Natural variation of water use and water productivity in *Arabidopsis thaliana*

John Nicholas Ferguson

A thesis submitted for the degree of Doctor of Philosophy

School of Biological Sciences

University of Essex

2016

Abstract

Plant performance under reduced water availability has traditionally been assessed as drought resistance and more recently as water use efficiency (WUE). An extensive body of work has been established over the past 15 years where the natural variation of water use efficiency has been studied in the model species *Arabidopsis thaliana* (Arabidopsis). At the same time, a substantial degree of criticism has arisen with respect to the use of drought resistance and WUE as measures of plant performance, due to the lack of relatedness of these parameters to reproductive performance, i.e. yield.

The work in this thesis is centered on understanding the physiological and genetic basis of water use and water productivity as alternative measures of plant performance under the context of reduced water availability. The first part of this study describes an extensive assessment of the natural variation of water use and water productivity in Arabidopsis in relation to numerous key physiological, phenological, and developmental parameters. Furthermore, this work concisely relates plasticity of key traits to historical climatic variation. A fundamental aspect of this work was the clarification that it is possible to estimate long term water use to a high degree of accuracy based on short term water use, i.e. soil drying rate, and flowering time. Flowering time was demonstrated to be the predominant driver of vegetative performance and water use, however it appeared to be genetically uncoupled from reproductive performance. This is in contrast to previous work that suggests WUE, measured as the ratio of C^{12} to C^{13} isotopes ($\delta^{13}C$), is positively associated with flowering time. Additionally, it was demonstrated that multiple commonly

employed proxies of reproductive performance including total biomass, WUE, and flowering time, were not sufficient at predicting seed yield in Arabidopsis across multiple environments.

The second part of this study involved the genetic dissection of water use and productivity related traits in Arabidopsis through a quantitative trait loci (QTL) mapping study and a genome wide association study (GWAS). QTL mapping using a recombinant inbred line (RIL) population developed from the ecotypes Col-0 and C24 revealed two key flowering time genes, *FLOWERING LOCUS C* (*FLC*) and *FRIGIDA* (*FRI*), as key regulators of water use. It was demonstrated that a combination of non-functional alleles of both *FLC* and *FRI* reduced long term water use via a shorted life cycle, which is again in contrast to previous work relating to the genetic dissection of WUE in Arabidopsis. Crucially, it was observed that reduced water use mediated in this fashion did not detrimentally impact upon reproductive performance. GWAS was employed subsequent to the QTL mapping in order to identify candidate genes underlying the variation for productivity as a unique trait and also as a factor of water use, i.e. water productivity. GWAS identified multiple promising candidate genes that potentially underlie the heritable genetic variation for flowering time, water use, and water productivity.

Acknowledgments

This project was funded through a BBSRC doctoral training partnership grant (Project reference: BB/J012564/1) where the University of Essex was the lead research organization and ATC Limited were project partners. The financial support from BBSRC and ATC is gratefully recognized.

My first and foremost thanks go to Uli, my Ph.D. supervisor. Her unwavering support, encouragement, and patience will forever be cherished. I thank you for constantly lifting my spirits at our weekly meetings and inspiring me to believe in our research. I leave Essex with a heavy heart knowing that I can no longer dedicate my work life solely to our natural variation research, especially since I believe it still has a lot of exiting potential for the future. Nevertheless, I look forward to continuing collaborating with you on this and other projects long into the future.

The School of Biological Sciences at the University of Essex was a fantastic environment to carry out my Ph.D. studies. I am as inspired today, as I was the day I started, by the plant science research community at Essex. I am indebted to all members of the plant science group, since all of you contributed to the happiness and productivity I enjoyed during my time at Essex. I must extend thanks to Dr Tracy Lawson for her role as my Advisor and also to Professor Phil Mullineaux for his advice and encouragement. Dr Phil Davey and Sue Corbett are gratefully recognized for helping me with my research on numerous occasions.

I am also hugely indebted to my supervisor at ATC, Matt, who taught me a huge part of everything I now understand about interpreting and performing quantitative genetics,

which I am sure will probably scare him slightly! I hugely enjoyed my time at ATC and also extend best wishes to Alan, Brian, Deirdre, Jen, Kieron, Tijs, and Pascaline.

My final acknowledgments are for my Mother and Father. I absolutely would not have been able to achieve this without your support and care. I will be forever thankful for the opportunities you have allowed for me and for encouraging me to follow a path that has allowed me to begin a career that I am sure I will continue to find rewarding and enjoyable.

Contents

Figures and Tables

Abbreviations

$\delta^{13}C$	Ratio of C^{12} to C^{13} natural isotopes of carbon
A	Photosynthetic CO_2 assimilation
ABA	Abscisic acid
Amax	Light saturated CO_2 assimilation
AMPRIL	Arabidopsis Multi Parent Recombinant Inbred Line
ANOVA	Analysis of variance
ASPG1	*ASPARTIC PROTEASE IN GUARD CELL 1*
ATBAG6	*BCL-2-ASSOCIATED ATHANOGENE 6*
BC	Backcross
BLUPs	Best Linear Unbiased Predictors
C_a	Extracellular CO_2 concentration
CHYR1	*CHY ZINC-FINGER AND RING PROTEIN 1*
C_i	Intracellular CO_2 concentration
CIM	Composite interval mapping
cM	Centimorgan
CPC	Climate principle component
COV_{AB}	Phenotypic covariance between traits A and B
C_{VG}	Coefficient of genetic variation
CW	Continuous watering
cWU	Calculated water use
cWUE	Calculated water use efficiency

cWP	Calculated water productivity
DEFL	Defensin-Like
DHL	Double haploid line
E	Evaporation
e-BIC	Extended Bayesian information criterion
EMMA	Efficient mixed-model association
ET	Evapotranspiration
FLC	*FLOWERING LOCUS C*
FRI	*FRIGIDA*
FT	*FLOWERING LOCUS T*
g_s	Stomatal conductance
GLMM	General linear mixed model
GWAS	Genome-wide association study
GxE	Genotypic by environment interaction
H^2	Broad sense heritability
iWUE	Leaf-level instantaneous water use efficiency
J_{max}	Maximum rate of electronic transport demand for RuBP regeneration
KW	Kruskal-Wallis
LD	Linkage disequilibrium
LOD	Logarithm (base 10) of odds
MAGIC	Multiple Advanced Generation InterCross Population
mBonf	Multiple-Bonferroni criterion
MLMM	Multi-locus mixed model

MQM	Multiple QTL mapping
mWU	Measured water use
mWUE	Measured water use efficiency
NASC	Nottingham Arabidopsis Stock Centre
NIL	Near isogenic line
PC	Principle component
PCA	Principle component analysis
PCD	Programmed cell death
PPFD	Photosynthetically active photon flux density
QTL	Quantitative trait loci
RH	Relative humidity
RIL	Recombinant inbred line
ROS	Reactive oxygen species
RTR	Relative Trait Range Index, i.e. plasticity
rSWC	Relative soil water content
SD	Short dehydration
SIAR1	*SILIQUES ARE RED* 1
SIM	Simple interval mapping
SNP	Single nucleotide polymorphism
SOC1	*SUPPRESSOR OF OVEREXPRESSION OF CONSTANS*
SSRP1	*STRUCTURE SPECIFIC RECOGNITION PROTEIN 1*
T	Transpiration
TAIR	The Arabidopsis Information Resource

TPC	Trait principle component
V_{cmax}	Maximum velocity of Rubisco for carboxylation
V_G	Genetic variation
V_P	Phenotypic variation
VPD	Vapor pressure deficit
WP	Water productivity
WUE	Water use efficiency

1 Introduction

1.1 *Arabidopsis thaliana*

1.1.2 Arabidopsis as a model organism for gene discovery

Arabidopsis thaliana (Arabidopsis) is a small annual herb belonging to the Brassicaceae family. It is the de facto model species for the global plant science community and is directly related to many Brassicaceae crop species, including *Brassica napus* (oilseed rape). Indeed, multiple studies have successfully demonstrated strong genomic relatedness between Arabidopsis and these crop species (Reviewed in: (Sharma *et al.*, 2014))

Johannes Thal provided the first taxonomic characterization of Arabidopsis in the 16th century in Northern Germany. It has subsequently undergone numerous name changes, however it was eventually renamed in honor of Thal (Meyerowitz, 2001). Friedrich Laibach is generally considered the father of Arabidopsis research. Laibach's seminal 1943 paper was the first to propose Arabidopsis as a model species, due to its short life cycle, fecundity, ease of crossing, and plausibility of generating mutants (Koornneef, 2013).

The information, methodologies, and resources that have followed the extensive employment of Arabidopsis as an experimental model have rendered it the ideal species for rapid and precise gene discovery (Bevan & Walsh, 2004; Provart *et al.*, 2016). The elucidation of the genetic basis of traits of interest in Arabidopsis can have important implications in terms of furthering basic biological understanding. Despite this, it is important to consider the impact gene discovery in Arabidopsis can have from a more applied perspective. To this end, the most obvious evaluation to make is that of the efficiency of

translating such research into the improvement of crop species. Due to their relatively recent divergence, the function of orthologous genes between Arabidopsis and *Brassica* crop species is often conserved. Indeed multiple studies have demonstrated that certain *Brassica* genes afford the same function as their respective Arabidopsis orthologs (Greco *et al.*, 2012; Xu *et al.*, 2014; Dong *et al.*, 2016), thereby demonstrating the effectiveness of comparative genomics studies for *Brassica* crop improvement when based upon genes discovered for traits of interest in Arabidopsis. Additionally, it is worth noting that directly introducing Arabidopsis genes of known functionality into *Brassica* species via transgenic approaches, has been somewhat effective in terms of improving or introducing target traits of interest (Bechtold *et al.*, 2013; Lannenpaa, 2014; Narusaka *et al.*, 2014).

As well as providing a crop improvement incentive, delineating the genetic basis of traits of interest in Arabidopsis has been vital for furthering our understanding of environmental adaptation and evolutionary biology (Reviewed in: Alonso-Blanco *et al.*, 2009). This of vital important in order to predict ecological trends under the context of global change (Méndez-Vigo *et al.*, 2011; Picó, 2012; Agren *et al.*, 2013; Oakley *et al.*, 2014; Wolfe & Tonsor, 2014; Brennan *et al.*, 2014). Furthermore, genes that facilitate adaptation are often parallel to those that cause variation for traits that are important for agronomic objectives, such as freezing tolerance (Gery *et al.*, 2011; Agren *et al.*, 2013; Oakley *et al.*, 2014) or resource use efficiency (Kobayashi & Koyama, 2002; Reymond *et al.*, 2006; Vasseur *et al.*, 2014).

1.1.3 Natural variation within Arabidopsis

Despite long standing interest in the natural variation of Arabidopsis, it was only in 2002 that the first detailed biogeographical account of the species was completed

(Hoffmann, 2002). The present range of Arabidopsis is incredibly vast, with genetically and physiologically characterized populations existing as far west as Canada and as far east as Japan, as well as most places in-between. Arabidopsis is believed to have been introduced to North America from Eurasia somewhere between 150 and 200 years ago (Samis *et al.*, 2012). Although there are pockets of populations in the Southern Hemisphere, which are likely lab escapes or artificial introductions, Arabidopsis exists almost exclusively in the Northern Hemisphere. Here, it encompasses a broad latitudinal range that spreads well into the Arctic Circle and as far south as Northern Africa and the islands of Cape Verde (Hoffmann, 2002).

Due to its global distribution, the various populations of Arabidopsis that constitute the species are subject to differential abiotic and biotic perturbations. These populations are typically referred to as ecotypes and their names often, but not always, reflect their site of collection, for example Col-0 was collected in Columbia, Missouri. The environmental conditions that these populations have historically been subjected to have placed selective pressures on them to adapt. For this reason, genetic mutations that spontaneously arise and that confer a clear adaptive advantage to site specific and recurring conditions are maintained via natural selection. It is for this reason that we see such enormous biological variation between the ecotypes of Arabidopsis (Weigel, 2012).

Natural variation is present within all plant species and has been exploited since plants were first domesticated thousands of years ago (Alonso-Blanco *et al.*, 2009). The majority of this variation is quantitative, as is the variation that exists for adaptation-related traits in Arabidopsis. It is termed quantitative because it is influenced by multiple genes, known as quantitative trait loci (QTLs). There are two primary methodologies through which

QTLs may be dissected, both of which are detailed later as part of this chapter. In brief, the first of these strategies is known as linkage disequilibrium mapping, or QTL mapping, and requires the development of a mapping population that segregates for the trait of interest. The success of QTL mapping was first demonstrated in crop-based studies (Paterson *et al.*, 1990), but its successful application using Arabidopsis shortly followed (Kowalski *et al.*, 1994). The second of these methodologies is association mapping, commonly referred to as genome-wide association studies (GWAS). First pioneered in Human genetic research (Tomfohrde *et al.*, 1994), GWAS is employed to directly identify genetic polymorphisms that are precisely related to variation of interest (Atwell *et al.*, 2010; Trontin *et al.*, 2011; Horton *et al.*, 2012).

Due to its importance for gene discovery, Arabidopsis natural variation is a hugely important resource for the plant science community. Despite this, the popularity and direct nature of performing gene discovery studies in Arabidopsis based on natural variation has raised concerns amongst evolutionary and functional ecologists. These concerns relate to the amount of variation identified genes underlying traits of interest actually control and also the knock-on effects manipulation of said genes has on other traits, primarily fitness related traits (Reviewed in: Tonsor *et al.*, 2005). To this end it is worth noting the vast majority of Arabidopsis-based QTL and GWAS studies are based on phenotyping performed under strictly controlled environmental conditions, where the effect of QTLs or genes are not validated in natural or agronomic environments (E.g. Masle *et al.*, 2005; Filiault & Maloof, 2012; Gnan *et al.*, 2014; Verslues *et al.*, 2014; Morrison & Linder, 2014; van Rooijen *et al.*, 2015). Furthermore, and with respect to additional effects of identified loci, QTL underlying freezing tolerance and water use efficiency (WUE) have both recently been demonstrated to

show clear trade off interactions with reproductive performance in Arabidopsis (Agren *et al.*, 2013; Oakley *et al.*, 2014; Mojica *et al.*, 2016). These cautions highlight how the optimal iteration of available genetic variation may be desirable one year, depending on the trait of interest, but not the next year. In the same vain, it may be important in the presence of stress, but a hindrance in its absence (Bechtold *et al.*, 2010; Alsdurf *et al.*, 2013). For this reason, it is imperative that holistic approaches are employed in order to fully understand trait variation of interest; thereby ensuring relevant loci are elucidated upon genetic mapping.

1.2 Water use in agriculture

Less than 5% of the total volume of water that occupies the surface area of the Earth is considered freshwater; as such the vast majority of water on Earth is not suitable for use by the agricultural sector. Furthermore, most freshwater is unobtainable since it is contained with glaciers and ice caps. Estimates suggest that we presently have access to less than 30% of the total freshwater, either from lakes, rivers, or aquifers (Rijsberman & Mohammed, 2003; Hess *et al.*, 2014)

The finite nature of freshwater resources has meant that large water withdrawals have had enormous environmental and ecological consequences. For example, abstractions from the Yellow River in China have resulted in regular instances of zero flow, where the river runs dry before reaching the sea. This has detrimentally impacted upon both estuarine and costal ecosystems (Cong *et al.*, 2009). The need to address this and other related concerns has been recognized by the Chinese government, who have set a 20% reduction target for nationwide agricultural water use by 2020 (Morison *et al.*, 2008). However, given the importance of this river and other freshwater resource for supporting agriculture in China, it

is vital that any reductions in water use do not penalize food security. This is a concern that humanity has regularly been made aware of due to disastrous experiences in the past. There are numerous historical instances where the strong positive relationship between water availability and food production has been hugely damaging to civilizations (Reviewed in: Lamb *et al.*, 1995)), for example it is now clear that the fall of the Roman Empire was to a large degree a result of poor grain yields following short-term climate variability and concurrent low freshwater availability (Dermody *et al.*, 2014). Centuries on, we are still all too familiar with this relationship as evident from the simultaneous rates of drought and famine in Eastern and Central Africa (Morison *et al.*, 2008).

Current rates of abstraction of freshwater are unsustainable and will likely yield continued societal and environmental impacts (Rijsberman & Mohammed, 2003). Improving our ability to extract further sources of water would only serve as a short-term fix and would likely amplify the problem in the future, especially given projected global population increases (Cohen, 2003) and the increasing threat of climatic incidences of drought in the future (Patz *et al.*, 2014; Ray *et al.*, 2015). As our population size continues to grow, there will be increasing demands for water from the municipal and industrial sectors as well as from agriculture. These demands will be even greater as developing countries continue to seek improved quality of life for their citizens. In order to achieve both water use sustainability and food security, it is vital that we begin to increase the efficiency of water use within agricultural systems. This is an essential priority for regions where freshwater is in scarce supply and where anthropogenic climate change is set to exasperate the problem, however it should unquestionably be treated as a global problem. Parry *et al.*, (2005) succinctly describes how this problem must be addressed through a dynamic approach where

expertise in agronomy, engineering, crop breeding, plant physiology, genetics, and molecular biology are harnessed in a cooperative fashion and are combined to improve both crop and soil management and to develop new crop varieties through breeding and biotechnology.

Agriculture is enormously dependent on consistent and reliable supplies of water for crop production. Despite the imperative nature of water, it has been estimated that approximately 90% of the water required by many crops is not biochemically harnessed for growth or productivity (Morison *et al.*, 2008). That is not to say that the ~90% that is not employed for carbon fixation is wasted, since it is still important for other processes, such as canopy cooling (Kim *et al.*, 2006). Instead it suggests that there is room for improvement in terms of advancing the economy of water use in agriculture. Alongside improvements to general agronomic practice, the development of elite crop varieties that are able to produce improved and stabilized yields with reduced water inputs is a clear avenue through which this target may be reached. Achieving this goal will require cooperation between scientists working at the molecular and whole-plant level, as well from crop physiologists and breeders (Parry *et al.*, 2005; Morison *et al.*, 2008).

1.3 Assessing plant performance under the context of reduced water availability

1.3.1 Plant water use

Photosynthesis necessitates the exchange of water for carbon dioxide in order to synthesize sugars, which are essential for growth via the process of respiration. It is in this manner through which water is most commonly associated as a limiting factor of growth and

productivity (Passioura, 2010). Additionally, the pivotal role water plays in the maintenance of turgor pressure is also a means through which it is considered a limiting factor. This is due to turgor pressure controlling multiple growth and productivity dependent processes such as stomatal opening, phloem transport, and cell elongation (Pritchard, 2001). In order to appropriately understand how reduced water availability impacts upon plant growth and productivity, and consequently identify the most efficacious target traits for gene discovery, it is important to first briefly consider water movement through plants.

The morphology of plants reflects their need to obtain and transport water, whilst also balancing the loss of this water in exchange for the absorption of carbon dioxide for photosynthesis. This is demonstrated by exploratory roots systems which facilitate the abstraction of soil water, by xylem vessels which have virtually no resistance and thus permit the translocation of abstracted water, and by stomatal apertures which enable the evaporation of water out of photosynthetically active leaves in exchange for carbon dioxide (Kramer & Boyer, 1995). Plants rely on physical forces for the transport of water from roots to the atmosphere, thereby circumventing the need for active energy expenditure. However, associated processes, such as the development and maintenance of morphological structures necessary for water movement, do require a source of energy. The well characterized mechanisms of diffusion, bulk flow, and osmosis combine to achieve the movement of water through the plant, and represent an organismal process that is analogous to a continuous hydraulic system (Steudle, 2000).

The above-described hydraulic system connects the water surrounding the root system with the atmospheric water vapor surrounding the leaves. The loss of water from the leaves via transpiration through stomata is regulated by guard cells, which function to adjust

the size of stomatal apertures. This adjustment is driven by the photosynthetic demand for carbon dioxide and is achieved via the active regulation of guard cell turgidity (Vatén & Bergmann, 2012). As water evaporates out of the leaves via stomatal apertures from mesophyll cell walls, negative pressures are generated within the apoplasitc water. These pressures are transmitted through the continuous hydraulic system to draw further water up through the xylem, which is replaced by further soil water via root absorption (Kim *et al.*, 2014). If there is not enough soil water available to keep up with photosynthetic demand, plants must elicit appropriate responses, these are described in the preceding sub-chapter.

1.3.2 Plant responses to drought stress

One of the fundamental issues with assessing the response of plants to drought stress is the definition and significance of drought itself. For example, a farmer is likely to consider drought stress in terms of growing seasons, and will be concerned with how yield is limited by water availability. Conversely, a molecular physiologist may be more interested in transcriptome changes of plants that are not watered for a few days compared to those that are (Passioura, 2004). These sorts of disparities in timescales, interests, and target goals are likely the crux of the relative lack of cooperation between crop breeders and plant scientists. However, this deficiency in collaboration does not necessarily reflect a lack of usefulness in basic plant science research, rather it hints at the necessity to ensure that studies of the response of plants to drought simulate conditions that ensue in the field. This will help to ensure that traits that are measured and/or genes that are discovered can contribute to improved agronomic performance (Parry *et al.*, 2005; Morison *et al.*, 2008).

For the purposes of this discussion, drought is referred to as a situation where a soil-water deficit exists, which consequently causes a plant-water deficit (Kramer & Boyer, 1995).

Plant drought responses are dependent on multiple factors, such as the intensity of the soil-water deficit, the length of time for which a deficit exists, and the plant developmental stage at which the deficit occurs and/or persists (Chaves *et al.*, 2009). Regardless of these dependent variables, it is stomatal activity that is predominantly affected by drought stress (Osakabe *et al.*, 2014) When a soil-water deficit exists the previously described continuous hydraulic system is disrupted, since water lost through transpiration will not be replaced (Kim *et al.*, 2014). In response to such deficits, the turgor pressure of guard cells is actively reduced so that stomatal apertures close to reduce transpiration (Osakabe *et al.*, 2014; Dodd & Ryan, 2016) This reaction allows for the maintenance of turgor pressure, which is essential for multiple processes underlying growth and productivity (Pritchard, 2001).

The maintenance of plant water status through a reduction in stomatal conductance results in a very important trade-off, where the amount of CO_2 that can be obtained is diminished, which in turn reduces the rate of photosynthetic assimilation. If photosynthesis is constrained for long enough, growth and productivity will eventually be impacted upon (Chaves *et al.*, 2009; Osakabe *et al.*, 2014; Dodd & Ryan, 2016). Moreover, if drought-induced stomatal closure co-occurs with sustained periods of elevated irradiance, leaves will be subject to elevated levels of incident energy compared to the CO_2 available for photosynthetic assimilation, which can lead to photoinhibition (He *et al.*, 1996; Demmig-Adams & Adams, 2006; Chaves *et al.*, 2009).

In general, drought-induced reductions in stomatal conductance are facilitated by associated regulatory events, which involve abscisic acid (ABA) signaling and ion transport. Ion transport systems function to adjust turgor pressure in guard cells, which in turn open (ion influx) and close stomata (ion efflux). Adjustments in ion transport systems that facilitate

ion effluxes are a result of a cascade of physiological changes that are typically triggered by the rapid production of endogenous ABA upon the detection of drought (Pinheiro & Chaves, 2011; Osakabe *et al.*, 2014).

With respect to Arabidopsis, we presently have a good understanding of the molecular mechanisms and signaling pathways that underlie the above described physiological responses to drought stress. For example, the *9-cis-epoxycarotenoid dioxygenase* (*NCED3*) gene is rapidly induced by drought stress and is involved in ABA biosynthesis in a vascular tissue-specific manner (Jensen *et al.*, 2013). Synthesized ABA is transported to guard cells via passive diffusion and by specific transporters, such as those belong to the localized ABC transporter family (Kang *et al.*, 2011) The accumulation of ABA induces pathways that produce reactive oxygen species (ROS), which in turn increase the concentration of cytosolic Ca^{2+} ions (Kwak *et al.*, 2003). This activates anion channels and the associated anion efflux depolarizes the plasma membrane, which reduces the activity of inward K+ channels, but increases that of outward K^+ channels, such as *GUARD CELL OUTWARD RECTIFYING* K^+ *CHANNEL* (*GORK*; (Hosy *et al.*, 2003)). The efflux of anions and K^+ from guard cells reduces their turgor, thereby inducing stomatal closure and preventing water loss via transpiration (Osakabe *et al.*, 2013).

1.3.3 Drought resistance

Plant performance under the context of reduced water availability has traditionally been assessed under the context of drought resistance. Assessing plant performance in this manner is complex due to the disparity that exists amongst different practitioners with respect to what drought is and why it is important (Passioura, 2004). Further convolution exists here due to the constantly evolving definition of drought resistance, however multiple

prominent figures concerned with understanding plant performance under the context of reduced water availability have called for the adoption of Jacob Levitt's original explanation of drought resistance (Blum, 2005, 2009; Morison *et al.*, 2008; Juenger, 2013). Levit (1972) defines drought resistance as the ability of a plant to resist drought through dehydration avoidance (drought avoidance) or desiccation tolerance (drought tolerance), both of which are described in brief below.

1.3.3.1 Drought avoidance

Drought avoidance is an umbrella term that defines mechanisms that facilitate the ability of a plant to maintain water status, i.e. turgor pressure, or cellular hydration during a period of reduced water availability (Blum, 2005). As previously described, the maintenance of water status facilitates the continuation of photosynthetic dependent processes, such as growth and productivity, albeit it a reduced rate compared to optimal conditions. In general, plants achieve dehydration avoidance by enhancing their ability to capture soil moisture, by reducing water loss, and by osmotic adjustment (Blum, 2005; Dodd & Ryan, 2016).

Variation in root systems almost exclusively defines variation in the ability to capture soil water. In general, long and more expansive root systems are able to penetrate deeper and gain access to a greater volume of soil water (Blum, 2005; Zhan *et al.*, 2015). Despite the initial attractiveness of utilizing deep root systems as a target trait for crop improvement, there are certain agronomic settings where root length is irrelevant or even productivity limiting (Reviewed in: Lynch, 2015). For example, cereal varieties that have root types that would normally grow deeper, i.e. those possessing crown roots, are often not able to penetrate through dry top soil; as such shallower root types are preferred here and often grow deeper (Asseng *et al.*, 1998). Furthermore, there is also evidence to suggest that a

trade off with reproductive biomass accumulation can occur when diverting resources toward root growth (Maire *et al.*, 2009). Additionally, the acquisition of key nutrients, such as phosphorous, is massively reduced with deeper roots as they tend to exist in shallower soils (Ge *et al.*, 2000; Bishopp & Lynch, 2015). Deep root systems can undoubtedly improve drought avoidance, given appropriate soil conditions; however, this can have important consequences for above-ground biomass accumulation.

As well as being achieved by improved acquisition of water, elevated drought avoidance may also be achieved by reducing water use. Reduced drought avoidance in this sense is predominantly achieved through restraining stomatal conductance (Pinheiro & Chaves, 2011). Variation in water use is also a factor of variation in associated morphological structures (Blum, 2005, 2009). For example, only a very small proportion of the net radiation that loads onto plants is actually used to facilitate photosynthesis, with excess radiation increasing plant temperature. This in turn requires plant cooling via transpiration (Kim *et al.*, 2006). It is possible for plants to reduce water expenditure for cooling purpose by increasing plant albedo, which is the general term for describing the spectral reflectance capability of plants, or more commonly crop canopies (Singarayer *et al.*, 2009). Numerous surface-based structures enable albedo, with the most effective for reflection of un-harnessable radiation being epicuticular wax and plant glaucousness (Holmes & Keiller, 2002). Water use can also be curbed via reductions in plant size and leaf area. Reductions in total size and in leaf areas are often associated with smaller leaves, early flowering, and reduced tillering in cereal species, which are not necessarily desirable agronomic characteristics, but do improve drought avoidance nonetheless (Mitchell *et al.*, 1998; Blum, 2005).

