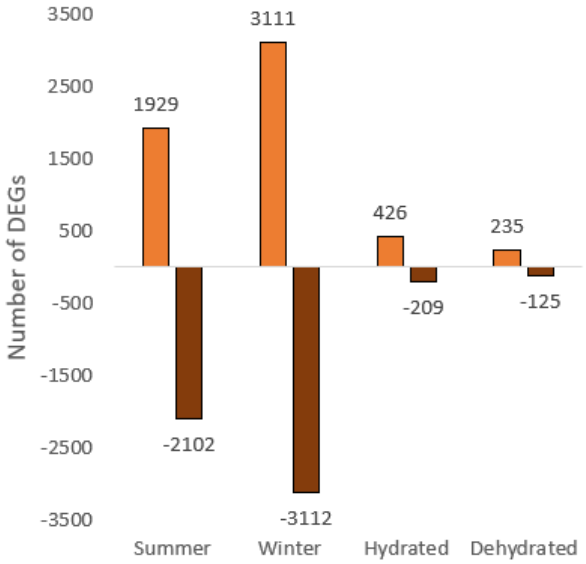

Figure 3: Number of differentially expressed genes (DEGs) observed in rhizomes via pairwise comparison.

The DEGs that were up- and down-regulated in each comparison were subjected to GO enrichment analysis. The profiles of enriched processes in the summer and winter comparisons were very similar, further supporting the hypothesis that the rhizome is likely tolerant in both seasons (Figure 4). Figure 4 demonstrates that both rhizomes achieve desiccation tolerance via similar mechanisms but also have responses that are unique to each season. Broadly, desiccated summer (DT) and winter (DS) rhizomes launch an antioxidant response, increase efforts to maintain cellular ion homeostasis, modulate lipid metabolism and show a prominent response

governed by the phytohormone abscisic acid (ABA) (Figure 4). Both desiccated rhizomes appear to offset the high energy requirements of the desiccation response by sacrificing normal growth and development, modifications of structural elements such as lateral roots and cell walls, and downregulating transcription (Figure 4). The number of enriched transcriptional responses in summer were noticeably lower than in winter. Presumably the added stressors of high light, temperature and limited carbohydrate stores reduced the ability of rhizomes to respond to desiccation only. It is difficult to delineate whether minor differences in the responses of each rhizome were the result of a heat- or light-specific response. It is important to remember that plants were collected directly from the field where they would have fully experienced summer and winter conditions but were acclimated to the same artificial conditions to minimize variability in the study. This was because it was not the direct aim of this study to explore season-specific responses, but rather to first characterise the desiccation response in rhizomes. While this may certainly have affected the responses, the main purpose of this study is to determine the overall mechanisms of desiccation tolerance in rhizomes. Therefore, commentary on the season-specific effects will mostly be limited.

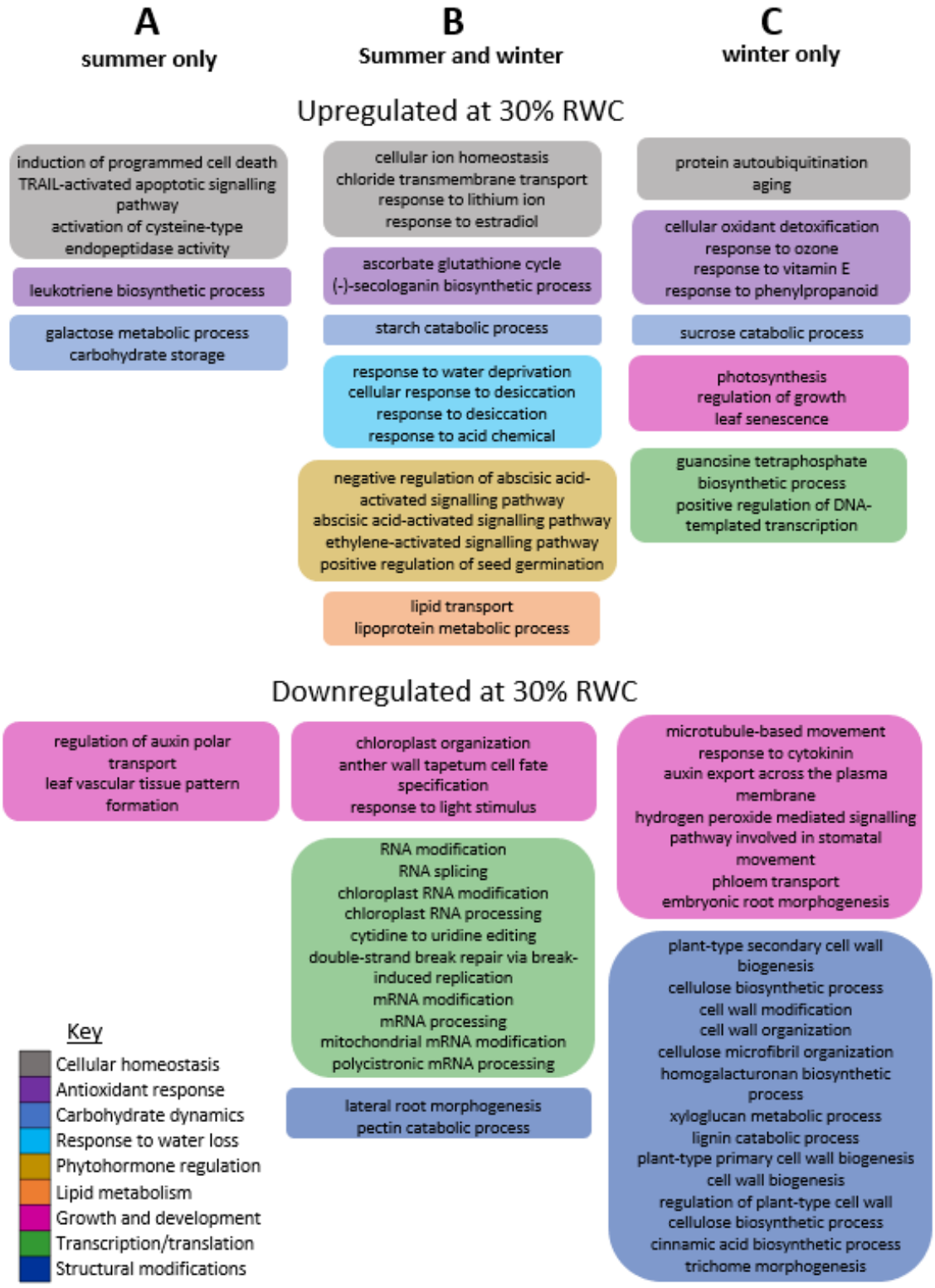

Figure 4: Processes that are up- and down-regulated in desiccated rhizomes (30% RWC) relative to fully hydrated (100% RWC) rhizomes. Summer rhizomes represent the desiccation tolerant (DT) phenotype. Winter rhizomes represent the desiccation sensitive (DS) phenotype.

9.2.1. Vitrification is promoted and photosynthetic centres protected

The rhizomes of *A. caffrorum* at 30% RWC showed a marked increase in transcripts encoding desiccation-induced proteins in other DT systems (Figure 5) which have been implicated in protection of cellular components (together with sugars) in the glassy state. The most widely studied are the late embryogenesis (LEA) proteins and heat shock proteins (HSPs) (Oliver *et al.*, 2020). LEAs form a hydration shell around import enzymes, lipid bilayers and DNA molecules as to prevent their denaturation during dehydration. LEAs, in the presence of water, have disordered confirmations which become ordered during water loss (Tolleter *et al.*, 2007; Tolleter, Hincha and Macherel, 2010; Gechev *et al.*, 2012; Banerjee and Roychoudhury, 2016). LEAs also interact with sugars in the cytoplasmic glass formation process (Shimizu *et al.*, 2010). Not all LEA proteins have a universal function, and as such the specific LEA genes expressed during desiccation are of interest *LEA-31*, *LEA-47* and *LEA 47-LIKE*, *LEA D-34-LIKE* and *LEA-14A* were significantly upregulated in DT rhizomes compared to DS rhizomes (LFC $\geq$ 3.90 for each gene) (Figure 5). *LEA-14* genes have shown an increase in transcript number in response to drying in the resurrection angiosperms *Craterostigma plantagineum* (*C. plantagineum*) and *Sporobolus stapfianus* (*S. stapfianus*) (Costa *et al.*, 2016). *LEA 47-LIKE* and *LEA D-34-LIKE* have only been observed in the pollen of cotton (*Gossypium hirsutum*) subjected to heat stress. *LEA 47-LIKE* was upregulated, while *LEA D-34-LIKE* was downregulated in response to heat. *LEA D-34-LIKE* (Masoomi-Aladizgeh *et al.*, 2021). Literature of the potential roles of *LEA-31* and *LEA-47* proteins was not available. Uncharacterised LEAs presented here may be unique to ferns, or rhizomes. HSPs are also molecular chaperones that play a fundamental role in the response of plants to abiotic stress. They play roles in protein stabilization, refolding and the prevention of protein aggregation and unwanted enzymatic reactions during water loss (Lee, Pokala and Vierling, 1995; Richter, Haslbeck and Buchner, 2010; Gechev *et al.*, 2013). *HSP70* was preferentially upregulated in the desiccated summer and winter rhizomes of *A. caffrorum* (LFC $\geq$ 2.52) (Figure 5). *HSP70* has been shown to confer desiccation tolerance to a range of plant species, but the exact role of *HSP70* in the desiccation response has not yet been elucidated (Alvim *et al.*, 2001; Su and Li, 2008; Guo *et al.*, 2016; Yu *et al.*, 2021).

Another category of proteins highly induced during desiccation were early light inducible proteins (ELIPs). Their accumulation has been observed in the leaves of several angiosperm resurrection plant species (Zeng, Chen and Wood, 2002; Gechev *et al.*, 2013; Yobi *et al.*, 2017; Van Buren *et al.*, 2019) where they are thought to bind chlorophyll and anthocyanins, thereby

protecting the photosystem in homoiochlorophyllous DT plants (Adamska *et al.*, 1999; Alamillo and Bartels, 2001). Despite not being directly related to rhizome physiology, ELIPs were the class of protectants that had the greatest fold expression changes in desiccated rhizomes (Figure S1 and S2) and were significantly upregulated in desiccated rhizomes ($14.01 \leq LFC \geq 2.62$) (Figure 5). As one may not expect active chloroplasts in the rhizome, it is possible that these massively accumulated transcripts are not translated in the rhizome. Although photosynthesis does appear to be upregulated in the desiccated winter rhizome (Figure 4C), and this will be explored later, ELIPs are upregulated in both summer and winter rhizomes. It is therefore more likely that these transcripts may be transported to fronds where they are converted into functional proteins.

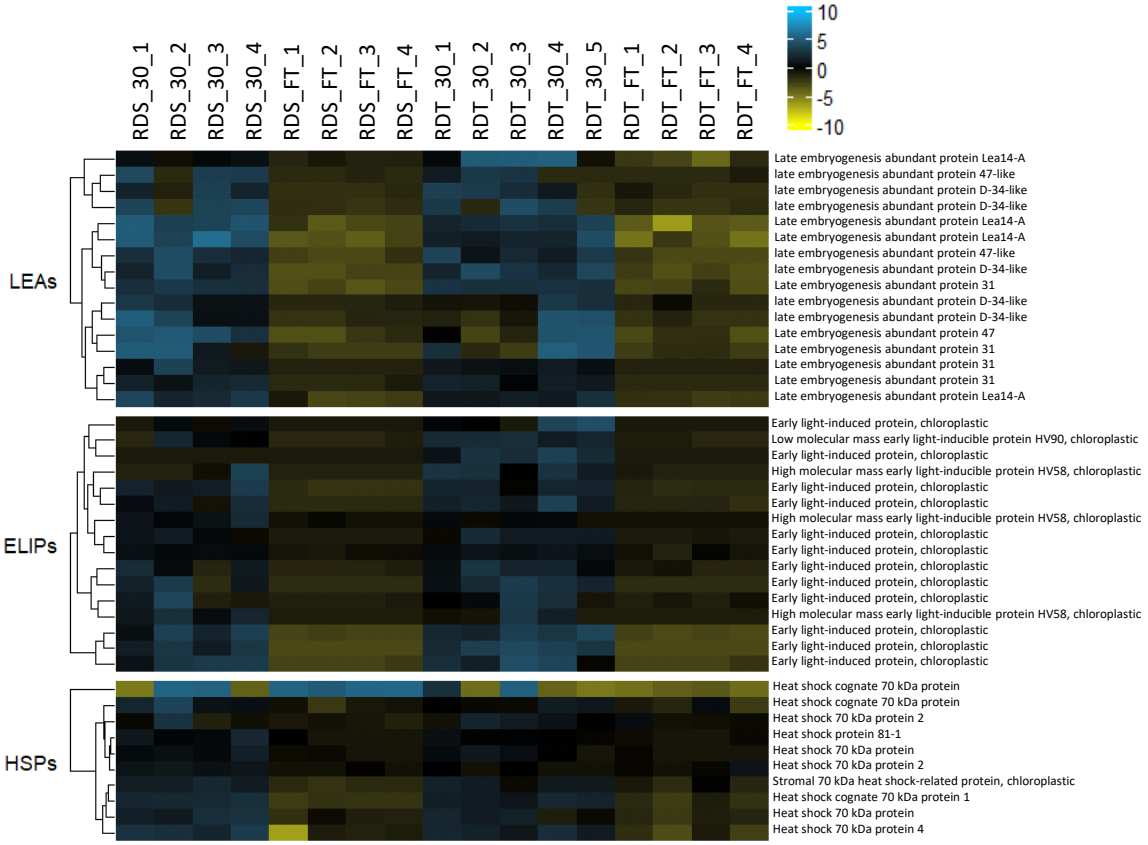

Figure 5: Upregulated LEAs, ELIPs and HSPs in the desiccated rhizomes (30% RWC) compared the fully hydrated (FT) rhizomes in summer and winter. RDS = rhizome desiccation sensitive (winter), RDT = rhizome desiccation tolerant (summer).

The rapid accumulation of non-reducing sugars in the process of vitrification has been well documented. Sucrose is ubiquitously increased in resurrection plants upon drying, and raffinose oligosaccharides (RFOs), stachyose and sugar alcohols are increased in a species-dependent manner (Zhang, Song and Bartels, 2016). Metabolomic analyses of the rhizome samples as employed for transcriptomics (Figure 1) demonstrated an increase in sucrose in desiccated rhizomes (Wittenberg, 2021; Honours dissertation). Although RFOs were not detected by metabolomic investigation, the genes required for the conversion of sucrose to RFOs, *GALACTINOL SYNTHASE* (LFC = 7.12 in summer rhizomes, and LFC = 10.69 in winter rhizomes) and *PROBABLE GALACTINOL--SUCROSE GALACTOSYLTRANSFERASE* (LFC = 4.83 and 11.01 in summer and winter respectively), were upregulated in desiccated rhizomes (Table S4). The co-product of raffinose, myo-inositol, was significantly accumulated in rhizomes during drying (Sengupta et al., 2015; Wittenberg, 2021; Honours dissertation). The apparent discrepancy in RFO presence could be due to fluxing through the system, not necessarily captured by the sampling methods employed. As described in Figure 1, rhizome meristems were used for transcriptomic work, while the region vertically adjacent to the meristem was used for metabolomic measurements. It is possible that RFO accumulation is prioritized in the meristems to improve the survivability of the whole organism during desiccation.

The enzymes typically required for sucrose production were not observed in the upregulated DEGs, suggesting alternate means of synthesis of sucrose or possible transport of sucrose from fronds (source) to rhizomes (sink) during drying (Tauzin and Giardina, 2014). Supporting this, the sucrose catabolizing enzyme *SUCROSE SYNTHASE 3* (LFC = 6.00 and 2.32 in winter and summer respectively) and *SUGAR TRANSPORT PROTEIN 7* (LFC = 10.50 and 6.10 in summer and winter respectively) were upregulated in desiccated rhizomes. Increased SUS3 transcripts have been reported during the late stages of desiccation of the angiosperm resurrection plants *Xerophyta schlechteri* (Radermacher, 2021) and *Eragrostis nindensis* (Madden, 2020), when there is notable accumulation of sucrose. This could suggest selection of the reverse reaction that SUS3 catalyses. Alternatively, the presence of enzymes that degrade sucrose, where there are confirmed high levels of sucrose, could suggest a role in the recovery mechanism. Upon rewetting, stored *SUCROSE SYNTHASE* and *SUGAR TRANSPORT PROTEIN 7* could facilitate the rapid degradation of sucrose to simpler sugars and their mobilization to regions that require additional energetic input during recovery (Tauzin and Giardina, 2014; Stein and Granot, 2019).

Protective sugars are also believed to play a role in membrane protection. They maintain membranes by replacing the water lost during water stress and act as osmotic spacers which prevent the fusion of unrelated membranous structures (Shimizu *et al.*, 2010; Oliver *et al.*, 2020). Electron microscopic images demonstrated increased vacuolation in rhizome cortical cells during desiccation (Shoko, 2007). While there were no prominent transcript expression patterns related to vacuolation in the 'lipid metabolism' enrichment set (Figure 4), there was evidence for protection of membrane contact sites. The proteins related to membrane contact site protection *NUCLEUS-VACUOLE JUNCTION PROTEIN 2* (LFC = 1.91 and 2.00 in winter and summer respectively), *SYNAPTOTGAMIN-2* (LFC = 2.71 and 2.23 in winter and summer respectively) and *SYNAPTOGAMIN-5* (LFC = 5.75 and 4.51 in winter and summer respectively) were significantly upregulated in desiccated rhizomes (Jeong *et al.*, 2017; Ruiz-Lopez *et al.*, 2021).

Lastly, the spores of the ferns *Osmunda japonica* (*O. japonica*) and *Adiantum capillus-veneris* (*A. capillus-veneris*) accumulate a fern-specific protein called *22 KDA PROTEIN OF CHLOROPLASTS* in thylakoid membranes prior to germination. The role of *22 KDA PROTEIN OF CHLOROPLASTS* has not been elucidated but the degradation of this protein during germination suggests a role for this protein in chloroplast maintenance in the desiccated state (Minamikawa, Koshiba and Wada, 1984; Haupt, 1990; Inoue *et al.*, 2000). A log2foldchange of 9.85 was observed for this protein in desiccated rhizomes. Taken together, rhizome cells prioritize the accumulation of protective proteins, sugars and undergo lipid alterations to maintain cellular structure in the glassy state.

9.2.2. Antioxidant response is upregulated

In addition to the upregulation of photoprotective ELIPs, other canonical responses to oxidative stress were observed in the desiccated rhizomes of *A. caffrorum*. Enhanced activity in the antioxidant enzymes glutathione reductase (GR), superoxide dismutase (SOD), catalases (CATs), peroxidases (PODs), and non-protein antioxidants such as carotenoids, ascorbate, tocopherols, flavonoids, a range of phenolic compounds and glutathione are common in response to dehydration/desiccation stress in resurrection plants (Pandey *et al.*, 2010; Morse, Rafudeen and Farrant, 2011). The ascorbate-glutathione cycle is one of the more prominent, ubiquitous antioxidant systems in resurrection plants (Du and Rennenberg, 2018; Mihailova *et al.*, 2022). An upregulation of this system was implied via the increased transcript abundance of *GLUTATHIONE S-TRANSFERASE* (GST; LFC = 11.06 and 5.29 in winter and summer

respectively) and its related enzymes (probable *GST DEHYDROASCORBATE REDUCTASE* (*DHAR1*) (LFC = 14.05 and 12.40 in winter and summer respectively) and *ABC TRANSPORT C FAMILY MEMBER 4* (*ABCC4*) (LFC = 2.70 in winter only) in desiccated rhizomes (Table S1 and S2). GST DHAR1 is directly involved in the cycle, while *ABCC4* actively removes glutathione conjugates from the cellular space (Hwang *et al.*, 2016; Krattinger *et al.*, 2019). In desiccated summer (DT) rhizomes only, the GO term 'leukotriene biosynthetic process' contained the genes for microsomal *GST 3* enzymes and in the desiccated, winter (DS) rhizomes the terms associated with the antioxidant system were all related to the increase of cytosolic and peroxisomal catalases (Figure 4). There was evidence of additional photoprotective proteins such the chloroplastic *ACTIVITY OF BC1 COMPLEX KINASE 1* (LFC = 7.51 and 6.22 in summer and winter respectively) which maintain the photosystem via its association with plastoglobules. Plastoglobules are lipoproteins that reside in chloroplasts and are upregulated during oxidative stress (Martinis *et al.*, 2014; Charuvi *et al.*, 2019; Espinoza-Corral, Schwenkert and Lundquist, 2021). Enzymes related to 'secologanin biosynthesis' were observed in the set of genes related to 'response to water deprivation' (Figure 4) and were therefore likely induced by water stress. An accumulation of the substrate required to synthesize secologanin – loganin – was also observed in desiccated rhizomes (Wittenberg, 2021; Honours dissertation). Secologanin is an iridoid that is required in the production of monoterpene indole alkaloids (MIAs), isoprenoids in particular. These are are known to act as thylakoid membrane strengtheners with antioxidant potential (Tattini et al., 2014). A potential role for MIAs as an antioxidant and their induction via secondary messengers such as Ca^{2+} and ROS has also been determined (Asada *et al.*, 2013; Matsuura, Rau and Fett-Neto, 2014; Courdavault and Besseau, 2022).