Plants may also achieve drought avoidance through osmotic adjustment, which is a term to describe the cellular uptake of water in order to recover turgor pressure. If a soil-water deficit is severe enough to disrupt plant-water relations, there will be a substantial decrease in the extracellular water potential. This means water within cells will leave via osmosis, thus turgor pressure and all the processes that rely on it will be disrupted (Buckley, 2005). Osmotic adjustment can help to alleviate this disruption by allowing for the intracellular accumulation of solutes, mainly proline and sugars, called osmolytes. Crucially, these solutes reduce intracellular water potential, so that water refluxes into the cell and turgor pressure is readjusted (Pritchard, 2001; Dodd & Ryan, 2016).. There have been no documented cases of yield penalties existing due to increased osmotic adjustment capabilities (Morgan, 2000; Chimenti *et al.*, 2002; Moinuddin & Khanna-Chopra, 2004). For this reason, it could be argued that improving drought avoidance through osmotic adjustment may be the most efficacious manner through which to improve drought resistance in crops.

1.3.3.2 Drought tolerance

Drought resistance may also be achieved through desiccation tolerance, i.e. drought tolerance. Drought tolerance is the capacity of a plant to sustain vital plant functions whilst in a completely dehydrated, seemingly dormant state (Blum, 2005). Apart from its importance for seed storage (Blum, 2005; Dekkers *et al.*, 2015), drought tolerance in its most commonly associated term is somewhat impractical as a target trait for crop improvement, indeed I can find no evidence of its successful application in an agronomic setting. To this end, it appears that crop plant species do not retain their drought tolerance capacities post-germination (Blum, 2005).

It is perhaps from an ecological standpoint that drought tolerance is most important, since it enables a number of plant species to persist in extreme environments and to survive extend periods of serve drought (Gechev *et al.*, 2012). For example, the desert annual *Anastatica hierochuntica* is a prime example of species capable of demonstrating drought tolerance. Commonly referred to as the Rose of Jericho, *A. hierochuntica* drops all leaves at the onset of the dry season in its native North African range, it then proceeds to curl into a tight ball. Within the ball, the fruits remain attached and protected, thereby preventing the premature dispersal of seeds (Hegazy *et al.*, 2006). Species that demonstrate this sort of extreme drought tolerance are typically referred to as resurrection plants. There has been a movement within plant research to try and understand the molecular basis of drought tolerance, with the overarching aim being to translate this into crop species; however, no tangible gains appear to have been made thus far (Farrant *et al.*, 2015).

An extensive literature review conducted by Abraham Blum (Available at: http://www.plantstress.com/File/Abiotic-stress_gene.htm) suggests that both breeding and natural selection for drought resistance has favored avoidance when adapting plants to recurrent periods of soil-water deficits. The predominant concessions to this are desiccation tolerance in seeds and resurrection plants. Despite this, a number of studies have demonstrated substantial variation in cereals with respect to the ability to utilize vegetative carbon stores for grain filling during drought stress (Reviewed in: Yang & Zhang, 2006), which is a phenomenon that is often considered an aspect of drought tolerance (Asseng *et al.*, 1998; Blum, 2005). Undoubtedly, the capacity to rely on vegetative carbohydrate stores for grain filling when rates of photosynthetic assimilation are disrupted due to drought stress is an effective means of supporting yield. Furthermore, it is understandable as to why it might

be considered an avenue for achieving agronomically relevant drought tolerance in crops, because it allows for the stabilization of reproductive biomass accumulation whilst under a dehydrated state (Blum, 2005, 2009).

1.3.4 Water use efficiency

Assessments of drought resistance tend to provide information relating to the ability of a plant to survive a period of drought stress. From an ecological perspective this is of great importance, because if a plant does not complete its life cycle it will not reproduce (Passioura, 2004). In the same vein, if a crop plant does not survive a period of drought stress it will not produce any harvestable yield. Conversely, a crop plant that is able to survive a period of drought stress is not guaranteed to produce yields that are close to maximal potential yields, indeed this rarely the case (Passioura, 2004; Blum, 2005, 2009).

Since drought resistance does not provide any indication of the eventual reproductive performance of a plant, many research groups choose to focus on the idea of WUE, since it takes photosynthetic performance or biomass accumulation into consideration (Vadez *et al.*, 2014). Like drought resistance and drought itself, WUE is a term whose definition and importance is varied. Fundamentally, WUE either assesses the ratio between transpiration and photosynthetic assimilation or between biomass (whole or harvestable yield) and water use (evapotranspiration or supplied water). For the purposes of this discussion, I will focus primarily on the ratio of the two physiological entities, i.e. transpiration (*T*) and assimilation (*A*), which receive far more research attention in the Arabidopsis community than the ratio of water use and biomass. WUE in this sense is typically measured using infra-red gas exchange analysis as leaf-level instantaneous WUE (iWUE), which is described as the ratio of carbon assimilated through carbon fixation to water lost through transpiration (Penman &

Schofield, 1951). Stomatal conductance (g_s) is also commonly employed as the denominator for iWUE, however since g_s and T are tightly linked, iWUE calculated using either is appropriate and produce strongly correlated estimates of WUE. iWUE is typically assessed by measuring gas exchange on a single leaf, however it is worth noting that platforms for assessing whole-plant gas exchange are starting to be developed (Easlon *et al.*, 2014). Such platforms should facilitate the estimation of more accurate measures of iWUE, given the heterogeneous nature of gas exchange rates over the surface of individual leaves and therefore the whole plant (Seemann *et al.*, 1987; Nardini *et al.*, 2008).

WUE is also commonly assessed as the ratio of naturally occurring carbon isotopes in plant tissues. This assessment of WUE is typically termed delta^{13}C ($\Delta^{13}C$ or $\delta^{13}C$) and was first described by Farquhar *et al.*, (1982). $\delta^{13}C$ provides an integrated assessment of WUE over the lifetime of a plant, because the RuBisCo enzyme, which is essential for catalyzing carbon fixation, discriminates against the heavier C^{13} isotope, showing preference for the lighter C^{12} isotope. However, under drought conditions this discrimination typically recedes as plants are pressed into using both isotopic variants to meet photosynthetic demands, therefore $\delta^{13}C$ becomes less negative under drought and it those plants with the least negative values that are the most water use efficient (Farquhar *et al.*, 1982; Araus *et al.*, 2002). It has recently been demonstrated that WUE can also be instantaneously assessed by combining chlorophyll fluorescence imaging and thermal imaging (McAusland *et al.*, 2013). The distinct advantage of assessing WUE through the combined imaging approach is that it is possible to perform dynamic measurements under changing environmental conditions.

Both iWUE and $\delta^{13}C$ allow for a measure of the maintenance of photosynthesis under the context of the stomatal regulation of water loss. Since photosynthesis likely defines a substantial proportion of the variation that exists for growth and productivity, advances toward understanding the molecular and physiological basis of variation for WUE (Condon *et al.*, 2004) has facilitated a more improved understanding of plant performance under the context of reduced water availability. This is especially true since drought resistance does not relate to growth or productivity (Skirycz *et al.*, 2011). Despite the variable subjective meaning of drought resistance and WUE, the two terms are often used interchangeably (Blum, 2009). To this end, the associations between drought resistance, WUE, water use, and biomass accumulation have been explored to certain degree in numerous plant species, however the results have been striking in their inconsistency and appear to hugely dependent on environmental conditions (Condon *et al.*, 2004; Blum, 2009).

Many studies have assessed the relationship between WUE, predominately assessed via $\delta^{13}C$, and grain yield in field-grown cereal species (See Condon *et al.*, (2004) for a comprehensive list). The results of these projects have been considerably variable. However, for the most part it appears the association between WUE and yield is neutral, suggesting WUE is not a suitable proxy for productivity across multiple environments (Condon *et al.*, 1987, 1993; Fishcer *et al.*, 1998; Voltas *et al.*, 1999; Merah *et al.*, 2001; Royo *et al.*, 2002). Within the Arabidopsis research community, there are a number of researchers focusing almost exclusively on $\delta^{13}C$ as an assessment of plant performance under drought stress (E.g. (McKay *et al.*, 2003; Juenger & Mckay, 2005; Masle *et al.*, 2005; Kenney *et al.*, 2014). These researchers frequently highlight the isolated example of the Drysdale wheat cultivar and the improved yield it delivers given reduced water availability (Rebetzke *et al.*,

2002). Drysdale does indeed achieve marginally improved yield performances under drought and was developed following selection for improved WUE, measured as $\delta^{13}C$, under well-watered conditions. However, this particular and well cited study highlights the extensive efforts necessary to achieve only marginal yield gains under drought. Additionally, it should be noted that measuring $\delta^{13}C$ under well-watered conditions does not provide an accurate measure of WUE, since there will be no reduction in the enzymatic discrimination of C^{13}. In general, the association between WUE and yield accumulation is unsubstantiated and appears to be very much dependent on the environment and crop in question (Blum, 2009).

There is some evidence to suggest that plants that have deep root systems, and thus achieve elevated drought avoidance, have comparatively low WUE (Kobata *et al.*, 1996; Pinheiro & Chaves, 2011). These same plants often demonstrate elevated levels of biomass accumulation, suggesting that achievement of improved aboveground biomass under drought stress is to some extent facilitated through the maintenance of transpiration and not via improved WUE (Zhang *et al.*, 1997; Tolk & Howell, 2003). This suggestion is in concurrence with the mounting evidence that suggests that alterations to WUE are predominantly a result of changes to the denominator, that is to say improvements to WUE are achieved via reducing water use, *T*, and *Gs*, not by improving *A* or biomass accumulation (Menendez & Hall, 1995; Blum, 2005; Juenger & Mckay, 2005; Easlon *et al.*, 2014). As such, it is unsurprising that selection for high WUE under drought has produced crop variants that have traits associated with reduced water use, e.g. smaller leaf areas and early flowering (White *et al.*, 1990; Ngugi *et al.*, 1994; Menendez & Hall, 1995; Sayre *et al.*, 1995; Martin *et al.*, 1999).

1.3.5 Water productivity

Steduto *et al.*, (2007) describes how the exchange of water lost via transpiration for the assimilation of carbon dioxide is the most fundamental step in water use for crop production. The net carbon gain achieved via this exchange is likely of huge importance for determining the production of biomass (Long *et al.*, 2015), thus the popularity of measuring WUE. However, for the majority of commercially cultivated crop species, harvested yield is only a portion of this total biomass (FAO, 2016). Measures of total biomass or yield produced in relation to water use are typically assessed as the ratio of dry biomass to evapotranspiration (ET; Bernacchi *et al.*, 2007; Qiao *et al.*, 2014), where ET is the sum of transpiration and evaporation from soil. This reflects the idea of water productivity as first outlined by Passioura (2004). Water productivity is defined to this end as the amount of crop yield produced per volume of water supplied, typically quantified in units of kg m^{-3} mm^{-1}. The review and meta-analysis of Steduto *et al*., (2007) builds upon the ideas first detailed by Passioura (2004) and argues that the efficiency of water use by crops would be more relevant for variable environmental conditions if water productivity were defined and measured as the amount of crop yield produced per unit of water used by the plant or plants. This newer definition of water productivity is specifically focused with the water demands of the plant or plants as opposed to the agricultural and/or environmental system because it defines water use as transpiration rather than evapotranspiration (Steduto *et al*., 2007). Consequently, improvements in plant water use in relation to biomass production are more likely to be achieved via focusing on the later definition of water productivity, since it is concerned with plant-level not system-level water use.

With respect to Arabidopsis, very few studies have attempted to determine water productivity and its phenotypic or genetic basis (Bechtold *et al.*, 2010, 2013). This is presumably due to the relatively complex nature of manually phenotyping such a trait. Of these studies, Bechtold *et al* (2010) was the first to begin to characterize the natural variation that exists for this trait in Arabidopsis. This study demonstrated that there is substantial variation for water use and water productivity between four of the most common experimentally employed ecotypes, namely Col-0, C24, Ws-0, and Ws-2, at two consistent relative soil water contents (rSWCs). Additionally, this was the first Arabidopsis-based study to determine a protocol whereby water productivity can be precisely measured.

The idea of WUE and the research that has incorporated it as a trait of interest has built upon our understanding of plant performance under the context of drought stress (Kooyers, 2015). This has been of particular interest in recent times due to the evidence that suggests a lack of relationship between drought resistance and growth (Skirycz *et al.*, 2011; Ollas & Dodd, 2016). However, only a portion of photosynthetic assimilates are translocated to reproductive sinks (Nunes-Nesi *et al.*, 2016), which likely explains the neutral relationship between WUE and reproductive performance (Condon *et al.*, 2004). It is very plausible that the idea of water productivity can build upon WUE-based studies, in a similar manner the later has progressed prior research into plant performance under the context of reduced water inputs. There are no documented instances of water productivity as a target trait for assessment in crop species, although some previous research has assessed yield as a ratio of ET (Bernacchi *et al.*, 2007; Qiao *et al.*, 2014). Research to this end could help to determine the importance of water productivity and elucidate the mechanisms that underlie any variation. In Arabidopsis, we have a basic understanding of the biological basis of water

productivity, however this understanding has provided evidence to suggest it is a highly relevant trait to measure (Bechtold *et al.*, 2010). Future Arabidopsis-based studies should aim to determine the usefulness of proxy traits, such as photosynthesis, WUE, flowering time, and vegetative biomass for predicting water productivity. Furthermore, incorporating more genotypically distinct lines may facilitate the use of forward genetic methodologies to determine the genetic basis of water productivity.

1.4 Using forward genetics to elucidate the genetic basis of natural variation

Improving the sustainability of agricultural production in the face of climate change and population growth is one of the greatest challenges of modern times. I have described how a comprehensive understanding of the physiological basis of improved water use is of utmost importance for improving economy of water use in agriculture. Beyond this, understanding the genetic basis of traits that underlie variation for economy of water use will facilitate the development of crops that help to address the challenge of agricultural sustainability (Blum, 2009). Information pertaining to genetic loci that define this variation can be harnessed for crop improvement via biotechnological approaches, e.g. through marker assisted selection (He *et al.*, 2014) or through the development of genetically modified crops (Hiwasa-Tanase & Ezura, 2016). Dissecting the genetic basis of variation of interest is achieved through forward genetics. QTL mapping and GWAS are two of the most utilized forward genetics methodologies and both are employed in this present study. The use of QTL mapping and GWAS are described in the two following sub-sections.

1.4.1 Linkage mapping

The methodological pathway to successful QTL mapping is now well established and has been succinctly reviewed multiple times (Semagn *et al.*, 2006; Cooper *et al.*, 2009; Zhang & Gai, 2009; El-Soda *et al.*, 2014; Wijnen & Keurentjes, 2014). Although specific variations do exist, the general pathway required for successful QTL mapping can be divided into six fundamental steps as follows:

1. Appropriate parental lines that demonstrate significant differences in the trait of interest are selected.
2. Parental lines are crossed to produce the F_1 generation. Desired mapping population developed through appropriate rounds of crossing and/or selfing.
3. Assessment of variation for trait of interest in the mapping population.
4. Genotyping of mapping population using molecular markers that are polymorphic between the two parental lines. Molecular makers must be uniformly distributed across the genome and provide adequate coverage.
5. Generation of linkage map based on genotypic data.
6. Combine phenotypic and genotypic data with linkage map to perform statistical analyses to detect QTL by assessing which markers and pseudo-markers are linked to the assessed variation.

1.4.1.1 Mapping Populations

There are several different types of mapping populations, which each have specific advantages and disadvantages, all of which are very well reviewed by both Zou (2009) and Würschum (2012). The simplest populations to develop are F_2 and backcross (BC)

populations, since they only require two rounds of selfing or crossing. Therefore, it is possible to rapidly progress through to genotyping, phenotyping, and QTL detection, which is especially important for species with extended generation times. However, these population types are characterized by substantial heterozygosity, which limits the ability to perform multiple rounds of phenotyping or genotyping with new markers. Double haploid lines (DHLs) and recombinant inbred lines (RILs) are popular population types that alleviate this problem since they represent true-breeding, homozygous, lines. These populations are somewhat more demanding in terms of development however, with DHLs requiring diploidization of pollen through tissue culture and RILs typically needing seven rounds of single seed descent to achieve ~100% homozygosity.

One of the fundamental issues with bi-parental mapping populations is the limited degree of variation they represent. When developing such a population, parental lines are typically selected based on one particular trait, e.g. drought resistance. Consequently, such populations are not necessarily suitable for dissecting the genomic location of QTLs for additional traits, e.g. disease resistance. Multiple groups have tried to address this population by developing QTL mapping populations that are derived from multiple parental lines. In Arabidopsis, two such populations have been developed: the multiple advanced generation intercross population (MAGIC; Kover *et al.*, (2009)) and the Arabidopsis multiparent RIL population (AMPRIL; Huang *et al.*, (2011)). The MAGIC population is more comprehensive then the AMPRIL population in the sense that it incorporates more recombination events across the population, however this is a trade off since the founders are less represented in each individual line. The detection of QTL in these populations is more complex than with traditional bi-parental populations and is explained in the associated

citations; however as with traditional QTL mapping it is possible to ascertain local founder genotypes. Additionally, it is worth noting that these population types are an advance upon bi-parental populations, and indeed GWAS population sets, in the sense that individual lines represent allelic combinations that are highly unlikely to exist in the wild (Weigel, 2012).

1.4.1.2 Detecting QTLs

Once a population has been genotyped a linkage map can be created. The linkage map acts as a road map for the mapping population and spaces the molecular markers according to genetic distances, which are estimated from recombination frequencies. The genotypic data and linkage map can be combined with phenotypic data to scan for QTLs underlying the variance of interest using specific statistical analyses.

As with the mapping population types, there are numerous methods of QTL detection which range from single marker analysis, which involves marker-by-marker *t-tests*, to relatively complex multiple QTL mapping (MQM), which involves data augmentation and movement of QTLs along sliding windows in order to select the QTL model that explains the most variation (Arends *et al.*, 2010). For the most part, the most commonly employed methods of QTL detection through linkage mapping are simple interval mapping (SIM) and composite interval mapping (CIM). SIM improves upon single marker analyses by scanning for QTLs at pseudo-markers, which reside between actual markers at user-defined intervals, typically every 2cM. The significance of SIM QTL models is testable via regression testing, which is analogous to the single-marker method in that phenotypes are regressed on QTL models to test goodness of fit (Arends *et al.*, 2010). It is also possible to test model significance via maximum likelihood, however this has long been perceived to be

unnecessarily complex and often computationally slow (Haley & Knott, 1992; Arends *et al.*, 2010).

1.4.1.2 QTL mapping for drought resistance and WUE

QTL mapping has been consistently employed by crop scientists and breeders as means of identifying genes underlying variation for drought resistance and WUE (Section 1.3.4). The first published instance QTL mapping for drought resistance is the study of (Quarrie *et al.*, 1997), where QTL for ABA accumulation in rice leaves were mapped using an F2 population derived from drought resistant and susceptible genotypes. Subsequently there have been hundreds of drought resistance mapping studies in multiple crop species, such as wheat (Bennett *et al.*, 2012; Christopher *et al.*, 2013; Zhang *et al.*, 2013; Yang *et al.*, 2016), barley (Honsdorf *et al.*, 2014; Wehner *et al.*, 2015), oilseed (Fletcher *et al.*, 2016), potato (Anithakumari *et al.*, 2012), and tomato (Foolad *et al.*, 2003; Arms *et al.*, 2015).

WUE is a more recent concept than drought resistance; additionally it is often a more complex trait to phenotype accurately. Consequently, WUE has received less attention from those interested in dissecting the genetic location of QTLs underlying crop performance under the context of water stress. Thumma *et al* (2001) were the first to publish an account of successful QTL mapping for WUE, where they identified multiple QTL underlying WUE measured as $\delta^{13}C$ in the forage legume Stylosanthes. As with drought resistance, there have latterly been many more studies delineating genetic regions associated with WUE in crops through QTL mapping (Chen *et al.*, 2012; Aprile *et al.*, 2013; Adiredjo *et al.*, 2014; Marguerit *et al.*, 2014; Kaminski *et al.*, 2015).

Arabidopsis has served as a highly useful model for QTL mapping for drought resistance and WUE. Indeed, multiple specific genes have been elucidated as responsible for many of these important QTL (Lefebvre *et al.*, 2009; Kooyers, 2015). Furthermore, the contribution of these QTL to adaptive plasticity has also been achieved through reciprocal transplant experiments (Mojica *et al.*, 2016) and through testing different watering regimes (El-Soda *et al.*, 2014).

1.4.2 GWAS

QTL mapping using bi-parental mapping populations can be limited in terms of allelic diversity and genomic resolution (Borevitz & Nordborg, 2003). GWAS alleviates these restrictions by providing gene-level resolution and by using populations in which commonly occurring genotypic variation can be linked to phenotypic variation. The continually falling cost of whole-genome sequencing and high-density genotyping with single nucleotide polymorphism (SNP) markers has spearheaded the use of GWAS in human disease based studies over the past decade (Ngeow & Eng, 2015). The plant research community is following this lead and taking advantage of these increasingly affordable resources, consequently GWAS is fast becoming the method of choice for forward genetic studies in plants (Brachi *et al.*, 2011; Huang & Han, 2014).

GWAS differs from traditional QTL mapping based on bi-parental populations in the sense that it uses individual representatives of multiple naturally occurring populations and/or accessions, often termed cohorts. Like QTL mapping, the fundamental aspects of GWAS have now been well reviewed (Hayes, 2013; Gupta *et al.*, 2014; Ogura & Busch, 2015). At the most basic level, GWAS are carried out using single-locus tests to identify

associations between polymorphisms and traits of interest using one-way analysis of variance (ANOVA) testing. Since GWAS is based on multiple populations/cohorts, this statistical approach can yield false positive results that arise due to population structure. Multiple methods have been proposed and successfully implemented to control for inflations of test statistics that arise due to population structure (Reviewed in: Hayes, 2013). With respect to plant-based research, the most popular of these is the efficient mixed-model association (EMMA) statistical test developed by Kang *et al* (2008). EMMA applies mixed models which correct for both population structure and genetic relatedness using kinship matrices (Kang *et al.*, 2008).

Traditional GWAS tests, including EMMA, are based on single-locus testing. However, when working with complex traits that are controlled by many large-effect loci, such approaches have been described as inappropriate and may lead to inflation of test statistics, even if there is no population structure effect (Atwell *et al.*, 2010). Segura *et al.*, (2012) proposed the use of models that incorporate multiple markers as cofactors in order to address this statistical inflation effect. Their multi-locus mixed model approach (MLMM) to GWAS is much like CIM, however they argue that the case for using cofactors when performing GWAS is even greater than it is for linkage mapping, since the confounding effects of background loci may be present across the entire genome as opposed to only locally (Platt *et al.*, 2010; Segura *et al.*, 2012). The MLMM approach to GWAS has been successfully employed to precisely map metabolic genes in Humans (Segura *et al.*, 2012), as well as flowering time and sodium accumulation genes in Arabidopsis (Segura *et al.*, 2012), and loci associated with metabolite accumulation in Tomato (Sauvage *et al.*, 2014).

1.4.2.1 GWAS for drought resistance and WUE

The increased utilization of whole genome sequencing and high-density genotyping has seen an explosion of GWAS studies in crop species, many of which have focused on drought resistance and WUE related traits. Hao *et al.*, (2011) was the first to publish the implementation of GWAS in crops, with their study that identified polymorphisms associated with drought tolerance in maize. GWAS has since been implemented in multiple studies to identify the genetic basis of drought resistance; e.g. deep root systems in rice (Courtois *et al.,* (2013) and induced early flowering in rapeseed (Wang *et al.*, 2016). There has only been one published instance so far of the use of association mapping to dissect polymorphisms related to WUE. Dhanapal *et al.*, (2015) used $\delta^{13}C$ as physical marker of WUE in soybean and identified multiple polymorphic loci associated with variation for $\delta^{13}C$ in a diverse set of soybean genotypes.

With specific reference to Arabidopsis, GWAS has thus far been twice utilized to identify loci linked to variation for drought resistance. Verslues *et al.*, (2014) published a comprehensive study whereby variation in proline accumulation in Arabidopsis was attributed to multiple genes. Indeed, this is the only GWAS study in Arabidopsis where genes have actually been validated through mutation and overexpression, providing an excellent example of a combination of forward and reverse genetics for gene discovery. Bac-Molenaar *et al.*, (2015) recently described the identification of loci liked to differential growth responses under mild drought stress. Although this is not strictly a measure of drought resistance, it has provided us with an improved understanding of the physiological and genetic basis of the maintenance of growth under drought stress, which is arguably more agronomically relevant

that drought resistance *per se.* There have been no instances of association mapping being used to identify genetic regions associated with variation for WUE in Arabidopsis.

1.5 Objectives of present study

This present study is primarily concerned with understanding the phenotypic and genetic basis of water use and water productivity in Arabidopsis using the previously described natural variation that exists within the model species. Drought resistance and WUE have received substantial research attention within the Arabidopsis community, as such we have a broad understanding of the physiological and genetic mechanisms that underlie variation for these traits. Water use and water productivity arguably pertain more agronomical relevance than the aforementioned traits, however they are comparatively understudied in Arabidopsis, as well as other model and crop species.

In light of the above, the initial aim of this research is to develop an improved understanding of the phenotypic variance that exists within Arabidopsis for water use and water productivity and whether commonly employed proxies for performance, such as flowering time and WUE, are suitable predictors of these pseudo novel traits. Additionally, this study aims to understand whether performance related to these traits as measured in controlled environmental conditions is reflected when Arabidopsis is grown in field-style environments. Based on these comprehensive phenotypic assessments of natural variation in Arabidopsis, this present study will move forward to genetic dissection of traits identified as key in relation to water use and water productivity in order to establish a better understanding of the genetic basis of improved performance under the context of reduced water availability.

2. Natural variation of water use and productivity traits in Arabidopsis subjected to two different watering regimes

2.1 Introduction

Efforts to understand plant performance under water-limited or drought scenarios have traditionally focused on studying species or genotypes that originate from dry environments and consequently possess mechanisms that confer drought resistance (Levit, 1972; Morison *et al.*, 2008). Arabidopsis ecotypes displaying responses most associated with drought resistance tend to originate from areas characterized by low precipitation (Wolfe & Tonsor, 2014). It has been argued, however, that drought resistance is not necessarily a useful target for crop improvement. This is because in situations where water deficits elicit the initiation of drought resistance mechanisms to enable plants to avoid or tolerate reduced water availability, yield is certain to be severely impacted upon regardless of the ability of plants to survive the period of drought stress (Passioura, 2004; Blum, 2005, 2009). Despite its evident facilitation of plant survival, drought resistance is now widely accepted to neither contribute to the maintenance of productivity following drought stress, nor is it understood to be associated with elevated productivity under water-replete conditions (Passioura, 2004; Blum, 2005, 2009). In addition, drought resistance conferred through numerous tolerance mechanisms has been demonstrated to be unrelated to growth under water-limited conditions in Arabidopsis (Skirycz *et al.*, 2011).

Today, WUE receives a greater proportional share of research attention than drought resistance for assessing plant performance under water limited conditions. The fundamental reason for this is that WUE provides a measure of the maintenance of photosynthesis during the active regulation of water loss (Adiredjo *et al.*, 2014; Easlon *et al.*, 2014). Since photosynthesis is the primary determinant of plant dry weight biomass and is also believed to contribute significantly to crop yield (Long *et al.*, 2006), it is unsurprising that WUE has become an increasingly popular research focus. The availability of resources, such as water, critical for plant bioprocesses is central in determining the reproductive allocation of all organisms. Optimal availability of important resources enables plants to compensate maximal reproductive fitness (i.e. seed yield) through increasing uptake of resources. Conversely, if resource availability is poor, survival trade-off costs will result in reduced reproductive fitness (Von Euler *et al.*, 2014; Sletvold & Agren, 2015).

WUE has been successfully utilized as a target for artificial selection to optimize wheat yield in water-limited agricultural environments (Rebetzke *et al.*, 2002). More commonly, however, the relationship between WUE and grain yield is often neutral or inconsistent between trials and environmental conditions (Condon *et al.*, 2004; Sinclair & Purcell, 2005; Blum, 2009). It has been suggested that this is most likely due to the substantial variation in traits other than WUE that also strongly influence yield, such as flowering time and plant height (Condon *et al.*, 2004). Existing genotypic variation for WUE is thought to be predominantly a result of variation in water use rather than variation in photosynthetic assimilation or biomass accumulation (Blum, 2005; Monclus *et al.*, 2006; Monneveux *et al.*, 2006). Perhaps complimentary to this notion is the development of crop varieties with early flowering and smaller leaf areas following selection for improved WUE (White *et al.*, 1990;

Ngugi *et al.*, 1994; Menendez & Hall, 1995; Sayre *et al.*, 1995; Martin *et al.*, 1999). It should be noted, however, that this is in stark contrast to more recent Arabidopsis-based studies where significant variation for WUE has been positively correlated with flowering time, where late flowering ecotypes exhibit the greatest WUE (McKay *et al.*, 2003; Kenney *et al.*, 2014; Easlon *et al.*, 2014). This phenotypic and genetic correlation between WUE and flowering time is indicative of pleiotropy, which is further supported by increased WUE observed in later flowering time mutants (McKay *et al.*, 2003) and by the identification of QTLs that affect flowering time following linkage disequibrlium mapping for WUE (Juenger & Mckay, 2005; Hausmann *et al.*, 2005; Juenger *et al.*, 2010).

Substantial natural genetic variation has been demonstrated for traits associated with drought resistance and iWUE (Bouchabke *et al.*, 2008; Verslues & Juenger, 2011; Kenney *et al.*, 2014). However, the assessment of long term water use, plant performance, and water productivity has received little attention, with only a few ecotypes or mutants studied at a time (Bechtold *et al.*, 2010, 2013). Both vegetative and reproductive performance under demonstrate a heritable basis in Arabidopsis under optimal conditions (Meyer *et al.*, 2010; Gnan *et al.*, 2014). However, the extent to which variation or long term water use is driven by genetic variation is unclear. Understanding the impact of long term water use on plant performance will act as a precursor for the determination of the genetic basis of these agronomically relevant traits, and will consequently allow for reductions in long term water use without penalizing plant productivity. This present study was undertaken to obtain a quantitative genetic measure of the variation that exists for long time water use and water productivity and to assess their relationship with developmental and physiological parameters, including iWUE.

2.2 Materials and Methods

2.2.1 Plant material

Seed for all ecotypes comprising this study were obtained from the Nottingham Arabidopsis Stock Centre (NASC; Scholl *et al.*, 2000). This present study included 46 Arabidopsis ecotypes, chosen for their wide distribution across the Northern Hemisphere (Figure 2.1; Table 1).

2.2.2 Growth conditions

Plants were grown in peat-based compost (Levington F2+S, The Scotts Company, Ipswich, UK) in two different environments. In the controlled environment room, plants were kept in an 8/16h light/dark cycle at a photosynthetically active photon flux density (PPFD) of 120 μmol $m^{-2}s^{-1}$, 60% relative humidity (RH), and 23^0C. Within the glasshouse, the environmental conditions were variable, as temperature and external light cycles fluctuated during the experimental period. Supplemental lighting was maintained at a minimum PPFD threshold of ~200 μmol $m^{-2}s^{-1}$ at plant level for 12h day. Plants were kept fully saturated except during the different watering and drying regimes described below. In both environments and at all experiment stages, plant positions were randomly changed every three days.

2.2.3 Trait parameters assessed and watering regimes

For the determination of short term water use I performed *short dehydration* (SD) experiments. Here, plants were grown in 6cm diameter pots filled with the same volume of soil. Additionally, three control pots were used to determine the mass of fully saturated and

fully dried soil. These masses were used to calculate the percentage relative soil water content (% rSWC) of all pots based on fully saturated (100% rSWC) and fully dried (0% rSWC) control pots. Fifteen biological replicates of each ecotypes were grown for each SD experiment.

At 54 day's all plants were watered until fully saturated (~100% rSWC). At 55 days, excess water was withdrawn and pot weight was recorded daily, to calculate rSWC. Plants were left to progressively dry to the point where the average rSWC of all plants of each ecotype was 20%, at which point all plants of that ecotype were re-watered and transferred from the controlled environment room to the glasshouse. Within the glasshouse all plants were kept well-watered and allowed to set seed, during which time flowering time and the number of rosette leaves at bud initiation was monitored. Upon the opening of a plant's final flower it was bagged to capture all seeds. Once bagged, plants could dry down for harvesting. The mass of the reproductive biomass components (seed yield and chaff (stalks and silique pods)) was assessed as was the mass of the predominant vegetative biomass component (rosette) following a drying down period of one week within a 60ºC oven. Flowering time and biomass components were determined for all 15 biological repeats of each ecotype.

Prior to the initiation of the SD experimental period, the rosette area of four biological repeats of each ecotype was determined using a digital image of individual plants with the software package ImageJ (http://www.imagej.nih.gov/ij/) image analysis software.

Photosynthetic snapshots were taken through infra-red gas exchange analysis using portable infra-red gas exchange systems (CIRAS-2; PP Systems, Amesbury, MA, USA), to obtain an operational measure of the photosynthetic performance of ecotypes during the SD

period. All snapshots were performed on fully expanded upper rosette leaves. Snapshot photosynthetic gas exchange measurements were taken at three points during the SD period of four randomly selected biological repeats of each ecotype; when the average rSWC of a particular ecotype was 100% (start of SD for all ecotypes), 50%, and 20%. These measurements were taken within the above-described controlled growth room, at steady-state rates of A, g_s, and at current atmospheric CO_2 and a PPFD of 150 µmmol m^{-2} s^{-1}. As well as A and g_s, these measurements allowed us to determine operational values for T, also termed evaporation (E), iWUE (calculated as A/E), and intracellular CO_2 concentration (C_i).

For the determination of long term water use I performed a *continuous watering* (CW) experiment (This experiment was performed with assistance from Dr Ulrike Bechtold, Professor Philip Mullineaux, and Mrs. Sue Corbett). Here, plants were grown in 8cm diameter pots. All pots were filled with the same volume of soil and a known mass of plastic beads were placed on top of the soil to minimize evaporation from the soil. Additionally, a 5ml pipette was inserted into the soil to allow for the addition of precise volumes of water daily, the mass of the pipettes was also recorded. As with the above-mentioned SD experiment, three control pots were again setup in the same manner as the experimental pots to allow for the determination of the pot weights at 100% and 0% rSWC. Plants were initially germinated in the controlled environment room and transferred into individual experimental pots two weeks after sowing. All plants were kept well-watered for a further four days in the controlled environment room. After these four days, plants were transferred to the glasshouse where they were soaked up to ~100% rSWC. Following soaking, all excess water was withdrawn and pot weight was monitored daily. Each pot weight was maintained at 40% rSWC by adding the volume of water that was calculated to be required to reach 40%

once rSWC fell below this level. Pots were maintained at 40% rSWC because this sufficient for detecting variation in water use and also it does not initiate drought resistance mechanisms, nor does it impact photosynthetic capacity or leaf water potential (Bechtold *et al*., 2010, 2016) Three control pots with no plants were also monitored and watered daily in the same fashion during this period and allowed for the determination of average soil evaporation which was used to adjust the long term water use calculations. During the period of daily watering, flowering time and the number of rosette leaves at the point of bud initiation was recorded. Long term water use (ml $plant^{-1}$) was calculated as the total volume of water added to each pot until the opening of the final flower. At this point the plant was bagged and daily watering ceased. Biomass components were harvested and weighed as with the SD experiment. Water productivity (WP) was determined as the amount of seed produced per unit of water used (mg ml^{-1}), additionally measured water use efficiency (mWUE) was determined as the total amount of above ground biomass produced per unit of water used (mg mL^{-1}).

For the CW, 15 biological repeats of each ecotype were watered daily. Additionally, flowering traits and the biomass components traits were determined for all biological repeats.

As well as measuring long term water use as part of the CW experiment, I also calculated it as part of the SD experiment. Here, the short-term water use of individual plants was multiplied by the point of bud initiation, i.e. flowering time, to give calculated water use (cWU). Calculated estimates of WP (cWP) and WUE (cWUE) were also made by dividing cWU from the amount of seed produced per plant and the amount of above ground biomass.

2.2.4 BIOCLIM climatic data

The point of origin of all 46 ecotypes was determined as the latitudinal and longitudinal coordinates from where they were collected according to The Arabidopsis Information Resource (TAIR; Rhee *et al.*, 2003). These coordinates were employed to extract the 19 biologically relevant climatic variables corresponding to these points from the BIOCLIM database (Hijmans *et al.*, 2005). Climatic parameters were extracted at a 2.5 minute resolution using the getData() function from the R package 'raster' (Hijmans & van Etten, 2012).

2.2.5 Data analysis

Unless stated, all statistical analyses were performed within the R software environment for statistical computing and graphics (R Development Core Team, 2008). Due to both space limitations and because of the complexity of measuring many of the phenotypic parameters assessed, this study was temporally divided into seven experimental blocks. Every experimental block contained the Col-0 and C24 ecotypes and between three and nine additional ecotypes. Shapiro-Wilks tests were performed for all parameters to test for normal distributions. For all parameters either parametric one-way analysis of variance (ANOVA) or non-parametric Kruskal-Wallis (KW) tests were performed (depending on the distribution of the parameter of interest) for means comparisons of all ecotypes across experimental blocks, as well as for just the Col-0 and C24 ecotypes across experimental blocks.

From the above comparison of means testing we detected significant experimental block effects. In order to account for these effects we obtained best linear unbiased

predictors (BLUPs) of the ecotype means for each parameter by using the function ranef() from the 'lme4' R package (Bates & Maechler, 2009) on appropriate general linear mixed models (GLMMs) as described in Merk *et al* (2012). BLUPs are typically used as breeding values and provide robust estimates of the genotype effect on a particular parameter, whilst accounting for random effects (Lynch & Walsh, 1998). To make BLUP values more understandable, I calculated the estimated mean values of each phenotypic parameter by adding the BLUP value to the population mean of each trait. Estimated means were utilized for all subsequent comparison and correlation/regression analyses.

I estimated the broad-sense heritability (H^2) of all phenotypic parameters, where H^2 is equivalent to the ratio of genetic variation (V_G) to phenotypic variation (V_P), i.e. V_G/V_P. The following GLMM was employed using lmer(): Y = E + B + Residual (Error) variance, where Y represents the parameter of interest and both E (Ecotype) and B (Experimental block) are treated as random effects. The model allowed for the determination of both V_G and V_P for each trait. To test the significance of heritability for every trait an analysis of variance (deviance) table was computed for a GLMM including ecotype as a random effect, as in the model described above, and a GLMM not including ecotype, i.e. Y = B + Residual (Error) variance. A significant difference between the models is indicative of significant H^2.

To further test the effect of genetic variance on observed phenotypic variance, I calculated the coefficient of genetic variation (C_{VG}), often termed and herein referred to as evolvability. Evolvability was determined for each trait parameter as $\left(100 \times \sqrt{\mathrm{V_G}}\right)/\bar{\mathrm{x}}$, where $\bar{\mathrm{x}}$ represents the mean of the parameter of interest. The definition and theoretical basis of C_{VG} is contested (Pigliucci, 2008), however it principally provides a measure for the potential for any given trait to respond to selection.

To test for associations between phenotypic traits we calculated genetic correlations between all pairwise traits as Pearson's product moment correlation coefficients between the estimated means of all pairwise trait interactions for inference of statistical significance. A sequential Bonferroni correction method was adopted for the multiple Pearson's product moment correlation analyses (Bonferroni, 1936). This estimate of genetic correlation displayed a highly significant positive correlation ($P < 0.001$, $R2 = 0.76$) with the "rG" method which is also commonly employed (McKay et al 2003; 2008). rG was calculated as: $\frac{\mathrm{Cov_{AB}}}{\sqrt{\mathrm{V_{GA}} \times \mathrm{V_{GB}}}}$, where COV_{AB} represents the phenotypic covariance between any two traits (A and B) and $V_{GA/B}$ represents the among-ecotype genetic variance for those individual traits from which phenotypic covariance was estimated.

In order to reduce the dimensionality of the trait data set, principal component analysis (PCA) was performed. Due to technical issues certain trait parameters were not assessed for all of the ecotypes. PCA requires complete data for all lines (ecotypes) and all parameters (traits) incorporated into the analysis, for this reason those ecotypes with missing data for certain parameters where not incorporated into the PCA. 14 of the 46 ecotypes had missing data at certain points, as such these ecotypes were removed from the PCA. Additionally, the dimensionality of the climatic dataset was also reduced through PCA. For consistency and comparison purposes, only the climatic data from the 32 ecotypes used in the trait PCA were used in the climatic PCA. For both PCAs, the Kaiser-Guttmann criterion was adopted for determining significant principle components (PCs; Kaiser, 1960). This method is based on the eigenvalues of each PC, with only those PCs with eigenvalues above the mean of all PCs being considered significant.

To select a subset of the 46 ecotypes comprising the SD experiment to bring forward for the CW experiment, hierarchical clustering was performed using Ward's minimized variance as implemented using the hclust() function within the R base code (R Development Core Team, 2008).

The phenotypic plasticity of selected traits was calculated as the relative trait range index (RTR; Valladares *et al.*, 2006) across the two watering regimes. Briefly, RTR was calculated as the difference between the mean in watering regime 1 (SD) – mean in watering regime 2 (CW), divided by the maximum observed mean values. RTR consequently ranges between -1 and 1, with positive values indicating a trait difference between the two watering regimes in the expected direction, as water use and productivity were expected to be higher in the SD experiment.

2.3 Results

Phenotypic variation associated with water use and productivity was examined for a set of Arabidopsis ecotypes originating from a range of different habitats across the Northern Hemisphere. This range in sites of origin ensured that the ecotypes of interest have been subject to varying climatic histories during their evolutionary histories (Fig 2.1; Table 2.1).

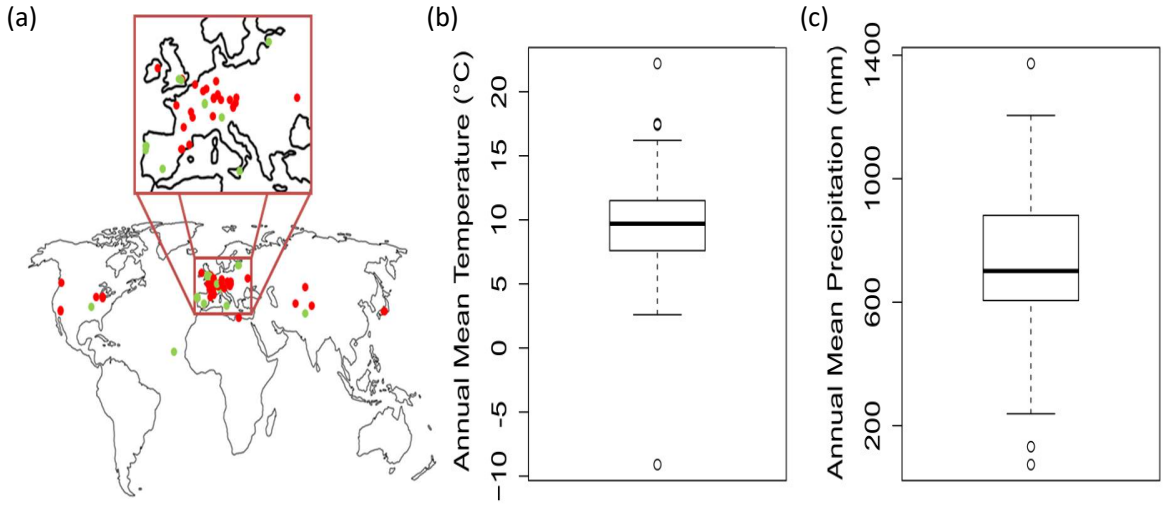

Figure 2.1. Geographic distribution and climatic history variation of the 46 Arabidopsis ecotypes comprising the present study. (a) Geographic site of origin all ecotypes comprising this study. Ecotypes indicated with green dots were included as part of both the short dehydration and continuous watering experiments. Those indicated by red dots were included as part of the short dehydration experiment only. **(b)** The variation in annual mean temperature at the site of origin of the 46 ecotypes. **(c)** The variation in annual mean precipitation at the site of origin of the 46 ecotypes.

Table 2.1. The native names (Ecotype), geographic site of origin (Location), TAIR germplasm ID, and the experimental blocks containing each ecotype comprising this study

Ecotype	Location	Germplasm ID	Experimental block(s)
An-1	Anderlecht, Belgium	CS944	4
Bay-0	Bayreuth, Germany	CS954	4
Bor-4	Borky, Czech Republic	CS28093	8
Br-0	Bruno, Czech Republic	CS994	8
Bur-0	Burren, Ireland	CS1028	8
C24 *	Coimbra, Portugal	CS906	All blocks
CIBC-5 *	Ascot, UK	CS78894	2
Col-0 *	Columbia, USA	CS1092	All blocks
Ct-1 *	Catania, Italy	CS1094	8
Cvi-0 *	Cape Verde Islands	CS902	2
Ei-2 *	Eifel, Germany	CS1124	1
En-2	Frankfurt, Germany	CS1138	6
Est-1 *	Vagli, Estonia	CS28244	3
Fei-0 *	St. Maria d. Feiria, Portugal	CS28250	7
Got-22	Gottingen, Germany	CS76884	3
Gu-0	Guckingen, Germany	CS28330	5
HR-5 *	Ascot, UK	CS76514	8
Kas-1	Kashmir, India	CS903	8
Knox-10	Knox, USA	CS22410	4
Kondara *	Kondara, Tajikistan	CS916	3
Kz-9	Karagandy, Kazakhstan	CS76537	3
Ler-0 *	Landsberg am Lech, Germany	CS20	6
LL-0	Llagostera, Spain	CS799513	5
Lp2-6	Lipovec, Czech Republic	CS28480	6
Lz-0	Lezoux, France	CS28482	3
Mrk-0	Markt-Baden, Germany	CS28498	5
Mt-0	Martuba/Cyrenaika, Libya	CS28502	6
NFA-10	Ascot, UK	CS28533	4
NFA-8	Ascot, UK	CS28532	6
Pna-17	Benton Harbor, USA	CS28647	2
Pu2-23	Prudka, Czech Republic	CS28655	3
Pu2-7	Prudka, Czech Republic	CS28654	3
Ra-0	Randan, France	CS28665	4
Ren-11	Rennes, France	CS77211	3
RRS-7	North Liberty, USA	CS28713	4
Shakdara	Shakdara, Tajikistan	CS929	4
Se-0 *	San Eleno, Spain	CS28726	3
Sq-1	Ascot, UK	CS28746	6
Ts-5	Tossa del Mar, Spain	CS1558	3
Tsu-1	Tsushima, Japan	CS1640	8
Van-0	Vancouver, Canada	CS1584	6
Wei-0	Weiningen, Switzerland	CS3110	2
Ws-0	Wassilwskija, Russia	CS1602	3
Ws-2	Wassilwskija, Russia	CS2360	3
Wt-5	Wietze, Germany	CS6896	8
Yo-0	Yosemite, USA	CS6901	4

An SD experiment, where pots could dry from 100% to 20% rSWC, was performed using all 46 ecotypes. This facilitated the calculation of short term water use as the slope of the linear regression of the rate of drying. Plants were kept well-watered before and after the SD experimental period. Traits evaluated as part of this experiment included seed yield, chaff (stalk and silique biomass), rosette biomass and area, flowering time, and leaf number at flowering (Fig 2.2).

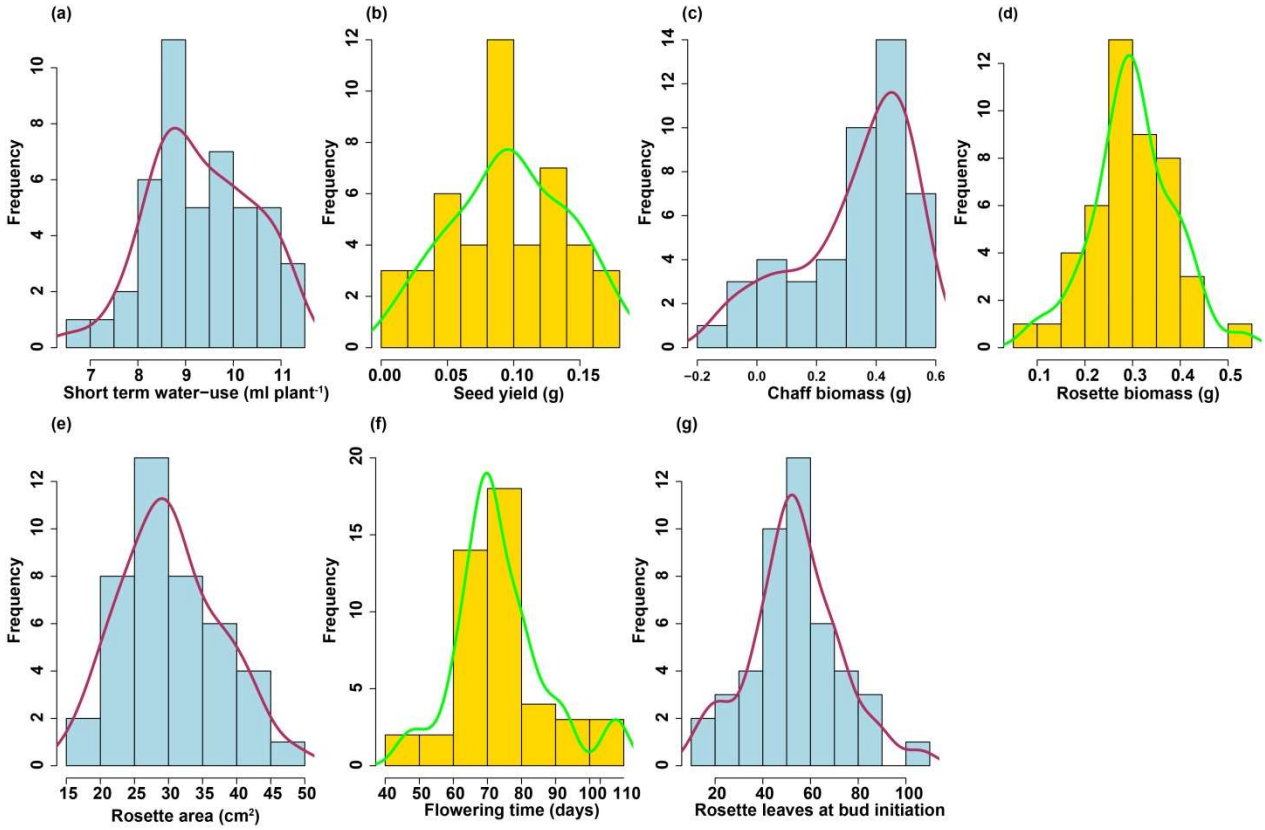

Figure 2.2. Natural variation for key physiological and developmental traits assessed during the short dehydration experiment. (a) Short term-water use. **(b)** Seed yield. **(c)** Chaff biomass. **(d)** Rosette biomass. **(e)** Rosette area. **(f)** Flowering time. **(g)** Number of rosette leaves at bud initiation. The variation displayed in each histogram is the estimated mean of all ecotypes included as part of this study.

In addition to the SD experiment, I performed a CW experiment on 14 specifically selected ecotypes. With this experiment, rSWC was maintained at moderate drought levels (~40–45% rSWC; Bechtold et al 2010, 2013) throughout the lifecycle of the plant. These ecotypes were selected following hierarchical clustering of the 46 ecotypes from the SD experiment based on Euclidean distances computed on a combination of the variation for short term water use and seed yield (Fig 2.3). The associated dendrogram was divided into 5 distinct clusters and two or three representative ecotypes were selected from each cluster for the CW experiment. This proceeding experiment allowed us to accurately measure and compare short term and long term water use, as well as other plant performance parameters.

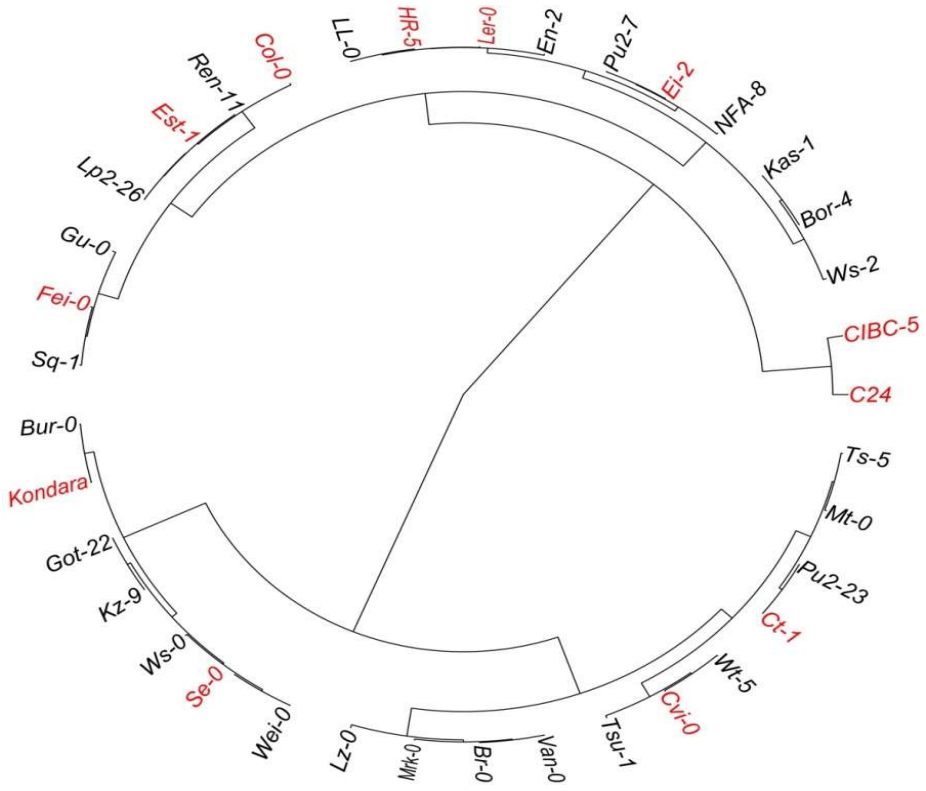

Figure 2.3. Fan dendrogram displaying the hierarchical clustering of the 46 ecotypes. Clustering is based on short term water use and seed yield. Those ecotypes highlighted in

red comprising each of the five main clusters were incorporated as part of the CW experiment (with two additional ecotypes; Kin-0 and Br-0)

The 46 ecotypes comprising the SD experiment were grown in the same growth room over a period of two years. Environmental conditions were kept constant throughout all temporally divided blocks. In total, seven temporally unique experimental blocks were analyzed, with Col-0 and C24 included as part of each to assess variation due to experimental differences. We identified significant block affects for all traits when comparing all ecotypes across all blocks, but also when just comparing Col-0 and C24 across all blocks (Table 2.2).

Table 2.2. Comparisons of means testing for all traits assessed as part of the short dehydration experiment. Depending on the distribution of the parameter of interest, either one-way ANOVA or Kruskal-Wallis tests were performed. Comparisons are made based on the means of all ecotypes, just Col-0, and just C24. Where relevant, the rSWC at the time of the assessment of a trait is given as a percentage before the trait. Significance is indicated at *** $p < 0.0001$, ** $p < 0.01$, and * $p < 0.05$. N.s. denotes that the means were not significantly different between experimental blocks.