9.2.3. Growth and protein translation are downregulated in rhizomes at 30% RWC

Desiccated rhizomes displayed prominent patterns of downregulation in nuclear, mitochondrial and chloroplastic transcription (Figure 4). The gene sets related to these processes were enriched with pentatricopeptides (PPRs). PPR proteins constitute one of the largest protein families in plants, so it is not surprising that so many PPRs were identified in *A. caffrorum.* PPRs have been implicated in the post-transcriptional regulation of chloroplast and/or mitochondrial genes. This includes stabilization of mRNA transcripts and initiation of translation (Wang *et al.*, 2021). PPRs are massively reduced in both summer and winter desiccated rhizomes (Figure 6). A decrease in the PPR population suggests that post-transcriptional modification and translation have been downregulated in the rhizomes by 30% RWC. This could mean that transcripts were mass accumulated but will only be translated into functional proteins upon rewetting of the rhizome. Interestingly, similar mass changes in PPRs were displayed in the fronds of *A. caffrorum* but with different patterns of expression (Figure 6). In DS fronds, PPRs were mostly downregulated through the course of drying and during rehydration, and only spiked at 10% RWC (Figure 6). As DS fronds are effectively dead at 10% RWC, an apparent increase in PPRs at 10% RWC could simply be due to a reduction in other transcripts that may be reabsorbed into the rhizome. Remobilization of molecules into the rhizome at 10% RWC suggests that although frond tissue has effectively died, the necessary channels in the stipe are intact and can still be used for transport. In DT fronds, PPRs were downregulated at full turgor, but have increased in number at 55%, 30% and 10% RWC. Elevated PPR content at these water contents implies considerable regulation of transcription through the course of drying in DT fronds which is totally absent from DS fronds (Figure 6). A relative decrease once upon rewetting suggests that less regulation occurred once water become available, but after 24-hours had not yet returned to full turgor levels. Presumably the regulation of transcription occurring at this stage is more geared towards recovery processes (Figure 6). Altogether, the data suggests that rhizomes are, for the most part, transcriptionally inactive by 30% RWC, presumably because they are in a state of quiescence. Since transcription, and therefore translation, have stopped at this stage, it is not surprising that the rhizomes also show evidence that normal growth and development, and structural modifications are downregulated (Figure 4). If the rhizome has reached a state of quiescence by 30% RWC, regulatory elements that are present in this state might either 1) have accumulated and contributed to the entry into quiescence or 2) were accumulated and are important for resumption of metabolic activity upon rehydration. In addition, this is the state

from which new fronds will develop and may therefore contain elements related to the control of DT/DS frond formation.

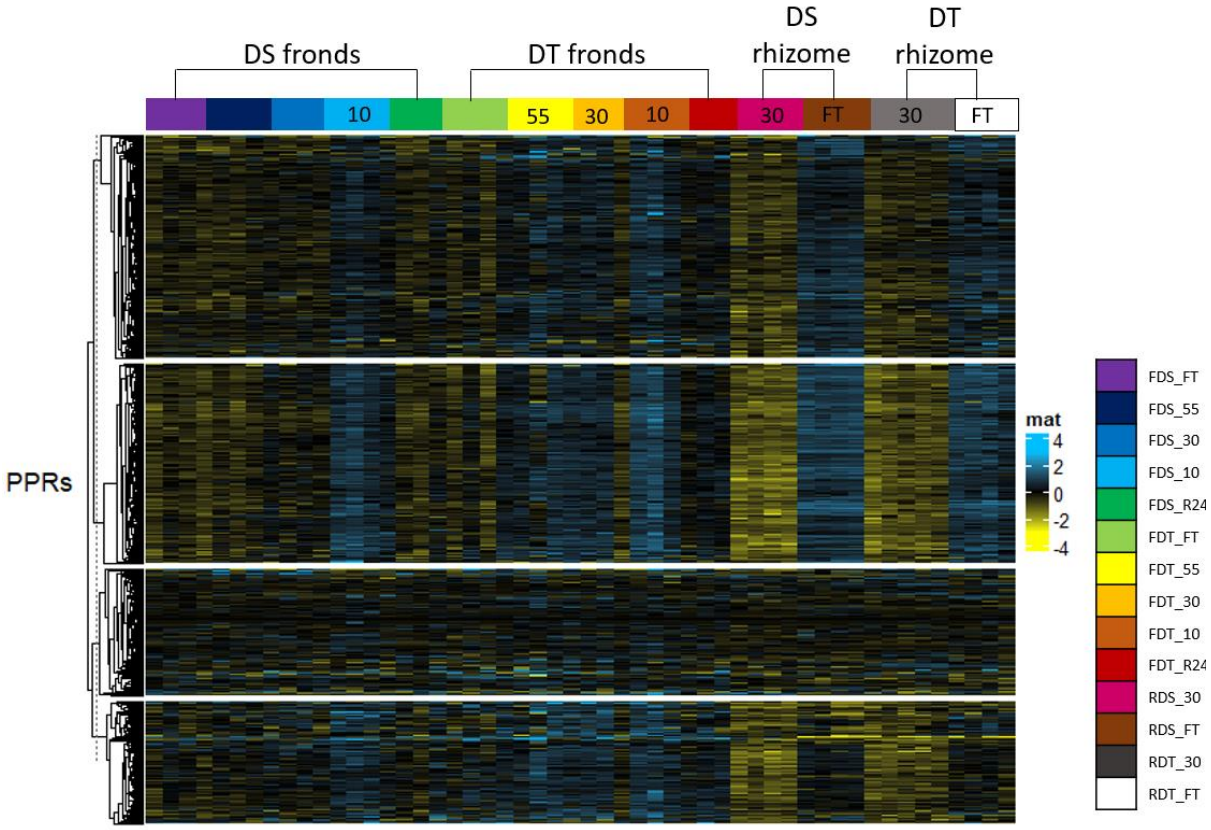

Figure 6: Mass changes in pentatricopeptides (PPRs) were highly correlated with the control of transcription and translation in fronds and rhizomes.

4.3. ABA regulation in desiccated rhizomes

Numerous studies have demonstrated that the overexpression of a single transcription factor (TF) implicated in the plant stress response can confer resistance to abiotic and/or biotic stressors (Park *et al.*, 2001; Singh, Foley and Oñate-Sánchez, 2002; Pitzschke *et al.*, 2009; Polizel *et al.*, 2011; Li *et al.*, 2012; Tsugama, Liu and Takano, 2012; Hsu *et al.*, 2013; Ravikumar *et al.*, 2014). Water-stress induced responses to certain phytohormones have also shown regulatory roles in tolerant phenotypes. More specifically, salicylic acid (SA), cytokinin and auxin, gibberellins (GAs), abscisic acid (ABA), ethylene and jasmonic acid (JA) have been implicated in drought stress, as reviewed by Wahab et al., 2022. ABA is the most widely studied hormone in the desiccation response. ABA levels tend to increase in response to water deficit stress in the vegetative tissues of resurrection plants, seeds and even DS angiosperm leaves (Fernando and Schroeder, 2016; Gutierrez et al., 2007; Khandelwal et al., 2010). The GO terms 'negative regulation of abscisic acid-activated signalling pathway', 'abscisic acid-activated signalling pathway', 'ethylene-activated signalling pathway', 'positive regulation of seed germination' are all highly enriched in transcription factors (TFs) that have been implicated in several abiotic stress responses (Figure 4). It is worth noting that this gene set was subjected to pathway analysis, but the results were not fruitful. Of the total set of genes used for pathway analysis, less than 25% of the genes contained annotations in the relevant databases. Enriched pathways were evaluated and compared to a manual exploration of all genes in the set. Pathway analysis results were comparatively highly reduced, and the concern was that important regulatory elements were omitted owing to their lack of annotation. Therefore, a manual exploration of the gene set was used for further exploration.

The identified TFs and their roles have been summarized in Table S3. Genome-wide expression studies of the vegetative tissues of several desiccation tolerant species have highlighted an increase in transcript abundance related to seed-specific genes and proteins (Hilhorst and Farrant, 2018), so it is not unusual to have gene sets enriched in apparent seed-specific genes. Several key components of the ABA-signalling pathway were upregulated in desiccated rhizomes. *PYRABACTIN RESISTANCE (PYR)/PYRABACTIN RESISTANCE-LIKE (PYL)/REGULATORY COMPONENT OF ABA RECEPTORS (RCAR), PROTEIN PHOSPHATASE 2C (PP2C) and SUCROSE NON-FERMENTING 1-RELATED PROTEIN KINASE 2* (SnRK2) are three major components of ABA signalling. PP2Cs act as negative regulators and SnRK2s act as positive regulators of PYR/PYL/RCAR, respectively (Fidler *et al.*, 2022) (Figure 6A). ABA-activated SnRK2s induce the expression of several TFs, ion

channels, and other metabolic components. The ABA signalling pathway and the associated TFs are part of a highly complex, interconnected system which has not yet been fully elucidated. The complex nature of the phytohormone regulation presents a current and monumental challenge in the development of climate smart crops. For this reason, a basic model that reflects our current understanding of the system, and individual focus on the possible key players of ABA signalling in the rhizome will be detailed here (Figure 7B). All elements and their functions are detailed in Table S3.

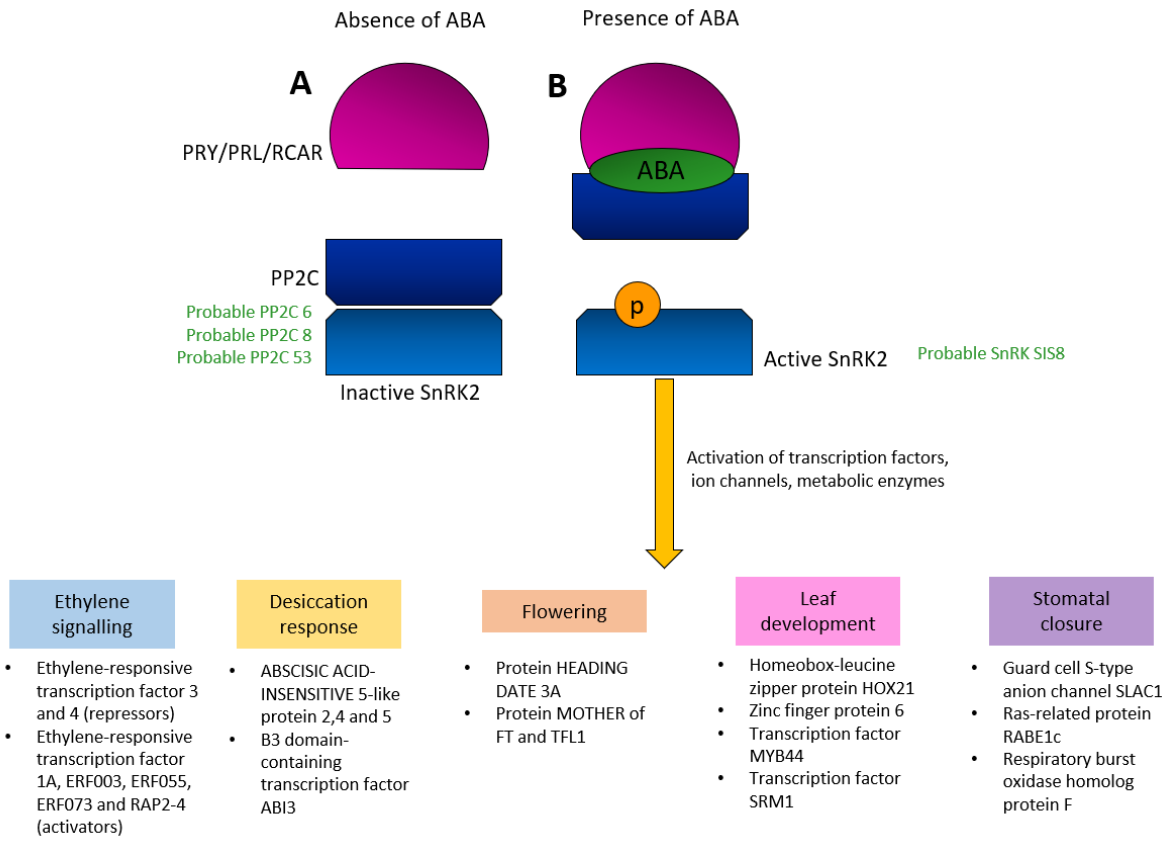

Figure 7: Observed components of the ABA signalling pathway in the desiccated rhizomes of *A. caffrorum*. A: in the absence of ABA, PP2C genes prevent activation of ABA signalling by SnRK2s. The PP2C genes observed in desiccated rhizomes are presented in green. B: in the presence of ABA, repression of SnRK2s is removed and active SnRK2s activate several downstream elements. The observed SnRK2s in desiccated rhizomes are presented in green. The observed upregualted downstream elements of ABA signalling observed in desiccated rhizomes were all related ethylene signalling, the desiccation response, flower timing, leaf development or stomatal control.

Activators and repressors

Genes that resulted in the enrichment of the 'negative regulation of ABA signalling pathway' (Figure 4) were mainly PP2Cs, which are known negative regulators of ABA signalling. Probable *PROTEIN PHOSPHATASE* 2C 6, 53 and 8 were observed in the desiccated rhizomes with LFC values of 5.37, 3.01 and 1.81 respectively (Figure 7A). Probable serine/threonine-protein kinase SIS8 was the most highly expressed gene in winter desiccated rhizomes (LFC = 10.65) but was also significantly upregulated in summer desiccated rhizomes (LFC = 1.50). SIS8 has been shown to mediate sugar responses during early seedling development in *A. thaliana* and may play a role in salt tolerance (Gao and Xiang, 2008; Chanwala *et al.*, 2020). In the presence of ABA, SIS8 could activate the necessary pathways for the accumulation of sugars required for rhizome vitrification (i.e., raffinose-producing genes, since sucrose is more likely transported from the fronds).

Abscisic acid insensitive-proteins and ethylene signalling

Downstream elements of an active ABA-signalling pathway included genes related to ethylene signalling and genes that have been highly implicated in the desiccation response (Figure 6B). *ABSCISIC ACID-INSENSITIVE 5-LIKE PROTEIN 5* (*ABI5-LIKE PROTEIN 5*) was the most highly upregulated regulatory element in desiccated winter rhizomes, with an LFC = 12.62. *B3 DOMAIN-CONTAINING TRANSCRIPTION FACTOR ABI3* was also present (LFC = 5.99). Both components were also upregulated in summer desiccated rhizomes but with lower LFC values (*ABI5-LIKE PROTEIN 5* LFC = 4.74 and *B3 DOMAIN-CONTAINING PROTEIN* VP1 LFC = 3.32). ABI5 is believed to function in the core of ABA signalling. ABI-5 has been shown to participate in the regulation of processes such as inhibition of germination, abiotic stress adaptions, chlorophyll catabolism, inhibition of photosynthesis and leaf senescence (as reviewed by Skubacz et al., 2016). More recently, *ABI5-LIKE PROTEIN 5* in maize has been shown to regulate ABA biosynthesis and the drought response, and is active during rewatering (Cao *et al.*, 2021). As previously discussed, ABA signalling pathways are highly complex and not fully understood just yet, so the exact role that the ABI5-like protein plays in rhizomes would need to be validated by other methods. ABI3, ABI4 and certain MYB and WRKY transcription factors are believed to modulate the expression of ABI5 (Skubacz, Daszkowska-Golec and Szarejko, 2016). Beyond its role as regulatory element of ABI5, ABI3 has also been implicated in seed development and maturation, and lateral root formation (Kumar *et al.*, 2019). Kumar et al., (2019) also detail the crosstalk between ABA signalling and other signalling

pathways. Enriched in this dataset, were several ethylene-responsive transcription factors (ERFs) suggesting possible crosstalk between ABA and ethylene in the desiccated rhizomes of *A. caffrorum*. Broadly speaking, ERFs mediate plant response to drought, fruit ripening and induction of secondary metabolite production. Figure 7B provides a list of the ERFs that were significantly upregulated in desiccated rhizomes and whether they were identified as an activator or repressor of the ABA signalling pathway (Fujimoto *et al.*, 2000; Müller and Munné-Bosch, 2015; Tao *et al.*, 2018; Srivastava and Kumar, 2019). The roles of many of these ERFs is yet to be determined but the protein ENHANCED DISEASE RESISTANCE 2 (LFC = 8.53 in winter rhizomes and LFC = 4.6 in summer rhizomes) has been shown to prevent ethylene-induced senescence in *A. thaliana* plants experiencing pathogen stress (Vorwerke *et al.*, 2007). Protein SABRE was upregulated in summer, desiccated rhizomes only (LFC = 7.94), and has been described as muting ethylene-induced radial expansion in plant cells (Aeschbacher et al., 1995). Together, these results imply a blocking of the usual functions of ethylene during desiccation by the ABA signalling pathway. This phenomenon has been observed for the phytohormones SA and JA as well (Kumar et al., 2019).

Regulators related to leaf phenotype

The main hypothesis that drove this study was that the rhizome is regulating the phenotype of the fronds that are produced. Within the ABA signalling network (Figure 4) there are several regulatory elements that seemingly support this hypothesis. Firstly, several TFs have been implicated in the process of leaf development (Figure 7B). Homeobox-leucine zipper protein HOX21 (LFC = 2.12 and LFC = 2.23 in summer and winter rhizomes respectively) is a microRNA activator that is involved in the metabolism and nutrition of rhizome shoots in bamboo. It has also been shown to play a role in leaf development (Jin *et al.*, 2016; Chen *et al.*, 2020). Transcription factor SRM1 (LFC = 1.77) has also been implicated in the development of leaves in tomato plants. Loss-of-function studies of SRM1 demonstrated that absence of this gene resulted in abnormal leaf morphology (Tang *et al.*, 2022), suggesting that SRM1 may not necessarily play a role in determining the phenotype of the fronds that are produced from the rhizome. Similarly, *MYB44* (LFC = 3.61 and 2.51 in winter and summer respectively) may play a role in leaf senescence in response to wounding (Kobayashi *et al.*, 2004), but does not imply a role in determining the phenotype of emerging fronds. Finally, zinc finger protein 6 (*ZFP6*) is believed to be a possible TF that initiates the production of trichomes in response to GA and cytokinin signalling (Zhou *et al.*, 2013). Several *ZFP6* genes were observed in the dataset, and

all had much higher LFC values than the other proteins involved in leaf development. *ZFP6* had an LFC of 6.08 in summer desiccated rhizomes and an LFC of 6.50 in winter desiccated rhizomes. The upregulation of this gene in both seasons implies that it is required in both season for trichome production. Zhou et al., (2013) have detailed a trichome initiation cascade where *ZFP6*, after activation via GA and/or cytokinin, is the first element in the signalling cascade. The downstream element GLABRA2 expression modulator was also highly upregulated in winter, desiccated rhizomes (LFC = 8.68). Together these findings suggest that the rhizome may be preparing transcripts that play a fundamental role in trichome production, and as trichomes are one of the key features in DT fronds, it might have some role in frond phenotype regulation.

The second set of regulatory elements was related to the modulation of stomata (Figure 7B). *GUARD CELL S-TYPE ANION CHANNEL* SLAC1 (SLAC1) was upregulated in summer (LFC = 4.46) and winter (LFC = 4.95) desiccated rhizomes. SLAC1 is preferentially expressed in guard cells and is a stress-induced anion efflux transporter that mediates stomatal closure (Li *et al.*, 2022). *RAS-RELATED PROTEIN RABE1C* was the most highly expressed gene in the summer and winter desiccated rhizomes (LFC = 10.05 and LFC = 10.66 respectively) and was also associated with stomatal control during the drought response (Chen *et al.*, 2021). Finally, respiratory burst oxidase homolog protein F (RBOH) is implicated in the closing of stomata in response to ABA and/or ethylene signalling. It is an NADPH oxidase that generates reactive oxygen species (ROS), suggesting a possible, additional role in ROS signalling (Desikan *et al.*, 2006). RBOH was upregulated on summer rhizomes only with an LFC of 2.22.

Control of flowering

The upregulation of two elements suggested a role for the rhizome in controlling flowering time. *HEADING DATE 3A* (*H3DA*) was upregulated in the desiccated rhizomes of both seasons (LFC = 6.47 and LFC = 8.76 in summer and winter respectively). *H3DA* acts as a flower-promoting signal that is transported from the leaf to the shoot apical meristem (SAM) in rice. In the SAM it is responsible for inducing flowering under short day conditions (Komiya, Yokoi and Shimamoto, 2009). *MOTHER of FT* and *TFL1* (LFC = 3.94 in summer only) has also been shown to induce early reproduction in the fern *Adiantum capillus-veneris* (*A. capillus-veneris*) when overexpressed (Hou and Yang, 2016; Yu *et al.*, 2019). The presence of these transcripts in desiccated rhizomes also suggests that the rhizome 1) controls reproductive behaviour in fronds and 2) may be storing such transcripts in response to a desiccation event to exert a finer

control of reproduction depending on the environmental conditions upon rewatering. In the field one might expect summer plants to experience desiccation, and hence accumulation of these transcripts are in preparation for the short-day light conditions of winter, which is when reproductive fronds are observed to appear.

4.4. Unique season-specific responses

Rhizomes promote the senescence of DS fronds

Winter, desiccated rhizomes exclusively upregulated processes related to regulation of growth, leaf senescence and photosynthesis (Figure 4). *NAC DOMAIN-CONTAINING PROTEIN 53* is a promoter of genes which encode for ROS biosynthetic enzymes. ROS accumulation triggers leaf senescence via programmed cell death (PCD) (Lee *et al.*, 2012). It was upregulated in winter, desiccated rhizomes with an LFC = 7.59. Overexpression of *PLASTOGLOBULE-LOCALIZED METALLOPEPTIDASE 48* (*PGM48*) in *A. thaliana* accelerated leaf senescence (Bhuiyan *et al.*, 2016). PGM48 was also upregulated in winter, desiccated rhizomes (LFC = 1.65). The presence of leaf senescence elements in the rhizomes suggests that DS fronds are programmed for death by the rhizome. Presumably this process is underway once the fronds are at least 55% RWC, as this is where gene expression profiles for DS fronds begin to decline (Chapter 3, figure 5).

Winter rhizomes increase transcription related to photosynthesis

The enrichment of photosynthesis in desiccated winter rhizomes (Figure 4) was a result of several photosynthetic genes. Upregulation of *PHOTOSYSTEM I P700 CHLOROPHYLL A APOPROTEIN A2* and *PHOTOSYSTEM II 10 KDA POLYPEPTIDE* (LFC = 2.92 and 11.86 respectively) suggests that both photosystems are actively transcribed in desiccated, winter rhizomes. Evidence for photosynthesis was not present in summer, desiccated rhizomes. Studies have demonstrated that when roots are detached from photosynthetic shoots, they can develop chloroplasts and several genes associated with photosynthesis become upregulated (Kobayashi *et al.*, 2017). Rhizome meristems sampled for this study (Figure 1) did show evidence of greening, but chloroplasts were not observed in the cortical cells of rhizomes as evaluated by transmission electron microscopy (TEM) studies (Shoko, 2007). More likely then, these transcripts were accumulated and stored for the reinstatement of photosynthesis in emerging fronds. In summer, DT fronds survive the desiccation event and upon rewetting, uncurl. However, DS fronds do not survive a desiccation event and therefore upon rehydration,

the plant would need to produce new fronds. This further extends the role of the rhizome in frond dynamics. However, cortical cells evaluated by TEM may have been sampled further along the rhizome i.e., away from the meristem, which may misleadingly indicate a lack of chloroplasts. Additional TEM studies of the rhizome tissue in the meristem specifically would shed light on which hypothesis is more likely.