Trait Parameter	P value		
	Between experimental blocks (All ecotypes)	Between experimental blocks (Col-0)	Between experimental blocks (C24)
Rosette area (mm^2)	< .0001 ***	< .05 *	< .0001 ***
Drying rate (ml H_2O day^{-1})	< .0001 ***	< .0001 ***	< .0001 ***
~100% A (µmol m-2 s-1)	< .0001 ***	< .0001 ***	< .05 *
~100% E (mmol m-2 s-1)	< .0001 ***	< .0001 ***	< .0001 ***
~100% WUE (µmol/mmol m-2 s-1)	< .0001 ***	< .0001 ***	< .01 **
~100% Gs (mmol m-2 s-1)	< .0001 ***	< .0001 ***	< .01 **
~100% Ci (µmol mol-1)	< .0001 ***	< .01 **	0.13 n.s.
~50% A (µmol m-2 s-1)	< .0001 ***	< .0001 ***	< .01 **
~50% E (mmol m-2 s-1)	< .0001 ***	< .0001 ***	< .01 **
~50% WUE (µmol/mmol m-2 s-1)	< .0001 ***	< .0001 ***	< .01 **
~50% Gs(mmol m-2 s-1)	< .0001 ***	< .01 **	< .01 **
~50% Ci (µmol mol-1)	< .0001 ***	< .01 **	< .01 **
~20% A (µmol m-2 s-1)	< .0001 ***	< .0001 ***	< .0001 ***
~20% E (mmol m-2 s-1)	< .0001 ***	< .0001 ***	< .0001 ***
~20% WUE (µmol/mmol m-2 s-1)	< .0001 ***	< .01 **	< .01 **
~20% Gs (mmol m-2 s-1)	< .0001 ***	< .0001 ***	< .0001 ***
~20% Ci (µmol mol-1)	< .0001 ***	< .0001 ***	0.22 n.s.
Flowering time (days)	< .0001 ***	< .0001 ***	< .0001 ***
Rosette leaves flowering	< .0001 ***	< .0001 ***	< .0001 ***
Rosette biomass (g)	< .0001 ***	< .0001 ***	< .0001 ***
Chaff biomass (g)	< .0001 ***	< .0001 ***	< .0001 ***
Seed yield (g)	< .0001 ***	< .0001 ***	< .0001 ***
Calculated water-use	< .0001 ***	< .0001 ***	< .0001 ***
Calculated water productivity	< .0001 ***	< .0001 ***	< .0001 ***

To control for random experimental block effects, BLUPs and estimated means were obtained and calculated respectively for all parameters. Estimated means were subsequently used in regression, correlation, and principle component analyses. Pearson

product-moment correlations were calculated between arithmetic means and estimated means and demonstrated a significant positive correlation (Table 2.3)

Table 2.3. Association of estimated means, obtained from BLUPs, and true (arithmetic) means. Associations testing as achieved via Pearson's regression analysis on estimated and true means of all ecotypes for each trait assessed as part of the short dehydration experiment. Where relevant, the rSWC is given as a percentage before the trait. Pearson's correlation coefficients and p-values are given.

Trait	Pearson's correlation coefficient	P value
Rosette area (mm^2)	0.915	0.000
Short term water-use (ml H_2O day^{-1})	0.809	0.000
~100% A (µmol m-2 s-1)	0.566	0.000
~100% Ci (µmol mol-1)	0.535	0.000
~100% E (mmol m-2 s-1)	0.549	0.000
~100% Gs (mmol m-2 s-1)	0.873	0.000
~100% WUE (µmol/mmol m-2 s-1)	0.402	0.003
~50% A (µmol m-2 s-1)	0.459	0.000
~50% Ci (µmol mol-1)	0.605	0.000
~50% E (mmol m-2 s-1)	0.877	0.000
~50% Gs(mmol m-2 s-1)	0.898	0.000
~50% WUE (µmol/mmol m-2 s-1)	0.664	0.000
~20% A (µmol m-2 s-1)	0.520	0.000
~20% Ci (µmol mol-1)	0.959	0.000
~20% E (mmol m-2 s-1)	0.426	0.001
~20% Gs (mmol m-2 s-1)	0.506	0.000
~20% WUE (µmol/mmol m-2 s-1)	0.818	0.000
Flowering time (days)	0.937	0.000
Rosette leaves flowering	0.852	0.000
Rosette biomass (g)	0.716	0.000
Chaff biomass (g)	0.416	0.004
Seed yield (g)	0.723	0.000
Calculated water-use	0.873	0.000
Calculated water productivity	0.798	0.000

2.3.1 Phenotypic variation and plasticity for photosynthesis, water use, flowering time, and biomass accumulation

Snapshot gas exchange measurements were performed to assess operational rates of photosynthesis and water use in all the Arabidopsis ecotypes comprising the SD experiment. To benchmark the experiment, I checked for genetic correlations between known functionally related traits. Positive correlations between functionally related traits, such as g_s and A were observed at all measured rSWCs (Fig. 2.4), while negative correlations occurred between traits in evolutionary constraint of one another, such as iWUE and g_s (Fig. 2.4). Despite the strong genetic correlations that exists between g_s and A, and between g_s and iWUE, the relationship between A and iWUE was neutral (Fig. 2.4). This suggests that alterations in WUE are more likely to be brought about by changes in water use, as opposed to changes in the rate of photosynthetic assimilation (Figs 2.4, 2.5). It was also observed that A, g_s, and E followed a general trend of decreasing as the short dehydration period progresses, whilst iWUE increased concurrently (Figure 2.5). It is interesting to note that the reduction in variation for E and g_s that occurs at 20% rSWC was not matched by a reduction in variation for A, which suggests that many ecotypes were able to maintain relatively high levels of A despite the dynamic reduction of g_s. Furthermore, it should also be noted that this reduction in variation for E and g_s was also met by an increase in variation for iWUE, which provides further evidence to suggest that changes in iWUE were brought about predominantly by changes to g_s, and consequently E, but not by changes in A (Fig 2.5).

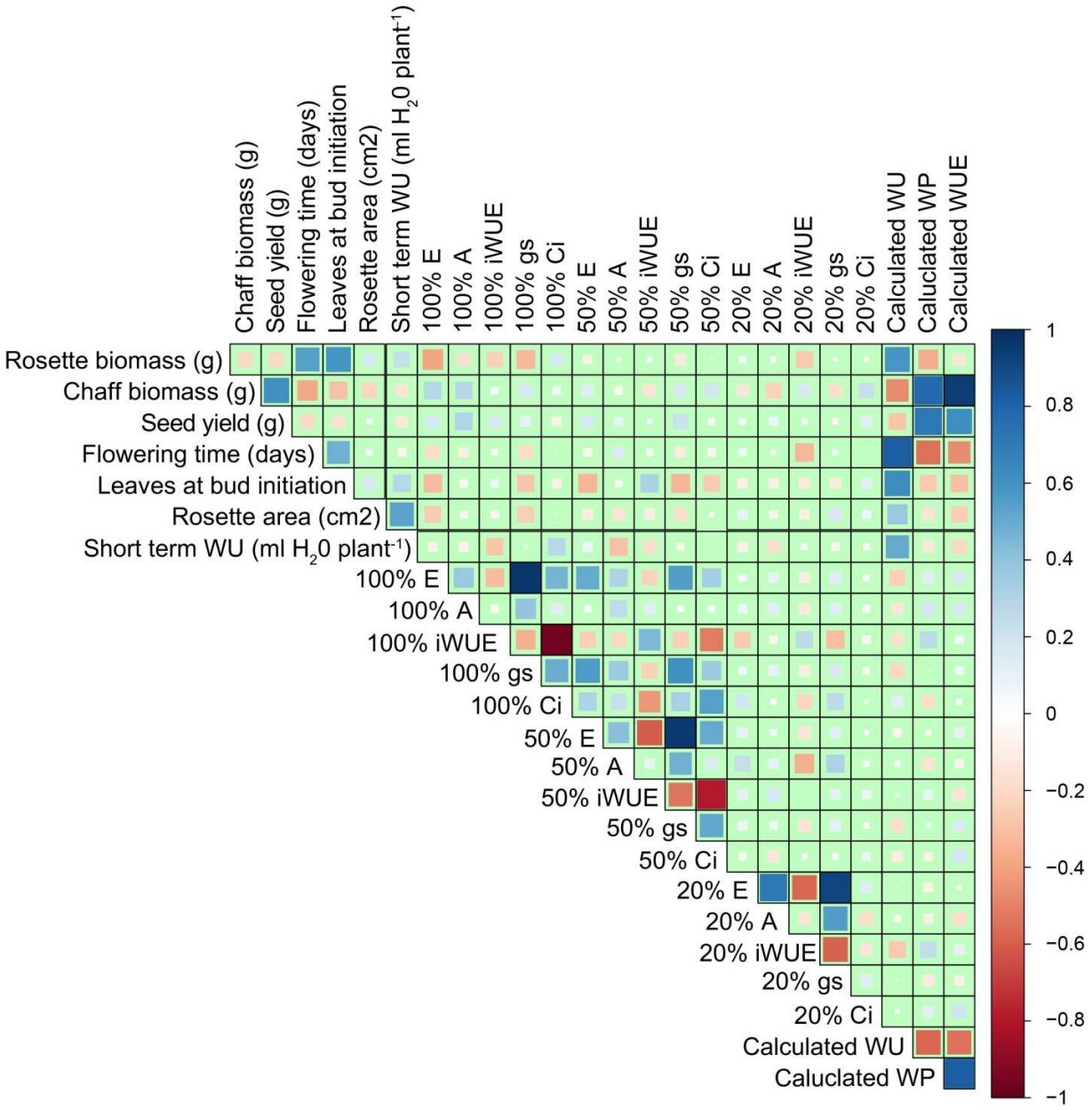

Figure 2.4. Correlation plot displaying the genetic correlations between all parameters assessed as part of the short dehydration (SD) experiment. The significance of any particular pair-wise genetic correlation is indicated by the size of the individual squares, where larger squares denote a lower p-value. The color of the square indicates the direction of the Pearson correlation coefficient (see heat bar to the right), where red denotes a

negative genetic correlation and blue denotes a positive genetic correlation. Where relevant the rSWC is indicated by a percentage at the beginning of trait names.

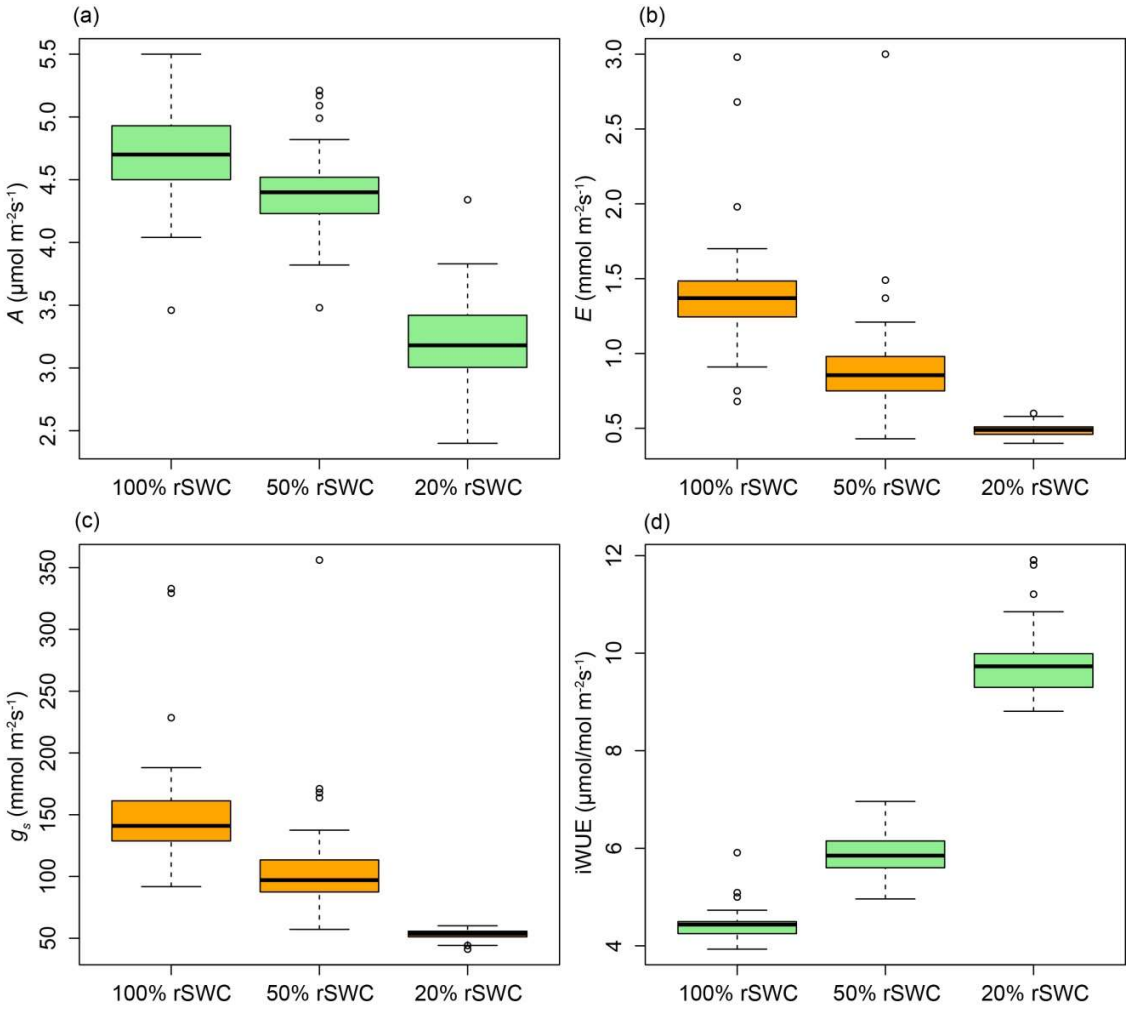

Figure 2.5. Variation in photosynthesis during the short dehydration (SD) period. (a) Variation in photosynthetic assimilation (*A*) at 100%, 50%, and 20% relative soil water content (rSWC). **(b)** Variation in evaporation (*E*) at 100%, 50%, and 20% rSWC. **(c)** Variation in stomatal conductance (g_s) at 100%, 50%, and 20% rSWC. **(d)** Variation in instantaneous water use efficiency (iWUE) at 100%, 50%, and 20% rSWC. For all plots, the bottom and top boxes denote the 25th and 75th percentiles respectively. The central band is the 50th

percentile. Whiskers extend to the most extreme data points which are no more than 1.5x the length of the upper or lower segment away from the respective segment. Circles that lie away from the whiskers denote extreme outliers. $N = 4$.

Genetic and phenotypic variances were calculated using GLMMs. These facilitated the estimation of H^2 and evolvability, also termed the C_{VG}. I observed significant H^2 for 20 of the 22 phenotypic traits measured as part of the SD experiment (Table 2.4). The H^2 of the primary fitness related traits, namely seed yield and chaff, were comparatively lower than traits known to have a strong genetic basis, such as flowering time. Both seed yield and chaff, however, displayed the greatest evolvability (Table 2.4), which indicates the ability of a population/ecotype to respond to environmental or artificial selection with respect to a particular trait (Houle, 1992; Wagner & Altenberg, 1996). This suggests that reproductive performance in Arabidopsis responded much more strongly to environmental selection, compared to the response of short term water use or flowering time for example (Table 2.4). As a continuation, I compared the effect of both watering regimes on reproductive performance, flowering time, and water use using both RTR plasticity indices (Valladares *et al.*, 2006) and the relative performance of these traits for the 12 ecotypes included in both experiments. The plasticity indices were most variable for fitness associated traits, such as seed yield and chaff, while flowering time and water use appeared to be much more stable traits across the different watering regimes (Fig 2.6). Clear genotype by environment (GxE) interactions were observed, where four ecotypes showed differences in the deviations from the average yield for specific watering regimes. Conversely, 8 ecotypes demonstrated consistently higher or lower than average yields across both watering regimes (Fig. 2.7a). Similarly, seven ecotypes showed consistently higher or lower than average water use

across both watering regimes (2.7b). It is interesting to note that those ecotypes with higher than average seed yields did not exhibit higher than average water use (Fig. 2.7).

Table 2.4. Genetic and phenotypic variation for the 24 phenotypic traits assessed as part of the short dehydration experiment. The true mean, standard error (SE), genetic variance (V_G), phenotypic variance (V_P), broad sense heritability (H^2), coefficient of genetic variance (C_{VG}), and significance of H^2 are displayed for all traits. Significance is indicated by *** = p < 0.0001, ** = p < 0.01, and * = p < 0.05. N.s. denotes that the H^2 of the associated trait was not significant.

Trait	Mean	SE	V_G	V_P	H^2	C_{VG}	Sig.
Rosette area (mm^2)	27.40	0.71	73.36	136.57	0.54	31.26	***
Short term water-use (ml H_2O day^{-1})	9.19	0.06	1.22	2.99	0.41	12.04	***
~100% A (µmol m-2 s-1)	4.50	0.80	0.25	2.72	0.09	11.01	***
~100% E (mmol m-2 s-1)	1.35	0.06	0.17	1.15	0.15	30.65	***
~100% Gs (mmol m-2 s-1)	143.66	4.80	2660.00	8581.00	0.31	35.90	***
~100% iWUE (µmol/mmol m-2 s-1)	4.65	0.17	0.83	10.67	0.08	19.64	**
~100% Ci (µmol mol-1)	297.76	2.39	99.74	1959.47	0.05	3.35	***
~50% A (µmol m-2 s-1)	4.35	0.09	0.23	3.22	0.07	11.05	***
~50% E (mmol m-2 s-1)	0.87	0.29	0.15	1.12	0.14	45.06	***
~50% Gs(mmol m-2 s-1)	102.59	3.49	2120.10	4397.70	0.48	44.88	***
~50% iWUE (µmol/mmol m-2 s-1)	6.18	0.18	0.57	11.18	0.05	12.24	*
~50% Ci (µmol mol-1)	291.84	2.51	137.00	2190.80	0.06	4.01	***
~20% A (µmol m-2 s-1)	3.09	0.07	0.24	2.31	0.10	15.68	***
~20% E (mmol m-2 s-1)	0.45	0.02	0.00	0.10	0.04	15.09	**
~20% Gs (mmol m-2 s-1)	49.43	1.89	54.37	1450.53	0.04	14.92	**
~20% iWUE (µmol/mmol m-2 s-1)	10.14	0.53	0.53	11.36	0.05	7.20	n.s.
~20% Ci (µmol mol-1)	266.59	5.02	1381.70	13979.00	0.10	13.94	***
Flowering time (days)	72.10	0.61	199.48	355.39	0.56	19.59	***
Rosette leaves at bud initiation	56.55	0.78	396.90	705.40	0.56	35.23	***
Rosette biomass (g)	0.30	0.04	0.00	1.51	0.00	0.00	n.s.
Chaff biomass (g)	0.40	0.01	0.04	0.10	0.39	49.01	***
Seed yield (g)	0.10	0.00	0.00	0.01	0.19	48.67	***
Calculated water-use	667.20	21.64	14719.00	33063.00	0.45	18.18	***
Calculated water productivity	0.12	0.01	0.01	0.32	0.32	66.40	***

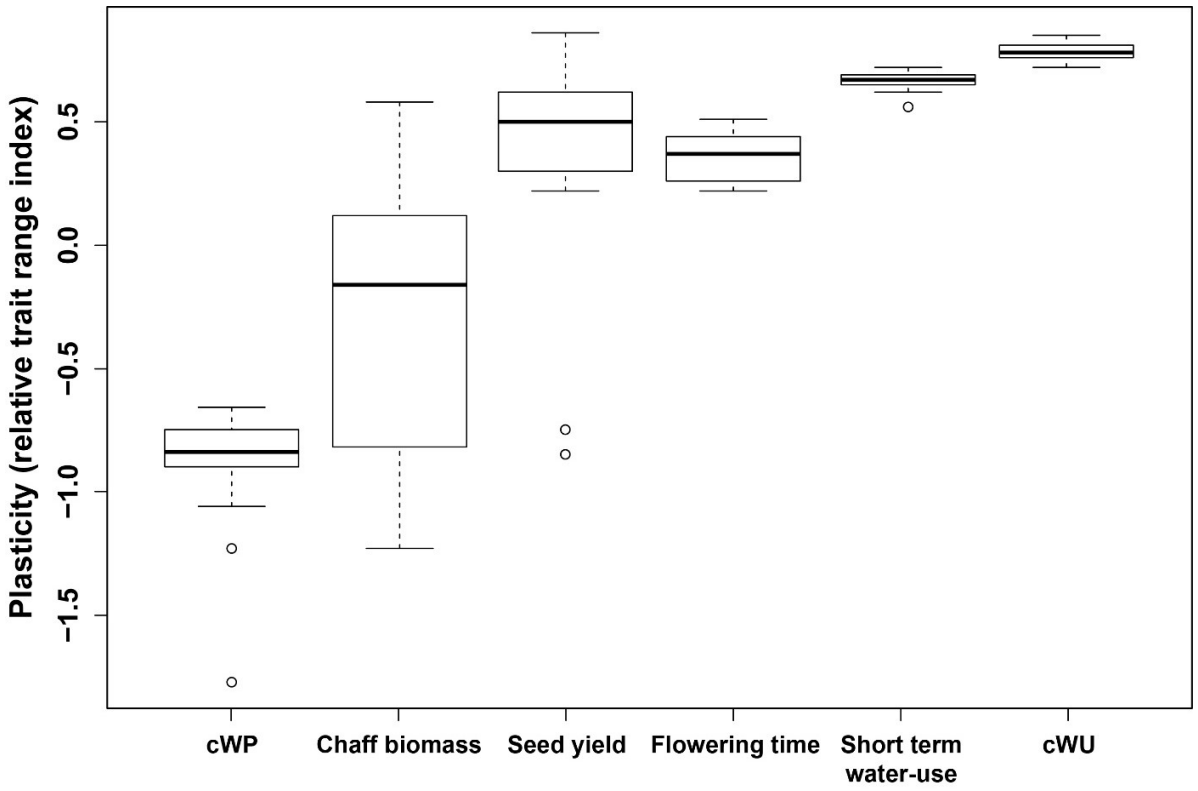

Figure 2.6. Box-and-whisker plots displaying the variation in plasticity. Plasticity is calculated as relative trait range (RTR) indices for calculated water productivity, chaff biomass, seed yield, flowering time, short term water use, and calculated water use. The bottom and the top boxes denote the 25th and 75th percentiles respectively. The central band is the 50th percentile. Whiskers extend to the most extreme data points which are no more than 1.5x the length of the upper or lower segment away from the respective segment. Circles that lie away from the whiskers denote extreme outliers.

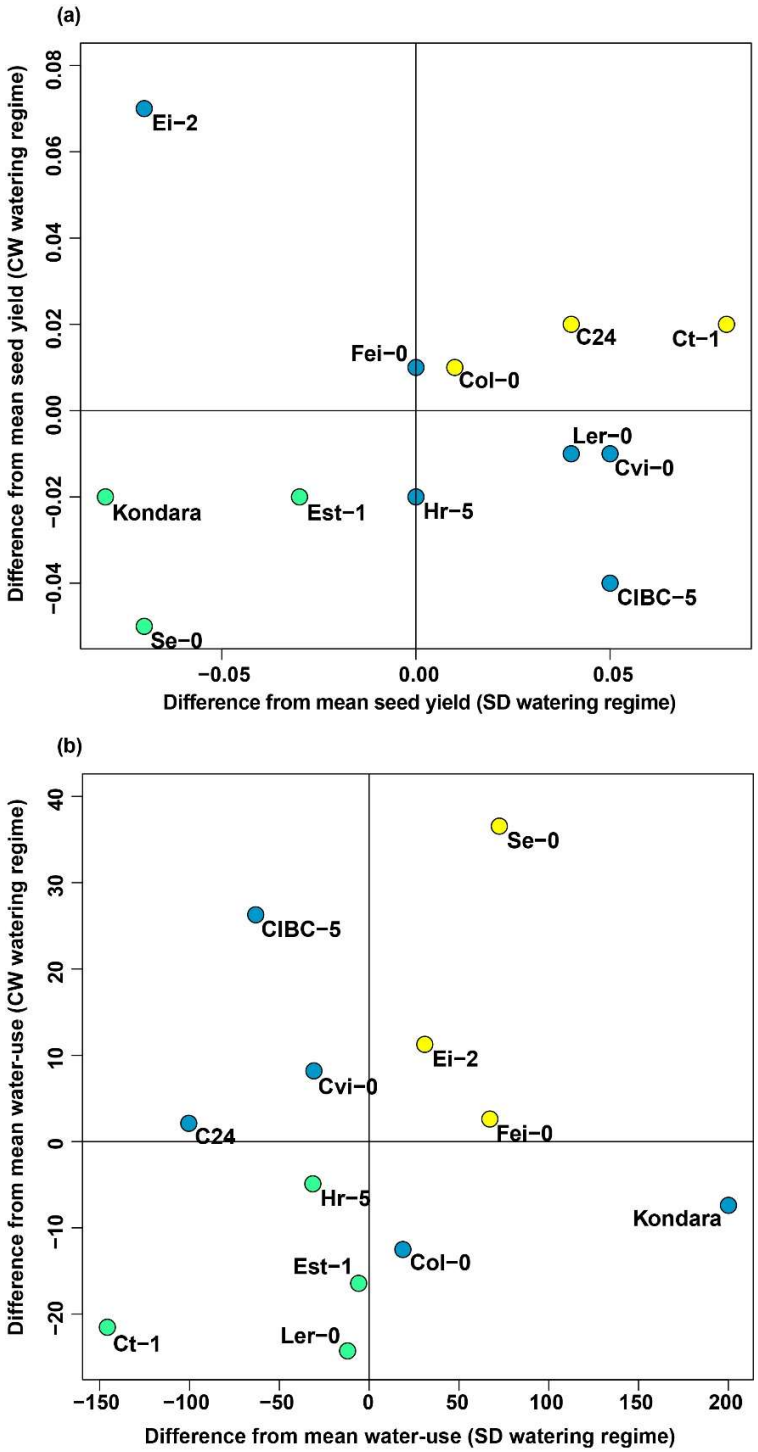

Figure 2.7. Phenotypic plasticity for seed yield and water use. (a) Association of the difference from mean seed yield of the 12 ecotypes common to the short dehydration (SD) and continuous watering (CW) experiments. **(b)** Association of the difference from the mean

calculated water use of the 12 ecotypes common to the SD and CW experiment. Green dots (non-adaptive plasticity), yellow dots (adaptive plasticity), and blue dots (genotype by environment (GxE) interactions) describe the different response to the watering regimes of the 12 ecotypes.

2.3.2 Short term and long term water use is driven by vegetative performance and flowering time

Due to the intensive nature of the CW experiment, I was unable to measure long term water use for all 46 ecotypes comprising the SD experiment. I therefore estimated long term water use by multiplying short term water use by flowering time to obtain a calculated water use (cWU) parameter. The CW experiment facilitated the accurate determination of long term as well as short term water use, and consequently allowed for the comparison of measured long term water use (mWU) and cWU parameters. A highly significant correlation was observed between both parameters, suggesting that cWU is a highly appropriate approximation of mWU for ecotypes of Arabidopsis (Fig. 2.8a). Additionally, a further highly significant positive correlation was observed between cWU from the SD experiment and mWU from the CW experiment (Fig. 2.8b), even though these were different sets of plants grown under different condition (SD – short days, CW – long days). This provides yet further support for using cWU as a proxy for mWU. I observed similar significant associations for calculated water productivity (cWP) and measured water productivity (mWP) also (Fig. 2.9).

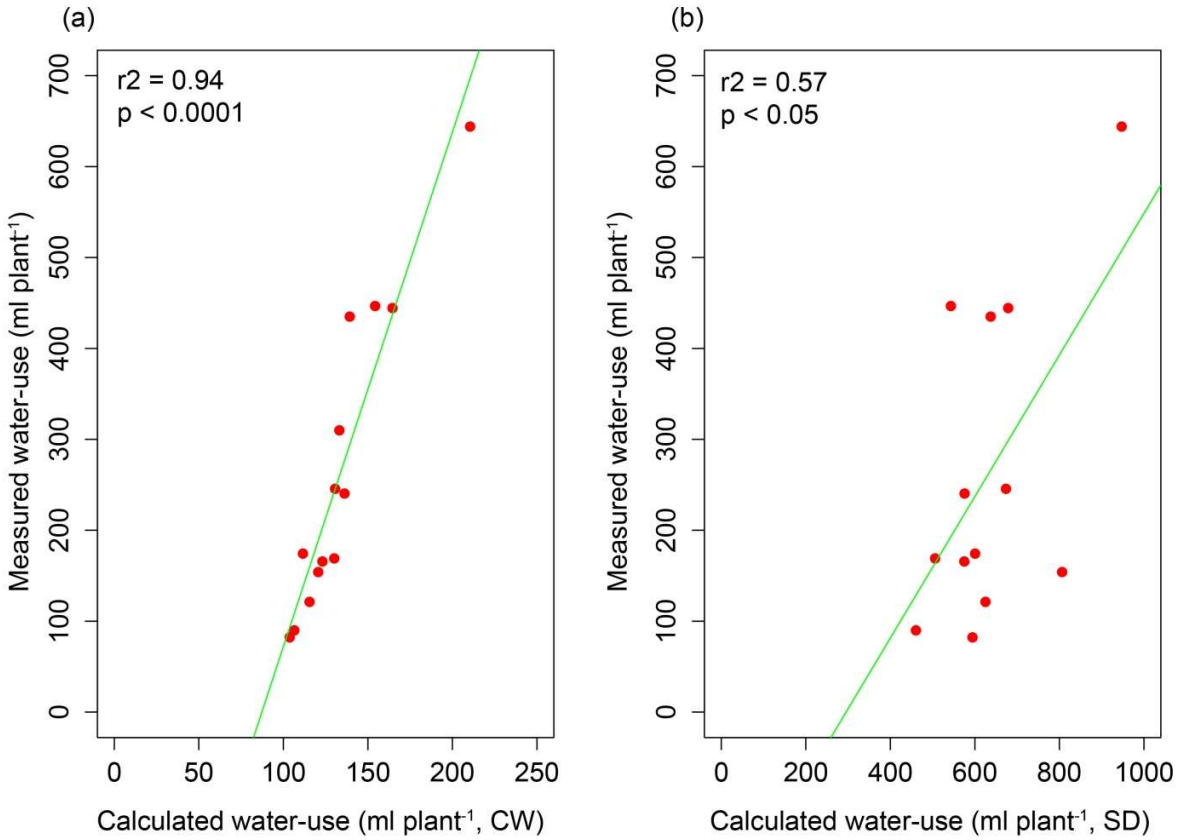

Figure 2.8. Association of calculated water use (cWU) and measured water use (mWU). **(a)** Significant positive association between calculated water use from the continuous watering experiment (CW) and measured water used from the CW experiment. **(b)** Significant positive association between calculated water use from the short dehydration (SD) experiment and measured water use from the CW experiment.. R-squared and significance thresholds are provided for both associations.

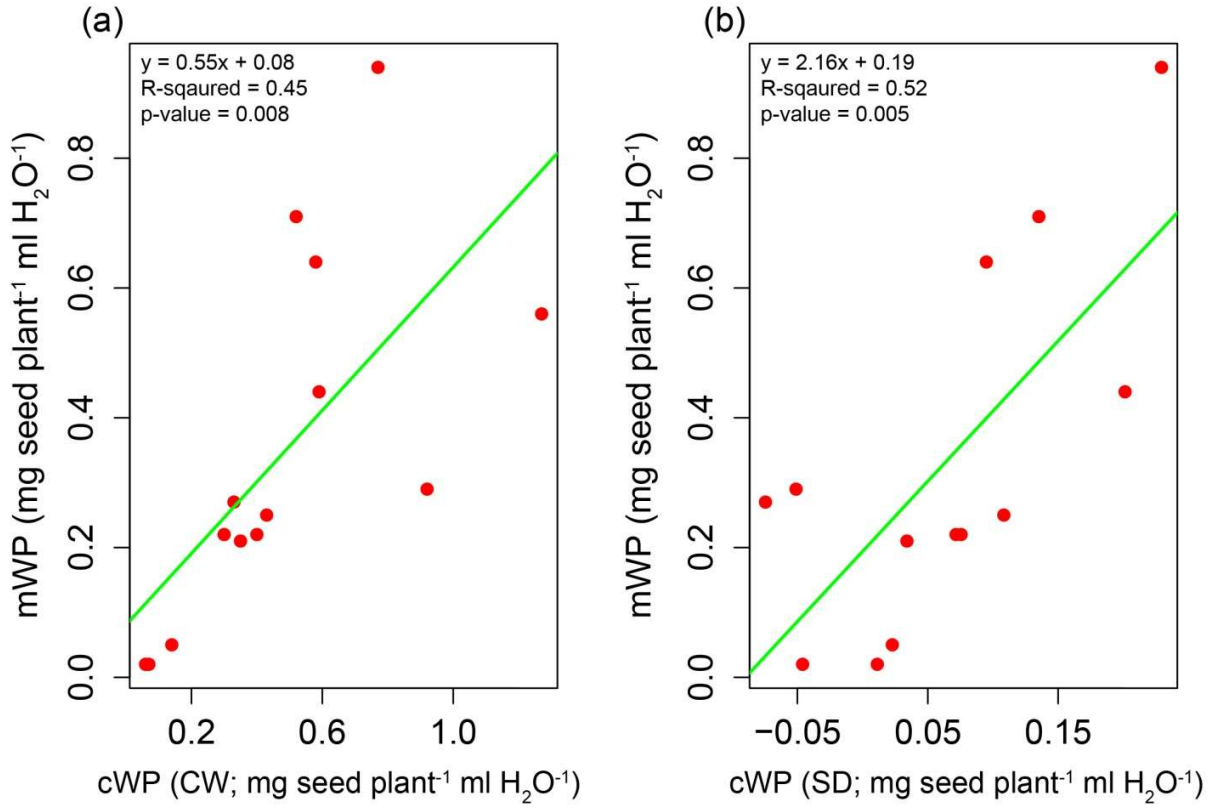

Figure 2.9. Association between calculated water productivity (cWP) and measured water productivity (mWP). (a) Significant positive association between cWP from the continuous watering (CW) experiment and water productivity from continuous watering experiment (mWP). **(b)** Significant positive association between cWP from the short dehydration (SD) experiment and water productivity from the continuous watering experiment (cWP). Linear model regression equations, R-squared values, and p-values are given for both associations.

Neither cWU nor short term water use were correlated with g_s or iWUE (Fig. 2.4), suggesting that iWUE is not an appropriate measure for water use, and that long term water use strategies in Arabidopsis do not rely on the present physiological status of the plant. Similarly, there was relationship between reproductive performance and total water use in either of the watering regimes (Fig. 2.13), suggesting that reproductive strategies of these

ecotypes were independent of their water use strategies across the two different watering regimes employed here.

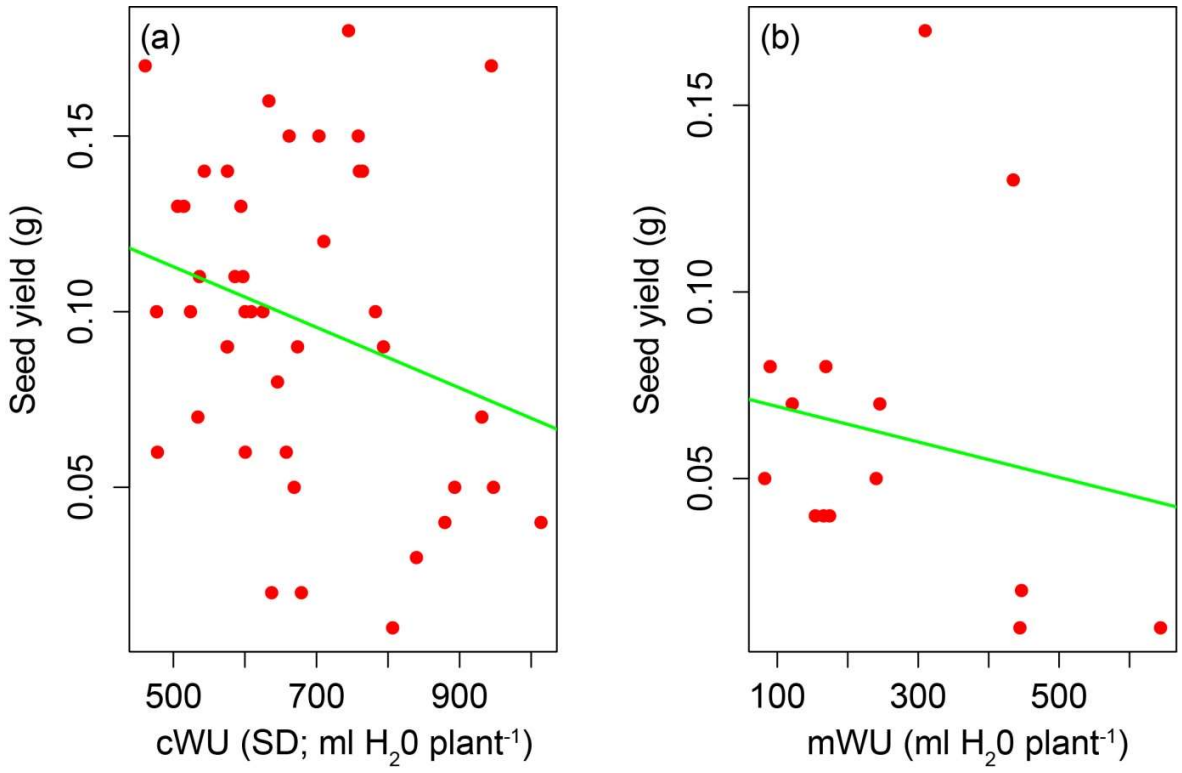

Figure 2.10. Associations between water use and productivity (seed yield). (a) Neutral association between calculated water use (cWU) and seed yield from the short dehydration (SD). P-value > 0.05. **(b)** Neutral association between measured water use (mWU) and seed yield from the continuous watering (CW) experiment. P-value > 0.05.

Water use was predominantly determined by flowering time (Figs. 2.4, 2.11) and vegetative biomass (Fig. 2.4), where bigger plants that flowered later used the most water. Flowering time was also observed to have neutral relationship with iWUE at 100% and 50% rSWC, but a significantly negative genetic correlation with iWUE at 20% rSWC (Fig. 2.4), which suggests that plants that flower later, use more water but do not compliment this with

enough of an increase in photosynthetic activity to generate improved reproductive biomass, consequently their WUE was reduced.

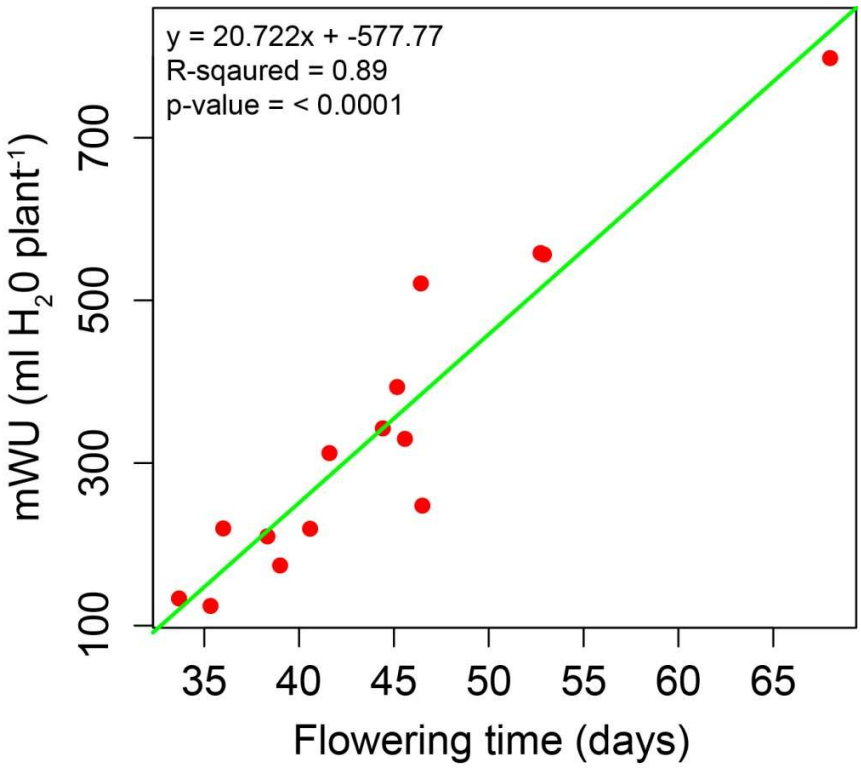

Figure 2.11. Genetic correlation between flowering time and measured water use (mWU) of the 14 ecotypes comprising the continuous watering (CW) experiment. The green line represents the linear model whose equation is given, along with the associated r^2 value and p-value.

The increase in long term water use that appears to be met with increasing flowering time (Fig. 2.11) might be predicted since a plant that lives for a longer period is likely to use more water. For this reason, I tested the variation in daily water use between the 14 ecotypes comprising the CW experiment and assess whether this was related to long term water use (Fig. 2.12). Those ecotype that use more water use daily appear to use more water in the

long term independent of flowering time, in other words to the later flowering ecotypes also appear to use more water on a day-by-day basis as well as over an extended period (Fig. 2.12.). Substantial variation was detected for daily water use, which mirrored long term water use. (Fig. 2.12b)

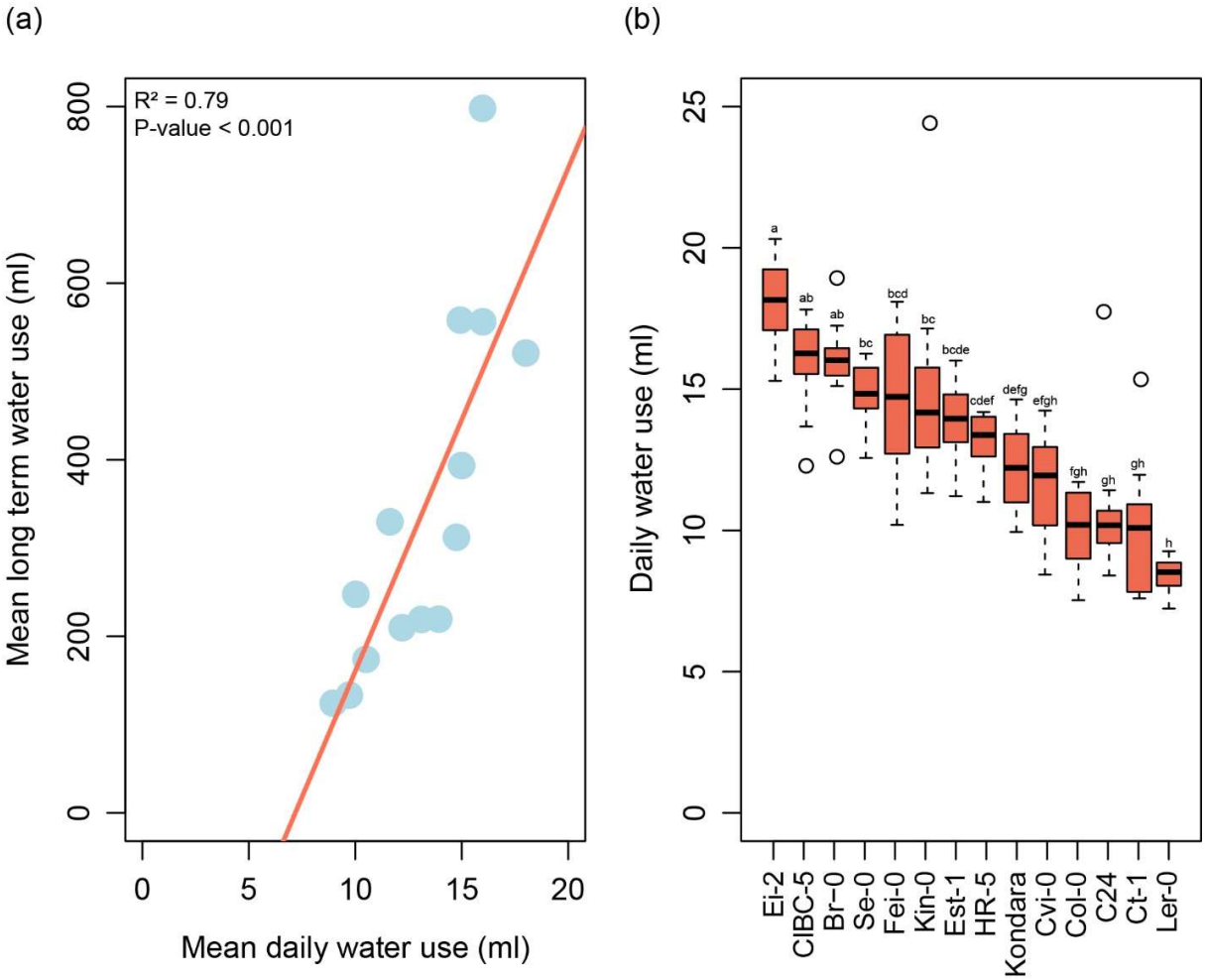

Figure 2.12. Variation for daily water use and its relationship with long term water use. (a) Relationship between daily water use and long term water use of the 14 ecotypes comprising the continuous watering experiment. The linear model of the

relationship between mean long term water use and mean daily water use is provided as the orange fit line. R^2 and P values are provided. (b) Boxplots describing the variation observed in daily water use between the 14 ecotypes comprising the continuous watering experiment. *Post-hoc* Tukey groups are provided above the boxplot of each ecotype. Those ecotypes with common letters, i.e. *post-hoc* Tukey groups, are not statistically different from one another, whereas those with no common letters are statistically different. For each boxplot, the bold line in center of the boxplots represents the median, the box edges represent the 25^{th} (lower) and 75^{th} (upper) percentiles, the whiskers extend to the most extreme data points that are no more than 1.5x the length of the upper or lower segment, and points away from the whiskers indicate extreme outliers. $N = 15$.

These results emphasize the importance of assessing long term water use. They also highlight the importance of determining water productivity, which I define as the ratio of the mass of seed produced over the amount of water used per plant. As mentioned above, this can be either measured (mWP) or calculated (cWP; Fig. 2.9). As well as water productivity, I also calculated long term water use efficiency (WUE) from the CW experiment as the total above ground biomass divided by mWU. WUE was also calculated from the SD experiment, by dividing total above ground biomass by cWU from this experiment, i.e. cWUE. The effectiveness of utilizing iWUE as a predictor of WUE or WP was explored by testing the associations of these traits at the three rSWC points at which iWUE was measured (Fig. 2.14). iWUE at 100% and 50% rSWC appeared a precise proxy for WUE from the CW experiment, but this association dissipated for iWUE at 20% *rSWC* (Fig. 2.12). Conversely, iWUE had a neutral relationship with WP from the CW experiment at all rSWC measurement

points (Fig. 2.12), suggesting that transpiration may determine total biomass, however resource allocation into yield is not supported.

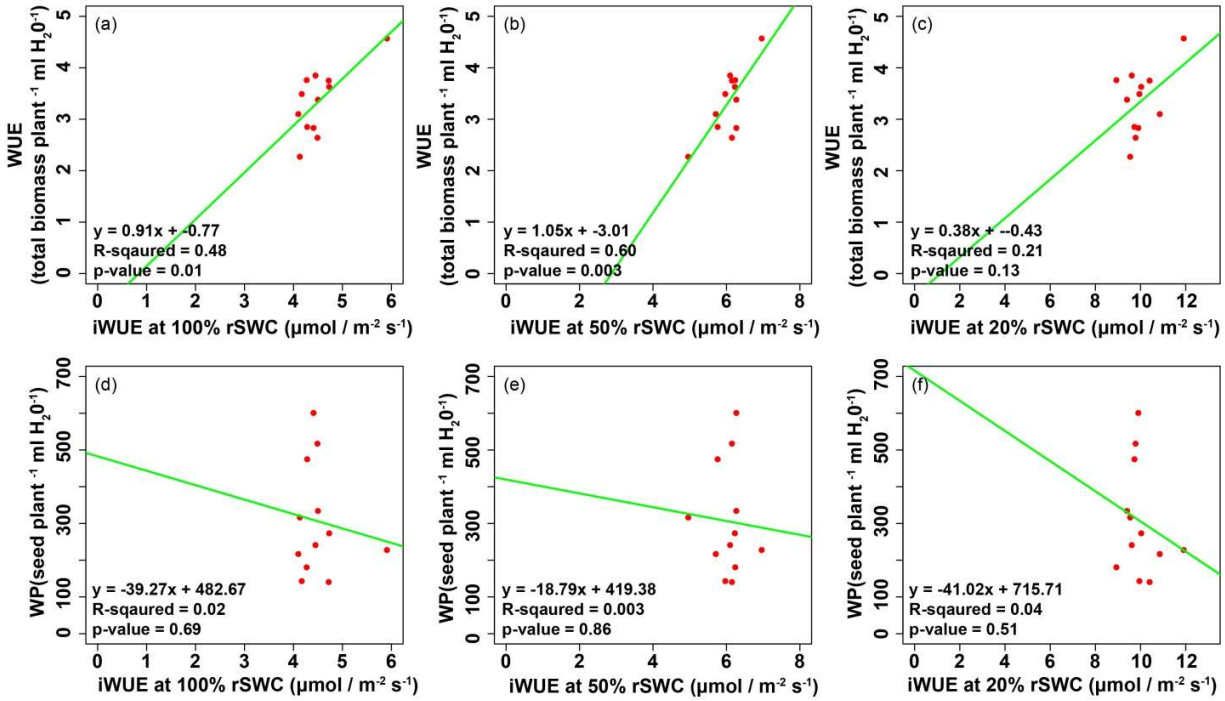

Figure 2.13. Associations between instantaneous water use efficiency (iWUE) and biomass accumulation as a factor of water use. (a) Genetic correlation between iWUE at 100% relative soil water content (*rSWC*) and long term water use efficiency (WUE). **(b)** Genetic correlation between iWUE at 50% *rSWC* and WUE. **(c)** Genetic correlation between iWUE at 20% *rSWC* and WUE. **(d)** Association between iWUE at 100% rSWC and water productivity (WP). **(e)** Association between iWUE at 50% rSWC and WP. (f) Association between iWUE at 20% rSWC and WP. For all plots, the green lines represent the associated linear models. The equations of the linear models are provided along with the associated r^2 values and p-values.

When comparing the WP of the 12 ecotypes comprising both of the experimental watering regimes there was an obvious split in terms of ecotypes displaying adaptive

plasticity, non-adaptive plasticity, and distinct GxE interactions. Four ecotypes displayed consistently higher than average WP across both environments (adaptive plasticity), seven displayed consistently lower than average WP across both environments (non-adaptive plasticity), and one ecotype displayed opposing differences away from the average WP across the environments, which is characteristic of a GxE interaction for WP (Fig 2.13).

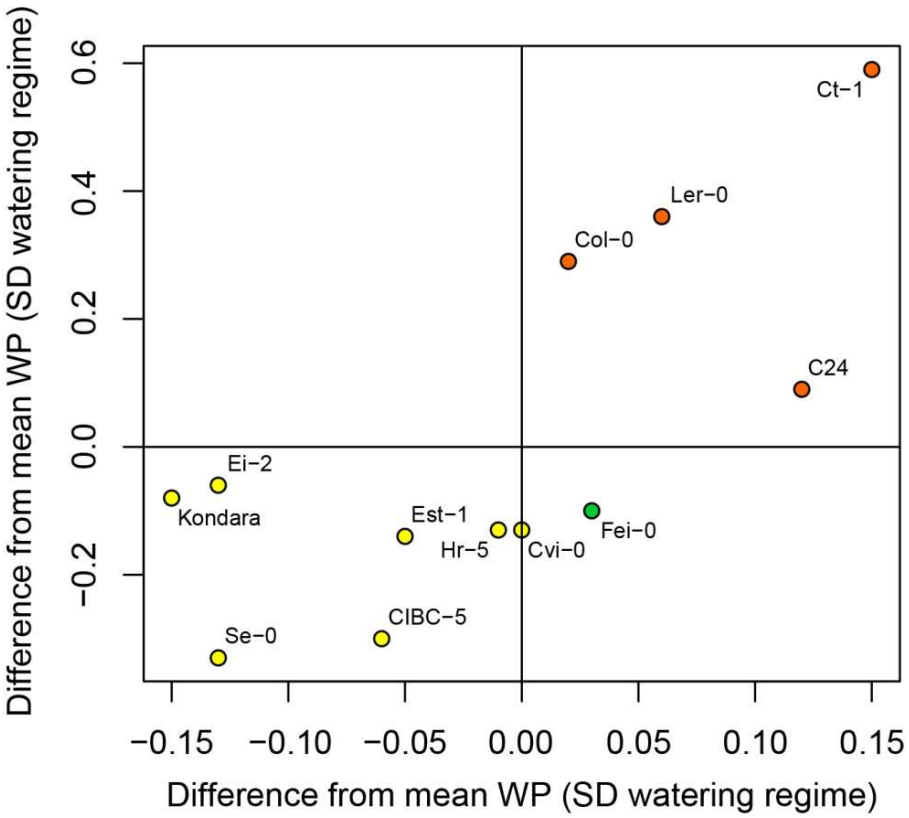

Figure 2.14. Adaptive plasticity, non-adaptive plasticity and genotype by environment (GxE) interactions for water productivity. Association between the difference from the mean water productivity (WP) of the 12 ecotypes across the short dehydration (SD) and continuous watering (CW) experimental regimes. The yellow dots denote ecotypes displaying non-adaptive plasticity for WP, the orange dots denote ecotypes displaying

adaptive plasticity for WP, and the green dot denotes the one ecotype displaying a distinct genotype by environment (GxE) interactive effect for WP.

2.3.3 Dimensionality reduction of the climate and trait space highlights the variability of life history traits associated with water use and water productivity

I measured and calculated a host of phenotypic traits that pertain to short term and long term water use and productivity, including flowering time, vegetative and reproductive biomass accumulation, and operational rates of photosynthetic parameters. No physiological or growth parameters were observed to be effective in predicting the variation observed for reproductive fitness, i.e. seed yield (Fig. 2.4). In concurrence to this, reproductive fitness displayed the greatest variation in phenotypic plasticity (Figure 2.6). Additionally, it demonstrated elevated levels of evolvability and reduced levels of heritability (Table 2.4), suggesting that reproductive fitness is much more sensitive to the environment than other physiological and/or developmental traits.

Due to the above, I hypothesized that a link may exist between climatic history at the point of origin of all ecotypes and trait performance. To this end, I obtained biologically relevant climatic data relevant to all ecotypes. To identify whether adaptive gradients existed for traits relating to water use and productivity and climatic parameters, the dimensionality of both the trait and climatic datasets were reduced through PCA. PCA of the climatic dataset resulted in five significantly relevant climatic principle component (CPCs; Fig. 2.14). Inspection of each component based on loading values facilitated the classification of

individual components with regards to the main climatic parameters contributing to each principle component (PC; Table 2.5). The first two climate principle components (CPCs) explained 56% of the total climatic variation (Fig 2.14b), where CPC1 was characterized predominantly by temperature and CPC2 by precipitation levels during the driest periods of the year (Table 2.5)

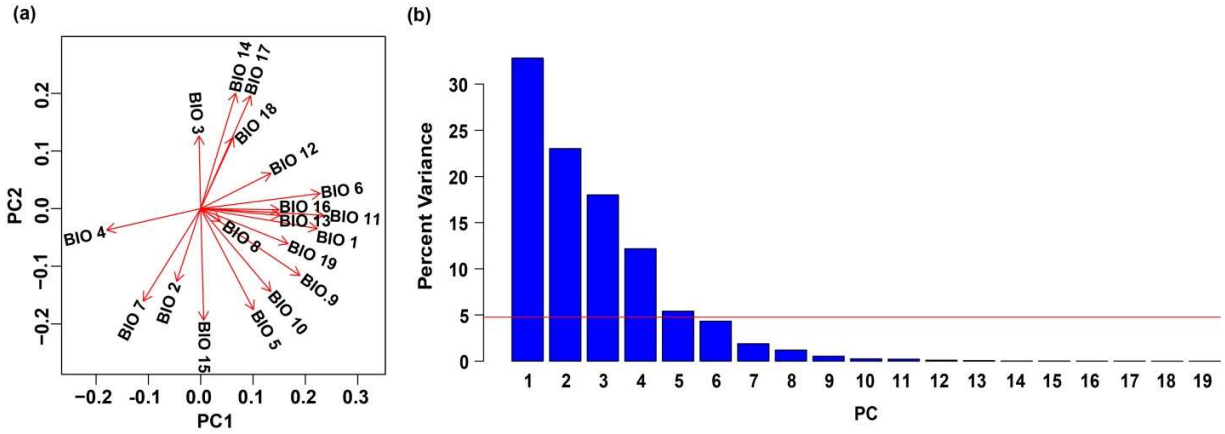

Figure 2.15. Principal component analysis of the climatic dataset. (a) Biplot displaying the loading onto climate principal component (CPC) 1 and 2 of the 19 biologically relevant climatic parameters (BIO 1-19) from the BIOCLIM dataset that correspond to the point of origin of the 46 ecotypes comprising this study. The direction of the arrow represents the association of any particular climatic parameter to both climate PC1 and climate PC2 and the length of the arrow represents the strength of that relationship. A glossary and list of definitions of all the BIOCLIM parameters is available at: www.worldclim.org/bioclim **(b)** Scree plot displaying the percentage of the total climatic variation explained by each climatic PC. The horizontal red line denotes the Kaiser-Guttmann significance criterion, where all PCs that explain variation greater than that line being considered significant.