Carbon metabolism dynamics across seasons

Sucrose is produced during photosynthesis and transported to sink tissues such as rhizomes via the phloem (Tauzin and Giardina, 2014). Mass accumulation of sucrose in the desiccated rhizomes in both seasons suggests that both rhizomes enter a glassy state at 30% RWC (Wittenberg, 2021; Honours dissertation). The absence of enzymes related to sucrose anabolism in both seasons support the fronds as the source of sucrose. The upregulation of 'starch catabolism' in the desiccated rhizomes of both seasons (Figure 4) were related mainly to the breakdown of starch to simpler sugars. Carbohydrate metabolism that was upregulated and defined as relating to the 'response to water deprivation' (Figure 4) included the upregulation of enzymes (*GALACTINOL SYNTHASE*, LFC = 10.69 and *GALACTINOL--SUCROSE GALACTOSYLTRANSFERASE 5*, LFC = 11.04) that convert sucrose to raffinose. The upregulation of galactinol synthase was also responsible for the enrichment of the processes 'galactose metabolic process' and 'carbohydrate storage' in summer (Figure 4). It was only in the profile of carbohydrate enzymes produced exclusively in winter that these enzymes were not observed. In winter, 'sucrose catabolic process' contains genes related to the catabolism of starch (amylases) and of the hydrolysis of sucrose to glucose and fructose (invertases). Overall, it seems that the conversion of sucrose to raffinose is important for vitrification and is prioritized in both seasons. In winter rhizomes, that are seemingly photosynthetically active owing to the loss of the DS fronds, sucrose catabolism is completely likely. Alternatively, during winter when DS fronds are more photosynthetically productive, they may have built up the carbon reserves in the rhizome which may also act as additional sources for more advanced carbohydrate metabolism.

10. Supplementary material

Table S1: Most up- and downregulated genes in 30% RWC DT rhizomes (Control : 100% DT rhizomes)

	gene name	log2FoldChange	pvalue
PB_9218	Early light-induced protein, chloroplastic	14.00874373	1.27462E-22
PB_25229	Probable inorganic phosphate transporter 1-4	13.48411025	4.31017E-20
PB_21611	Probable glutathione S-transferase DHAR1, cytosolic	12.39746429	1.0237E-09
PB_8313	Senescence domain-containing protein	12.113411	8.72295E-20
PB_18445	phospholipase	11.92643023	4.35799E-09
PB_19404	Phosphomethylpyrimidine synthase, chloroplastic	11.75062927	4.31656E-07
PB_17010	ACT domain-containing protein ACR3	11.74213138	3.40977E-11
PB_19516	Putative hydrolase YtaP	11.5973763	5.01871E-11
PB_32848	Predicted protein	11.36398242	2.08924E-15
PB_9235	Early light-induced protein, chloroplastic	11.14400128	2.39414E-07
PB_31898	Senescence domain-containing protein	11.10842623	0.000914997
PB_14724	Long chain acyl-CoA synthetase 8	10.97697329	9.49414E-07
PB_9216	Early light-induced protein, chloroplastic	10.77234698	1.09145E-07
PB_971	7-deoxyloganetin glucosyltransferase	10.75493276	8.96128E-07
PB_41469	Early light-induced protein, chloroplastic	10.69515443	7.14723E-22
PB_9538	Eukaryotic translation initiation factor 5A-2	10.55569355	6.02852E-05
PB_16765	Sugar transport protein 7	10.5444329	1.36226E-10
PB_15776	1-Cys peroxiredoxin PER1	10.50432032	9.8255E-07
PB_19297	Phragmoplastin DRP1E	10.47198967	2.79778E-12
PB_367	Phragmoplastin DRP1E	10.44014145	3.78419E-06
PB_40720	Peroxidase 12	-8.155986634	4.50566E-11
PB_1565	Putative elongation factor TypA-like SVR3, chloroplastic	-8.230139706	1.20961E-26
PB_19187	Pentatricopeptide repeat-containing protein At4g18520, chloroplastic	-8.234408894	1.41247E-07
PB_42882	AAI domain-containing protein	-8.30303015	1.6389E-14
PB_41488	Expansin-A10	-8.499385573	5.17754E-11
PB_42741	glycine-rich protein DOT1-like isoform X1	-8.635193952	5.75285E-13
PB_32332	Subtilisin-like protease SBT5.5	-8.645533014	7.25438E-46
PB_30042	Chitinase-like protein 1	-8.80420709	8.24265E-06
PB_9873	Exosome complex component RRP41 homolog	-8.851512379	1.35966E-05
PB_10512	Probable iron/ascorbate oxidoreductase DDB_G0283291	-9.061932164	1.43348E-07
PB_3041	Desmethyl-yatein O-methyltransferase	-9.233648369	6.91261E-06
PB_45394	Subtilisin-like protease SBT2.3	-9.266286439	1.80227E-05
PB_38005	Cationic peroxidase SPC4	-9.383528961	8.81769E-10
PB_30178	Laccase-4	-9.438165615	0.000107745
PB_42465	Probable boron transporter 2	-9.468609985	1.26355E-10
PB_41109	Probable pectinesterase 53	-9.491141835	1.28994E-09
PB_19176	Pentatricopeptide repeat-containing protein At5g16860	-9.75514891	2.90952E-06

PB_1381	Probable pectinesterase 53	-9.775818955	5.21378E-06
PB_29372	Protein JOKA2	-10.03695085	0.00083175
PB_29494	Probable inactive receptor kinase At3g08680	-10.63267272	6.25129E-10

Table S2: Most up- and downregulated genes in 30% RWC DS rhizomes (Control : 100% DS rhizomes)

	gene name	log2FoldChange	pvalue
PB_523	Transmembrane protein 205	15.20853949	9.97E-12
PB_25229	Probable inorganic phosphate transporter 1-4	14.55731556	4.57E-23
PB_17826	Peroxisomal catalase	14.54140434	4.55E-24
PB_8104	Probable glutathione S-transferase DHAR1, cytosolic	14.05287045	7.74E-20
PB_43829	Polyubiquitin 11	13.83950763	2.01E-07
PB_31077	S-(hydroxymethyl)glutathione dehydrogenase	13.71620221	1.18E-07
PB_25233	Probable inorganic phosphate transporter 1-4	13.64755454	1.15E-08
PB_9218	Early light-induced protein, chloroplastic	13.51793615	9.78E-21
PB_14054	Anoctamin-like protein At1g73020	13.27425908	1.14E-06
PB_1776	Hva1_TUDOR domain-containing protein	12.92229313	1.25E-07
PB_29477	Protein JOKA2	12.66817168	4.04E-06
PB_29560	BAG family molecular chaperone regulator 6	12.65659538	1.88E-11
PB_22232	ABSCISIC ACID-INSENSITIVE 5-like protein 5	12.6215597	4.76E-24
PB_25232	Probable inorganic phosphate transporter 1-4	12.46981314	9.34E-09
PB_23826	Mediator of RNA polymerase II transcription subunit 13	12.29893949	2.03E-07
PB_21844	Casein kinase 1-like protein HD16	12.28970877	1.4E-31
PB_10073	Wall-associated receptor kinase-like 20	12.16765662	5.6E-21
PB_29840	MLO-like protein 5	12.09947783	1.33E-10
PB_32675	Peroxygenase	12.00787414	1.32E-14
PB_6030	Photosystem II 10 kDa polypeptide, chloroplastic	11.86002225	4.05E-09
PB_30665	Laccase-11	-9.697841922	1.08E-08
PB_36670	Protein GLUTELIN PRECURSOR ACCUMULATION 3	-9.814003654	9.18E-10
PB_19145	Pentatricopeptide repeat-containing protein At2g27610	-9.817341504	5.98E-12
PB_30519	Multicopper oxidase LPR1 homolog 1	-9.829560075	4.56E-08
PB_50748	Transcription factor MYB61	-9.846798228	1.65E-10
PB_49223	Probable pectate lyase 15	-9.859203908	2.2E-09
PB_3041	Desmethyl-yatein O-methyltransferase	-10.15974187	9.92E-07
PB_27873	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	-10.19185166	9.31E-06
PB_25243	Probable inorganic phosphate transporter 1-8	-10.21705238	2.13E-12
PB_20895	Pentatricopeptide repeat-containing protein At2g33680	-10.58288741	6.71E-08
PB_12393	Xyloglucan endotransglucosylase protein 6	-10.62834215	3.7E-07

PB_19926	Pentatricopeptide repeat-containing protein At4g18520, chloroplastic	-10.70968889	7.11E-08
PB_42882	AAI domain-containing protein	-10.76089309	4.26E-20
PB_16825	Phenylalanine ammonia-lyase	-10.8614327	8.47E-10
PB_49312	L-ascorbate oxidase homolog	-10.94786192	2.52E-23
PB_16	Putative 4-hydroxy-4-methyl-2-oxoglutarate aldolase 1	-11.04508721	5.95E-09
PB_36129	Putative pentatricopeptide repeat-containing protein At3g23330	-11.40381085	4.56E-13
PB_13124	Lipid phosphate phosphatase 2	-11.64972535	8.29E-10
PB_40609	Peroxidase 49	-12.15302341	2.38E-09
PB_9163	Peroxidase 4	-12.66227327	8.53E-20

Table S3: Possible regulatory elements of the desiccation response in the rhizomes of *Anemia caffrorum*

gene ID	Gene name	Function	Reference
Negative regulation of abscisic acid-activated signalling pathway			
PB_8932	Homeobox-leucine zipper protein HOX21	microRNA in bamboo rhizome involved in metabolism and nutrition; transcriptional activator involved in leaf development in the fern *Dryopteris fragrans*	(Chen et al., 2020; Jin et al., 2016)
PB_26154	Probable protein phosphatase 2C 6	negative regulation of ABA via deactivation of SnRKs	(Pandey, 2020; Smolikova et al., 2021; Umezawa et al., 2009)
PB_21424	Probable protein phosphatase 2C 8	negative regulation of ABA via deactivation of SnRKs (the specific role for PPC2C8 has not been determined)	(Pandey, 2020; Smolikova et al., 2021; Umezawa et al., 2009)
PB_30832	Probable serine/threonine-protein kinase SIS8	negative regulator of salt tolerance; mediator of sugar response during early seedling development in *A. thaliana*	(Chanwala et al., 2020; Gao and Xiang, 2008)
PB_21457	Probable WRKY transcription factor 4	regulation of plant stress response (the specific role for WRKY4 has not been determined)	(Chanwala et al., 2020)
PB_13840	Protein phosphatase 2C 53	negative regulation of ABA via deactivation of SnRKs (the specific role for PPC2C8 has not been determined)	(Pandey, 2020; Smolikova et al., 2021; Umezawa et al., 2009)
PB_13258	Receptor-like protein kinase FERONIA	Regulation of cell expansion, flowering, energy metabolism, stomatal movement and hormone signalling	(Ji et al., 2020)
PB_12762	RNA polymerase II C-terminal domain phosphatase-like 3	degrades transcripts with premature translation-termination codons	(Cui et al., 2016)
PB_11603	Zinc finger protein 6	Probable transcription factor required for the initiation of inflorescence trichomes in response to gibberellin and cytokinin	(Zhou et al., 2013)
Positive seed germination			
PB_30764	Protein HEADING DATE 3A	Probable mobile flower-promoting signal (florigen) that moves from the leaf to the shoot apical meristem (SAM) and induces flowering. Promotes the transition from vegetative growth to flowering downstream of HD1 and EHD1 under short day (SD) conditions	(Komiya et al., 2009)
Abscisic acid-activated signalling pathway			
PB_11504	3-ketoacyl-CoA thiolase 2, peroxisomal	breakdown of fatty acids, performed by the β-oxidation cycle, is crucial for plant germination and sustainability	(Pye et al., 2010)
PB_14841	ABSCISIC ACID-INSENSITIVE 5-like protein 2	Participates in ABA-regulated gene expression during seed development and subsequent vegetative stage by acting as the major mediator of ABA repression of growth. Binds to the embryo specification element and the ABA-responsive element (ABRE) of the Dc3 gene promoter and to the ABRE of the Em1 and Em6 genes promoters. Can also trans-activate its own promoter, suggesting that it is autoregulated. Plays a role in sugar-mediated senescence.	(Finkelstein et al., 2005; Nakashima et al., 2006)
PB_30677	ABSCISIC ACID-INSENSITIVE 5-like protein 4		
PB_22231	ABSCISIC ACID-INSENSITIVE 5-like protein 5		
PB_21716	B3 domain-containing protein VP1	ABA-responsive and involved in plant growth and seed germination	(Waltner et al., 2005)
PB_37575	B3 domain-containing transcription factor ABI3	ABA-responsive and involved in plant growth and seed germination	(Waltner et al., 2005)
PB_23039	Guard cell S-type anion channel SLAC1	preferentially expressed in guard cells, has been identified to be mainly responsible for the stress-induced anion efflux and thus playing a pivotal role mediating stomatal closure	(Li et al., 2022)
PB_23074	LanC-like protein GCL1	ABA-independent role in seed germination and early seedling development	(Guo et al., 2008)
PB_4519	Mitogen-activated protein kinase 1	Wide roles in phytohormone signalling (auxin, JA, SA, brassinosteroids, ethylene, ABA), ABA-gene regulation, ABA-regulation in guard cells, ABA in seed germination.	(Jagodzik et al., 2018)
PB_27732	Mitogen-activated protein kinase 15		
PB_27730	Mitogen-activated protein kinase 9		
PB_7082	Mitogen-activated protein kinase kinase kinase 17		
PB_774	Nuclear transcription factor Y subunit B-6	regulator of embryo development.	(Zhao et al., 2017)

PB_30764	Protein HEADING DATE 3A	Probable mobile flower-promoting signal (florigen) that moves from the leaf to the shoot apical meristem (SAM) and induces flowering. Promotes the transition from vegetative growth to flowering downstream of HD1 and EHD1 under short day (SD) conditions	(Komiya et al., 2009)
PB_30757	Protein MOTHER of FT and TFL1	Overexpression of MFT leads to slightly early flowering; *Adiantum capillus-veneris* MFT (AcMFT) accelerated the floral transition and partially rescued the late flowering phenotype of Arabidopsis ft-1 mutant.	(Hou and Yang, 2016; Yu et al., 2019)
PB_28845	Protein-serine/threonine phosphatase	Cell signalling	(Brautigan and Shenolikar, 2018)
PB_9179	Ras-related protein RABE1c	regulates stomatal movements and drought stress responses by mediating the interaction with ABA receptors	(Chen et al., 2021)
PB_19261	Respiratory burst oxidase homolog protein F	Calcium-dependent NADPH oxidase that generates superoxide. Generates reactive oxygen species (ROS) during incompatible interactions with pathogens and is important in the regulation of the hypersensitive response (HR). Involved in abscisic acid-induced stomatal closing and in UV-B and abscisic acid ROS-dependent signalling	(Desikan et al., 2006)
PB_28906	Serine/threonine-protein kinase SAPK10	osmotic stress/ABA–activated protein kinase	(Kalbina and Strid, 2006)
PB_26593	Transcription factor MYB44	expressed throughout the plant life cycle and especially in senescing and wounded leaves	(Kobayashi et al., 2004)
PB_16960	Transcription factor MYC2	Transcriptional activator. Common transcription factor of light, abscisic acid (ABA), and jasmonic acid (JA) signalling pathways. With MYC3 and MYC4, controls additively subsets of JA-dependent responses. In cooperation with MYB2 is involved in the regulation of ABA-inducible genes under drought stress conditions	(Abe et al., 2003; Fernández-Calvo et al., 2011; Niu et al., 2011)
PB_33904	Transcription factor SRM1	SlSRM1-like plays an important role in regulating leaf development in tomatoes	(Tang et al., 2022)

Ethylene-activated signalling pathway

Mediation of plant responses to drought, fruit ripening and the induction of secondary metabolite production. (Srivastava and Kumar, 2019)

PB_359	Ethylene-responsive transcription factor 1A	Acts as a transcriptional activator. Binds to the GCC-box pathogenesis-related promoter element.	(Fujimoto et al., 2000)
PB_1476	Ethylene-responsive transcription factor 3	Acts as a transcriptional repressor. Binds to the GCC-box pathogenesis-related promoter element.	(Fujimoto et al., 2000)
PB_2791	Ethylene-responsive transcription factor 4	Acts as a transcriptional repressor. Binds to the GCC-box pathogenesis-related promoter element.	(Fujimoto et al., 2000)
PB_34803	Ethylene-responsive transcription factor ERF003	Probably acts as a transcriptional activator. Binds to the GCC-box pathogenesis-related promoter element.	(Fujimoto et al., 2000)
PB_30341	Ethylene-responsive transcription factor ERF023	unknown	
PB_32184	Ethylene-responsive transcription factor ERF055	Probably acts as a transcriptional activator. Binds to the GCC-box pathogenesis-related promoter element.	(Fujimoto et al., 2000)
PB_25963	Ethylene-responsive transcription factor ERF073	Transcriptional activator involved in the hypoxic stress response; Acts as a transcriptional activator via the cis-acting element GCC box	(Fujimoto et al., 2000; Müller and Munné-Bosch, 2015; Tao et al., 2018)
PB_50425	Ethylene-responsive transcription factor ERF110	mediates ethylene-regulated transcription of a sex determination-related orthologous gene in two Cucumis species	(Tao et al., 2018)
PB_27609	Ethylene-responsive transcription factor RAP2-4	Probably acts as a transcriptional activator. Binds to the GCC-box pathogenesis-related promoter element.	(Fujimoto et al., 2000)
PB_1028	Protein ENHANCED DISEASE RESISTANCE 2	Prevents ethylene-induced senescence. Binds to phosphatidylinositol-4-phosphate (PtdIns4P) in vitro	(Vorwerke et al., 2007)

PB_26551	Protein SABRE	Required for cell expansion, especially in root cortex, probably by counteracting the action of ethylene in promoting cells radial expansion; Antagonistically interacts with ethylene signalling to regulate plant responses to Pi starvation	(Yu et al., 2012; Courtneidge et al., 1993)
PB_9179	Ras-related protein RABE1c	regulates stomatal movements and drought stress responses by mediating the interaction with ABA receptors	(Chen et al., 2021)
PB_19261	Respiratory burst oxidase homolog protein F	Calcium-dependent NADPH oxidase that generates superoxide. Generates reactive oxygen species (ROS) during incompatible interactions with pathogens and is important in the regulation of the hypersensitive response (HR). Involved in abscisic acid-induced stomatal closing and in UV-B and abscisic acid ROS-dependent signalling	(Desikan et al., 2006)

Final discussion and conclusions:

The overall aim of the study was to explore the mechanism of seasonal desiccation tolerance displayed by the fronds and rhizome of *A. caffrorum* at the level of transcription. The three key objectives were to 1) characterise desiccation tolerance in the fronds, 2) characterise desiccation tolerance in the rhizome and 3) to evaluate, if any, the key regulators of frond phenotype that may exist in the rhizome.

Mechanisms of desiccation tolerance in the fronds

Exploration of the fronds yielded a considerably lower number of differentially expressed genes (DEGs) than is common for similar gene expression studies conducted in angiosperm resurrection plants (Figure 3 in Chapter 3). A low transcriptomic response is characteristic of bryophytic-type desiccation tolerance in which constitutive protection is a key feature. Although fewer responses were recorded, both inducible and constitutive features were present. Together, this agrees with the widely accepted proposal that Pteridophytes represent a mixed model of desiccation tolerance owing to their evolutionary intermediate position between the bryophytes and angiosperm desiccation tolerant species (Alejo-Jacuinde and Herrera-Estrella, 2022; Figure 1 in Chapter 1).

Desiccation sensitive (DS) fronds showed a marked reduction in transcript abundance of all enriched processes at the 55% RWC stage, with a seeming increase after rehydration (Figure 5 in Chapter 3). Evidence of senescent elements in the transcriptome followed similar expression patterns, supporting the physiological data which indicates that DS fronds rapidly lose photosynthetic ability from approximately 70% RWC and do not recover following a hydration event (Nadal *et al.*, 2021). Furthermore, a considerable increase in pentatricopeptides (PPRs) in DS fronds was observed at 10% RWC i.e., in tissue that one would consider effectively dead (Figure 6 in Chapter 4). PPRs are associated with post-transcriptional regulation, mRNA stabilization and the initiation of translation (Wang *et al.*, 2021), so it is possible that the system has been flooded with transcripts that drive senescence. Transport of mRNA across organs to distant tissues has been reported in *A. thaliana* (Thieme *et al.*, 2015). Here, and in other cases where mass transport of transcripts was suspected, transport of transcripts to areas of need is therefore feasible. Closer inspection of DEGs that were upregulated in DS fronds relative to DT fronds at 10% RWC support this theory. Two of the most highly upregulated transcripts in DS fronds at 10% RWC were an ATP-dependent protease ATPase subunit HslU (LFC = 9.20; $p = 9.98 \times 10^{-7}$) and peptidyl-prolyl cis-trans isomerase FKBP16-4 (LFC = 6.45; $p = 7.08 \times 10^{-}$

[3]). The former protein plays a role in protein degradation (Kanemori, Yanagi and Yura, 1999; Seong *et al.*, 1999) while the latter accelerates the folding of proteins (Dominguez-Solis *et al.*, 2008; Park *et al.*, 2013). This implies that some form of targeted translation was still occurring and that perhaps it was geared to towards senescent processes. Indeed, recent studies on angiosperm resurrection plants in which distinction is made between non-senescent tissues and those destined to senesce during desiccation have shown that senescence processes are induced in mid to late stages during dehydration, but senescence *per se* is only realised on rehydration (Radermacher et al., 2019; 2021; du Toit et al., 2022). The relative increase in transcripts related to photosynthesis, carbon fixation, and photorespiration observed after rehydration in DS fronds (Figure 5; Chapter 3) suggests that the transcripts housed within chloroplasts are relatively more protected from senescent processes than those occurring in the cytoplasm and other organelles. It is possible that other organelles, stress granules etc. are also well protected, but were not detected via enrichment studies. Overall, DS fronds do not survive desiccation and considerable senescent response could indicate the remobilization of nutrients back into the rhizome.