Table 2.5. Association of each climatic parameter to the five significant climatic principle components (PCs) from the climate principal component analysis (PCA). Climatic parameters load onto each PC on a scale from -1 to 1, where loadings above 0 indicate a positive correlation and a loading below 0 indicate a negative correlation. The loading values are colored according to their association, where dark blue denotes a highly positive correlation and dark red indicate a highly negative correlation.

Climatic parameter	PC1	PC2	PC3	PC4	PC5
BIO 1 - Annual mean temperature	0.36	-0.07	-0.01	-0.02	0.31
BIO 2 - Mean diurnal range	-0.07	-0.25	-0.15	0.35	0.03
BIO 3 - Isothermality	-0.01	0.24	-0.31	0.29	0.35
BIO 4 - Temperature seasonality	-0.29	-0.07	0.16	0.24	0.13
BIO 5 - Max temperature of the warmest month	0.16	-0.34	0.26	-0.17	-0.03
BIO 6 - Min temperature of coldest month	0.37	0.05	-0.11	-0.11	0.23
BIO 7 - Temperature annual range	-0.18	-0.31	0.30	-0.05	-0.21
BIO 8 - Mean temperature of wettest quarter	0.06	-0.04	0.45	-0.06	0.21
BIO 9 - Mean temperature of driest quarter	0.31	-0.23	-0.17	-0.07	0.12
BIO 10 - Mean temperature of warmest quarter	0.22	-0.28	0.27	-0.18	0.05
BIO 11 - Mean temperature of coldest quarter	0.38	-0.02	-0.08	-0.10	0.19
BIO 12 - Annual precipitation	0.22	0.12	-0.22	-0.07	-0.43
BIO 13 - Precipitation of wettest month	0.25	-0.02	0.17	0.46	-0.07
BIO 14 - Precipitation of driest month	0.11	0.39	0.24	-0.03	-0.12
BIO 15 - Precipitaton seasonality	0.01	-0.38	-0.12	0.31	0.08
BIO 16 - Precipitation of wettest quarter	0.24	0.00	0.17	0.46	-0.15
BIO 17 - Precipitation of driest quarter	0.16	0.38	0.21	-0.04	-0.13
BIO 18 - Precipitation of warmest quarter	0.10	0.24	0.35	0.31	0.10
BIO 19 - Precipitation of coldest quarter	0.27	-0.12	-0.17	0.13	-0.56

Additionally, PCA was performed on the 23 trait parameters, again to reduce dimensionality. The trait space was reduced to seven statistically significant trait PCs (TPCs; Fig 2.15). The first two TPCs explained 40% of the total observable variation for all traits. TPC1 was primarily associated with life history traits, including water use, seed yield, water

productivity, and flowering time, whereas TPC2 was characterized principally by strong correlations with short term physiological responses to drought, i.e. g_s and iWUE (Table 2.6).

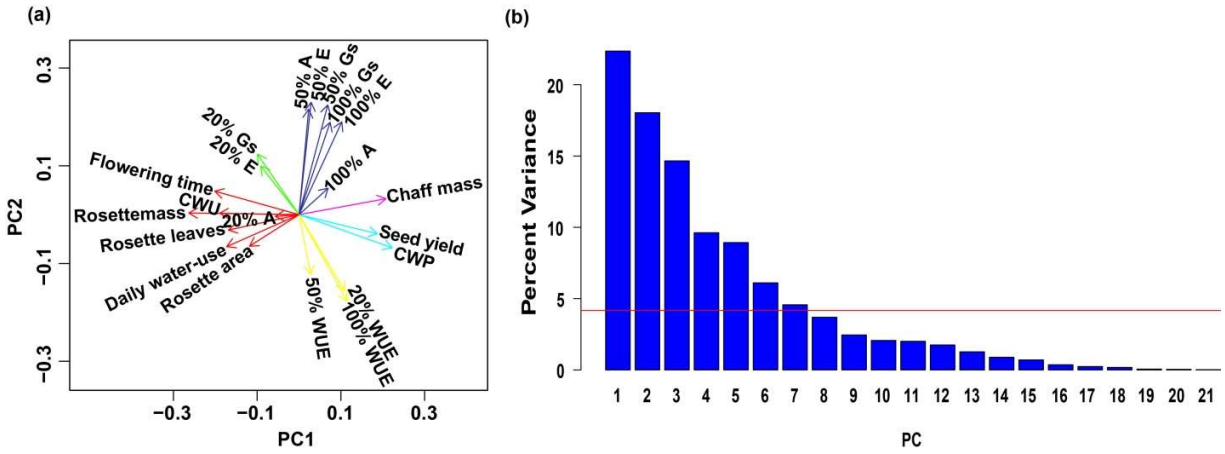

Figure 2.16. Principal component analysis of the phenotypic trait dataset. (a) Biplot displaying the loading onto trait principal component (PC) 1 and 2 of 23 key phenotypic traits. The direction of the arrow represents the association of any particular trait parameter to both trait PC1 and trait PC2 and the length of the arrow represents the strength of that relationship. **(b)** Scree plot displaying the percentage of the total climatic variation explained by each trait PC. The horizontal red line denotes the Kaiser-Guttmann significance criterion, where all PCs that explain variation above this line being considered significant.

Table 2.6. Association of each phenotypic trait parameter to the seven significant trait principle components (PCs) from the trait principal component analysis (PCA). Phenotypic trait parameters load onto each PC on a scale from -1 to 1, where loadings above 0 indicate a positive correlation and a loading below 0 indicate a negative correlation. The loading values are colored according to their association, where dark blue denotes a highly positive correlation and dark red indicate a highly negative correlation.

Phenotypic trait parameter	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Rosettte biomass	-0.29	0.01	0.08	-0.22	0.26	-0.32	-0.09
Chaff biomass	0.32	0.06	-0.10	-0.23	0.21	-0.28	-0.20
Seed yield	0.29	-0.07	-0.21	-0.24	0.29	-0.16	-0.10
Flowering time	-0.31	0.08	0.17	0.12	0.33	0.06	-0.31
Rosette leaves at bud initiation	-0.26	-0.05	0.06	0.05	0.39	-0.20	0.01
Rosette area	-0.18	-0.11	-0.24	-0.17	-0.32	-0.18	0.32
Short term water-use	-0.27	-0.11	-0.21	-0.06	-0.21	-0.36	0.10
E at 100% rSWC	0.16	0.33	-0.12	0.36	-0.11	-0.12	-0.27
A at 100% rSWC	0.11	0.09	-0.21	0.31	0.17	-0.45	0.28
iWUE at 100% rSWC	0.18	-0.30	0.05	0.12	0.24	0.31	0.16
Gs at 100% rSWC	0.11	0.32	-0.09	0.43	-0.12	-0.17	-0.12
E at 50% rSWC	0.04	0.40	0.21	-0.24	0.00	0.10	-0.01
A at 50% rSWC	0.04	0.37	0.01	0.10	0.26	0.05	0.43
iWUE at 50% rSWC	0.04	-0.21	-0.22	0.30	0.35	0.20	0.27
Gs at 50% rSWC	0.11	0.39	0.23	-0.22	0.04	0.05	0.13
E at 20% rSWC	-0.14	0.17	-0.45	-0.19	0.06	0.20	-0.12
A at 20% rSWC	-0.09	-0.01	-0.39	0.11	0.01	0.27	-0.37
iWUE at 20% rSWC	0.17	-0.27	0.17	0.17	-0.18	-0.12	-0.24
Gs at 20% rSWC	-0.15	0.21	-0.42	-0.13	-0.02	0.21	0.07
Calculated water productivity	0.34	-0.12	-0.18	-0.24	0.17	-0.13	-0.16
Calculated water-use	-0.41	0.01	0.05	0.10	0.17	-0.09	-0.17

To test whether a combination of climatic parameters could be harnessed to predict multi-phenotypic trait performance, I compared the population (ecotype) scores between the significant CPCs and TPCs as per Wolfe & Tonsor (2014) through general linear regressions. Three significant general linear regressions were observed; these were between CPC1 and TPC6, CPC3 and TPC5, and CPC5 and TPC6 (Fig. 2.16). The association between CPC1 and TPC6 suggests that a reduction in temperature and an increase in temperature seasonality results in reduced biomass accumulation and water use and consequently an increase in WUE (Fig. 2.16). However, this postulation should be approached with caution since TPC6 explains very little of the total phenotypic variation (Fig. 2.15b).

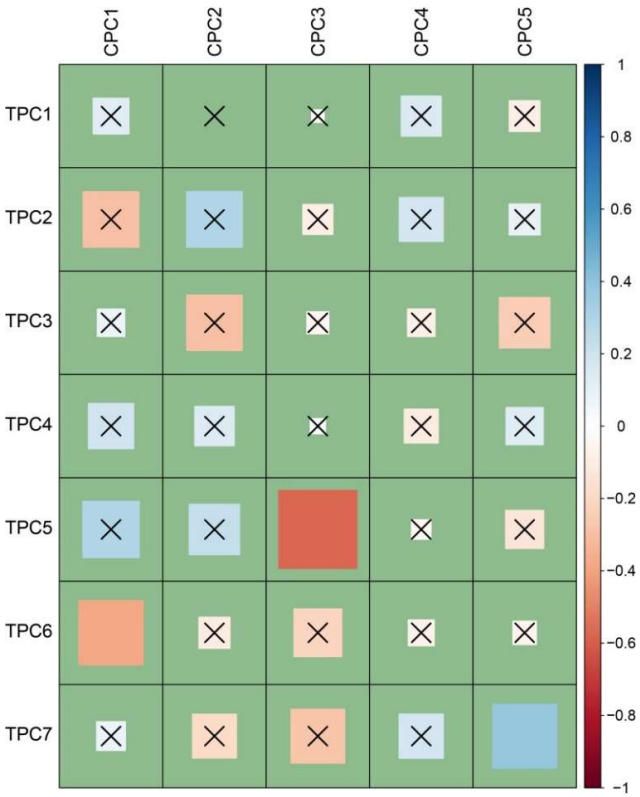

Figure 2.17. Associations between the significant climatic and trait principle components (CPCs and TPCs). Associations are based on linear regressions between the

population (ecotype) scores for the respective PCs. The size of the rectangle denotes the p-value, where large rectangles represent lower p-values, i.e. greater significance. Additionally, the presence of crosses in the rectangles denote non-significant associations. The color of the rectangle denotes the direction of the association, where dark red denotes a highly negative association and ark blue denotes a highly positive association.

To gain an insight into the influence of climatic exposure over evolutionary time on the ability of ecotypes to adjust trait performance according to present conditions, I tested the extent to which phenotypic plasticity is associated with historical climate. This was achieved by performing linear regression analysis of plasticity indices for multiple traits with biologically relevant climatic parameters for the 12 ecotypes comprising both the SD and CW experiments (Table 2.7).

Plasticity for short term water use was observed to negatively associated with latitude and precipitation of the driest quarter (Table 2.7), but it was observed to positively associate with precipitation seasonality (Fig. 2.17b). Plasticity for cWU demonstrated a strong positive association with temperature seasonality, where those ecotypes from areas of high temperature seasonality displaying the greatest plasticity for water use (Fig. 2.17c). Plasticity for flowering time was positively associated with annual temperature (Table 2.7.) and temperature seasonality (Figure 2.17a). It was also negatively associated with multiple temperature parameters (Table 2.7). The plasticity for the number of rosette leaves at bud initiation, which is essentially a proxy for flowering time, was also associated with temperature seasonality (positive; Table 2.7) and temperature parameters (negative; Table 2.7).

Table 2.7. Associations between climatic parameters and trait plasticity. Significant linear associations between climatic parameters and trait plasticity (RTR indices) for the 12 ecotypes common to the short dehydration (SD) and continuous watering (CW) experiments. The linear regression equation, r-squared value, and p-value of each significant association are given.

Climatic parameter	Trait plasticity	Equation	R-squared value	p-value
Latitude	Short term water-use	-0.00x + 0.75	0.36	0.04
BIO 15 - Precipitation seaonality	Short term water-use	0.00x + 0.63	0.41	0.03
BIO 17 - Precipitation of driest quarter	Short term water-use	-0.00x + 0.69	0.34	0.05
BIO 4 - Temperature seasonality	Calculated water-use	0.01x + 0.73	0.33	0.05
BIO 1 - Annual mean temperature	Flowering time	0.55x + -0.01	0.46	0.01
BIO 4 - Temperature seasonality	Flowering time	0.02x + 0.24	0.34	0.04
BIO 6 - Minimum temperature of coldest month	Flowering time	-0.01x + 0.40	0.57	0.004
BIO 9 - Mean temperature of driest quarter	Flowering time	-0.01x + 0.45	0.36	0.04
BIO 11 - Mean temperature of coldest quarter	Flowering time	-0.01x + 0.44	0.58	0.003
BIO 4 - Temperature seasonality	Rosette leaves at bud initiation	0.05x + 0.30	0.42	0.02
BIO 6 - Minimum temperature of coldest month	Rosette leaves at bud initiation	-0.02x + 0.62	0.37	0.03
BIO 11 - Mean temperature of coldest quarter	Rosette leaves at bud initiation	-0.02x + 0.68	0.35	0.04

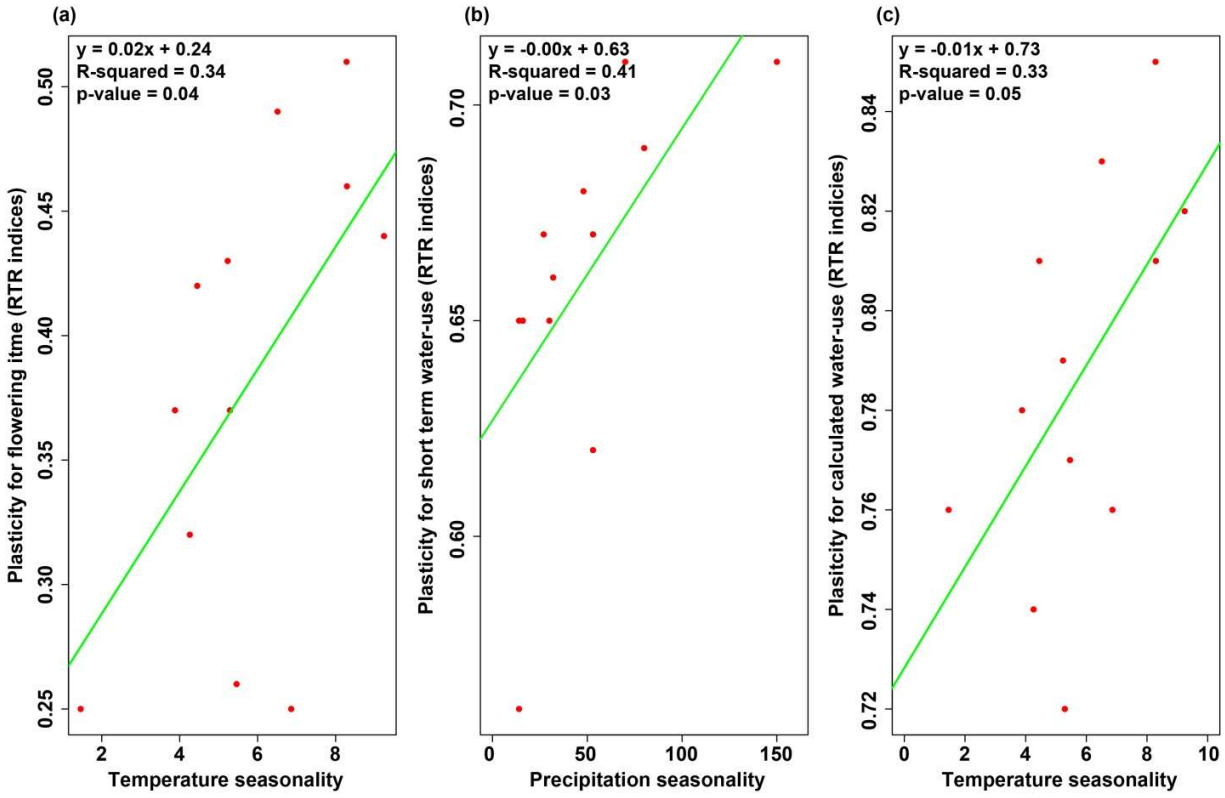

Figure 2.18. Associations between climate and phenotypic plasticity for flowering time and water use. (a) Significant positive linear relationship between temperature seasonality and plasticity for flowering time. **(b)** Significant positive linear relationship between precipitation seasonality and plasticity for short term water use. **(c)** Significant positive linear relationship between temperature seasonality and plasticity for calculated water use. For all plots, the green fit line represents the associated linear models. The equations of each linear model are provided along with appropriate r^2 values and p-values.

2.4 Discussion

2.4.1 Phenotypic plasticity along climatic gradients

Arabidopsis has evolved different life history strategies to adapt to a wide range of growing environments and seasons (Hoffmann, 2002, 2005; Mitchell-Olds & Schmitt, 2006; Banta *et al.*, 2012). As such, it has been suggested that local adaptation requires environment-dependent variation in fitness (Hancock *et al.*, 2011; Agren & Schemske, 2012; Easlon *et al.*, 2014), however little is known about the mechanisms that link genetic variation to fitness along global climatic environments. Understanding these biological mechanisms is vital, because phenotypic plasticity, either adaptive or non-adaptive, is essential for plants to adjust to changes in either or both natural habitats and environmental stress situations (Ghalambor *et al.*, 2007), as such it has both agricultural and ecological implications.

As part of this chapter, I investigated a set of ecotypes that represent a diverse range of climatic environments at their sites of origin (Fig. 2.1). Studying the relationship between climate and trait plasticity produced mixed results, with only plasticity for water use and flowering time showing significant associations with climatic history (Table 2.17). In general, as precipitation and temperature becomes more variable during the growing season, plasticity for flowering time and water use increased correspondingly. This suggests that the ability to alter performance of these traits (plasticity) is vital for continued persistence in environments that are characterized by inconsistent climatic conditions (Fig. 2.17). This idea is further supported by the plasticity observed for key traits, namely water use and productivity across the two experimental watering regimes (Fig. 2.6), which is discussed below.

With respect to fitness (seed yield), the performance of the 12 ecotypes across the SD and CW experiments was somewhat variable. Half of the ecotypes displayed distinct GxE interactions, with the remaining six ecotypes being split in half again, with three displaying adaptive plasticity and another three displaying non-adaptive plasticity (Fig. 2.7a). However, we observed no association between plasticity or performance of these traits with climate, suggesting these dynamics depend on a combination of allelic state as well as the prevailing environmental conditions (Hancock *et al.*, 2011; Agren & Schemske, 2012; Des Marais *et al.*, 2012; Agren *et al.*, 2013; Easlon *et al.*, 2014).

Traits pertaining to reproductive fitness or yield are often highly dynamic and sensitive to the environment. For this reason, traits such as flowering time and plant architecture are often employed as proxies of fitness, because they are far less sensitive to the environment. This has been documented previously by Jiaqin *et al* (2009) and it is further demonstrated as part of this study (Fig. 2.6). Due to their relative stability, genetic mapping for these proxy traits is much more likely to be successful in terms of basic gene and/or QTL discovery, as opposed to mapping for traits directly pertaining to yield. This is clearly a major advantage; however genetic loci identified in this manner are not guaranteed to contribute to yield stability or improvement in different environments (Fig. 2.4).

As part of this study, I have demonstrated the existence of highly significant genetic correlations between functionally related physiological traits. Despite this, I did not observe any significant relationships between traits that are typically considered proxies of seed yield in Arabidopsis, such as flowering time, photosynthesis, or vegetative biomass (Fig. 2.4). It has been suggested that the lack of association between photosynthetic parameters and fitness, or biomass accumulation in general, is because the *snapshot* parameters do not take

daily and seasonal variation into account, nor do they account for changeable environmental conditions, which may differ from measurement point to measurement point. Therefore, these kind of measures do not provide an integrated measure of photosynthetic performance over the life time of the plant (Long *et al.*, 2006; Lawson *et al.*, 2012; Driever *et al.*, 2014). Despite this explanation for the neutral photosynthetic association with yield, it is important to note the lack of association between flowering time, vegetative biomass, and iWUE. These traits are often used as indicators, or physical markers, of reproductive performance, especially within the Arabidopsis research community (Korves *et al.*, 2007; Nord & Lynch, 2008; Christman *et al.*, 2009; Todesco *et al.*, 2010; Verslues & Juenger, 2011; Ruts *et al.*, 2012; Suter & Widmer, 2013; Rosas *et al.*, 2014; Campitelli *et al.*, 2016) . Based on the observations described here however, it is apparent that under these two watering regimes, which are representative of conditions that would commonly be encountered by field grown plants, these proxies are ineffective. It is therefore vital that reproductive performance is assessed directly.

2.4.1 Relationship between water use and flowering time

To achieve greater yield under drought conditions or minimal water inputs, it is important to maximize the yield for each unit of water used by the plant. With this goal in mind, iWUE is often considered an important determinant of yield under water limited conditions, as well as being a key component of drought resistance (McKay *et al.*, 2003; Juenger & Mckay, 2005; Masle *et al.*, 2005; Kenney *et al.*, 2014; Easlon *et al.*, 2014).

I tested the relationship between iWUE, g_s, A, short term and long term water use, and productivity in two different experimental watering regimes. These traits are clearly

important in mediating drought resistance, almost exclusively through stomatal regulation of water loss, which can be presumed to affect the rates of all five traits. However, understanding the basis of drought resistance was consciously not the objective of this study. This was predominantly because drought resistance has been documented not to contribute to growth (Skirycz *et al.*, 2011). Growth is a key determinant of productivity, however survivability, as mediated by drought resistance, is regulated by different mechanisms (Skirycz *et al.*, 2011). Furthermore, drought conditions are known to divert resources away from growth and productivity and toward survival mechanisms (Levit, 1972; Blum, 2005).

The overarching aims of this study were to investigate the heritable basis of and relationship between water use and productivity trait parameters, as alternative assessment of plant performance. There has been a significant body of work which proposes the *effective use of water* as a far more important parameter to consider when looking at plant productivity under the context of reduced water availability (Reviewed in: Blum (2009). This idea is supported by my observations that iWUE correlated with g_s and *E* (Fig. 2.4), but not with *A* (Fig 2.4). Thereby suggesting that changes in WUE are driven primarily by changes in *E* that are in turn brought about by changes in g_s (Figs. 2.4, 2.5). Variation in *A* does not appear to genetically correlate with variation in iWUE (Fig. 2.4). This is complimentary to previous work where stomatal limitations were observed to be the main driver of reductions in carbon assimilation in Arabidopsis under well-watered and drought stress conditions, respectively (Easlon *et al.*, 2014; Bechtold *et al.*, 2016). Furthermore, and with respect to photosynthetic measurements performed at 100% and 50% rSWC, *A* showed substantially lower degrees of broad sense heritability than g_s (Table 2.4), which indicates that selecting for improved WUE in Arabidopsis largely depends on the genetic variation of the stomatal response to

water limitations (Lawson & Blatt, 2014; Easlon *et al.*, 2014). Similar observations have been documented in domesticated species, where breeding for enhanced WUE has led to impaired plant productivity, as high WUE is often achieved through stomatal closures, which improves water use but imposes constraints on biomass accumulation, be it vegetative or reproductive (Blum, 2005; Chaves *et al.*, 2009; Pinheiro & Chaves, 2011; Lawson & Blatt, 2014).

With respect to the above it is also worth remembering that the majority of elite crop varieties have been bred through programs that have selected for improved yields under conditions where there is a plentiful supply of water (Condon *et al.*, 2004; Fess *et al.*, 2011). Furthermore, they are commonly grown and harvested under such conditions also (Morison *et al.*, 2008). For this reason, they typically fail to optimize stomatal behavior under water limited conditions (Fischer *et al.*, 1998). It is difficult to directly translate results obtained from studies of this nature into crop improvement efforts, since Arabidopsis is an undomesticated, model species. However, these results do yield information regarding the importance of key traits for water use and productivity in a species that has not been *bred* to disregard water use in pursuit of maximal rates of photosynthesis and productivity. Additionally, elucidating the genetic basis of such key traits in Arabidopsis, or other model species, may form the foundations of crop improvement programs through multiple avenues, such as direct transgenic approaches.

Flowering time was positively associated with total plant water use (Fig. 2.11), but not iWUE, where the genetic correlation was neutral at 100% and 50% rSWC and significantly negative at 20% rSWC (Fig. 2.12). Previous studies have shown a positive correlation between integrated measures of WUE and flowering time, indicating that plants with longer

lifespans exhibit higher WUE, which would presumably be accompanied by reduced water consumption, or long term water use (McKay *et al.*, 2003, 2008; Kenney *et al.*, 2014). This is in contrast to much earlier studies on seasonal water use, which suggest that water use is increased in longer growing seasons (Penman & Schofield, 1951; Milthorpe, 1960). Therefore, selection for increased WUE under water limiting conditions would be expected to develop traits that limit plant water use. This is indeed the case, as demonstrated by the transition to early flowering and/or smaller leaf areas in crop plants that have been selected in this manner (White *et al.*, 1990; Ngugi *et al.*, 1994; Menendez & Hall, 1995; Sayre *et al.*, 1995; Martin *et al.*, 1999). In this study, genetic variation for flowering time was evident across the diverse selection of ecotypes, resulting in genetic variation for water use. In addition, flowering time as a life history or productivity associated trait, showed no correlation with reproductive performance (Fig 2.4), which suggests that flowering may be an important survival trait (Kenney *et al.*, 2014; Kooyers, 2015), but not a *maximizing productivity* trait, at least under the environmental conditions described and employed in this study.

It was highly interesting to note that water use measured on daily basis correlates positively with long term water use. Suggesting that although the latter is highly dependent on flowering time, ecotypes characterized by increased water use will use more water on both the short and long term (Fig. 2.12). Those ecotypes that used the least water daily were the smallest, e.g. C24 and Ct-1, and produced the greatest amount of reproductive product. This provides further demonstration that it is feasible to combine reduced water use and productivity and that the two are independent from one another.

The results constituting this chapter shed light on the inefficiency of using commonly employed proxy traits for predicting reproductive performance. It is vital that productivity is

assessed as a unique trait or as a factor of long term, i.e. WP, or calculated water use, i.e. cWP. These fitness related traits display much greater plasticity than traits such as flowering time, which appear far more stable and less sensitive to the environment. This stability is likely to have contributed to the widespread and continued implementation of non-plastic traits as proxies for yield in breeding programs and biological studies, especially within the Arabidopsis community.

3. A detailed survey of the diversity of water use and productivity related traits in 13 Arabidopsis ecotypes in outdoor and controlled environment conditions

3.1 Introduction

The preceding Chapter validated the importance of assessing or calculating long term measures of water use and productivity, compared to commonly employed proxies such as flowering time or operational measures of photosynthesis. The general conclusions pointed toward the erroneous nature of such proxies and suggested that water use and productivity are in fact more agronomically relevant traits of interest than drought resistance and/or WUE. Chapter Two provided an extensive assessment of the natural variation that exists for water use and productivity in Arabidopsis and the physiological parameters that do and do not relate to this variation. However, there are still unanswered questions surrounding the effectiveness of further proxy parameters, namely potential photosynthesis and $\delta^{13}C$. These parameters have been extensively employed as proxies of plant performance (Condon *et al.*, 2004; Parry *et al.*, 2011; Vadez *et al.*, 2014; Koester *et al.*, 2016), however there have been no previous assessments of their relatedness to reproductive performance or water use in Arabidopsis.

It is highly likely that the primary determinant of reproductive performance in domesticated and undomesticated species is the cumulative rate of photosynthetic

assimilation over the course of plants' lifecycle (Lawson *et al.*, 2012; Long *et al.*, 2015). As demonstrated in the previous Chapter, instantaneous or snapshot measures of operational photosynthesis at the per unit area of leaf level are in fact neutrally associated with biomass accumulation (Richards, 2000). This is very likely due to instantaneous measures of photosynthesis being unrepresentative of lifetime rates, a hypothesis supported by the demonstration that improved photosynthetic capacity can increase biomass accumulation (Kruger & Volin, 2006; Long *et al.*, 2006) and yield (Fischer *et al.*, 1998). Despite the empirical evidence, the natural variation that exists for potential photosynthetic rates is only beginning to be explored (Driever *et al.*, 2014). A more complete understanding of how genetic variation for maximal or potential rates of photosynthesis relates to biomass accumulation is required in order to make meaningful assertions as to how photosynthesis may be exploited to improve reproductive output under the context of abiotic stress.

The ratio of naturally occurring carbon isotopes in plant tissue, i.e. $\delta^{13}C$, has been extensively employed as a marker of WUE since it was first demonstrated as a highly accurate marker of the latter by Farquhar & Richards (1984). Yet, it should be noted that the nature of $\delta^{13}C$ as a marker for WUE is based on its relationship to iWUE, i.e. the ratio of A to g_s. The relationship between $\delta^{13}C$ and actual WUE, i.e. the ratio of biomass to plant water use or evapotranspiration, has been largely underexplored with inconsistent results obtained thus far (Condon *et al.*, 2004; Morison *et al.*, 2008) Regardless, $\delta^{13}C$ has been employed in numerous previous efforts to assess natural variation for WUE in Arabidopsis (E.g. Kenney *et al.*, 2014; Easlon *et al.*, 2014). One of the fundamental conclusions of these studies is the positive genetic correlation between WUE and flowering, where plants that flower later have improved WUE. This conclusion is further reflected through the identification of QTLs

controlling flowering time when performing genetic mapping for $\delta^{13}C$ (McKay *et al.*, 2003, 2008; Juenger & Mckay, 2005; Lovell *et al.*, 2015). Despite the extensive application of $\delta^{13}C$ as a trait of interest for assessing plant performance under the context of reduced water availability, no studies have thus far determined its usefulness for predicting reproductive and/or vegetative performance, or lifetime water use in Arabidopsis. It would be of significant interest to thoroughly assess the relationship between $\delta^{13}C$ and both water use and productivity in Arabidopsis in order to better inform future work for this field of study.

As part of both the Introduction and the preceding results Chapter, I have described the tendency for studies assessing natural variation in Arabidopsis to do so under controlled environment conditions that reflect exceptionally long days, often in excess of 16 hours of light. Such conditions naturally give rise to early flowering (E.g. McKay *et al.*, 2003; Juenger & Mckay, 2005; Hausmann *et al.*, 2005). This is a point worth noting since the majority of Arabidopsis ecotypes are winter annuals (Schmitz & Amasino, 2007), as such they are naturally subjected to very short photoperiods and flower in the spring when day length has begun to markedly increase. This progression is to some extent reflected by the short dehydration and continuous watering experiments employed as part of the present study. The winter annual lifestyle of the majority of Arabidopsis ecotypes could suggest that experimental subjection of such ecotypes to unnaturally long photoperiods may achieve conclusions that are biological questionable in terms of their translational significance for agricultural or ecological systems. It is likely that this is especially true when assessing mechanisms relating to water use or drought resistance, since these phenomena are often closely linked to flowering time (Verslues & Juenger, 2011). For example, Riboni *et al* (2013) recently demonstrated that the induced drought escape mechanism in Arabidopsis is

promoted by drought mediated upregulation of florigens in an ABA and photoperiod dependent manner, so that early flowering, i.e. drought escape, can only occur under long days.

As well as external environmental parameters, the biological composition of agricultural and ecological systems can have an enormous effect on individual plant performance. Numerous empirical ecological studies have demonstrated that increased degrees of intra- and interspecific genotypic diversity drive elevated above-ground productivity under optimal conditions and/or stabilize productivity following a period of abiotic and/or biotic stress (Hector *et al.*, 1999; Van Ruijven & Berendse, 2005; Roscher *et al.*, 2011). Stabilized productivity achieved in this manner is referred to as ecological stability and can arise through both resistance and resilience. Here, resistance refers to the ability of an ecological system to resist change in response to perturbations, whereas resilience refers to the ability of a system to retransition to its pre-perturbation state upon the reoccurrence of ideal conditions (Reviewed in: Tilman *et al.*, 1996). The ability to translate the ecological phenomena of diversity-driven stability to crop systems has been proposed as a means to improve yield outputs in the face of global climatic change (Li *et al.*, 2009). Despite this interest, no published studies have described attempts to understand the impact of intra-specific diversity in managed planting designs with respect to reduced water availability. However, in a similar vein Creissen *et al* (2013) recently compared the reproductive performance of monocultures and genotypic mixtures of Arabidopsis ecotypes in response to nutrient stress. This particular study demonstrated that polycultures achieve stabilized yields through compensation. That is to say ecotypes that are less susceptible to reduced nutrient availability or have improved nutrient use efficiency compensate for those that are

susceptible or have reduced nutrient use efficiency. The ability of genotypic diversity to achieve stabilization in Arabidopsis has only been demonstrated in controlled environment conditions as part of the Creissen *et al* (2013) study. Furthermore, there have been no prior attempts to understand the effect of this type of diversity in relation to reduced water availability. A broader understanding to this end may assist in predicting the usefulness of employing diversity-driven agroecological cropping systems in areas where climatic incidences of drought are frequent.

The research constituting this present chapter was undertaken in order to test the usefulness of $\delta^{13}C$ for pertaining information regarding eventual biomass accumulation and calculated water use using a short dehydration experiment. The experimental designs adopted for the previous chapter involved a very strict transition from short- to long-day conditions. In general, Arabidopsis-based natural variation studies of this nature are performed under long-day conditions (E.g. McKay *et al.*, 2003; Juenger & Mckay, 2005; Hausmann *et al.*, 2005). For this reason, I tested the effect of both long and short days on long term water use and water productivity using previously described continuous watering experiments. To further assess the impact of environmental conditions and in order to obtain a more applied understanding of the effect of water availability on plant performance, I performed an outdoor garden experiment using select ecotypes grown under covered, i.e. drought, and uncovered, i.e. well-watered, conditions. Furthermore, the garden experiment was sub-divided into ecotypes grown in monocultures and polycultures to address hypotheses regarding the potential for genotypic mixtures to stabilize plant performance during abiotic stress.

3.2 Materials & Methods

3.2.1 Plant Material and Growth Conditions

Seed for all ecotypes comprising this study were obtained from NASC (Scholl *et al.*, 2000). This present study included 13 ecotypes that were selected based on isolation following hierarchical clustering on key traits as described in the previous Chapter (Fig. 2.3). The 13 ecotypes studies were as follows: C24, Col-0, CIBC-5, Est-1, Ler-0, Fei-0, Kin-0, Ct-1, HR-5, Cvi-0, Se-0, Ei-2, and Kondara.

All plants were grown in peat-based compost (Levington F2+S, The Scotts Company, Ipswich, UK) and were subjected to a three-day period of stratification to promote germination. Following stratification, all plants were grown in the controlled environment room, where conditions were exactly as described in Chapter Two (Section 2.2). Following set periods of time and experimental conditions, plants comprising the short dehydration and continuous watering experiments were transferred into the glasshouse, where conditions were also exactly as described in Chapter Two. Those plants comprising the garden experiment were transferred to an outdoor garden environmental setup, where climate was somewhat variable, as described below.

3.2.2 Short dehydration experiment

The short dehydration experiment was exactly as described in Chapter Two. Briefly, all plants were grown in pots containing the same volume of soil. Pots were dried down from ~100% to ~20% rSWC, at which point they were re-watered and transferred to the glasshouse, where they were kept well-watered. Two short dehydration experiments were

performed. The first of these was performed to assess the significance of a short dehydration period in terms of reproductive performance; as such it was performed alongside a controlled experiment where plants were kept continuously well-watered. The ecotype Est-1 failed to germinate as part of this experiment, so this ecotype was not included in this experiment. For both short dehydration experiments and the control well-watered experiment, 15 biological repeats of each ecotype were grown.

With respect to the second short dehydration experiment, short term water use, flowering traits, and biomass accumulation were recorded for all ecotypes (n = 15). These parameters were used to calculate water use (cWU) and water productivity (cWP). Additionally, the ratio of total above ground biomass to cWU was also estimated as calculated WUE (cWUE). Operational rates of photosynthesis were assessed for every ecotype (n = 4) at the same time every day during the short dehydration period through infra-red gas exchange analysis as described in Chapter Two.

The ratio of the natural isotopes of carbon, i.e. $^{13}C/^{12}C$ or $\delta^{13}C$, was assessed for all ecotypes (n =4) comprising the second short dehydration experiment at the point at which their respective rSWCs were approximately 40%. The ratio of naturally occurring carbon isotopes acts as a proxy for WUE to which it is positively linked to, i.e. the biological sample with the least negative value for $\delta^{13}C$ represents the plant with the greatest WUE. $\delta^{13}C$ provides a lifetime integrated measure of WUE, as opposed to the snapshot measures obtained from leaf level gas exchange measurements. To reduce the effect of noise from soil respiration, two fully developed upper rosette leaves were harvested for $\delta^{13}C$ analysis. For each biological repeat, $\delta^{13}C$ was measured as per Roussel *et al* (2009). Briefly, 1mg of lyophilized, pulverized rosette leaf tissue were measured for $\delta^{13}C$ using a continuous flow

isotope ratio mass spectrometer coupled to an elemental analyzer as described in. $\delta^{13}C$ was calculated as: $(R_s - R_b)/R_b$ x 1000, where R_s and R_b represent the $^{13}C/^{12}C$ ratio in the samples and in the Pee Dee Belemnite standard respectively (Craig, 1957). Additionally, the percent content of carbon and nitrogen were also assessed as part of this analysis.

3.2.3 Continuous watering experiment

Two continuous watering experiments were performed, one under long-day conditions and one under short-day conditions. Both were performed in the same manner as the continuous watering experiment described in the previous Chapter. Briefly, all plants were grown in pots containing the same volume of soil. The soil of all pots was covered with transparent plastic beads to minimize transpiration from the soil. A 5ml pipette tip was inserted into the soil to facilitate watering with precise volumes of water. The rSWC of all plants was maintained at 40% daily from the point of initiation of the continuous water regime, which was different for the long day and short day experiments. With these experiments measured water use (mWU) and measured water productivity (mWP) were determined, as well as flowering and biomass parameters (n = 15).

All plants comprising both continuous watering experiments were germinated as per the short dehydration experiment in the same soil type. All plants remained in the controlled environment (short day conditions), where they were kept well-watered until transfer to the glasshouse (long day conditions). The continuous watering regimes were initiated immediately upon entering the glasshouse. Those plants comprising the short-day experiments were transferred to the glasshouse at 50 days old and those comprising the long day experiments were transferred at 25 days old. As with the previous continuous

watering experiment, watering was ceased on the day of opening of the final flower on a plant-by-plant basis.

3.2.4 Additional experiments and parameters assessed

Those plants that encompassed the first short dehydration experiment were used to assess potential rates of photosynthesis before commencement of the drying period (n = 4 for each ecotype). These measurements were made using the same portable infra-red gas exchange systems used for the previously described operational photosynthesis assessments. Potential photosynthesis measurements were performed on randomly selected, fully expanded, upper rosette leaves of plants between 50-54 days old. The response of CO_2 assimilation (A) to changes in the intracellular CO_2 concentration (C_i) were measured under a saturating red light irradiance of 576 µmol photons m^{-2} s^{-1}, a leaf temperature of 20.1ºC (± 1.3ºC), and a vapor pressure deficit (VPD) of 0.7 kPa (±0.01 kPa). To assess the response of A to changes in C_i, the extracellular CO_2 concentration (C_a) was incrementally decreased from its ambient starting point of 400 µmol mol^{-1} to 250, 150, and 50 µmol mol^{-1}. Subsequently, C_a was incrementally increased to 300, 400, 600, 800, 1200, and 1500 µmol mol^{-1}. Readings of A were logged after they had stabilized to the incremental C_a conditions, which typically took between 1.5-2 minutes. The data obtained from the A/C_i response measurements were used to determine the maximum velocity of Rubisco for carboxylation (V_{cmax}) and the maximum rate of electron transport demand for RuBP regeneration (J_{max}) through curve fitting as describe in Sharkey *et al* (2007). Furthermore, light saturated CO2 assimilation rates (A_{max}) were determined through calculating mean A at the two points where C_a was set at 400 µmol mol^{-1}.

3.2.5 Garden experiment

Six of the full set of ecotypes, namely Col-0, Ct-1, C24, Se-0, Est-1, and HR-5, were randomly selected to be incorporated as part of the garden experiment. Plants were grown within raised $1m^2$ troughs filled with the same soil type used for the short dehydration and continuous watering experiments. After stratification, plants were grown in the previously described controlled growth room for 35 days, at which point they were transferred to the garden experimental location.

The garden experiment consisted of four experimental treatments; Monoculture covered, monoculture uncovered, polyculture covered, and polyculture uncovered. The monoculture treatments consisted of Col-0 plants only (n = 20), whereas the polyculture treatments consisted of all six ecotypes (n = 15). For all treatments, plants were grown in rows at a consistent density of one plant every 8cm as per a randomized design, where the rows were also 8cm apart.

The two covered treatments employed raised, static rainout shelters constructed from UV treated polythene sheets to minimize rainfed water, thereby simulating *natural* drought conditions. These shelters completely prevented any rainwater reaching the soil during the experimental drought period. Additionally, they did not significantly impact upon the wavelengths of light penetrating through to plant level, nor did they alter the air temperature at plant level. The uncovered treatment troughs did not have associated rainout shelters.

The soil for both the covered and uncovered treatments was kept well soaked until the plants were 50 days old, at which point a drought period was initiated for the covered treatments where watering was completely stopped and never reinitiated. The soil of the

uncovered treatments remained well watered during this time. The weather was continually monitored and recorded from two weeks before the drought period and until the end of the experiment. Weather was monitored and recorded using an Aercus Instruments WS2083 Professional Wireless Weather Station (http://www.aercusinstruments.com), which was placed adjacent to the four troughs.

Flowering time (bud emergence) and the number of rosette leaves at the point of flowering was recorded. Upon the opening of the final flower, plants were bagged to prevent seed loss through silique shattering. Upon bagging, plants were cut below the rosette and allowed to completely dry down. The biomass components (Rosette, chaff, seed yield) were then separated and weighed. Additionally, the average plant height of all ecotypes from all treatments was measured just before bagging.

3.2.6 Statistical analyses

All statistical analyses relevant to this study were performed with the R software environment for statistical computing and graphics (R Development Core Team, 2008). As an initial investigation of the associations of parameters, the Pearson's r rank correlation coefficients were computed between all possible pairs of trait parameters. This was achieved with using the rcorr() function from the Hmisc package (Harrell, 2006). Subsequently, significant associations (P-value < 0.05) of interest between parameters were investigated further through fitting simple linear regression models using the lm() function in the R base code. The relationship between all the traits assessed as part of the garden experiment were also investigated through PCA using the prcomp() function, also part of the R base code. -

One-way ANOVA comparison of means tests were performed to test for differences between ecotypes for multiple traits and to tests for differences between the treatments employed as part of the garden experiment. One-way ANOVAs were performed using the aov() R base code function and concurrent *post-hoc* Tukey tests. This was achieved via Tukey Honest Significant Difference (HSD) tests using the HSD.test() function from the 'agricolae' R package (de Mendiburu, 2016). Two-way and Three-way ANOVA comparison of mean statistical tests were performed in unison with one-way ANOVA tests to detect experiment-wide error where appropriate. This was also achieved using the aov() function in R.

3.3 Results and Discussion

3.3.1 The effect of short dehydration on flowering time and seed yield

Flowering time and reproductive performance were assessed for 13 ecotypes that represent a substantial proportion of the natural variation observed in Chapter Two (Fig. 2.1; 2.3). The difference in phenotypic traits was assessed for all ecotypes grown under both well-watered and short dehydration conditions. The well-watered experimental conditions represented a control to assess the effect of a short dehydration period on these key parameters. Only three ecotypes, namely Kondara, Ler-0, and C24, demonstrated significantly different flowering times as part of the short dehydration experiment. Kondara and Ler-0 appeared to delay flowering in response to the short dehydration period, whereas C24 appeared to initiate early flowering indicating a drought escape strategy as determined through one-way ANOVA testing (Fig. 3.1). The results from a two-way ANOVA comparison

of means test where the effect of ecotype and treatment, and their interaction, on flowering time was tested confirmed the statistical significance of variation in flowering time between the ecotypes but did not detect an effect of treatment, since the effect was null for the majority of the ecotypes and opposing for the aforementioned ecotypes (Table 3.1) Neither Kondara, Ler-0, nor C24 demonstrated significantly different reproductive performance in response to short dehydration. Indeed, only the Kin-0 ecotype appeared to have altered reproductive biomass accumulation, where seed yield increased in response to the short dehydration (Fig. 3.2). A two-way ANOVA for seed yield demonstrated the statistically significant difference in reproductive output of the various ecotypes (Table 3.2). An effect of seed yield due to the watering treatment was also detected, but this was only due to the difference in seed yield for the Kin-0 ecotype between the treatments (Table 3.2, Fig. 3.2).

The above suggests that the short dehydration period does not represent a severe drought stress. It is objectively visible that Arabidopsis does respond to the extended period of no water input, since wilting occurs. Wilting is more severe for those ecotypes that have elevated drying rates, i.e. increased short term water use. Additionally, results from the present and previous Chapters demonstrated that physiological changes occur as the short dehydration period persists and these changes are what would be expected from a drought response (Figs. 2.5, 3.7, 3.8). However, upon re-watering all ecotypes appear to visibly recover in less than one hour. Despite this, the relative lack of effect of short dehydration on either flowering time or productivity provides encouragement that calculated estimates of water use and water productivity are not influenced by drought-induced early flowering and/or diverting of resources away from reproductive output and toward defensive mechanisms.

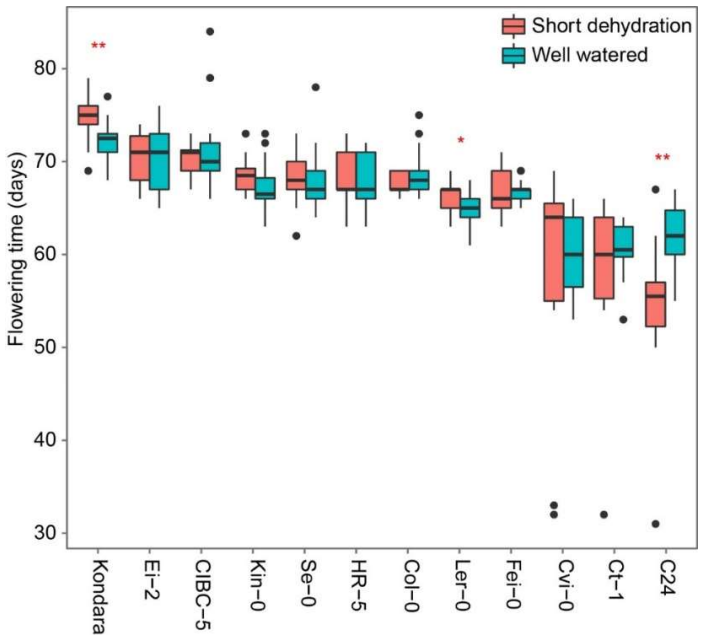

Figure 3.1. Boxplots describing the variation for flowering time demonstrated by the ecotypes grown as part of parallel short dehydration (red) and well-watered experiments (blue). Red asterisks denote significant differences between the same ecotype grown under short dehydration and well-watered conditions, where *** = P < 0.001, ** = P < 0.01, and * = P < 0.05. The bold line in the center of the boxplots represents the median, the box edges represent the 25th (lower) and 75th (upper) percentiles, the whiskers extend to the most extreme data points that are no more than 1.5x the length of the upper or lower segment, and points away from the whiskers indicate extreme outliers. *N* = 15.

Table 3.1. Results from a two-way ANOVA to test for effects of ecotype and treatment (well-watered and short dehydration) on flowering time of the 13 ecotypes comprising the initial short dehydration experiment.

Source	Sum of squares	Degrees of freedom	Mean square	F-value	P-value
Ecotype	7583	11	689.4	40.552	< 0.0001
Treatment	45	1	44.8	2.636	0.1053
Ecotype:Treatment	611	11	55.6	3.269	< 0.0001
Residuals	5950	350	17		

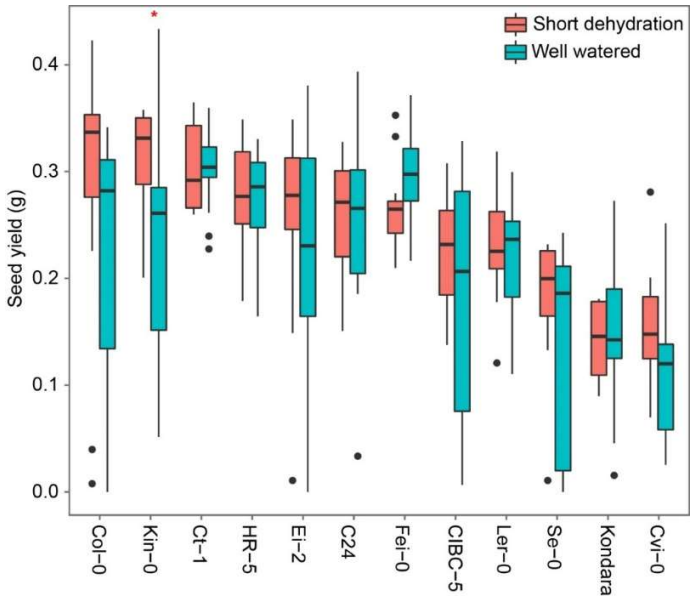

Figure 3.2. Boxplots describing the variation for seed yield demonstrated by the ecotypes grown as part of parallel short dehydration (red) and well-watered experiments (blue). Red asterisks denote significant difference between the same ecotype grown under short dehydration and well-watered conditions, where *** = P < 0.001, ** = P < 0.01, and * = P < 0.05. The bold line in the center of the boxplots represents the median, the box edges represent the 25th (lower) and 75th (upper) percentiles, the whiskers extend to the most extreme data points that are no more than 1.5x the length of the upper or lower segment, and points away from the whiskers indicate extreme outliers. *N* = 15.

Table 3.2. Results from a two-way ANOVA to test for effects of ecotype and treatment (well-watered and short dehydration) on seed yield of the 13 ecotypes comprising the initial short dehydration experiment.

Source	Sum of squares	Degrees of freedom	Mean square	F-value	P-value
Ecotype	0.9912	11	0.9011	15.589	< 0.0001
Treatment	0.46	1	0.046	7.959	< 0.01
Ecotype:Treatment	0.0735	11	0.00668	1.156	0.3174
Residuals	1.7341	300	0.00578		

3.3.2 Natural variation for potential photosynthesis

The response of *A* to changes in *Ci* was assessed to estimate photosynthetic capacity parameters, i.e. A_{max}, J_{max}, and V_{cmax}. *A*/*Ci* response curves were performed between 45-50 days before the onset of the short dehydration period and allowed for direct comparisons between the parameters and multiple other assessed phenotypic traits in order to ascertain the effectiveness of utilizing potential photosynthetic rates as a means of predicting plant performance. There was only a very small degree of variation for photosynthetic capacity between the ecotypes. No significant differences were detected between any two ecotypes for A_{max}. For J_{max}, only C24 and Fei-0 were significantly different from one another, with all other ecotypes being comparable with each other and these two ecotypes. V_{cmax} represented the potential photosynthetic parameter with the most natural variation, but there were still only two significantly different post-hoc groups identified (Fig. 3.3).

Despite the lack of variation for potential photosynthetic capacity, significant positive associations were demonstrated between V_{cmax} and both the accumulation of chaff biomass and cWUE (Fig. 3.4). This initially suggests that those ecotypes that demonstrate rapid rates of Rubisco carboxylation also accumulate elevated levels of biomass. However, it should be

noted that significant associations between V_{cmax} and other biomass parameters, namely rosette biomass and seed yield, were not detected. This perhaps suggests that V_{cmax} is particularly related to chaff and total biomass as a function of water use and not biomass accumulation in general.

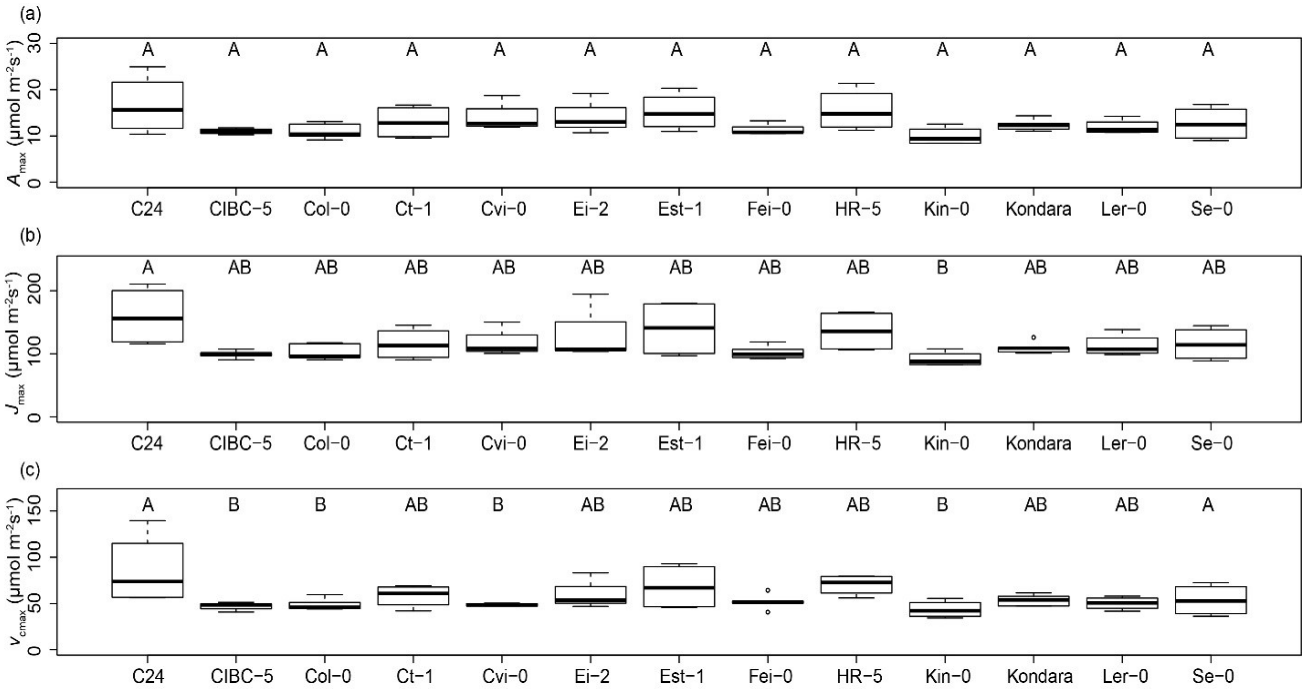

Figure 3.3. Boxplots describing the natural variation for (a) A_{max}, (b) J_{max}, and (c) V_{cmax} before the initiation of a short dehydration period. The letters above the individual boxplots denote post-hoc Tukey groupings. Those ecotypes with the same letters above boxplots for specific traits are not significantly different (P > 0.05) from one another. The bold line the center of the boxplots represents the median, the box edges represent the 25th (lower) and 75th (upper) percentiles, the whiskers extend to the most extreme data points that are no more than 1.5x the length of the upper or lower segment, and points away from the whiskers indicate extreme outliers. *N* = 4.