Desiccation tolerant (DT) fronds and their key inducible and constitutive features of tolerance were described in Chapter 3. Typical inducible and constitutive antioxidant responses to desiccation were observed in DT fronds. Notable in its relative uniqueness was the increase in production of the antioxidant secologanin which was observed as a prominently activated antioxidant pathway in both the fronds and rhizomes. Constitutively maintained features of desiccation tolerance in DT fronds at higher water contents included the production and maintenance of anthocyanins, presumably for the maintenance orange, trichome-like scales on the abaxial surface of fronds. These orange trichome-like scales protect the fronds from excess photodamage and may assist in slowing dehydration (Farrant *et al.*, 2009; John and Hasenstein, 2017). Evidence for H_2O_2 priming was also observed at higher water contents and seemingly lost thereafter. Early-stage responses included inducible protection of the photosystem via a transition to cyclic electron flow, prevention of toxic iron accumulation and production of photoprotective proteins. Related to this, but as a constitutive feature, regulation of cellular respiration was observed from the early stage through to 10% RWC. Also constitutively maintained during this period was a greater proportion of unsaturated fatty acids. Inducible lipid modulation included the conversion of toxic storage lipids such as diacylglycerol to a non-toxic forms and increased production in very long chain fatty acids. In the mid-stages (i.e., entry into desiccation) ongoing transcription and translation were present and an inducible

response in root signalling occurred. More specifically, signals that promoted the recruitment of root endophytes and increased waterproofing of the roots was observed. In the late stage, evidence for cellular glass formation included the mass transport of sucrose out of chloroplasts into the cytoplasm and subsequent conversion to trehalose. Concurrently, mass upregulation of HSP70 transcripts occurred. While a noticeable increase in elements related to ABA-regulation occurred at 55% RWC, the data did not provide sufficient evidence to explain the full extent to which ABA is involved in the desiccation response. ABA has been shown to play significant roles in the desiccation response in other ferns (as summarized in Chapter 1), as well as several other resurrection plants from the bryophytes and angiosperm groupings (Gutierrez *et al.*, 2007; Khandelwal *et al.*, 2010; Fernando and Schroeder, 2016). Additional studies directly measuring ABA content in DT fronds during drying, desiccation and rehydration is therefore worthy of further probing by alternative methods. Taken together, the transcriptional mechanisms of desiccation tolerance in the fronds of *A. caffrorum* showed greater similarity to those observed in angiosperms; a finding that was mirrored in the initial biochemical characterisation of *A. caffrorum* (Farrant *et al.*, 2009).

Mechanisms of desiccation tolerance in the rhizome

Evaluation of the rhizomes led to several novel insights about the interplay between this perennial, underground organ, and the transient vegetative tissues. However, it was more difficult to characterise this organ since most studies of desiccation tolerance are centred around photosynthetic tissues. Furthermore, such organs have even more rarely been studied via a systems biology approach. In terms of the life strategies adopted by ferns, Holmlund *et al* (2019, 2020) have recently described two highly useful insights to rhizome biology in the context of desiccation tolerance. Firstly, they demonstrated that several DT ferns of the terrestrial variety embolise xylem tracheids in the stipe as to protect the rhizome from water loss during desiccation, thereby protecting the survival of the whole organism. Furthermore, Holmlund *et al* (2019, 2020) demonstrated that frond turgor and recovery only occurs when the rhizome has been rehydrated in the terrestrial ferns *Pentagramma triangularis* and *Pellaea andromedifolia*, strongly implicating the organ in re-establishing normal metabolism after desiccation. Recovery via foliar uptake has only been demonstrated in epiphytic ferns thus far, presumably as their rhizomes are more sensitive to water loss in their niche environments (Prats and Brodersen, 2021). Stipe embolism and whole organ rehydration dynamics of *A. caffrorum* have shown similarities to what has been described for these other terrestrial DT ferns, but only anecdotally observed. As a priority, future studies should include a characterisation of *A.*

caffrorum rhizome biology and hydraulic measurements that is more up to date with current knowledge. This will be necessary to better contextualise the omics studies conducted to date i.e., proteomics (Shoko, 2007), transcriptomics presented here and the metabolomics research that was recently completed (Wittenberg 2021; Honours dissertation). An unexpected finding that requires explanation to contextualize the concluding remarks is the prominence of chloroplast-exclusive transcripts in rhizomes. This is very likely a result of sampling the rhizome meristem which showed clear signs of greening during experimentation (Figure 1 in Chapter 4). The meristem was specifically sampled as isolations of RNA from the whole rhizome were not producing sufficient yields for NGS experimentation. More concentrated RNA in this region implied a greater transcriptional role for the meristem. Additionally, new growth occurs from this area, so it was assumed that regulatory control of frond phenotype would be strongly associated with the meristem. Sampling of this region therefore aligned with the goals of this study. However, future physiological studies could include basic TEM studies of the meristematic zone to confirm the assertion that this region contains chloroplasts and may be photosynthetically active as will be discussed shortly.

Previous proteomic studies demonstrated that the rhizome of *A. caffrorum* is tolerant in both seasons and positioned this organ as a possible regulator of frond phenotype (Shoko, 2007). Rhizomes occurring in both seasons showed signatures related to common desiccation responses observed in the vegetative tissues of other resurrection plants, as well as cross-tolerances to heat and pathogen stress in summer and winter respectively. The initial aim of this portion of the study was to first characterise transcriptional mechanisms of desiccation tolerance in the rhizomes, and secondly to ascertain whether there is evidence for control of frond phenotype that is governed by the perennial organ. The total number of DEGs yielded when comparing desiccated rhizomes with their hydrated counterparts was considerably higher than the number observed when comparing the rhizomes across seasons (Figure 3 in Chapter 4). In addition, the total number of DEGs was also considerably lower compared to other studies on the vegetative tissues of angiosperm resurrection plants. However, it is difficult to conclude whether this suggested a low transcriptional response which is characteristic of desiccation tolerance in bryophytes or simply characteristic for this organ. Desiccated rhizomes featured several responses to desiccation that were akin to those observed in the vegetative tissues of angiosperm resurrection plants. The trade-off for an upregulation of vitrification, launching of an antioxidant response and protection of the photosystems, was a concurrent downregulation of normal growth, transcription, and translation (Figure 4 in Chapter 4).

Although rhizomes were showing evidence for the vitrified state at 30% RWC as there was a prominent upregulation in the late embryogenesis (LEA) proteins, heat shock proteins (HSPs) and raffinose oligosaccharide (RFO) producing enzymes, there was still too much water present to conclude this. Together these protective proteins and sugars are strongly implicated in cellular glass formation (Tolleter *et al.*, 2007; Tolleter, Hincha and Macherel, 2010; Gechev *et al.*, 2012; Oliver *et al.*, 2020), but more likely the rhizome was fully prepared for vitrification should water decrease further. The rhizomes were more likely in an intermediate rubbery state at this RWC (Vertucci and Farrant, 1995). Note that rhizomes will continue to be referred to as desiccated to maintain persistent terminology throughout the dissertation, but it is now hypothesized that they were approaching desiccation and not fully desiccated. Protection of the photosystem in both desiccated rhizomes was implicated via the mass accumulation of transcripts encoding ELIPs and the fern-specific protein - 22 kDa protein of chloroplasts. ELIPs were some of the most highly abundant transcripts in desiccated rhizomes. In such abundance, in an organ with limited photosynthetic potential, it is possible that these are mass accumulated to protect chloroplasts in new fronds that emerge following rehydration of the rhizome. The degradation of 22 kDa protein of chloroplasts during spore germination led to the implied role for this protein in chloroplast maintenance in the desiccated state (Minamikawa, Koshiba and Wada, 1984; Haupt, 1990; Inoue et al., 2000). This protein was induced and accumulated in DT fronds at 55% RWC through to desiccation and rehydration and was accumulated in desiccated rhizomes. This strongly agrees with a role for this protein in the protection of chloroplasts during desiccation and extends its function beyond fern spores, to the sporophyte generation and the rhizome. In addition to preparation of cellular glass formation and protection of the photosystems, evidence for the active maintenance of membrane contact sites and a canonical antioxidant system response were also observed (as detailed in Chapter 4). Together, these findings agree with the hypothesis that the rhizome is constitutively tolerant and that it displays features of desiccation tolerance that are common to vegetative tissues of other resurrection plants.

This study did not find similar heat and pathogen cross-tolerances in the summer and winter plants respectively. This is likely because the proteomics study was able to maintain plants in their respective phenotypes by growing the plants in conditions that mimic summer and winter respectively. A robust and reproducible method has not been established since, and so cross-tolerances were not detected. However, a lower number of DEGs and less enrichment of process (Figure 1 and 2 in Chapter 4) in summer collected plants may imply that summer

rhizomes are simply less equipped for a wider range of transcriptional activities. Since both phenotypes had been acclimated to mesic conditions in this study, it is more likely reflective of summer-collected plants having depleted storage reserves in their rhizomes at the time of collection. Winter-grown plants, owing to the greater photosynthetic potential of DS fronds, likely had more stored nutrients at the start of experimentation. The drawback to this advantage in plants producing DS fronds is that when rhizomes producing DS fronds experience a drought event, they lose their DS fronds and therefore, most likely, cannot rely on these source tissues (fronds) for key metabolites. Although, they may regain some nutrients from senescent fronds that are transported back to the rhizome.

Sucrose was measured in high abundance in the desiccated rhizomes of both phenotypes (Wittenberg 2021; Honours dissertation), yet sucrose-producing enzymes were not upregulated in the meristem (discussed in Chapter 4). This implies that sucrose in the rhizome is received from source tissues (fronds) at some point during drying and/or desiccation, and presumably transported before DT fronds enter a state of quiescence (below 10% RWC) and before DS fronds lose viability (which begins at approximately 70% RWC) (Farrant *et al.*, 2009). If sucrose is being sourced by the fronds in both seasons, this suggests that mass transport thereof and also amino acids via the phloem and is enabled and during drying. In alternative explanation, upregulation of photosynthesis in winter rhizomes may be an adaptation to source sucrose which is essential to protect the rhizome during desiccation. Again, this is because winter rhizomes produce desiccation sensitive fronds which cannot tolerate desiccation. In fact, studies have demonstrated that when roots are detached from photosynthetic shoots, they can develop chloroplasts and several genes associated with photosynthesis become upregulated (Kobayashi *et al.*, 2017). Omics studies to date have clearly demonstrated that source-sink dynamics play a fundamental role in the life strategy of *A. caffrorum* and should be considered in future work. Radiolabelling key biomolecules and tracking their movement during a desiccation event would be enormously useful in this regard. In addition, it would be worthwhile to measure the RWC of fronds and rhizomes in future studies as to determine what stage of desiccation each organ is in relative to one another during a desiccation event. To facilitate this, a protocol for growing adult plants from spore should be generated as to protect the natural populations.

Regulation of frond phenotype in the rhizome

The final goal of the study was to determine whether the rhizome is regulating frond phenotype and to identify the key elements/pathways featured in this process. Functional enrichment analysis drew attention to a possible role of abscisic acid (ABA) in this regard (Figure 7 in Chapter 4). The set of genes included activators and repressors of ABA signalling pathway, genes that play a role in frond morphology, frond senescence and those that control the timing of spore formation in fronds. Pathway analysis was attempted to identify genes that are key drivers in this process. Of the genes studied, less than 25% had an annotation in the *A. thaliana* databases required for pathway analysis (*results not shown*). The resulting analyses were compared to a manual evaluation of key genes. Although pathway analysis was useful in identifying enriched pathways and the most upregulated genes in each process, the findings omitted many interesting genes that were not included because a lack of annotation. This is a common complication when working with non-model organisms and having to rely on databases that highly favour popular organisms such as *A. thaliana*. The relatively low number of DEGs identified in this study most likely exacerbated the issue. On the other hand, a more robust manual evaluation was possible only because the gene set was smaller. The results, summarized in Chapter 4, highlight several key players which should be further explored. Of particular interest was the presence of zinc finger protein 6 (ZFP6) which is a transcription factor that initiates the production of trichomes (Zhou *et al.*, 2013). As discussed in Chapter 3, the maintenance and production of anthocyanins in orange, trichome-like scales is a key distinguisher between DT and DS fronds under conditions of higher water availability, implying that it plays a role in determining the phenotype of the fronds. Its presence in the desiccated rhizome is interesting as one would naturally expect that it must be mobilized to emerging fronds where trichomes-like scales are produced. The activation/or repression of the trichome production cascade in emerging DT and DS fronds respectively, presumably plays a role in the initial determination of frond phenotype which is later maintained by elements such as ABC transporter B family member 19 and UDP-glycosyltransferases (as discussed in Chapter 3). Future work should focus on a comparison of emerging DS and DT fronds, which would be coupled with rhizome samples following rehydration. Although we initially intended to capture this time-point our efforts were disrupted by the pandemic and subsequently a wildfire that eradicated the population at our site of study. This further underlines the need for a system in which adult plants can be grown from spore, and the development of a reproducible method to maintain each phenotype under artificial conditions. Alternatively, future frond samples could be collected from the field. Despite the very low number of identified DEGs, the comparison of full turgor summer and winter rhizomes was conducted as it may contain

elements regulatory elements that contribute to the maintenance of DT and DS fronds (Figure 3 in Chapter 4). Several elements that were upregulated in full turgor rhizomes producing DT fronds relative to full turgor rhizomes producing DS fronds were identified and are worth brief discussion. Noticeably, ABSCISIC ACID-INSENSITIVE 5-like protein 5 (ABI5-like protein 5) was comparatively upregulated in hydrated DT rhizomes (LFC = 8.89, $p = 1.44 \times 10^{-9}$). ABI5-like protein 5 was therefore upregulated in rhizomes producing DT fronds and in both desiccated rhizomes which strongly implies a role for this protein as a key regulated of desiccation tolerance in the whole organism. Nuclear speckle splicing regulatory protein 1 modulates alternative splicing of pre-mRNA (Kim *et al.*, 2011; Galganski, Urbanek and Krzyzosiak, 2017) and was also highly upregulated in hydrated DT rhizomes compared to hydrated DS rhizomes (LFC = 10.71, $p = 3.89 \times 10^{-5}$). This protein may therefore be implicated in the alternative splicing events that are associated with the maintenance of tolerant fronds. The chloroplastic chaperone protein ClpB3 was upregulated in hydrated DT rhizomes (LFC = 3.99, $p = 1.81 \times 10^{-4}$). Overexpression of ClpB3 has been shown to inhibit chloroplast development, result in a pale-green phenotype and confers thermotolerance to chloroplasts during heat stress (Myouga *et al.*, 2006; Pan *et al.*, 2008). ClpB3 is therefore likely a key element in limiting photosynthetic potential in DT fronds. DT fronds are also characteristically much smaller than DS fronds, and this may be accredited to the upregulation of stem-specific protein TSJT1 in hydrated DT rhizomes (LFC = 4.65, $p = 8.90 \times 10^{-7}$). TSJT1 was more highly upregulated in a dwarf variety of castor bean and was postulated to be a negative regulator of internode development (Hu *et al.*, 2016). The exact function of this protein has yet to elucidated but the upregulation of this gene in rhizomes producing DT fronds certainly supports its role in maintaining vegetative tissues in their dwarf state. These findings further extend the role of the rhizome in maintaining the DT phenotype.

It is worth noting that the full hydration state of DT fronds is more likely an artificial condition brought about through experimentation, rather than a likely scenario in the field. So, the continued maintenance of anthocyanins in orange, trichome-like scales, dwarfism and limited photosynthetic output in DT fronds could imply that, even under conditions of high-water availability, DT fronds would not ‘change their strategy’, so to speak. Anecdotal evidence has led to the hypothesis that DT fronds cannot revert to the DS phenotype, and this finding supports that notion. It is however still possible that instead of fronds that were predetermined to be DT or DS, DS fronds are gradually hardened during the winter-to-summer transition. Although the rhizome has control over the production of protective structures such as the

orange, trichome-like scales on DT fronds it does not answer the question of whether an emerging frond is programmed at an early stage for its irreversible phenotype. As an initial exploration of this question, fronds emerging in the field during summer and winter could be subjected to scanning electron microscopy (SEM) studies. This could at least provide insight as to whether emerging fronds are already equipped with orange, trichome-like scales or if they are gradually developed as the season continues. The belief is that all the elements that are required for DT in seeds, must still be present in the vegetative tissue of non-DT plants. As a result, a lot of research focuses on how to reactivate these pathways in sensitive plants. In the case of *A. caffrorum* this could be a reactivation of DT in DS fronds. If this is the case, collection of DS fronds undergoing the transition from DS to DT could very likely provide a strong list of candidates for further study in the production of climate smart, transgenic crops. Work from this study suggests that fronds are predetermined to be DT or DS, but this may simply be reflective of the experimental design that captured tissues in these states. Altogether, the study has provided a framework for the mechanism of desiccation tolerance in fronds and rhizomes and provided insight into frond regulation but has likely not captured a transition of DS to DT which could easily be studied now that a high-quality transcriptome is available.

REFERENCE LIST

‘A RE-EXAMINATION OF THE RELATIVE TURGIDITY TECHNIQUE FOR ESTIMATING WATER DEFICITS IN LEAVES By H. D. BARRs* and P. E. WEATHERLEyt’ (1962).

Adamska, I. *et al.* (1999) ‘Isolation of pigment-binding early light-inducible proteins from pea’, *European Journal of Biochemistry*, 260(2), pp. 453–460. doi: 10.1046/j.1432-1327.1999.00178.x.

Alamillo, J. M. and Bartels, D. (2001) ‘Effects of desiccation on photosynthesis pigments and the ELIP-like dsp 22 protein complexes in the resurrection plant Craterostigma plantagineum’, *Plant Science*. doi: 10.1016/S0168-9452(01)00356-9.

Alejo-Jacuinde, G. *et al.* (2020) ‘Comparative transcriptome analysis suggests convergent evolution of desiccation tolerance in Selaginella species’, *BMC Plant Biology*. BMC Plant Biology, 20(1), pp. 1–18. doi: 10.1186/s12870-020-02638-3.

Alejo-Jacuinde, G. and Herrera-Estrella, L. (2022) ‘Exploring the High Variability of Vegetative Desiccation Tolerance in Pteridophytes’, *Plants*, 11(9), pp. 1–22. doi: 10.3390/plants11091222.

All, U. T. C. (2016) ‘Physiological Benefits of Stem Curling for Resurrection Plants in the Field Author (s): Jefferson G . Lebkuecher and William G . Eickmeier Published by : Wiley Stable URL : http://www.jstor.org/stable/1940477 REFERENCES Linked references are available o’, 74(4), pp. 1073–1080.

Alpert, P. (2000) ‘The discovery, scope, and puzzle of desiccation tolerance in plants’, *Plant Ecology*, 151(1), pp. 5–17. doi: 10.1023/A:1026513800380.

Alpert, P. (2005) ‘The limits and frontiers of desiccation-tolerant life’, *Integrative and Comparative Biology*, 45(5), pp. 685–695. doi: 10.1093/icb/45.5.685.

Alpert, P. (2006) ‘Constraints of tolerance: Why are desiccation-tolerant organisms so small or rare?’, *Journal of Experimental Biology*. doi: 10.1242/jeb.02179.

Alvim, F. C. *et al.* (2001) ‘Enhanced accumulation of BiP in transgenic plants confers tolerance to water stress’, *Plant Physiology*, 126(3), pp. 1042–1054. doi: 10.1104/pp.126.3.1042.

Andersen, G. R., Nissen, P. and Nyborg, J. (2003) ‘Elongation factors in protein biosynthesis’, *Trends in Biochemical Sciences*, 28(8), pp. 434–441. doi: 10.1016/S0968-0004(03)00162-2.

Asada, K. *et al.* (2013) ‘A 7-Deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in madagascar periwinkle’, *Plant Cell*, 25(10), pp. 4123–4134. doi: 10.1105/tpc.113.115154.

Aubourg, S. *et al.* (2000) ‘In Arabidopsis thaliana, 1% of the genome codes for a novel protein family unique to plants’, *Plant Molecular Biology*, 42(4), pp. 603–613. doi:

10.1023/A:1006352315928.

Bai, Y. *et al.* (2015) 'X-ray structure of a mammalian stearoyl-CoA desaturase', *Nature*, 524(7564), pp. 252–256. doi: 10.1038/nature14549.

Baiphethi, M. N. and Jacobs, P. T. (2009) 'The contribution of subsistence farming to food security in South Africa', *Agrekon*. doi: 10.1080/03031853.2009.9523836.

Ballesteros, D., Hill, L. M. and Walters, C. (2017) 'Variation of desiccation tolerance and longevity in fern spores', *Journal of Plant Physiology*. Elsevier GmbH., 211, pp. 53–62. doi: 10.1016/j.jplph.2017.01.003.

Balsamo, R. A. *et al.* (2006) 'Drought tolerance of selected Eragrostis species correlates with leaf tensile properties', *Handbook of Environmental Chemistry, Volume 5: Water Pollution*, 97(6), pp. 985–991. doi: 10.1093/aob/mcl068.

Banerjee, A. and Roychoudhury, A. (2016) 'Group II late embryogenesis abundant (LEA) proteins: structural and functional aspects in plant abiotic stress', *Plant Growth Regulation*. Springer Netherlands, 79(1), pp. 1–17. doi: 10.1007/s10725-015-0113-3.

Bateman, A. *et al.* (2021) 'UniProt: the universal protein knowledgebase in 2021', *Nucleic Acids Research*. Oxford University Press, 49(D1), pp. D480–D489. doi: 10.1093/nar/gkaa1100.

Benko, Z. *et al.* (2002) 'Desiccation survival times in different desiccation-tolerant plants', *Acta Biologica Szegediensis*, 46(3–4), pp. 231–233.

Berestovoy, M. A., Pavlenko, O. S. and Goldenkova-Pavlova, I. V. (2020) 'Plant Fatty Acid Desaturases: Role in the Life of Plants and Biotechnological Potential', *Biology Bulletin Reviews*, 10(2), pp. 127–139. doi: 10.1134/s2079086420020024.

Bersuker, K. *et al.* (2020) 'HHS Public Access', 575(7784), pp. 688–692. doi: 10.1038/s41586-019-1705-2.The.

Bhuiyan, N. H. *et al.* (2016) 'The plastoglobule-localized metallopeptidase PGM48 is a positive regulator of senescence in arabidopsis thaliana', *Plant Cell*, 28(12), pp. 3020–3037. doi: 10.1105/tpc.16.00745.

Blomstedt, C. K. *et al.* (2018) 'Plant desiccation tolerance and its regulation in the foliage of resurrection "flowering-plant" species', *Agronomy*, 8(8), pp. 1–33. doi: 10.3390/agronomy8080146.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

Braymer, J. J. and Lill, R. (2017) 'Iron–sulfur cluster biogenesis and trafficking in mitochondria', *Journal of Biological Chemistry*, 292(31), pp. 12754–12763. doi: 10.1074/jbc.R117.787101.

Brulé, V. *et al.* (2019) 'Three-dimensional functional gradients direct stem curling in the resurrection plant Selaginella lepidophylla', *Journal of the Royal Society Interface*, 16(159).

doi: 10.1098/rsif.2019.0454.

Bryant, D. M. *et al.* (2017) 'A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors', *Cell Reports*. ElsevierCompany., 18(3), pp. 762–776. doi: 10.1016/j.celrep.2016.12.063.

Bryceson, D. F. (2002) 'The scramble in Africa: Reorienting rural livelihoods', *World Development*. doi: 10.1016/S0305-750X(02)00006-2.

Buchfink, B., Reuter, K. and Drost, H. G. (2021) 'Sensitive protein alignments at tree-of-life scale using DIAMOND', *Nature Methods*. Springer US, 18(4), pp. 366–368. doi: 10.1038/s41592-021-01101-x.

Buitink, J. and Leprince, O. (2008) 'Intracellular glasses and seed survival in the dry state', *Comptes Rendus - Biologies*. Elsevier Masson SAS, 331(10), pp. 788–795. doi: 10.1016/j.crvi.2008.08.002.

Van Buren, R. *et al.* (2019) 'Massive tandem proliferation of elips supports convergent evolution of desiccation tolerance across land plants', *Plant Physiology*, 179(3), pp. 1040–1049. doi: 10.1104/pp.18.01420.

Bushmanova, E. *et al.* (2019) 'RnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data', *GigaScience*. Oxford University Press, 8(9), pp. 1–13. doi: 10.1093/gigascience/giz100.

Byrne, A. *et al.* (2017) 'Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells', *Nature Communications*. Nature Publishing Group, 8(May), pp. 1–11. doi: 10.1038/ncomms16027.

Byrne, A. *et al.* (2019) 'Realizing the potential of full-length transcriptome sequencing', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1786). doi: 10.1098/rstb.2019.0097.

Cai, Z. *et al.* (2022) 'Full-length transcriptome analysis of Adiantum flabellulatum gametophyte', *PeerJ*, 10. doi: 10.7717/peerj.13079.

Caiveau, O. *et al.* (2001) 'Consequences of ω-6-Oleate Desaturase Deficiency on Lipid Dynamics and Functional Properties of Mitochondrial Membranes of Arabidopsis thaliana', *Journal of Biological Chemistry*, 276(8), pp. 5788–5794. doi: 10.1074/jbc.M006231200.

Canzar, S. *et al.* (2016) 'CIDANE: Comprehensive isoform discovery and abundance estimation', *Genome Biology*. Genome Biology, 17(1), pp. 1–18. doi: 10.1186/s13059-015-0865-0.

Cao, L. *et al.* (2021) 'Maize ZmbZIP33 Is Involved in Drought Resistance and Recovery Ability Through an Abscisic Acid-Dependent Signaling Pathway', *Frontiers in Plant Science*, 12(April), pp. 1–16. doi: 10.3389/fpls.2021.629903.

Cavallaro, M. *et al.* (2021) '3 ′-5 ′ Crosstalk Contributes To Transcriptional Bursting', *Genome Biology*. Genome Biology, 22(1), pp. 1–20. doi: 10.1186/s13059-020-02227-5.

Cea, M. G. *et al.* (2014) 'Desiccation tolerance of Hymenophyllacea filmy ferns is mediated

by constitutive and non-inducible cellular mechanisms', *Comptes Rendus - Biologies*, 337(4), pp. 235–243. doi: 10.1016/j.crvi.2014.02.002.

Challabathula, D., Puthur, J. T. and Bartels, D. (2016) 'Surviving metabolic arrest: Photosynthesis during desiccation and rehydration in resurrection plants', *Annals of the New York Academy of Sciences*, 1365(1), pp. 89–99. doi: 10.1111/nyas.12884.

Chanwala, J. *et al.* (2020) 'Genome-wide identification and expression analysis of WRKY transcription factors in pearl millet (Pennisetum glaucum) under dehydration and salinity stress', *BMC Genomics*. BMC Genomics, 21(1), pp. 1–16. doi: 10.1186/s12864-020-6622-0.

Charuvi, D. *et al.* (2019) 'Chloroplast breakdown during dehydration of a homoiochlorophyllous resurrection plant proceeds via senescence-like processes', *Environmental and Experimental Botany*, 157(September 2018), pp. 100–111. doi: 10.1016/j.envexpbot.2018.09.027.

Chávez Montes, R. A. *et al.* (2022) 'A comparative genomics examination of desiccation tolerance and sensitivity in two sister grass species', *Proceedings of the National Academy of Sciences of the United States of America*, 119(5). doi: 10.1073/pnas.2118886119.

Chen, D. *et al.* (2021) 'A ras-related small GTP-binding protein, RabE1c, regulates stomatal movements and drought stress responses by mediating the interaction with ABA receptors', *Plant Science*. Elsevier B.V., 306(February), p. 110858. doi: 10.1016/j.plantsci.2021.110858.

Chen, J. *et al.* (2020) 'Ribose-5-phosphate isomerases: characteristics, structural features, and applications', *Applied Microbiology and Biotechnology*. Applied Microbiology and Biotechnology, 104(15), pp. 6429–6441. doi: 10.1007/s00253-020-10735-4.

Chen, L. *et al.* (2020) 'Transcriptomic analysis and specific expression of transcription factor genes in the root and sporophyll of dryopteris fragrans (L.) schott', *International Journal of Molecular Sciences*, 21(19), pp. 1–20. doi: 10.3390/ijms21197296.

Chen, Q., Wan, Y., Lei, Y., Zobel, J. and Verspoor, K. (2017) 'Evaluation of CD-HIT for constructing non-redundant databases', *Proceedings - 2016 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2016*. IEEE, pp. 703–706. doi: 10.1109/BIBM.2016.7822604.

Chen, Q., Wan, Y., Lei, Y., Zobel, J., Verspoor, K., *et al.* (2017) 'Minimap2: Pairwise alignment for nucleotide sequences', *Bioinformatics*. IEEE, 34(18), pp. 3150–3152. doi: 10.1093/bioinformatics/bts565.

Chin, C. S. *et al.* (2013) 'Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data', *Nature Methods*. Nature Publishing Group, 10(6), pp. 563–569. doi: 10.1038/nmeth.2474.

Chini, A. *et al.* (2007) 'The JAZ family of repressors is the missing link in jasmonate signalling', *Nature*, 448(7154), pp. 666–671. doi: 10.1038/nature06006.

CIMMYT (2013) 'Drought tolerant maize for africa project', *Cimmyt*.

Coelho, P. S. R. *et al.* (2002) 'A novel mitochondrial protein, Tar1p, is encoded on the antisense strand of the nuclear 25S rDNA', *Genes and Development*, 16(21), pp. 2755–2760.

doi: 10.1101/gad.1035002.

Cooper, K. and Farrant, J. M. (2002) 'Recovery of the resurrection plant Craterostigma wilmsii from desiccation: Protection versus repair', *Journal of Experimental Botany*. doi: 10.1093/jxb/erf028.

Costa, M. C. D. *et al.* (2016) 'Key genes involved in desiccation tolerance and dormancy across life forms', *Plant Science*. Elsevier Ireland Ltd, 251, pp. 162–168. doi: 10.1016/j.plantsci.2016.02.001.

Costa, M. C. D. *et al.* (2017) 'Orthodox seeds and resurrection plants: Two of a kind?', *Plant Physiology*, 175(2), pp. 589–599. doi: 10.1104/pp.17.00760.

Courdavault, V. and Besseau, S. (2022) *Catharanthus roseus Methods and Protocols Methods in Molecular Biology 2505*. Available at: http://www.springer.com/series/7651.

Daniell, H. *et al.* (2005) 'Chloroplast-derived vaccine antigens and other therapeutic proteins', *Vaccine*, 23(15 SPEC. ISS.), pp. 1779–1783. doi: 10.1016/j.vaccine.2004.11.004.

Danquah, A. *et al.* (2015) 'Identification and characterization of an ABA-activated MAP kinase cascade in Arabidopsis thaliana', *Plant Journal*, 82(2), pp. 232–244. doi: 10.1111/tpj.12808.

Davidson, N. M. and Oshlack, A. (2014) 'Corset: Enabling differential gene expression analysis for de novo assembled transcriptomes', *Genome Biology*, 15(7), pp. 1–14. doi: 10.1186/s13059-014-0410-6.

Deeba, F., Pandey, A. K. and Pandey, V. (2016) 'Organ specific proteomic dissection of selaginella bryopteris undergoing dehydration and rehydration', *Frontiers in Plant Science*, 7(APR2016), pp. 1–20. doi: 10.3389/fpls.2016.00425.

Dehghan Nayeri, F. and Yarizade, K. (2014) 'Bioinformatics study of delta-12 fatty acid desaturase 2 (FAD2) gene in oilseeds', *Molecular Biology Reports*, 41(8), pp. 5077–5087. doi: 10.1007/s11033-014-3373-5.

Desikan, R. *et al.* (2006) 'Ethylene-induced stomatal closure in Arabidopsis occurs via AtrbohF-mediated hydrogen peroxide synthesis', *Plant Journal*, 47(6), pp. 907–916. doi: 10.1111/j.1365-313X.2006.02842.x.

Van Dijk, E. L., Jaszczyszyn, Y. and Thermes, C. (2014) 'Library preparation methods for next-generation sequencing: Tone down the bias', *Experimental Cell Research*. Elsevier, 322(1), pp. 12–20. doi: 10.1016/j.yexcr.2014.01.008.

Dinakar, C. and Bartels, D. (2013) 'Desiccation tolerance in resurrection plants: New insights from transcriptome, proteome, and metabolome analysis', *Frontiers in Plant Science*, 4(NOV), pp. 1–14. doi: 10.3389/fpls.2013.00482.

Ding, Y. *et al.* (2012) 'Identification of the major functional proteins of prokaryotic lipid droplets', *Journal of Lipid Research*, 53(3), pp. 399–411. doi: 10.1194/jlr.M021899.

Dobin, A. *et al.* (2016) 'Mapping RNA-seq with STAR', *Curr Protoc Bioinformatics*, 51(4), pp. 586–597. doi: 10.1002/0471250953.bi1114s51.Mapping.

Doll, S. *et al.* (2019) 'FSP1 is a glutathione-independent ferroptosis suppressor', *Nature*, 575(7784), pp. 693–698. doi: 10.1038/s41586-019-1707-0.

Dominguez-Solis, J. R. *et al.* (2008) 'A cyclophilin links redox and light signals to cysteine biosynthesis and stress responses in chloroplasts', *Proceedings of the National Academy of Sciences of the United States of America*, 105(42), pp. 16386–16391. doi: 10.1073/pnas.0808204105.

Dong, N. Q. *et al.* (2020) 'UDP-glucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice', *Nature Communications*, 11(1), pp. 1–16. doi: 10.1038/s41467-020-16403-5.

Du, B. and Rennenberg, H. (2018) 'Physiological responses of lavender (Lavandula angustifolia Mill.) towater deficit and recovery', *South African Journal of Botany*, 119, pp. 212–218. doi: 10.1016/j.sajb.2018.09.002.

Eddy, S. R. (1998) 'Profile hidden Markov models', *Bioinformatics*, 14(9), pp. 755–763. doi: 10.1093/bioinformatics/14.9.755.

Eddy, S. R. (2011) 'Accelerated profile HMM searches', *PLoS Computational Biology*, 7(10). doi: 10.1371/journal.pcbi.1002195.

Espinoza-Corral, R., Schwenkert, S. and Lundquist, P. K. (2021) 'Molecular changes of Arabidopsis thaliana plastoglobules facilitate thylakoid membrane remodeling under high light stress', *Plant Journal*, 106(6), pp. 1571–1587. doi: 10.1111/tpj.15253.

Ewels, P. *et al.* (2016) 'MultiQC: Summarize analysis results for multiple tools and samples in a single report', *Bioinformatics*, 32(19), pp. 3047–3048. doi: 10.1093/bioinformatics/btw354.

Fait, A. *et al.* (2008) 'Highway or byway: the metabolic role of the GABA shunt in plants', *Trends in Plant Science*, 13(1), pp. 14–19. doi: 10.1016/j.tplants.2007.10.005.

Fallard, A. *et al.* (2018) 'Compatible solutes and metabolites accumulation does not explain partial desiccation tolerance in Hymenoglossum cruentum and Hymenophyllum dentatum (Hymenophyllaceae) two filmy ferns with contrasting vertical distribution', *Environmental and Experimental Botany*. Elsevier B.V., 150, pp. 272–279. doi: 10.1016/j.envexpbot.2018.02.002.

Farrant, J. M. *et al.* (1999) 'The effect of drying rate on the survival of three desiccation-tolerant angiosperm species', *Annals of Botany*, 84(3), pp. 371–379. doi: 10.1006/anbo.1999.0927.

Farrant, J. M. (2000) 'A comparison of mechanisms of desiccation tolerance among three angiosperm resurrection plant species', *Plant Ecology*. doi: 10.1023/A:1026534305831.

Farrant, J. M. *et al.* (2003) 'An investigation into the role of light during desiccation of three angiosperm resurrection plants', *Plant, Cell and Environment*, 26(8), pp. 1275–1286. doi: 10.1046/j.0016-8025.2003.01052.x.

Farrant, J. M. (2008) 'Mechanisms of Desiccation Tolerance in Angiosperm Resurrection Plants', in *Plant Desiccation Tolerance*. doi: 10.1002/9780470376881.ch3.

Farrant, Jill M. *et al.* (2009) 'Desiccation tolerance in the vegetative tissues of the fern Mohria caffrorum is seasonally regulated', *Plant Journal*, 57(1), pp. 65–79. doi: 10.1111/j.1365-313X.2008.03673.x.

Farrant, Jill M *et al.* (2009) 'Desiccation tolerance in the vegetative tissues of the fern Mohria caffrorum is seasonally regulated', pp. 65–79. doi: 10.1111/j.1365-313X.2008.03673.x.

Farrant, J. M. and Hilhorst, H. (2022) 'Crops for dry environments', *Current Opinion in Biotechnology*. Elsevier Ltd, 74, pp. 84–91. doi: 10.1016/j.copbio.2021.10.026.

Farrant, J. M. and Hilhorst, H. W. M. (2021) 'What is dry? Exploring metabolism and molecular mobility at extremely low water contents', *Journal of Experimental Botany*, 72(5), pp. 1507–1510. doi: 10.1093/jxb/eraa579.

Farrant, J. M. and Moore, J. P. (2011) 'Programming desiccation-tolerance: From plants to seeds to resurrection plants', *Current Opinion in Plant Biology*, 14(3), pp. 340–345. doi: 10.1016/j.pbi.2011.03.018.

Fernández-Marín, B. *et al.* (2013) 'Evidence for the absence of enzymatic reactions in the glassy state. A case study of xanthophyll cycle pigments in the desiccation-tolerant moss Syntrichia ruralis', *Journal of Experimental Botany*, 64(10), pp. 3033–3043. doi: 10.1093/jxb/ert145.

Fernández-Marín, B., Holzinger, A. and García-Plazaola, J. I. (2016) 'Photosynthetic strategies of desiccation-tolerant organisms', *Handbook of Photosynthesis, Third Edition*, (March), pp. 663–681. doi: 10.1201/9781315372136-36.

Fernando, V. C. D. and Schroeder, D. F. (2016) 'Role of ABA in Arabidopsis Salt, Drought, and Desiccation Tolerance', *Abiotic and Biotic Stress in Plants - Recent Advances and Future Perspectives*. doi: 10.5772/61957.

Fidler, J. *et al.* (2022) 'PYR/PYL/RCAR Receptors Play a Vital Role in the Abscisic-Acid-Dependent Responses of Plants to External or Internal Stimuli', *Cells*, 11(8). doi: 10.3390/cells11081352.

Fisher, M. *et al.* (2015) 'Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa: Determinants of adoption in eastern and southern Africa', *Climatic Change*. doi: 10.1007/s10584-015-1459-2.

Foka, I. C. K. *et al.* (2020) 'The emerging roles of diacylglycerol kinase (DGK) in plant stress tolerance, growth, and development', *Agronomy*, 10(9), pp. 1–26. doi: 10.3390/agronomy10091375.

Freedman, A. H., Clamp, M. and Sackton, T. B. (2021) 'Error, noise and bias in de novo transcriptome assemblies', *Molecular Ecology Resources*, 21(1), pp. 18–29. doi: 10.1111/1755-0998.13156.

Fu, L. *et al.* (2012) 'CD-HIT: Accelerated for clustering the next-generation sequencing data', *Bioinformatics*, 28(23), pp. 3150–3152. doi: 10.1093/bioinformatics/bts565.

Fujimoto, S. Y. *et al.* (2000) 'Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression', *Plant Cell*,

12(3), pp. 393–404. doi: 10.1105/tpc.12.3.393.

Gabier, H. *et al.* (2021) 'A label-free proteomic and complementary metabolomic analysis of leaves of the resurrection plant xerophyta schlechteri during dehydration', *Life*, 11(11). doi: 10.3390/life11111242.

Gaff, D. F. and Loveys, B. R. (1984) 'Abscisic acid content and effects during dehydration of detached leaves of desiccation tolerant plants', *Journal of Experimental Botany*. doi: 10.1093/jxb/35.9.1350.

Gaff, D. F. and Oliver, M. (2013a) 'The evolution of desiccation tolerance in angiosperm plants: A rare yet common phenomenon', *Functional Plant Biology*, 40(4), pp. 315–328. doi: 10.1071/FP12321.

Gaff, D. F. and Oliver, M. (2013b) 'The evolution of desiccation tolerance in angiosperm plants: A rare yet common phenomenon', *Functional Plant Biology*. doi: 10.1071/FP12321.

Galganski, L., Urbanek, M. O. and Krzyzosiak, W. J. (2017) 'Nuclear speckles: Molecular organization, biological function and role in disease', *Nucleic Acids Research*. Oxford University Press, 45(18), pp. 10350–10368. doi: 10.1093/nar/gkx759.

Gao, L. and Xiang, C. Bin (2008) 'The genetic locus At1g73660 encodes a putative MAPKKK and negatively regulates salt tolerance in Arabidopsis', *Plant Molecular Biology*, 67(1–2), pp. 125–134. doi: 10.1007/s11103-008-9306-8.

Gasulla, F. *et al.* (2013) 'The role of lipid metabolism in the acquisition of desiccation tolerance in Craterostigma plantagineum: A comparative approach', *Plant Journal*, 75(5), pp. 726–741. doi: 10.1111/tpj.12241.

Ge, T. *et al.* (2020) 'The Role of the Pentose Phosphate Pathway in Diabetes and Cancer', *Frontiers in Endocrinology*, 11(June), pp. 1–11. doi: 10.3389/fendo.2020.00365.

Gechev, T. S. *et al.* (2012) 'Molecular mechanisms of desiccation tolerance in resurrection plants', *Cellular and Molecular Life Sciences*, 69(19), pp. 3175–3186. doi: 10.1007/s00018-012-1088-0.

Gechev, T. S. *et al.* (2013) 'Molecular mechanisms of desiccation tolerance in the resurrection glacial relic Haberlea rhodopensis', *Cellular and Molecular Life Sciences*, 70(4), pp. 689–709. doi: 10.1007/s00018-012-1155-6.

Georgieva, K. *et al.* (2017) 'Alterations in the sugar metabolism and in the vacuolar system of mesophyll cells contribute to the desiccation tolerance of Haberlea rhodopensis ecotypes', *Protoplasma*. Protoplasma, 254(1), pp. 193–201. doi: 10.1007/s00709-015-0932-0.

Giarola, V., Hou, Q. and Bartels, D. (2017) 'Angiosperm Plant Desiccation Tolerance: Hints from Transcriptomics and Genome Sequencing', *Trends in Plant Science*. doi: 10.1016/j.tplants.2017.05.007.

Gonc, C. *et al.* (2018) 'Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage', pp. 1–28.

Grabherr, M. G. *et al.* (2011) 'Full-length transcriptome assembly from RNA-Seq data

without a reference genome', (April). doi: 10.1038/nbt.1883.

Gribskov, M., McLachlan, A. D. and Eisenberg, D. (1987) 'Profile analysis: detection of distantly related proteins.', *Proceedings of the National Academy of Sciences of the United States of America*, 84(13), pp. 4355–4358. doi: 10.1073/pnas.84.13.4355.

Gu, Z. (2022) 'Complex heatmap visualization', *iMeta*, 1(3), pp. 1–15. doi: 10.1002/imt2.43.

Gu, Z., Eils, R. and Schlesner, M. (2016) 'Complex heatmaps reveal patterns and correlations in multidimensional genomic data', *Bioinformatics*, 32(18), pp. 2847–2849. doi: 10.1093/bioinformatics/btw313.

Gudjonsson, T. *et al.* (2012) 'TRIP12 and UBR5 suppress spreading of chromatin ubiquitylation at damaged chromosomes', *Cell*. Elsevier, 150(4), pp. 697–709. doi: 10.1016/j.cell.2012.06.039.

Guo, M., Liu, J. H., Ma, X., Zhai, Y. F., *et al.* (2016) 'Genome-wide analysis of the Hsp70 family genes in pepper (Capsicum annuum L.) and functional identification of CaHsp70-2 involvement in heat stress', *Plant Science*. Elsevier Ireland Ltd, 252, pp. 246–256. doi: 10.1016/j.plantsci.2016.07.001.

Guo, M., Liu, J. H., Ma, X., Luo, D. X., *et al.* (2016) 'The plant heat stress transcription factors (HSFS): Structure, regulation, and function in response to abiotic stresses', *Frontiers in Plant Science*, 7(FEB2016). doi: 10.3389/fpls.2016.00114.

Gupta, I. *et al.* (2018) 'Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells', *Nature Biotechnology*, 36(12), pp. 1197–1202. doi: 10.1038/nbt.4259.

Gutierrez, L. *et al.* (2007) 'Combined networks regulating seed maturation', *Trends in Plant Science*, 12(7), pp. 294–300. doi: 10.1016/j.tplants.2007.06.003.

Haas, B. J. *et al.* (no date) 'De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis'. doi: 10.1038/nprot.2013.084.

Hackl, T. *et al.* (2014) 'Proovread: Large-scale high-accuracy PacBio correction through iterative short read consensus', *Bioinformatics*, 30(21), pp. 3004–3011. doi: 10.1093/bioinformatics/btu392.

Han, J. D. *et al.* (2017) 'Evolutionary analysis of the LAFL genes involved in the land plant seed maturation program', *Frontiers in Plant Science*, 8(April), pp. 1–11. doi: 10.3389/fpls.2017.00439.

Han, Y. *et al.* (2015) 'Advanced applications of RNA sequencing and challenges', *Bioinformatics and Biology Insights*, 9, pp. 29–46. doi: 10.4137/BBI.S28991.

Hansen, K. D., Brenner, S. E. and Dudoit, S. (2010) 'Biases in Illumina transcriptome sequencing caused by random hexamer priming', *Nucleic Acids Research*, 38(12), pp. 1–7. doi: 10.1093/nar/gkq224.

Harvesting, L., Photosynthetic, A. and Duarte, C. M. (2016) 'Light Harvesting Among Photosynthetic Organisms Author (s): S . Agustí , S . Enríquez , H . Frost-Christensen , K .

Sand-Jensen and C . M . Duarte Published by : British Ecological Society Stable URL : http://www.jstor.org/stable/2389911 Accessed : 14-0', 8(2), pp. 273–279.

Haupt, W. (1990) 'Phytochrome-Mediated Fern-Spore Germination: Inhibition By Elevated Temperatures', *Photochemistry and Photobiology*, 52(1), pp. 57–59. doi: 10.1111/j.1751-1097.1990.tb01755.x.

Hietz, P. (2010) *Fern adaptations to xeric environments, Fern Ecology*. doi: 10.1017/CBO9780511844898.006.

Hilhorst, H. W. M. and Farrant, J. M. (2018) 'Plant Desiccation Tolerance: A Survival Strategy with Exceptional Prospects for Climate-Smart Agriculture', *Annual Plant Reviews online*, 1, pp. 327–354. doi: 10.1002/9781119312994.apr0637.

Holmlund, H. I. *et al.* (2019) 'High-resolution computed tomography reveals dynamics of desiccation and rehydration in fern petioles of a desiccation-tolerant fern', *New Phytologist*, 224(1), pp. 97–105. doi: 10.1111/nph.16067.

Holmlund, H. I. *et al.* (2020a) 'Positive root pressure is critical for whole-plant desiccation recovery in two species of terrestrial resurrection ferns', *Journal of Experimental Botany*, 71(3), pp. 1139–1150. doi: 10.1093/jxb/erz472.

Holmlund, H. I. *et al.* (2020b) 'Positive root pressure is critical for whole-plant desiccation recovery in two species of terrestrial resurrection ferns', *Journal of Experimental Botany*, 71(3), pp. 1139–1150. doi: 10.1093/jxb/erz472.

Hölzer, M. and Marz, M. (2019) 'De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers', *GigaScience*, 8(5), pp. 1–16. doi: 10.1093/gigascience/giz039.

Hong, Y. *et al.* (2022) 'First Multi-Organ Full-Length Transcriptome of Tree Fern Alsophila spinulosa Highlights the Stress-Resistant and Light-Adapted Genes', *Frontiers in Genetics*, 12(February), pp. 1–21. doi: 10.3389/fgene.2021.784546.

Hoo, S. C. and Howe, G. A. (2009) 'A critical role for the TIFY motif in repression of jasmonate signaling by a stabilized splice variant of the JASMONATE ZIM-domain protein JAZ10 in Arabidopsis', *Plant Cell*, 21(1), pp. 131–145. doi: 10.1105/tpc.108.064097.

Hossain, M. A. *et al.* (2015) 'Hydrogen peroxide priming modulates abiotic oxidative stress tolerance: Insights from ROS detoxification and scavenging', *Frontiers in Plant Science*, 6(June), pp. 1–19. doi: 10.3389/fpls.2015.00420.

Hou, C. J. and Yang, C. H. (2016) 'Comparative analysis of the pteridophyte Adiantum MFT ortholog reveals the specificity of combined FT/MFT C and N terminal interaction with FD for the regulation of the downstream gene AP1', *Plant Molecular Biology*. Springer Netherlands, 91(4–5), pp. 563–579. doi: 10.1007/s11103-016-0489-0.

Hrdlickova, R., Toloue, M. and Tian, B. (2017) 'RNA-Seq methods for transcriptome analysis', *Wiley Interdisciplinary Reviews: RNA*, 8(1). doi: 10.1002/wrna.1364.

Hsu, F. C. *et al.* (2013) 'Submergence confers immunity mediated by the WRKY22 transcription factor in Arabidopsis', *Plant Cell*, 25(7), pp. 2699–2713. doi:

10.1105/tpc.113.114447.

Hu, W. *et al.* (2016) 'Morphological, physiological and proteomic analyses provide insights into the improvement of castor bean productivity of a dwarf variety in comparing with a high-stalk variety', *Frontiers in Plant Science*, 7(September2016), pp. 1–12. doi: 10.3389/fpls.2016.01473.

Huang, H. Y. and Cheng, Y. S. (2019) 'Heterologous overexpression, purification and functional analysis of plant cellulose synthase from green bamboo', *Plant Methods*. BioMed Central, 15(1), pp. 1–12. doi: 10.1186/s13007-019-0466-0.

Huang, Y. *et al.* (2016) 'Arabidopsis WRKY6 Transcription Factor Acts as a Positive Regulator of Abscisic Acid Signaling during Seed Germination and Early Seedling Development', *PLoS Genetics*, 12(2), pp. 1–22. doi: 10.1371/journal.pgen.1005833.

Hwang, J. U. *et al.* (2016) 'Plant ABC Transporters Enable Many Unique Aspects of a Terrestrial Plant's Lifestyle', *Molecular Plant*. Elsevier Ltd, 9(3), pp. 338–355. doi: 10.1016/j.molp.2016.02.003.

Illing, N. *et al.* (2005) 'The signature of seeds in resurrection plants: A molecular and physiological comparison of desiccation tolerance in seeds and vegetative tissues', *Integrative and Comparative Biology*, 45(5), pp. 771–787. doi: 10.1093/icb/45.5.771.

Ingle, R. A. *et al.* (2007) 'Proteomic analysis of leaf proteins during dehydration of the resurrection plant Xerophyta viscosa', *Plant, Cell and Environment*. doi: 10.1111/j.1365-3040.2006.01631.x.

Inoue, H. *et al.* (2000) 'Characterization of a protease that acts specifically on the 22-kDa protein in thylakoid membranes from green spores of the fern Osmunda japonica', *Physiologia Plantarum*, 109(2), pp. 129–136. doi: 10.1034/j.1399-3054.2000.100204.x.

Jain, M. *et al.* (2016) 'The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community', *Genome Biology*. Genome Biology, 17(1), pp. 1–11. doi: 10.1186/s13059-016-1103-0.

Jeong, H. *et al.* (2017) 'Mechanistic insight into the nucleus-vacuole junction based on the Vac8p-Nvj1p crystal structure', *Proceedings of the National Academy of Sciences of the United States of America*, 114(23), pp. E4539–E4548. doi: 10.1073/pnas.1701030114.

Jin, Q. Y. *et al.* (2016) 'Identification and characterization of differentially expressed miRNAs between bamboo shoot and rhizome shoot', *Journal of Plant Biology*, 59(4), pp. 322–335. doi: 10.1007/s12374-015-0581-z.

John, S. P. and Hasenstein, K. H. (2017) 'The role of peltate scales in desiccation tolerance of Pleopeltis polypodioides', *Planta*. Springer Berlin Heidelberg, 245(1), pp. 207–220. doi: 10.1007/s00425-016-2631-2.

Kanemori, M., Yanagi, H. and Yura, T. (1999) 'Marked instability of the σ32 heat shock transcription factor at high temperature. Implications for heat shock regulation', *Journal of Biological Chemistry*. Â© 1999 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 274(31), pp. 22002–22007. doi: 10.1074/jbc.274.31.22002.

Kang, C. H. *et al.* (2006) 'AtBAG6, a novel calmodulin-binding protein, induces programmed cell death in yeast and plants', *Cell Death and Differentiation*, 13(1), pp. 84–95. doi: 10.1038/sj.cdd.4401712.

Kanitz, A. *et al.* (2015) 'Comparative assessment of methods for the computational inference of transcript isoform abundance from RNA-seq data', *Genome Biology*. Genome Biology, 16(1), pp. 1–26. doi: 10.1186/s13059-015-0702-5.

Kent, W. J. (2002) ' BLAT —The BLAST -Like Alignment Tool ', *Genome Research*, 12(4), pp. 656–664. doi: 10.1101/gr.229202.

Khandelwal, A. *et al.* (2010) 'Role of ABA and ABI3 in desiccation tolerance', *Science*, 327(5965), p. 546. doi: 10.1126/science.1183672.

Kim, J., Lee, S. B. and Suh, M. C. (2021) 'Arabidopsis 3-Ketoacyl-CoA Synthase 4 is Essential for Root and Pollen Tube Growth', *Journal of Plant Biology*. Springer Berlin Heidelberg, 64(2), pp. 155–165. doi: 10.1007/s12374-020-09288-w.

Kim, L. J. *et al.* (2021) 'Ferroportin 3 is a dual-targeted mitochondrial/chloroplast iron exporter necessary for iron homeostasis in Arabidopsis', *Plant Journal*, 107(1), pp. 215–236. doi: 10.1111/tpj.15286.

Kim, Y. D. *et al.* (2011) 'NSrp70 is a novel nuclear speckle-related protein that modulates alternative pre-mRNA splicing in vivo', *Nucleic Acids Research*, 39(10), pp. 4300–4314. doi: 10.1093/nar/gkq1267.

Kirch, H. H., Nair, A. and Bartels, D. (2001) 'Novel ABA- and dehydration-inducible aldehyde dehydrogenase genes isolated from the resurrection plant Craterostigma plantagineum and Arabidopsis thaliana', *Plant Journal*, 28(5), pp. 555–567. doi: 10.1046/j.1365-313X.2001.01176.x.

Knight, C. A. *et al.* (2006) 'Expression profiling and local adaptation of Boechera holboellii populations for water use efficiency across a naturally occurring water stress gradient', *Molecular Ecology*, 15(5), pp. 1229–1237. doi: 10.1111/j.1365-294X.2006.02818.x.

Kobayashi, K. *et al.* (2017) 'Shoot removal induces chloroplast development in roots via cytokinin signaling', *Plant Physiology*, 173(4), pp. 2340–2355. doi: 10.1104/pp.16.01368.

Kobayashi, Y. *et al.* (2004) 'Differential activation of the rice sucrose nonfermenting1-related protein kinase2 family by hyperosmotic stress and abscisic acid', *Plant Cell*, 16(5), pp. 1163–1177. doi: 10.1105/tpc.019943.

Komiya, R., Yokoi, S. and Shimamoto, K. (2009) 'A gene network for long-day flowering activates RFT1 encoding a mobile flowering signal in rice', *Development*, 136(20), pp. 3443–3450. doi: 10.1242/dev.040170.

Krattinger, S. G. *et al.* (2019) 'Abscisic acid is a substrate of the ABC transporter encoded by the durable wheat disease resistance gene Lr34', *New Phytologist*, 223(2), pp. 853–866. doi: 10.1111/nph.15815.

Kumar, M. *et al.* (2019) 'Integration of abscisic acid signaling with other signaling pathways in plant stress responses and development', *Plants*, 8(12), pp. 1–20. doi:

10.3390/plants8120592.

Kunkel, W. G. (2021) ‘De Novo Sporophyte Transcriptome Assembly and Functional Annotation in the Endangered Fern Species Vandenboschia’.

Kuo, R. I. *et al.* (2019) ‘Illuminating the dark side of the human transcriptome with TAMA Iso-Seq analysis’, *bioRxiv*. BMC Genomics, pp. 1–22. doi: 10.1101/780015.

Lahens, N. F. *et al.* (2014) ‘IVT-seq reveals extreme bias in RNA sequencing’, *Genome Biology*, 15(6), pp. 1–15. doi: 10.1186/gb-2014-15-6-r86.

Langfelder, P. and Horvath, S. (2008) ‘WGCNA: An R package for weighted correlation network analysis’, *BMC Bioinformatics*, 9. doi: 10.1186/1471-2105-9-559.

Langmead, B. *et al.* (2019) ‘Sequence analysis Scaling read aligners to hundreds of threads on general-purpose processors’, 35(July 2018), pp. 421–432. doi: 10.1093/bioinformatics/bty648.

Langmead, B. and Salzberg, S. L. (2012) ‘Fast gapped-read alignment with Bowtie 2’, 9(4), pp. 357–360. doi: 10.1038/nmeth.1923.

Lara, J. A. *et al.* (2018) ‘Identification and characterization of sterol acyltransferases responsible for steryl ester biosynthesis in tomato’, *Frontiers in Plant Science*, 9(May), pp. 1–18. doi: 10.3389/fpls.2018.00588.

Le, T. N. *et al.* (2007) ‘Desiccation-tolerance specific gene expression in leaf tissue of the resurrection plant Sporobolus stapfianus’, *Functional Plant Biology*, 34(7), pp. 589–600. doi: 10.1071/FP06231.

Le, T. N. and McQueen-Mason, S. J. (2006) ‘Desiccation-tolerant plants in dry environments’, *Reviews in Environmental Science and Biotechnology*, 5(2–3), pp. 269–279. doi: 10.1007/s11157-006-0015-y.

Lee, G. J., Pokala, N. and Vierling, E. (1995) ‘Structure and in vitro molecular chaperone activity of cytosolic small heat shock proteins from pea’, *Journal of Biological Chemistry*, 270(18), pp. 10432–10438. doi: 10.1074/jbc.270.18.10432.

Lee, J. W. *et al.* (2020) ‘Early Light-Inducible Protein (ELIP) Can Enhance Resistance to Cold-Induced Photooxidative Stress in Chlamydomonas reinhardtii’, *Frontiers in Physiology*, 11(August), pp. 1–15. doi: 10.3389/fphys.2020.01083.

Lee, S. *et al.* (2012) ‘A NAC transcription factor NTL4 promotes reactive oxygen species production during drought-induced leaf senescence in Arabidopsis’, *Plant Journal*, 70(5), pp. 831–844. doi: 10.1111/j.1365-313X.2012.04932.x.

Leprince, O. and Buitink, J. (2015) ‘Introduction to desiccation biology: from old borders to new frontiers’, *Planta*. Springer Berlin Heidelberg, 242(2), pp. 369–378. doi: 10.1007/s00425-015-2357-6.

Li, F. *et al.* (2014) ‘Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns’, 111(18). doi: 10.1073/pnas.1319929111.

Li, G. *et al.* (2012) 'Dual-level regulation of ACC synthase activity by MPK3/MPK6 cascade and its downstream WRKY transcription factor during ethylene induction in arabidopsis', *PLoS Genetics*, 8(6). doi: 10.1371/journal.pgen.1002767.

Li, H. (2018) 'Minimap2: Pairwise alignment for nucleotide sequences', *Bioinformatics*, 34(18), pp. 3094–3100. doi: 10.1093/bioinformatics/bty191.

Li, J. *et al.* (2020) 'Ferroptosis: past, present and future', *Cell Death and Disease*. Springer US, 11(2). doi: 10.1038/s41419-020-2298-2.

Li, N. *et al.* (2021) 'Plant Hormone-Mediated Regulation of Heat Tolerance in Response to Global Climate Change', *Frontiers in Plant Science*, 11(February), pp. 1–11. doi: 10.3389/fpls.2020.627969.

Li, Y. *et al.* (2022) 'Structure of the Arabidopsis guard cell anion channel SLAC1 suggests activation mechanism by phosphorylation', *Nature Communications*. Springer US, 13(1), pp. 1–9. doi: 10.1038/s41467-022-30253-3.

Liang, X. *et al.* (2013) 'Proline mechanisms of stress survival', *Antioxidants and Redox Signaling*, 19(9), pp. 998–1011. doi: 10.1089/ars.2012.5074.

Lin, R. and Wang, H. (2005) 'Two homologous ATP-binding cassette transporter proteins, AtMDR1 and AtPGP1, regulate Arabidopsis photomorphogenesis and root development by mediating polar auxin transport', *Plant Physiology*, 138(2), pp. 949–964. doi: 10.1104/pp.105.061572.

Liu, H. C., Liao, H. T. and Charng, Y. Y. (2011) 'The role of class A1 heat shock factors (HSFA1s) in response to heat and other stresses in Arabidopsis', *Plant, Cell and Environment*, 34(5), pp. 738–751. doi: 10.1111/j.1365-3040.2011.02278.x.

Liu, M. Sen, Chien, C. Te and Lin, T. P. (2008) 'Constitutive components and induced gene expression are involved in the desiccation tolerance of Selaginella tamariscina', *Plant and Cell Physiology*, 49(4), pp. 653–663. doi: 10.1093/pcp/pcn040.

Lobell, D. B., Schlenker, W. and Costa-Roberts, J. (2011) 'Climate trends and global crop production since 1980', *Science*. doi: 10.1126/science.1204531.

Logsdon, G. A., Vollger, M. R. and Eichler, E. E. (2020) 'Long-read human genome sequencing and its applications', *Nature Reviews Genetics*. Springer US, 21(10), pp. 597–614. doi: 10.1038/s41576-020-0236-x.

Love, M. I., Hogenesch, J. B. and Irizarry, R. A. (2015) 'Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation', pp. 1–31.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), pp. 1–21. doi: 10.1186/s13059-014-0550-8.

Lyall, R. and Gechev, T. (2020) *Multi-omics insights into the evolution of angiosperm resurrection plants, Annual Plant Reviews Online*. doi: 10.1002/9781119312994.apr0730.

Lyon, B. (2009) 'Southern Africa summer drought and heat waves: Observations and coupled

model behavior', *Journal of Climate*. doi: 10.1175/2009JCLI3101.1.

Malik, L., Almodaresi, F. and Patro, R. (2018) 'Grouper: Graph-based clustering and annotation for improved de novo transcriptome analysis', *Bioinformatics*, 34(19), pp. 3265–3272. doi: 10.1093/bioinformatics/bty378.

Mangisoni, J. H. (2008) 'Impact of treadle pump irrigation technology on smallholder poverty and food security in Malawi: A case study of Blantyre and Mchinji districts', *International Journal of Agricultural Sustainability*. doi: 10.3763/ijas.2008.0306.

Manna, S. (2015) 'An overview of pentatricopeptide repeat proteins and their applications', *Biochimie*. Elsevier B.V, 113, pp. 93–99. doi: 10.1016/j.biochi.2015.04.004.

Mantione, K. J. *et al.* (2014) 'Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq', *Medical science monitor basic research*, 20, pp. 138–142. doi: 10.12659/MSMBR.892101.

Marchant, D. B. *et al.* (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. doi: 10.1038/s41598-019-53968-8.

Mardis, E. R. (2013) 'Next-generation sequencing platforms', *Annual Review of Analytical Chemistry*, 6, pp. 287–303. doi: 10.1146/annurev-anchem-062012-092628.

Marengo, M. *et al.* (2022) 'Towards a metabolomic approach to investigate iron–sulfur cluster biogenesis', *IUBMB Life*, 74(7), pp. 715–722. doi: 10.1002/iub.2618.

Markham, J. E. *et al.* (2011) 'Sphingolipids containing very-long-chain fatty acids define a secretory pathway for specific polar plasma membrane protein targeting in arabidopsis', *Plant Cell*, 23(6), pp. 2362–2378. doi: 10.1105/tpc.110.080473.

Marks, R. A. *et al.* (2021) 'Unexplored dimensions of variability in vegetative desiccation tolerance', *American Journal of Botany*, 108(2), pp. 346–358. doi: 10.1002/ajb2.1588.

Martinelli, T. *et al.* (2007) 'Amino acid pattern and glutamate metabolism during dehydration stress in the "resurrection" plant Sporobolus stapfianus: A comparison between desiccation-sensitive and desiccation-tolerant leaves', *Journal of Experimental Botany*. doi: 10.1093/jxb/erm161.

Martínez-Vilalta, J. *et al.* (2016) 'Dynamics of non-structural carbohydrates in terrestrial plants: A global synthesis', *Ecological Monographs*, 86(4), pp. 495–516. doi: 10.1002/ecm.1231.

Martinis, J. *et al.* (2014) 'ABC1K1/PGR6 kinase: A regulatory link between photosynthetic activity and chloroplast metabolism', *Plant Journal*, 77(2), pp. 269–283. doi: 10.1111/tpj.12385.

Masih, I. *et al.* (2014) 'A review of droughts on the African continent: A geospatial and long-term perspective', *Hydrology and Earth System Sciences*. doi: 10.5194/hess-18-3635-2014.

Masoomi-Aladizgeh, F. *et al.* (2021) 'Pollen development in cotton (Gossypium hirsutum) is highly sensitive to heat exposure during the tetrad stage', *Plant Cell and Environment*, 44(7),

pp. 2150–2166. doi: 10.1111/pce.13908.

Matos, A. R. *et al.* (2007) 'Alternative oxidase involvement in cold stress response of Arabidopsis thaliana fad2 and FAD3+ cell suspensions altered in membrane lipid composition', *Plant and Cell Physiology*, 48(6), pp. 856–865. doi: 10.1093/pcp/pcm061.

Matsuura, H. N., Rau, M. R. and Fett-Neto, A. G. (2014) 'Oxidative stress and production of bioactive monoterpene indole alkaloids: Biotechnological implications', *Biotechnology Letters*, 36(2), pp. 191–200. doi: 10.1007/s10529-013-1348-6.

McCombie, W. R., McPherson, J. D. and Mardis, E. R. (2019) 'Next-generation sequencing technologies', *Cold Spring Harbor Perspectives in Medicine*, 9(11). doi: 10.1101/cshperspect.a036798.

McManus, C. J. and Graveley, B. R. (2011) 'RNA structure and the mechanisms of alternative splicing', *Current Opinion in Genetics and Development*. Elsevier Ltd, 21(4), pp. 373–379. doi: 10.1016/j.gde.2011.04.001.

Mihailova, G., Vasileva, I., *et al.* (2022) 'Antioxidant Defense during Recovery of Resurrection', *Plants*, 11(175).

Mihailova, G., Christov, N. K., *et al.* (2022) 'Reactivation of the Photosynthetic Apparatus of Resurrection Plant Haberlea rhodopensis during the Early Phase of Recovery from Drought- and Freezing-Induced Desiccation', *Plants*, 11(17). doi: 10.3390/plants11172185.

Minamikawa, T., Koshiba, T. and Wada, M. (1984) 'Compositional changes in germinating spores of Adiantum capillus-veneris L.', *The Botanical Magazine Tokyo*, 97(3), pp. 313–322. doi: 10.1007/BF02488664.

Mkhize, K. G. W., Minibayeva, F. and Beckett, R. P. (2020) 'Induction of desiccation tolerance mechanisms occurs in both the fast-drying filmy fern Crepidomanes inopinatum and the slow-drying fern Loxogramme abyssinica', *South African Journal of Botany*. doi: 10.1016/j.sajb.2020.02.014.

Moore, J. P. *et al.* (2005) 'The predominant polyphenol in the leaves of the resurrection plant Myrothamnus flabellifolius, 3,4,5 tri-O-galloylquinic acid, protects membranes against desiccation and free radical-induced oxidation', *Biochemical Journal*, 385(1), pp. 301–308. doi: 10.1042/BJ20040499.

Moore, J. P. *et al.* (2007) 'Desiccation-induced ultrastructural and biochemical changes in the leaves of the resurrection plant Myrothamnus flabellifolia', *Australian Journal of Botany*, 55(4), pp. 482–491. doi: 10.1071/BT06172.

Moore, J. P. *et al.* (2009) 'Towards a systems-based understanding of plant desiccation tolerance', *Trends in Plant Science*, 14(2), pp. 110–117. doi: 10.1016/j.tplants.2008.11.007.

Moore, J. P. *et al.* (2013) 'Arabinose-rich polymers as an evolutionary strategy to plasticize resurrection plant cell walls against desiccation', *Planta*, 237(3), pp. 739–754. doi: 10.1007/s00425-012-1785-9.

Morse, M., Rafudeen, M. S. and Farrant, J. M. (2011) *An Overview of the Current Understanding of Desiccation Tolerance in the Vegetative Tissues of Higher Plants*. 1st edn,

Advances in Botanical Research. 1st edn. Elsevier Ltd. doi: 10.1016/B978-0-12-387692-8.00009-6.

Mossion, V. *et al.* (2020a) 'A reference transcriptome for the early-branching fern Botrychium lunaria enables fine-grained resolution of population structure', *bioRxiv*.

Mossion, V. *et al.* (2020b) 'A transcriptome for the early-branching fern Botrychium lunaria enables fine-grained resolution of population structure', pp. 1–29.

Mühr, L. S. A. *et al.* (2020) 'De novo sequence assembly requires bioinformatic checking of chimeric sequences', *PLoS ONE*, 15(8 July), pp. 1–11. doi: 10.1371/journal.pone.0237455.

Müller, M. and Munné-Bosch, S. (2015) 'Ethylene response factors: A key regulatory hub in hormone and stress signaling', *Plant Physiology*, 169(1), pp. 32–41. doi: 10.1104/pp.15.00677.

Mundree, S. G. *et al.* (2002) 'Physiological and molecular insights into drought tolerance', *African Journal of Biotechnology*, 1(2), pp. 1–19.

Myouga, F. *et al.* (2006) 'An Arabidopsis chloroplast-targeted Hsp101 homologue, APG6, has an essential role in chloroplast development as well as heat-stress response', *Plant Journal*, 48(2), pp. 249–260. doi: 10.1111/j.1365-313X.2006.02873.x.

Nachtigall, P. G., Kashiwabara, A. Y. and Durham, A. M. (2021) 'CodAn: Predictive models for precise identification of coding regions in eukaryotic transcripts', *Briefings in Bioinformatics*, 22(3), pp. 1–11. doi: 10.1093/bib/bbaa045.

Nadal, M. *et al.* (2021) 'Differences in biochemical, gas exchange and hydraulic response to water stress in desiccation tolerant and sensitive fronds of the fern Anemia caffrorum', *New Phytologist*, 231(4), pp. 1415–1430. doi: 10.1111/nph.17445.

Nakamura, T. *et al.* (2003) 'RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA', *European Journal of Biochemistry*, 270(20), pp. 4070–4081. doi: 10.1046/j.1432-1033.2003.03796.x.

Nakayama, T. *et al.* (2017) 'A peptide hormone required for Casparian strip diffusion barrier formation in Arabidopsis roots', *Science*, 355(6322), pp. 284–286. doi: 10.1126/science.aai9057.

Navari-Izzo, F. *et al.* (1995) 'Unusual composition of thylakoid membranes of the resurrection plant Boea hygroscopica: Changes in lipids upon dehydration and rehydration', *Physiologia Plantarum*, 94(1), pp. 135–142. doi: 10.1111/j.1399-3054.1995.tb00794.x.

Neeragunda Shivaraj, Y. *et al.* (2018) 'Perspectives on structural, physiological, cellular, and molecular responses to desiccation in resurrection plants', *Scientifica*, 2018. doi: 10.1155/2018/9464592.

Nhemachena, C. and Hassan, R. (2007) *Micro-Level Analysis of Farmers Adaption to climate change in Southern Africa*, *booksgooglecom*.

Niggeweg, R. *et al.* (2000) 'Tobacco transcription factor TGA2.2 is the main component of as-1- binding factor ASF-1 and is involved in salicylic acid- and auxin-inducible expression

of as-1-containing target promoters', *Journal of Biological Chemistry*. Â© 2000 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 275(26), pp. 19897–19905. doi: 10.1074/jbc.M909267199.

Nikkanen, L. *et al.* (2018) 'Regulation of cyclic electron flow by chloroplast NADPH-dependent thioredoxin system', *Plant Direct*, 2(11). doi: 10.1002/pld3.93.

Nikkanen, L. and Rintamäki, E. (2019) 'Chloroplast thioredoxin systems dynamically regulate photosynthesis in plants', *Biochemical Journal*, 476(7), pp. 1159–1172. doi: 10.1042/BCJ20180707.

Noh, B., Murphy, A. S. and Spalding, E. P. (2001) 'Multidrug resistance-like genes of Arabidopsis required for auxin transport and auxin-mediated development', *Plant Cell*, 13(11), pp. 2441–2454. doi: 10.1105/tpc.13.11.2441.

Nováková, P. *et al.* (2014) 'SAC phosphoinositide phosphatases at the tonoplast mediate vacuolar function in Arabidopsis', *Proceedings of the National Academy of Sciences of the United States of America*, 111(7), pp. 2818–2823. doi: 10.1073/pnas.1324264111.

O'Neil, D., Glowatz, H. and Schlumpberge, M. (2013) 'Ribosomal RNA depletion for efficient use of RNA-seq capacity', *Current Protocols in Molecular Biology*, (SUPPL.103), pp. 1–8. doi: 10.1002/0471142727.mb0419s103.

Oliver, M. J., Guo, L., *et al.* (2011) 'A sister group contrast using untargeted global metabolomic analysis delineates the biochemical regulation underlying desiccation tolerance in Sporobolus stapfianus', *Plant Cell*. doi: 10.1105/tpc.110.082800.

Oliver, M. J., Jain, R., *et al.* (2011) 'Proteome analysis of leaves of the desiccation-tolerant grass, Sporobolus stapfianus, in response to dehydration', *Phytochemistry*. doi: 10.1016/j.phytochem.2010.10.020.

Oliver, M. J. *et al.* (2020) 'Desiccation Tolerance: Avoiding Cellular Damage during Drying and Rehydration', *Annual Review of Plant Biology*, 71, pp. 435–460. doi: 10.1146/annurev-arplant-071219-105542.

Oliver, M. J., Tuba, Z. and Mishler, B. D. (2000) 'The evolution of vegetative desiccation tolerance in land plants', *Plant Ecology*, 151(1), pp. 85–100. doi: 10.1023/A:1026550808557.

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?', in *Integrative and Comparative Biology*. doi: 10.1093/icb/45.5.788.

Oliver, M. J., Wood, A. J. and O'Mahony, P. (1998) '"To dryness and beyond" - Preparation for the dried state and rehydration in vegetative desiccation-tolerant plants', *Plant Growth Regulation*, 24(3), pp. 193–201. doi: 10.1023/A:1005863015130.

Olvera-Carrillo, Y., Reyes, J. L. and Covarrubias, A. A. (2011) 'Late embryogenesis abundant proteins: Versatile players in the plant adaptation to water limiting environments', *Plant Signaling and Behavior*, 6(4), pp. 586–589. doi: 10.4161/psb.6.4.15042.

Ondov, B. D. *et al.* (2016) 'Mash: Fast genome and metagenome distance estimation using

MinHash', *Genome Biology*. Genome Biology, 17(1), pp. 1–14. doi: 10.1186/s13059-016-0997-x.

Pan, Q. *et al.* (2008) 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing', *Nature Genetics*, 40(12), pp. 1413–1415. doi: 10.1038/ng.259.

Panaro, N. J. *et al.* (2000) 'Evaluation of DNA fragment sizing and quantification by the Agilent 2100 bioanalyzer', *Clinical Chemistry*, 46(11), pp. 1851–1853. doi: 10.1093/clinchem/46.11.1851.

Pandey, V. *et al.* (2010) 'Desiccation-induced physiological and biochemical changes in resurrection plant, Selaginella bryopteris', *Journal of Plant Physiology*. Elsevier GmbH., 167(16), pp. 1351–1359. doi: 10.1016/j.jplph.2010.05.001.

Park, J. M. *et al.* (2001) 'Overexpression of the Tobacco Tsi1 Gene Encoding an EREBP/AP2-Type Transcription Factor Enhances Resistance against Pathogen Attack and Osmotic Stress in Tobacco', *The Plant Cell*, 13(5), p. 1035. doi: 10.2307/3871362.

Park, S. W. *et al.* (2013) 'Cyclophilin 20-3 relays a 12-oxo-phytodienoic acid signal during stress responsive regulation of cellular redox homeostasis', *Proceedings of the National Academy of Sciences of the United States of America*, 110(23), pp. 9559–9564. doi: 10.1073/pnas.1218872110.

Patro, R. *et al.* (2017) 'Salmon provides fast and bias-aware quantification of transcript expression', *Nature Publishing Group*. Nature Publishing Group, (march). doi: 10.1038/nmeth.4197.

Patzke, K. *et al.* (2019) 'The plastidic sugar transporter pSuT influences flowering and affects cold responses 1[OPEN]', *Plant Physiology*, 179(2), pp. 569–587. doi: 10.1104/pp.18.01036.

Peña, M. J. *et al.* (2012) 'A galacturonic acid-containing xyloglucan is involved in arabidopsis root hair tip growthw', *Plant Cell*, 24(11), pp. 4511–4524. doi: 10.1105/tpc.112.103390.

Peters, S. *et al.* (2007) 'Protection mechanisms in the resurrection plant Xerophyta viscosa (Baker): Both sucrose and raffinose family oligosaccharides (RFOs) accumulate in leaves in response to water deficit', *Journal of Experimental Botany*, 58(8), pp. 1947–1956. doi: 10.1093/jxb/erm056.

Petersen, R. (2016) 'North American Fungi', (October). doi: 10.2509/naf2014.009.003.

Pfister, A. *et al.* (2014) 'A receptor-like kinase mutant with absent endodermal diffusion barrier displays selective nutrient homeostasis defects', *eLife*, 3, p. e03115. doi: 10.7554/eLife.03115.

Philippe, G. *et al.* (2020) 'Cutin and suberin: assembly and origins of specialized lipidic cell wall scaffolds', *Current Opinion in Plant Biology*, 55, pp. 11–20. doi: 10.1016/j.pbi.2020.01.008.

Phillips, J. R. *et al.* (2008) 'Lindernia brevidens: A novel desiccation-tolerant vascular plant, endemic to ancient tropical rainforests', *Plant Journal*, 54(5), pp. 938–948. doi:

10.1111/j.1365-313X.2008.03478.x.

Picelli, S. *et al.* (2014) 'Full-length RNA-seq from single cells using Smart-seq2', *Nature Protocols*, 9(1), pp. 171–181. doi: 10.1038/nprot.2014.006.

Pitzschke, A. *et al.* (2009) 'VIP1 response elements mediate mitogen-activated protein kinase 3-induced stress gene expression', *Proceedings of the National Academy of Sciences of the United States of America*, 106(43), pp. 18414–18419. doi: 10.1073/pnas.0905599106.

Plancot, B. *et al.* (2019) 'Desiccation tolerance in plants: Structural characterization of the cell wall hemicellulosic polysaccharides in three Selaginella species', *Carbohydrate Polymers*. Elsevier Ltd., 208, pp. 180–190. doi: 10.1016/j.carbpol.2018.12.051.

Platt, K. A., Oliver, M. J. and Thomson, W. W. (1994) 'Membranes and organelles of dehydrated Selaginella and Tortula retain their normal configuration and structural integrity - Freeze fracture evidence', *Protoplasma*, 178(1–2), pp. 57–65. doi: 10.1007/BF01404121.

Podobnik, M., Kisovec, M. and Anderluh, G. (2017) 'Molecular mechanism of pore formation by aerolysin-like proteins', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1726). doi: 10.1098/rstb.2016.0209.

Polizel, A. M. *et al.* (2011) 'Molecular, anatomical and physiological properties of a genetically modified soybean line transformed with rd29A:AtDREB1A for the improvement of drought tolerance', *Genetics and Molecular Research*, 10(4), pp. 3641–3656. doi: 10.4238/2011.October.21.4.

Porembski, S. (2011) 'Evolution, Diversity, and Habitats of Poikilohydrous Vascular Plants', in. doi: 10.1007/978-3-642-19106-0_8.

Porembski, S. and Barthlott, W. (2000) 'Granitic and gneissic outcrops (inselbergs) as centers of diversity for desiccation-tolerant vascular plants', *Plant Ecology*. doi: 10.1023/A:1026565817218.

Prats, K. A. and Brodersen, C. R. (2021) 'Desiccation and rehydration dynamics in the epiphytic resurrection fern Pleopeltis polypodioides', *Plant Physiology*, 187(3), pp. 1501–1518. doi: 10.1093/plphys/kiab361.

PROCTOR, M. C. F. (1990) 'The physiological basis of bryophyte production', *Botanical Journal of the Linnean Society*. doi: 10.1111/j.1095-8339.1990.tb02211.x.

Proctor, M. C. F. and Pence, V. C. (2009) 'Vegetative tissues: bryophytes, vascular resurrection plants and vegetative propagules.', in *Desiccation and survival in plants: drying without dying*. doi: 10.1079/9780851995342.0207.

Puglisi, R. and Pastore, A. (2018) 'The role of chaperones in iron–sulfur cluster biogenesis', *FEBS Letters*, 592(24), pp. 4011–4019. doi: 10.1002/1873-3468.13245.

Qing, T. *et al.* (2013) 'mRNA enrichment protocols determine the quantification characteristics of external RNA spike-in controls in RNA-Seq studies', *Science China Life Sciences*, 56(2), pp. 134–142. doi: 10.1007/s11427-013-4437-9.

Quartacci, M. F. *et al.* (1997) 'Desiccation-tolerant Sporobolus stapfianus: Lipid composition

and cellular ultrastructure during dehydration and rehydration', *Journal of Experimental Botany*, 48(311), pp. 1269–1279. doi: 10.1093/jxb/48.6.1269.

Racolta, A., Bryan, A. C. and Tax, F. E. (2014) 'The receptor-like kinases GSO1 and GSO2 together regulate root growth in arabidopsis through control of cell division and cell fate specification', *Developmental Dynamics*, 243(2), pp. 257–278. doi: 10.1002/dvdy.24066.

Radermacher, A. L., du Toit, S. F. and Farrant, J. M. (2019) 'Desiccation-Driven Senescence in the Resurrection Plant Xerophyta schlechteri (Baker) N.L. Menezes: Comparison of Anatomical, Ultrastructural, and Metabolic Responses Between Senescent and Non-Senescent Tissues', *Frontiers in Plant Science*, 10(October), pp. 1–16. doi: 10.3389/fpls.2019.01396.

Rafsanjani, A. *et al.* (2015) 'Hydro-responsive curling of the resurrection plant Selaginella lepidophylla', *Scientific Reports*, 5, p. 8064. doi: 10.1038/srep08064.

Raghavan, V. *et al.* (2022) 'A simple guide to de novo transcriptome assembly and annotation', *Briefings in Bioinformatics*, 23(2), pp. 1–30. doi: 10.1093/bib/bbab563.

Rascio, N. and La Rocca, N. (2005) 'Resurrection plants: The puzzle of surviving extreme vegetative desiccation', *Critical Reviews in Plant Sciences*, 24(3), pp. 209–225. doi: 10.1080/07352680591008583.

Ravikumar, G. *et al.* (2014) 'Stress-inducible expression of AtDREB1A transcription factor greatly improves drought stress tolerance in transgenic indica rice', *Transgenic Research*, 23(3), pp. 421–439. doi: 10.1007/s11248-013-9776-6.

Razo-Mendivil, F. G., Martínez, O. and Hayano-Kanashiro, C. (2020) 'Compacta: A fast contig clustering tool for de novo assembled transcriptomes', *BMC Genomics*. BMC Genomics, 21(1), pp. 1–13. doi: 10.1186/s12864-020-6528-x.

Reina-Bueno, M. *et al.* (2012) 'Role of trehalose in heat and desiccation tolerance in the soil bacterium Rhizobium etli.', *BMC microbiology*, 12, pp. 1–17. doi: 10.1186/1471-2180-12-207.

Rennie, E. A. *et al.* (2012) 'Three members of the Arabidopsis glycosyltransferase family 8 are xylan glucuronosyltransferases', *Plant Physiology*, 159(4), pp. 1408–1417. doi: 10.1104/pp.112.200964.

Richter, K., Haslbeck, M. and Buchner, J. (2010) 'The Heat Shock Response: Life on the Verge of Death', *Molecular Cell*. Elsevier Inc., 40(2), pp. 253–266. doi: 10.1016/j.molcel.2010.10.006.

Robatzek, S. and Somssich, I. E. (2002) 'Robatzek and Somssich (2002).pdf', *Genes and Development*, 16, pp. 1139–1149. doi: 10.1101/gad.222702.Results.

Roberts, A. *et al.* (2011) 'Improving RNA-Seq expression estimates by correcting for fragment bias', *Genome Biology*. BioMed Central Ltd, 12(3), p. R22. doi: 10.1186/gb-2011-12-3-r22.

Roberts, A. and Pachter, L. (2012) 'Streaming fragment assignment for real-time analysis of sequencing experiments', *Nature Methods*. Nature Publishing Group, (November), pp. 1–7.

doi: 10.1038/nmeth.2251.

Rodriguez, M. C. S. *et al.* (2010) 'Transcriptomes of the desiccation-tolerant resurrection plant Craterostigma plantagineum', *Plant Journal*, 63(2), pp. 212–228. doi: 10.1111/j.1365-313X.2010.04243.x.

Le Roy, J. *et al.* (2016) 'Glycosylation is a major regulator of phenylpropanoid availability and biological activity in plants', *Frontiers in Plant Science*, 7(MAY2016). doi: 10.3389/fpls.2016.00735.

Ruiz-Lopez, N. *et al.* (2021) 'Synaptotagmins at the endoplasmic reticulum–plasma membrane contact sites maintain diacylglycerol homeostasis during abiotic stress', *Plant Cell*, 33(7), pp. 2431–2453. doi: 10.1093/plcell/koab122.

Runge, C. *et al.* (2004) 'Ending hunger in Africa prospects for the small farmer', *Issue briefs*.

Salmela, L. and Rivals, E. (2014) 'LoRDEC: Accurate and efficient long read error correction', *Bioinformatics*, 30(24), pp. 3506–3514. doi: 10.1093/bioinformatics/btu538.

Salzberg, S. L. (2019) 'Next-generation genome annotation: We still struggle to get it right', *Genome Biology*. Genome Biology, 20(1), pp. 19–21. doi: 10.1186/s13059-019-1715-2.

Sanger, F. and Nicklen, S. (1977) 'DNA sequencing with chain-terminating', 74(12), pp. 5463–5467. doi: 10.1073/pnas.74.12.5463.

Sayers, E. W. *et al.* (2021) 'Database resources of the National Center for Biotechnology Information', *Nucleic Acids Research*. Oxford University Press, 49(D1), pp. D10–D17. doi: 10.1093/nar/gkaa892.

Scherl, A. *et al.* (2002) 'Functional Proteomic Analysis of Human Nucleolus □', *Molecular Biology of the Cell*, 13(November), pp. 4100–4109. doi: 10.1091/mbc.E02.

Schonbeck, M. W. and Bewley, J. D. (1981) ' Responses of the moss Tortula ruralis to desiccation treatments. II. Variations in desiccation tolerance ', *Canadian Journal of Botany*, 59(12), pp. 2707–2712. doi: 10.1139/b81-321.

Schroeder, A. *et al.* (2006) 'The RIN: An RNA integrity number for assigning integrity values to RNA measurements', *BMC Molecular Biology*, 7, pp. 1–14. doi: 10.1186/1471-2199-7-3.

Sengupta, S. *et al.* (2015) 'Significance of galactinol and raffinose family oligosaccharide synthesis in plants', *Frontiers in Plant Science*, 6(AUG), pp. 1–11. doi: 10.3389/fpls.2015.00656.

Seong, I. S. *et al.* (1999) 'ATP-dependent degradation of SulA, a cell division inhibitor, by the HslVU protease in Escherichia coli', *FEBS Letters*, 456(1), pp. 211–214. doi: 10.1016/S0014-5793(99)00935-7.

Shen, B. *et al.* (1999) 'Roles of sugar alcohols in osmotic stress adaptation. Replacement of glycerol by mannitol and sorbitol in yeast', *Plant Physiology*, 121(1), pp. 45–52. doi: 10.1104/pp.121.1.45.

Sherwin, H. W. and Farrant, J. M. (1998) 'Protection mechanisms against excess light in the resurrection plants Craterostigma wilmsii and Xerophyta viscosa', *Plant Growth Regulation*, 24(3), pp. 203–210. doi: 10.1023/A:1005801610891.

Shi, R. *et al.* (2021) 'Biogenesis of Iron–Sulfur Clusters and Their Role in DNA Metabolism', *Frontiers in Cell and Developmental Biology*, 9(September), pp. 1–13. doi: 10.3389/fcell.2021.735678.

Shi, W. Y. *et al.* (2018) 'The WRKY transcription factor GmWRKY12 confers drought and salt tolerance in soybean', *International Journal of Molecular Sciences*, 19(12). doi: 10.3390/ijms19124087.

Shimizu, T. *et al.* (2010) 'Desiccation-induced structuralization and glass formation of group 3 late embryogenesis abundant protein model peptides', *Biochemistry*, 49(6), pp. 1093–1104. doi: 10.1021/bi901745f.

Signal, B. and Kahlke, T. (2021) 'Borf: Improved ORF prediction in de-novo assembled transcriptome annotation', *Biorxiv*, pp. 1–11. Available at: https://doi.org/10.1101/2021.04.12.439551.

Silberg, T. R. *et al.* (2017) 'Maize-legume intercropping in central Malawi: determinants of practice', *International Journal of Agricultural Sustainability*. doi: 10.1080/14735903.2017.1375070.

Silva Artur, M. A. *et al.* (2019) 'Genome-level responses to the environment: plant desiccation tolerance', *Emerging Topics in Life Sciences*, 3(2), pp. 153–163. doi: 10.1042/etls20180139.

Simão, F. A. *et al.* (2015) 'BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs', *Bioinformatics*, 31(19), pp. 3210–3212. doi: 10.1093/bioinformatics/btv351.

Sims, D. *et al.* (2014) 'Sequencing depth and coverage: Key considerations in genomic analyses', *Nature Reviews Genetics*. Nature Publishing Group, 15(2), pp. 121–132. doi: 10.1038/nrg3642.

Singh, K. B., Foley, R. C. and Oñate-Sánchez, L. (2002) 'Transcription factors in plant defense and stress responses', *Current Opinion in Plant Biology*, 5(5), pp. 430–436. doi: 10.1016/S1369-5266(02)00289-3.

Skubacz, A., Daszkowska-Golec, A. and Szarejko, I. (2016) 'The role and regulation of ABI5 (ABA-insensitive 5) in plant development, abiotic stress responses and phytohormone crosstalk', *Frontiers in Plant Science*, 7(DECEMBER2016), pp. 1–17. doi: 10.3389/fpls.2016.01884.

Smirnoff, N. (1992) 'The carbohydrates of bryophytes in relation to desiccation tolerance', *Journal of Bryology*, 17(2), pp. 185–191. doi: 10.1179/jbr.1992.17.2.185.

Song, L. and Florea, L. (2015) 'Rcorrector : efficient and accurate error correction for Illumina RNA-seq reads', pp. 1–8. doi: 10.1186/s13742-015-0089-y.

Song, L., Florea, L. and Langmead, B. (2014) 'Lighter: fast and memory-efficient sequencing

error correction without counting', *Genome biology*, 15(11), p. 509. doi: 10.1186/s13059-014-0509-9.

Sripinyowanich, S. *et al.* (2021) 'De Novo Transcriptome Assembly of Two Microsorum Fern Species Identifies Enzymes Required for Two Upstream Pathways of Phytoecdysteroids'.

Srivastava, A. *et al.* (2016) 'Accurate, Fast and Lightweight Clustering of de novo Transcriptomes using Fragment Equivalence Classes', (22), pp. 1–12. Available at: http://arxiv.org/abs/1604.03250.

Srivastava, R. and Kumar, R. (2019) 'The expanding roles of APETALA2/Ethylene Responsive Factors and their potential applications in crop improvement', *Briefings in Functional Genomics*, 18(4), pp. 240–254. doi: 10.1093/bfgp/elz001.

Stark, L. R. (2017) 'Ecology of desiccation tolerance in bryophytes: A conceptual framework and methodology', *Bryologist*, 120(2), pp. 130–165. doi: 10.1639/0007-2745-120.2.130.

Stark, R., Grzelak, M. and Hadfield, J. (2019) 'RNA sequencing: the teenage years', *Nature Reviews Genetics*. Springer US, 20(11), pp. 631–656. doi: 10.1038/s41576-019-0150-2.

Stein, O. and Granot, D. (2019) 'An overview of sucrose synthases in plants', *Frontiers in Plant Science*, 10(February), pp. 1–14. doi: 10.3389/fpls.2019.00095.

Stoler, N. and Nekrutenko, A. (2021) 'Sequencing error profiles of Illumina sequencing instruments', *NAR Genomics and Bioinformatics*. Oxford University Press, 3(1), pp. 1–9. doi: 10.1093/nargab/lqab019.

Su, P. H. and Li, H. M. (2008) 'Arabidopsis stromal 70-kD heat shock proteins are essential for plant development and important for thermotolerance of germinating seeds', *Plant Physiology*, 146(3), pp. 1231–1241. doi: 10.1104/pp.107.114496.

Sun, M. *et al.* (2020) 'Exogenous phosphatidylcholine treatment alleviates drought stress and maintains the integrity of root cell membranes in peach', *Scientia Horticulturae*. Elsevier, 259(July 2019), p. 108821. doi: 10.1016/j.scienta.2019.108821.

Suzek, B. E. *et al.* (2007) 'UniRef: Comprehensive and non-redundant UniProt reference clusters', *Bioinformatics*, 23(10), pp. 1282–1288. doi: 10.1093/bioinformatics/btm098.

Suzek, B. E. *et al.* (2015) 'UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches', *Bioinformatics*, 31(6), pp. 926–932. doi: 10.1093/bioinformatics/btu739.

Suzuki, H. and Kasahara, M. (2018) 'Introducing difference recurrence relations for faster semi-global alignment of long sequences', *BMC Bioinformatics*. BMC Bioinformatics, 19(Suppl 1). doi: 10.1186/s12859-018-2014-8.

Tang, Y. *et al.* (2022) 'CRISPR-Cas9-mediated mutagenesis of the SlSRM1-like gene leads to abnormal leaf development in tomatoes', *BMC Plant Biology*. BioMed Central, 22(1), pp. 1–14. doi: 10.1186/s12870-021-03397-5.

Tao, Q. *et al.* (2018) 'Ethylene responsive factor ERF110 mediates ethylene-regulated

transcription of a sex determination-related orthologous gene in two Cucumis species', *Journal of Experimental Botany*, 69(12), pp. 2953–2965. doi: 10.1093/jxb/ery128.

Tardaguila, M. *et al.* (2018) 'Corrigendum: SQANTI: Extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification (Genome Research (2018) 28 (396-411) DOI: 10.1101/gr.222976.117)', *Genome Research*, 28(7), p. 1096. doi: 10.1101/gr.239137.118.

Tauzin, A. S. and Giardina, T. (2014) 'Sucrose and invertases, a part of the plant defense response to the biotic stresses', *Frontiers in Plant Science*, 5(JUN), pp. 1–8. doi: 10.3389/fpls.2014.00293.

Thines, B. *et al.* (2007) 'JAZ repressor proteins are targets of the SCFCOI1 complex during jasmonate signalling', *Nature*, 448(7154), pp. 661–665. doi: 10.1038/nature05960.

Tilgner, H. *et al.* (2014) 'Defining a personal, allele-specific, and single-molecule long-read transcriptome', *Proceedings of the National Academy of Sciences of the United States of America*, 111(27), pp. 9869–9874. doi: 10.1073/pnas.1400447111.

du Toit, S. F., Bentley, J. and Farrant, J. M. (2021) 'NADES formation in vegetative desiccation tolerance: Prospects and challenges', *Advances in Botanical Research*, 97, pp. 225–252. doi: 10.1016/bs.abr.2020.09.007.

Toldi, O., Tuba, Z. and Scott, P. (2009) 'Vegetative desiccation tolerance: Is it a goldmine for bioengineering crops?', *Plant Science*. doi: 10.1016/j.plantsci.2008.10.002.

Tolleter, D. *et al.* (2007) 'Structure and function of a mitochondrial late embryogenesis abundant protein are revealed by desiccation', *Plant Cell*, 19(5), pp. 1580–1589. doi: 10.1105/tpc.107.050104.

Tolleter, D., Hincha, D. K. and Macherel, D. (2010) 'A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state', *Biochimica et Biophysica Acta - Biomembranes*. Elsevier B.V., 1798(10), pp. 1926–1933. doi: 10.1016/j.bbamem.2010.06.029.

Tseng, H.-H. E. (2014) 'Cogent: Reconstructing the coding genome using full-length transcriptome sequences without a reference genome'. Available at: https://github.com/Magdoll/Cogent.

Tshabuse, F. *et al.* (2018) 'Glycerolipid analysis during desiccation and recovery of the resurrection plant Xerophyta humilis (Bak) Dur and Schinz', *Plant Cell and Environment*, 41(3), pp. 533–547. doi: 10.1111/pce.13063.

Tsugama, D., Liu, S. and Takano, T. (2012) 'A bZIP protein, VIP1, is a regulator of osmosensory signaling in Arabidopsis', *Plant Physiology*, 159(1), pp. 144–155. doi: 10.1104/pp.112.197020.

Tsuwamoto, R., Fukuoka, H. and Takahata, Y. (2008) 'GASSHO1 and GASSHO2 encoding a putative leucine-rich repeat transmembrane-type receptor kinase are essential for the normal development of the epidermal surface in Arabidopsis embryos', *Plant Journal*, 54(1), pp. 30–42. doi: 10.1111/j.1365-313X.2007.03395.x.

Tuba, Z. and Lichtenthaler, H. K. (2011) 'Ecophysiology of Homoiochlorophyllous and Poikilochlorophyllous Desiccation-Tolerant Plants and Vegetations', pp. 157–183. doi: 10.1007/978-3-642-19106-0_9.

Tuba, Z., Protor, M. C. F. and Csintalan, Z. (1998) 'Ecophysiological responses of homoiochlorophyllous and poikilochlorophyllous desiccation tolerant plants: A comparison and an ecological perspective', in *Plant Growth Regulation*. doi: 10.1023/A:1005951908229.

Tucker, T., Marra, M. and Friedman, J. M. (2009) 'Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine', *American Journal of Human Genetics*. The American Society of Human Genetics, 85(2), pp. 142–154. doi: 10.1016/j.ajhg.2009.06.022.

Ukai, Y. *et al.* (2020) 'De novo transcriptome analysis reveals an unperturbed transcriptome under high cadmium conditions in the Cd-hypertolerant fern Athyrium yokoscense', pp. 65–74.

Valle, M. (2017) *"Pyruvate carboxylase, structure and function"*, *Sub-Cellular Biochemistry*. doi: 10.1007/978-3-319-46503-6_11.

Vanburen, R. *et al.* (2018) 'Extreme haplotype variation in the desiccation-tolerant clubmoss Selaginella lepidophylla', *Nature Communications*. Springer US, 9(1), pp. 1–8. doi: 10.1038/s41467-017-02546-5.

Vigani, G., Tarantino, D. and Murgia, I. (2013) 'Mitochondrial ferritin is a functional iron-storage protein in cucumber (Cucumis sativus) roots', *Frontiers in Plant Science*, 4(AUG), pp. 1–8. doi: 10.3389/fpls.2013.00316.

Vorwerke, S. *et al.* (2007) 'EDR2 negatively regulates salicylic acid-based defenses and cell death during powdery mildew infections of Arabidopsis thaliana', *BMC Plant Biology*, 7, pp. 1–14. doi: 10.1186/1471-2229-7-35.

Wahab, A. *et al.* (2022) 'Plants' Physio-Biochemical and Phyto-Hormonal Responses to Alleviate the Adverse Effects of Drought Stress: A Comprehensive Review', *Plants*, 11(13). doi: 10.3390/plants11131620.

Walters, C. (2015) 'Orthodoxy, recalcitrance and in-between: describing variation in seed storage characteristics using threshold responses to water loss', *Planta*. Springer Berlin Heidelberg, 242(2), pp. 397–406. doi: 10.1007/s00425-015-2312-6.

Wan, C. Y. and Wilkins, T. A. (1994) 'A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (Gossypium hirsutum L.)', *Analytical Biochemistry*, pp. 7–12. doi: 10.1006/abio.1994.1538.

Wang, E. *et al.* (2012) 'A common signaling process that promotes mycorrhizal and oomycete colonization of plants', *Current Biology*. Elsevier Ltd, 22(23), pp. 2242–2246. doi: 10.1016/j.cub.2012.09.043.

Wang, H. Sen *et al.* (2014) 'A tomato endoplasmic reticulum (ER)-type omega-3 fatty acid desaturase (LeFAD3) functions in early seedling tolerance to salinity stress', *Plant Cell Reports*, 33(1), pp. 131–142. doi: 10.1007/s00299-013-1517-z.

Wang, Q. and Huang, J. (2020) 'Fungal Genes in Plants: Impact and Potential Applications',

Trends in Plant Science. Elsevier Ltd, 25(11), pp. 1064–1067. doi: 10.1016/j.tplants.2020.07.005.

Wang, S. *et al.* (2019) 'The Structure and Function of Major Plant Metabolite Modifications', *Molecular Plant*. Elsevier Ltd, 12(7), pp. 899–919. doi: 10.1016/j.molp.2019.06.001.

Wang, X. *et al.* (2021) 'Functioning of PPR Proteins in Organelle RNA Metabolism and Chloroplast Biogenesis', *Frontiers in Plant Science*, 12(February), pp. 1–8. doi: 10.3389/fpls.2021.627501.

Waters, E. R. and Vierling, E. (2020) 'Plant small heat shock proteins – evolutionary and functional diversity', *New Phytologist*, 227(1), pp. 24–37. doi: 10.1111/nph.16536.

Wenger, A. M. *et al.* (2019) 'Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome', *Nature Biotechnology*, 37(10), pp. 1155–1162. doi: 10.1038/s41587-019-0217-9.

Wheeler, T. and Von Braun, J. (2013) 'Climate change impacts on global food security', *Science*. doi: 10.1126/science.1239402.

Whittaker, A. *et al.* (2004) 'Comparison of sucrose metabolism during the rehydration of desiccation-tolerant and desiccation-sensitve leaf material of Sporobolus stapfianus', in *Physiologia Plantarum*. doi: 10.1111/j.1399-3054.2004.00346.x.

Whittaker, M. M. *et al.* (1996) 'Glyoxal oxidase from Phanerochaete chrysosporium is a new radical-copper oxidase', *Journal of Biological Chemistry*. Â© 1996 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 271(2), pp. 681–687. doi: 10.1074/jbc.271.2.681.

Wickett, N. J. *et al.* (2014) 'Phylotranscriptomic analysis of the origin and early diversification of land plants', *Proceedings of the National Academy of Sciences of the United States of America*, 111(45), pp. E4859–E4868. doi: 10.1073/pnas.1323926111.

Wilfinger, W. W., Mackey, K. and Chomczynski, P. (1997) 'Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity', *BioTechniques*, 22(3), pp. 474–481. doi: 10.2144/97223st01.

Willigen, C. Vander *et al.* (2004) 'Mechanical stabilization of desiccated vegetative tissues of the resurrection grass Eragrostis nindensis: Does a TIP 3;1 and/or compartmentalization of subcellular components and metabolites play a role?', *Journal of Experimental Botany*, 55(397), pp. 651–661. doi: 10.1093/jxb/erh089.

Wisecaver, J. H. *et al.* (2016) 'Dynamic Evolution of Nitric Oxide Detoxifying Flavohemoglobins , a Family of Single-Protein Metabolic Modules in Bacteria and Eukaryotes', 33(8), pp. 1979–1987. doi: 10.1093/molbev/msw073.

Wood, A. J. (2007) 'The nature and distribution of vegetative desiccation-tolerance in hornworts, liverworts and mosses', *The Bryologist*. doi: 10.1639/0007-2745(2007)110[163:ienfib]2.0.co;2.

Workman, R. E. *et al.* (2019) 'Nanopore native RNA sequencing of a human poly(A) transcriptome', *Nature Methods*. Springer US, 16(12), pp. 1297–1305. doi: 10.1038/s41592-

019-0617-2.

Wu, T. D. and Watanabe, C. K. (2005) 'GMAP: A genomic mapping and alignment program for mRNA and EST sequences', *Bioinformatics*, 21(9), pp. 1859–1875. doi: 10.1093/bioinformatics/bti310.

Xing, D. H. *et al.* (2008) 'Stress- and pathogen-induced Arabidopsis WRKY48 is a transcriptional activator that represses plant basal defense', *Molecular Plant*. The Authors 2008, 1(3), pp. 459–470. doi: 10.1093/mp/ssn020.

Xu, X. *et al.* (2021) 'Molecular insights into plant desiccation tolerance: transcriptomics, proteomics and targeted metabolite profiling in Craterostigma plantagineum', *Plant Journal*, 107(2), pp. 377–398. doi: 10.1111/tpj.15294.

Xu, Y. *et al.* (2017) 'HOTHEAD-Like HTH1 is Involved in Anther Cutin Biosynthesis and is Required for Pollen Fertility in Rice', *Plant and Cell Physiology*, 58(7), pp. 1238–1248. doi: 10.1093/pcp/pcx063.

Xu, Z. *et al.* (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. doi: 10.1016/j.molp.2018.05.003.

Yang, X. D., Dong, C. J. and Liu, J. Y. (2006) 'A plant mitochondrial phospholipid hydroperoxide glutathione peroxidase: Its precise localization and higher enzymatic activity', *Plant Molecular Biology*, 62(6), pp. 951–962. doi: 10.1007/s11103-006-9068-0.

Yobi, A. *et al.* (2012) 'Comparative metabolic profiling between desiccation-sensitive and desiccation-tolerant species of Selaginella reveals insights into the resurrection trait', *Plant Journal*, 72(6), pp. 983–999. doi: 10.1111/tpj.12008.

Yobi, A. *et al.* (2013) 'Metabolomic profiling in selaginella lepidophylla at various hydration states provides new insights into the mechanistic basis of desiccation tolerance', *Molecular Plant*. Â© The Authors. All rights reserved., 6(2), pp. 369–385. doi: 10.1093/mp/sss155.

Yobi, A. *et al.* (2017) 'Sporobolus stapfianus: Insights into desiccation tolerance in the resurrection grasses from linking transcriptomics to metabolomics', *BMC Plant Biology*. BMC Plant Biology, 17(1), pp. 1–30. doi: 10.1186/s12870-017-1013-7.

Yu, X. *et al.* (2019) 'Identification of cotton mother of FT and TFL1 homologs, GhMFT1 and GhMFT2, involved in seed germination', *PLoS ONE*, 14(4), pp. 1–23. doi: 10.1371/journal.pone.0215771.

Yu, X. *et al.* (2021) 'Genome-wide analysis of HSP70 gene superfamily in Pyropia yezoensis (Bangiales, Rhodophyta): identification, characterization and expression profiles in response to dehydration stress', *BMC Plant Biology*. BioMed Central, 21(1), pp. 1–14. doi: 10.1186/s12870-021-03213-0.

Zdobnov, E. M. *et al.* (2021) 'OrthoDB in 2020: Evolutionary and functional annotations of orthologs', *Nucleic Acids Research*, 49(D1), pp. D389–D393. doi: 10.1093/nar/gkaa1009.

Zeng, Q., Chen, X. and Wood, A. J. (2002) 'Two early light-inducible protein (ELIP) cDNAs from the resurrection plant Tortula ruralis are differentially expressed in response to

desiccation, rehydration, salinity, and high light', *Journal of Experimental Botany*, 53(371), pp. 1197–1205. doi: 10.1093/jexbot/53.371.1197.

Zhang, H. D. *et al.* (2015) 'PPR protein PDM1/SEL1 is involved in RNA editing and splicing of plastid genes in Arabidopsis thaliana', *Photosynthesis Research*, 126(2–3), pp. 311–321. doi: 10.1007/s11120-015-0171-4.

Zhang, L. *et al.* (2015) 'Three WRKY transcription factors additively repress abscisic acid and gibberellin signaling in aleurone cells', *Plant Science*. Elsevier Ireland Ltd, 236, pp. 214–222. doi: 10.1016/j.plantsci.2015.04.014.

Zhang, Q. and Bartels, D. (2018) 'Molecular responses to dehydration and desiccation in desiccation-tolerant angiosperm plants', *Journal of Experimental Botany*, 69(13), pp. 3211–3222. doi: 10.1093/jxb/erx489.

Zhang, Q., Song, X. and Bartels, D. (2016) 'Enzymes and metabolites in carbohydrate metabolism of desiccation tolerant plants', *Proteomes*, 4(4). doi: 10.3390/proteomes4040040.

Zhang, Y. *et al.* (2017) 'Inhibition of glutamate decarboxylase (GAD) by ethyl ketopentenoate (EKP) induces treatment-resistant epileptic seizures in zebrafish', *Scientific Reports*. Springer US, 7(1), pp. 1–13. doi: 10.1038/s41598-017-06294-w.

Zhang, Y. *et al.* (2021) 'Alternative splicing and cancer: a systematic review', *Signal Transduction and Targeted Therapy*. Springer US, 6(1). doi: 10.1038/s41392-021-00486-7.

Zhao, S. (2019) 'Alternative splicing, RNA-seq and drug discovery', *Drug Discovery Today*. Elsevier Ltd, 24(6), pp. 1258–1267. doi: 10.1016/j.drudis.2019.03.030.

Zhao, W. *et al.* (2014) 'Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling', *BMC Genomics*, 15(1), pp. 1–11. doi: 10.1186/1471-2164-15-419.

Zhou, Y. (2021) 'A De Novo Transcriptome Assembly of Ceratopteris richardii Gene Families in Land Plants', 13(March), pp. 1–14. doi: 10.1093/gbe/evab042.

Zhou, Z. *et al.* (2013) 'Zinc Finger Protein 6 (ZFP6) regulates trichome initiation by integrating gibberellin and cytokinin signaling in Arabidopsis thaliana', *New Phytologist*, 198(3), pp. 699–708. doi: 10.1111/nph.12211.

Zhu, A., Ibrahim, J. G. and Love, M. I. (2019) 'Heavy-Tailed prior distributions for sequence count data: Removing the noise and preserving large differences', *Bioinformatics*, 35(12), pp. 2084–2092. doi: 10.1093/bioinformatics/bty895.

Ziveri, J. *et al.* (2017) 'The metabolic enzyme fructose-1,6-bisphosphate aldolase acts as a transcriptional regulator in pathogenic Francisella', *Nature Communications*. Springer US, 8(1). doi: 10.1038/s41467-017-00889-7.