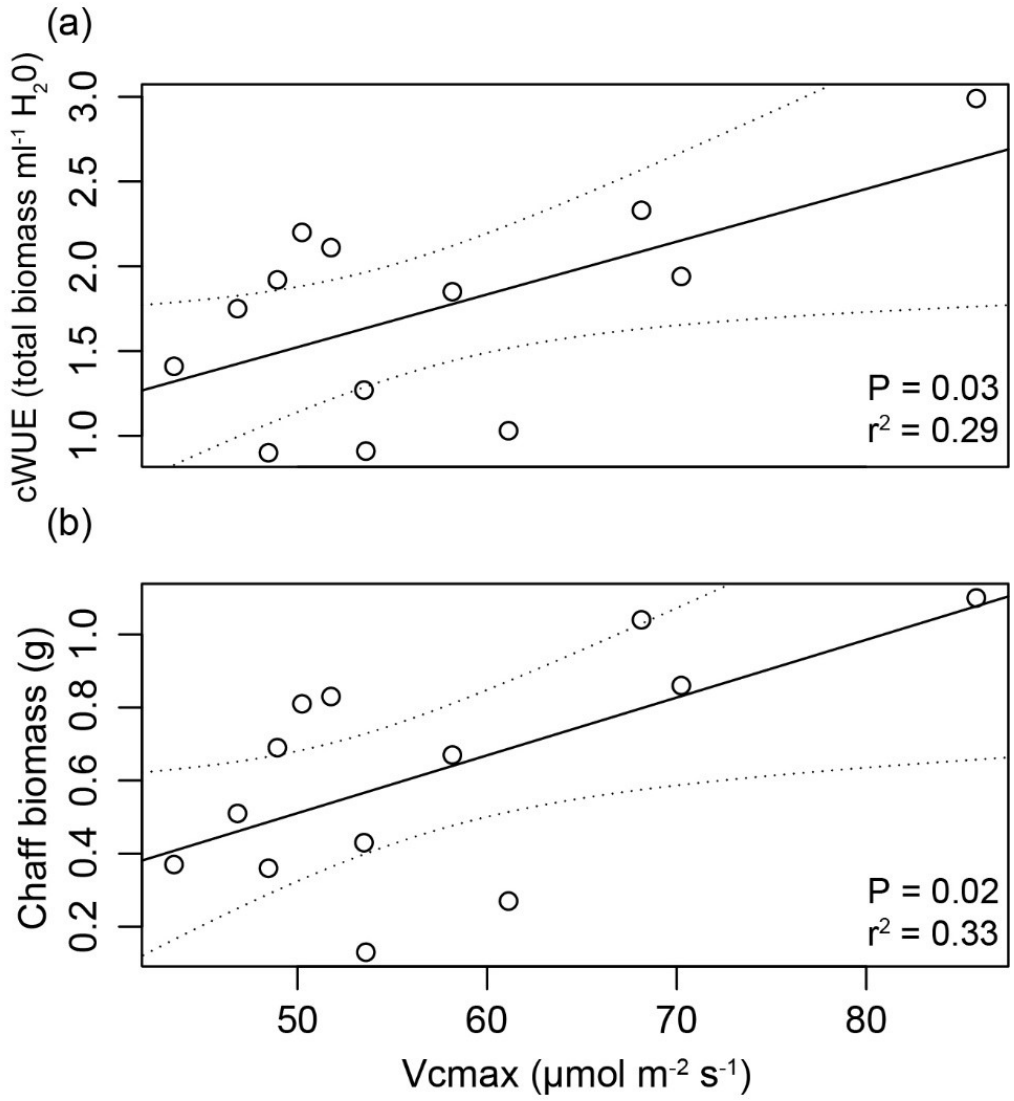

Figure 3.4. Relationship between V_{cmax} and both (a) calculated water use efficiency (cWUE) and (b) chaff biomass. The bold fit line represents the equation of the linear regression model. The adjacent dashed lines represent the lower and upper 95% confidence intervals. The P-value and adjusted r^2 value associated with the linear model are provided for each association.

3.3.3 Contribution of $\delta^{13}C$, %N, and %C to biomass accumulation and water use

During the short dehydration experiment tissue from all ecotypes was harvested at 40% rSWC to analyze the percent content of nitrogen and carbon in dry leaf matter. Additionally, $\delta^{13}C$ was also assessed as a proxy of WUE. Substantial variation was demonstrated for these parameters. Five and four *post-hoc* Tukey groups were identified post-ANOVA for percent nitrogen content and percent carbon content respectively (Fig 3.5). For nitrogen, C24 had and Se-0 the lowest percentage (Fig. 3.5a). Interestingly, these two ecotypes were similar for percent carbon content, where the most extreme ecotypes here were CIBC-5 and Cvi-0 (Fig. 3.5b). $\delta^{13}C$ also demonstrated significant genetic variation, where there were five post-hoc groups detected. The C24 ecotype is noted for demonstrating markedly reduced $\delta^{13}C$, i.e. high WUE. As with total carbon content, Cvi-0 also demonstrated the lowest $\delta^{13}C$ and was signficantly different from all other ecotypes (Fig. 3.6).

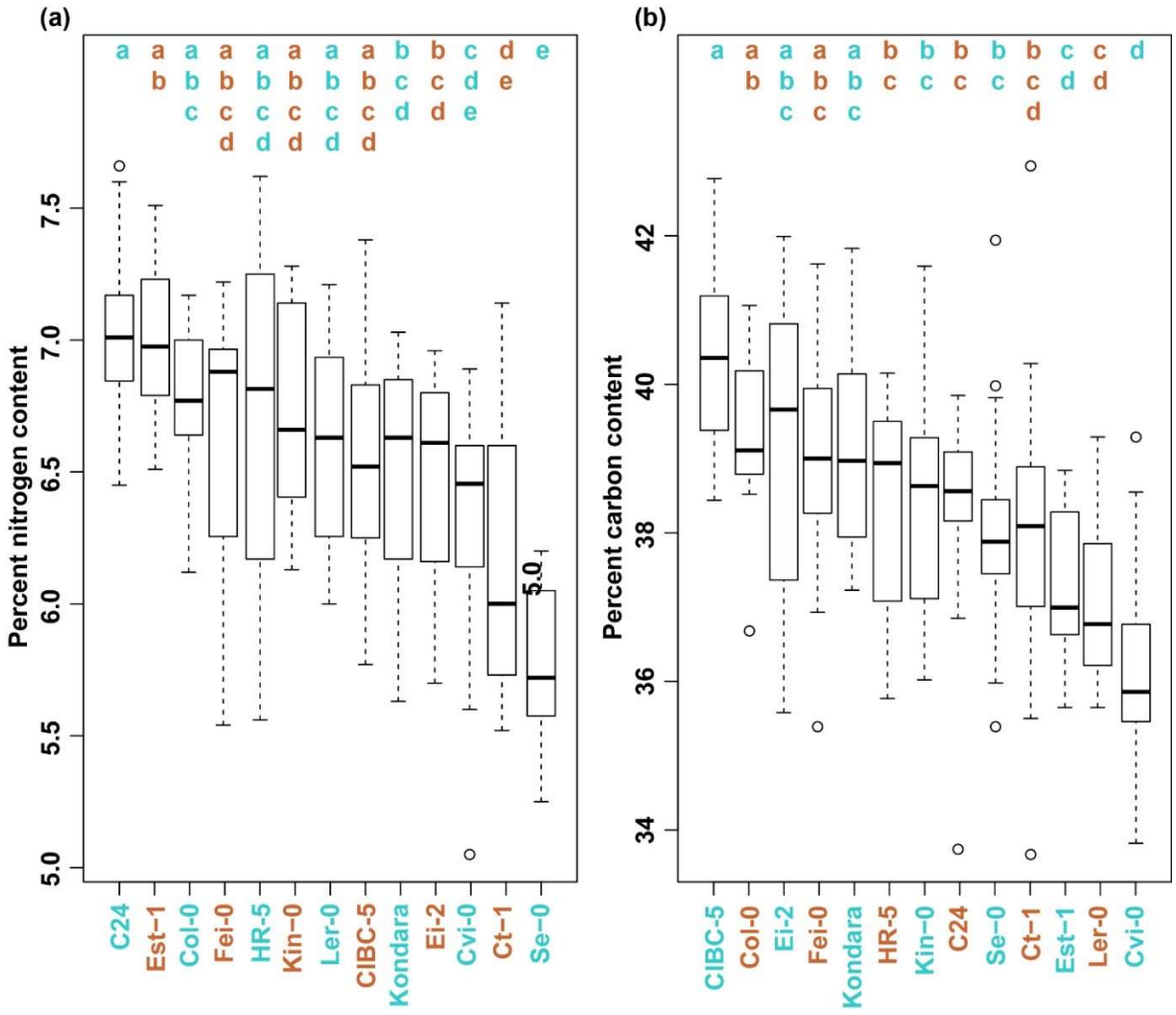

Figure 3.5. Boxplots describing the percentage content of (a) nitrogen and (b) carbon in leaf tissue of the 13 Arabidopsis ecotypes as measured at ~40% relative soil water content during a short dehydration period. The letters above the individual boxplots denote post-hoc Tukey groupings. Those ecotypes with the same letters above boxplots for either of the two traits are not significantly different ($P > 0.05$) from one another. The bold line in the center of the boxplots represent the median, the box edges represents the 25% (lower) and 75% (upper) percentiles, the whiskers extend to the most extreme data points that are no more than 1.5x the length of the upper and lower segment, and points away from the whiskers indicate extreme outliers. *N* = 4.

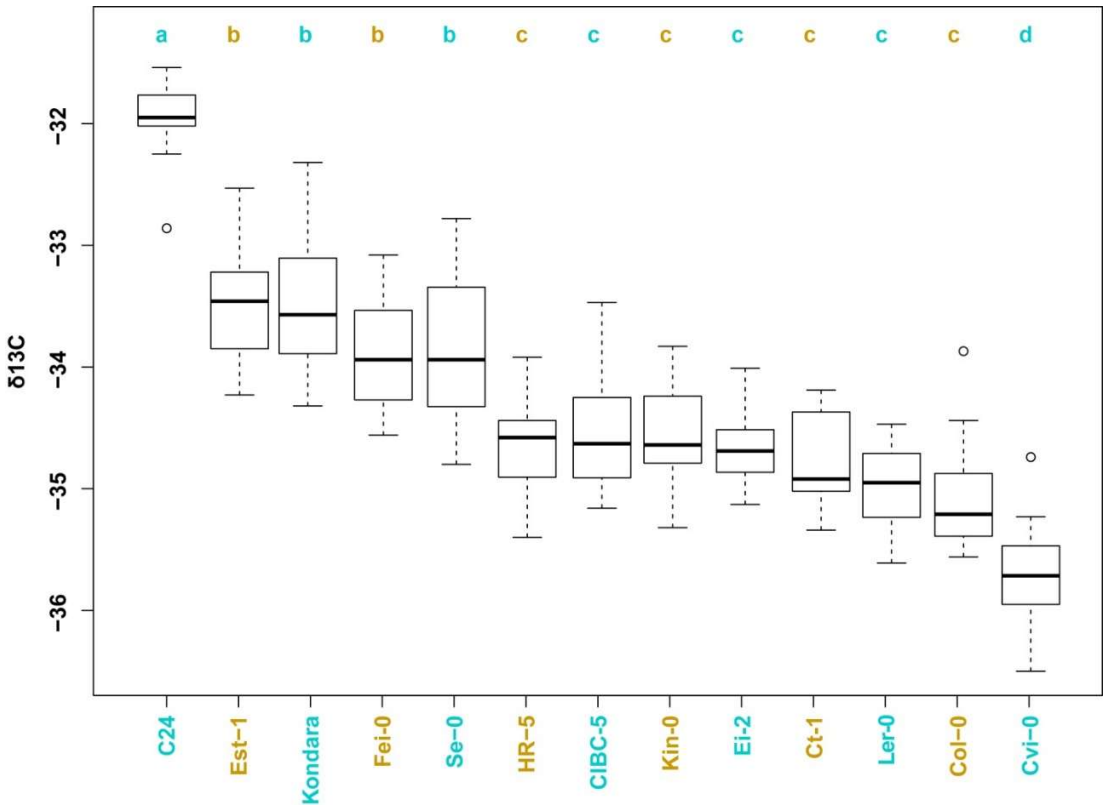

Figure 3.6. Boxplots describing the ratio of ^{13}C to ^{12}C isotopes ($\delta^{13}C$) in leaf tissue of the 13 ecotypes at ~40% relative soil water content during a short dehydration period. The letters above the individual boxplots denote post-hoc Tukey grouping. Those ecotypes with the same letters above boxplots for either of the two traits are not significantly different ($P > 0.05$) from one another. The bold line in the center of the boxplots represent the median, the box edges represents the 25% (lower) and 75% (upper) percentiles, the whiskers extend to the most extreme data points that are no more than 1.5x the length of the upper and lower segment, and points away from the whiskers indicate extreme outliers. $N = 4$.

Since $\delta^{13}C$ is used as a proxy for WUE, I tested the relationship of variation for $\delta^{13}C$ with variation for operational assessments of the rate of g_s, T, A, and iWUE at multiple rSWCs. $\delta^{13}C$ demonstrated significant negative relationships with both g_s and T at 100%

rSWC (Fig. 3.7). At 80% and 40% rSWCs these same associations were only marginally non-significant, where P-values were either 0.06 (G_s at 80% rSWC) or 0.07 (G_s at 40% rSWC and T at both 80% and 40% rSWC), but never below the standard 95% significance threshold. However, at 20% rSWC these associations entirely dissipated and were totally non-significant (P-values for g_s = 0.36, T = 0.33). Interestingly, significant relationships were not detected between $\delta^{13}C$ and either A or iWUE at any rSWCs. This further reflects the observations made between A, g_s, T, and iWUE as part of the previous chapter and suggests that alterations to $\delta^{13}C$ are achieved through changes in stomatal conductance not through changes to photosynthetic assimilation (Fig. 2.7). The breakdown of the relationship between $\delta^{13}C$ and both g_s and T at 20% rSWC may suggest that $\delta^{13}C$ only provides an integrated measure of WUE up to the point at which tissue for isotopic analyses is harvested and it is therefore not a suitable physical marker for predicting future photosynthetic and stomatal dynamics (Fig. 3.7).

The variation for $\delta^{13}C$ was also compared to all other phenotypic parameters assessed as part of the same short dehydration experiment. No significant associations were detected between parameters that could conceivably be envisaged to correlate with $\delta^{13}C$, e.g. rosette biomass accumulation and cWUE, thereby suggesting $\delta^{13}C$ is not a suitable predictor of biomass accumulation or actual water use. However, $\delta^{13}C$ was observed to share a significant positive association with V_{cmax} (Fig. 3.8), which suggests that those ecotypes that have the greatest potential photosynthetic capacity are also able to reduce g_s, and consequently T, in response to reduced water availability, thereby increasing WUE.

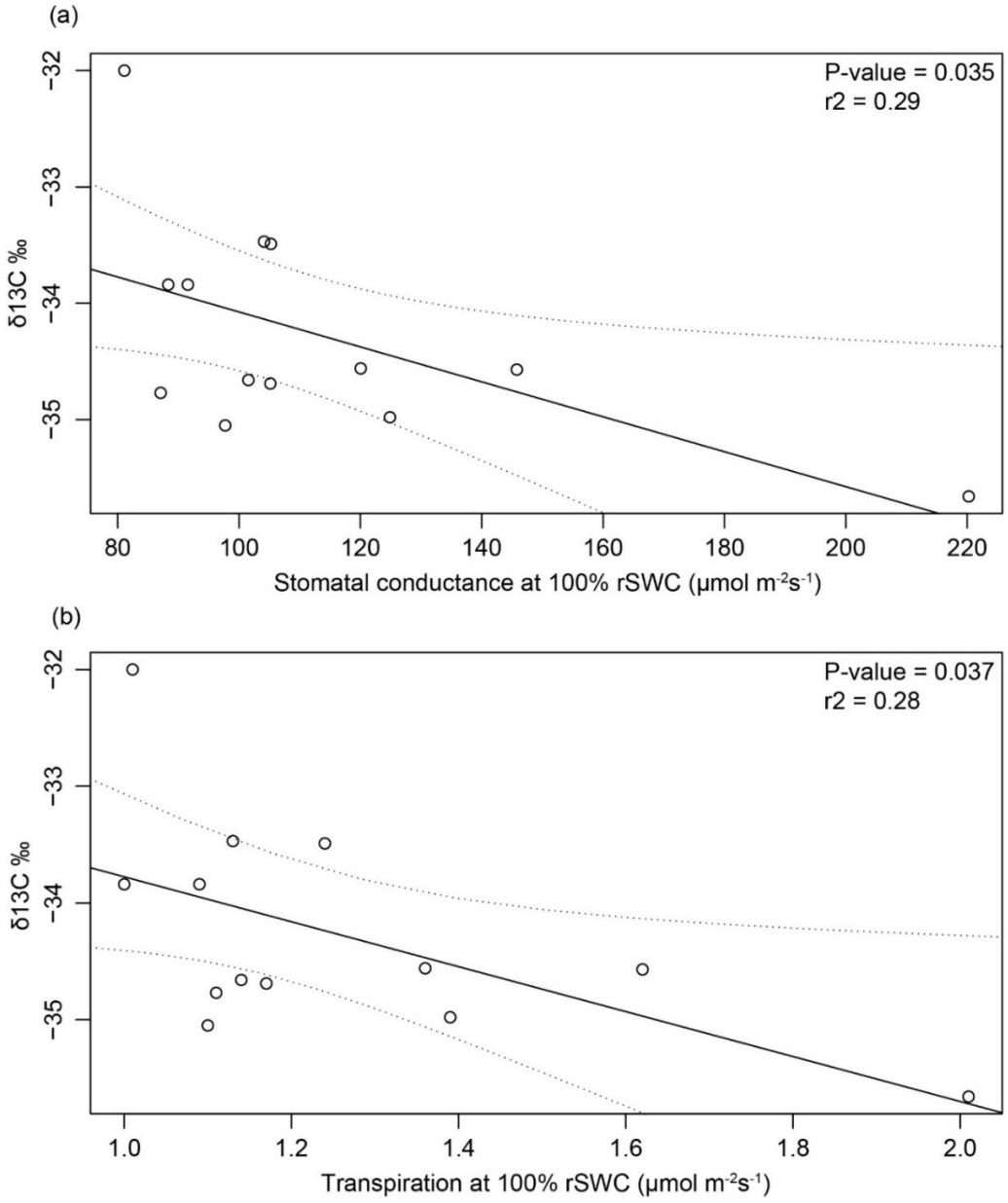

Fig. 3.7. (a) Significant negative relationship between $\delta^{13}C$ and stomatal conductance at 100% rSWC. (b) Significant negative relationship between $\delta^{13}C$ and transpiration at 100% rSWC. For both plots, the bold fit line represents the equation of the linear regression model. The adjacent dashed lines represent the lower and upper 95% confidence intervals. The p-value and adjusted r^2 value associated with the linear models are provided for each association.

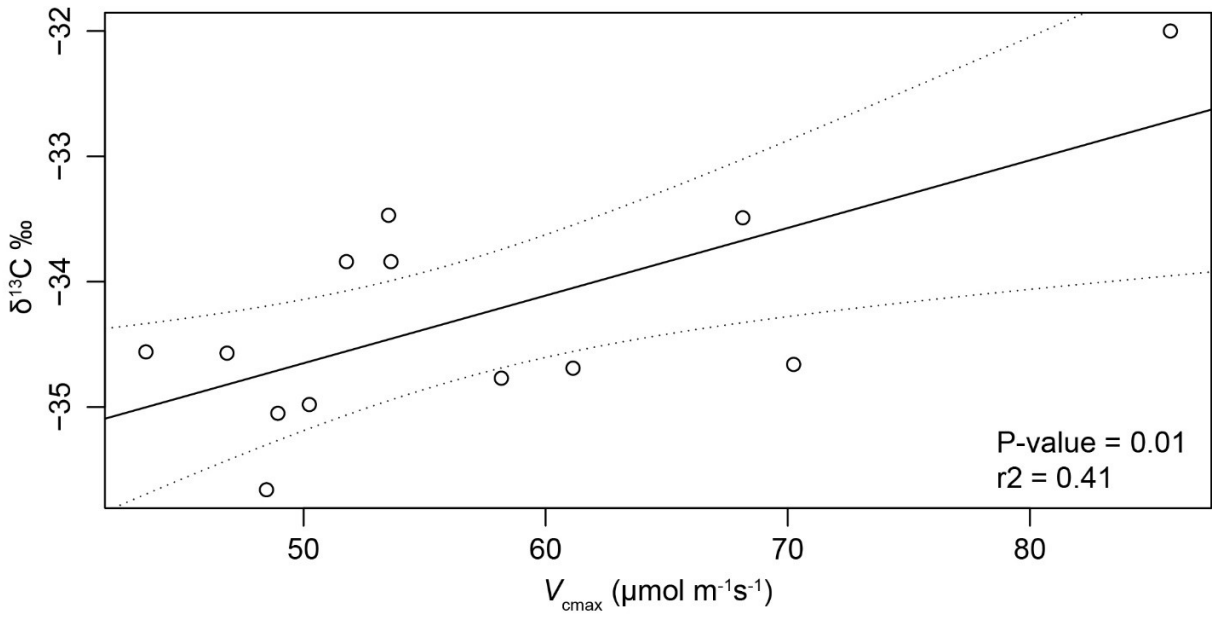

Figure 3.8. Significant positive relationship between V_{cmax} and $\delta^{13}C$ of the 13 Arabidopsis ecotypes. The bold fit line represents the equation of the linear regression model. The adjacent dashed lines represent the lower and upper 95% confidence intervals. The p-value and adjusted r^2 values associated with the linear model are provided for each association.

Significant positives associations were detected between flowering time and percent carbon content in dried leaf tissue (Fig. 3.9a). Since flowering time is closely linked to vegetative biomass accumulation, the same association was observed with the later (Fig. 3.9b). Those ecotypes that flowered later and consequently had elevated vegetative biomass also had a greater proportional content of carbon in their vegetative tissue than the smaller and earlier flowering ecotypes.

Flowering time is closely linked to senescence and the two share a common genetic basis (Wingler *et al.*, 2010). Additionally, delayed senescence has for some time been understood to drive improved crop productivity (Reviewed in: Gregersen *et al.*, 2013),

consequently multiple crop improvement programs have centered on selection for late flowering time and delayed leaf senescence for improved productivity (Borrell *et al.*, 2000). The existence of these physiological trends could lead to the hypothesis that the observed variation in leaf tissue carbon content is to some extent due to late flowering. However, it is important to remember that the tissue harvested for this analysis was done so at the same time during the short dehydration period, which occurred before flowering and senescence. Therefore, the detected association in question may not necessarily reflect improved carbon assimilation due to an increased life cycle or delayed senescence. Additionally, since sampling was performed before flowering it is perhaps unlikely that the differentiation for percent carbon content is due to changes to in source-sink relationships. Such an explanation could well be plausible if tissue harvesting had occurred post floral transitioning, since this phenomenon is well understood to alter nitrogen and carbon resource allocation (Chardon *et al.*, 2014).

Due to the above it is somewhat difficult to discern the biological basis of the association between flowering time and percent carbon content in vegetative tissue. Despite this, I would argue that it is likely due to dynamic partitioning of carbon, since this has recently been demonstrated to show both a strong a genetic basis and developmentally-associated variation even before the onset of floral transitioning (Kolling *et al.*, 2015). Furthermore, variation in carbon partitioning is understood to be independent of photosynthesis (Kolling *et al.*, 2015), which goes someway to explain the lack of detection of an association between leaf carbon content and either operational or maximal rates of photosynthesis. It is conceivable that the different developmental stages that are sure to be linked to the onset of flowering result in variation between the ecotypes in terms of carbon partitioning, which are

thus picked up and linked to flowering time. Although the initiation of the short dehydration period was specifically designed to occur before the onset of flowering, natural variation studies of this kind are very difficult to perform whilst controlling for biases that may arise due to different timings of transitions between developmental stages, thus artefact associations to this end may be detected.

The substantial variation observed for percent nitrogen content demonstrated a very strong association with reproductive performance and cWP to a lesser extent (Fig. 3.10). The association between leaf nitrogen content and productivity is relatively underexplored in Arabidopsis (Guan *et al.*, 2015), however the relationship between the former and both growth and development has been well characterized both at the genetic and physiological level in multiple species (Reviewed in: Gutiérrez, 2012). Upon seed filling, it is necessary for plants to redistribute nitrogen contained within vegetative tissue toward reproductive organs in a process known as nitrogen mobilization. The proportion of nitrogen required from the various sources of acquisition is not presently understood in Arabidopsis. In other species such as Soybean, in excess of 50% of the total nitrogen required for the synthesis of seed storage proteins is supplied via the mobilization of vegetative nitrogen stores as opposed to being supplied through the direct assimilation of atmospheric or soil nitrogen, as such a strong positive association exists between leaf nitrogen content and seed yield in field-grown soybean (Shibles & Sundberg, 1998). To this end, it likely that those ecotypes that demonstrated elevated leaf nitrogen content also demonstrated improved reproductive performance and cWP because of the increased availability of nitrogen for eventual seed filling.

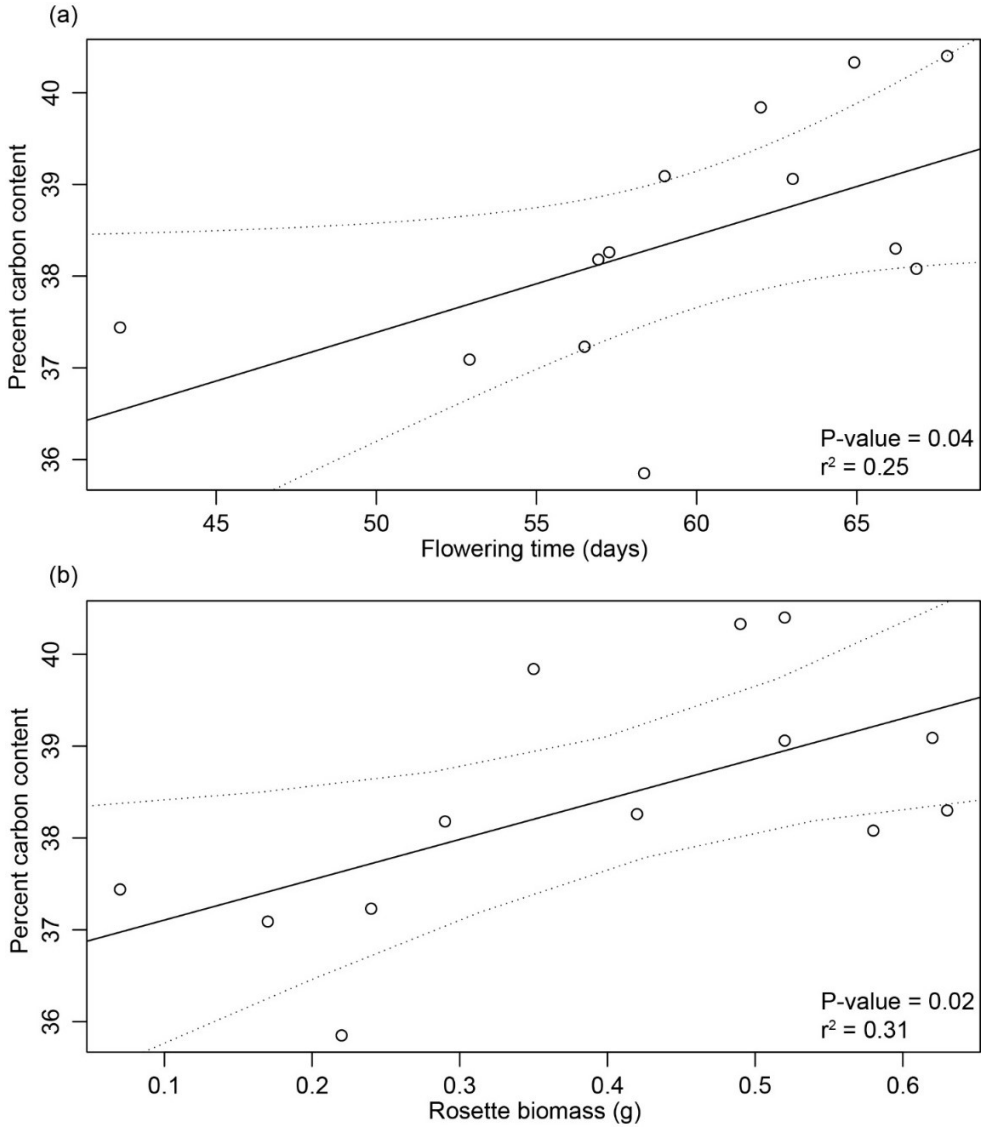

Figure 3.9. Relationship between the percent carbon content of leaf tissue and (a) flowering time and (b) rosette biomass of the 13 Arabidopsis ecotypes. The bold fit line represents the equation of the linear regression model. The adjacent dashed lines represent the lower and upper 95% confidence intervals. The p-value and adjusted r^2 value associated with the linear models are provided for each association.

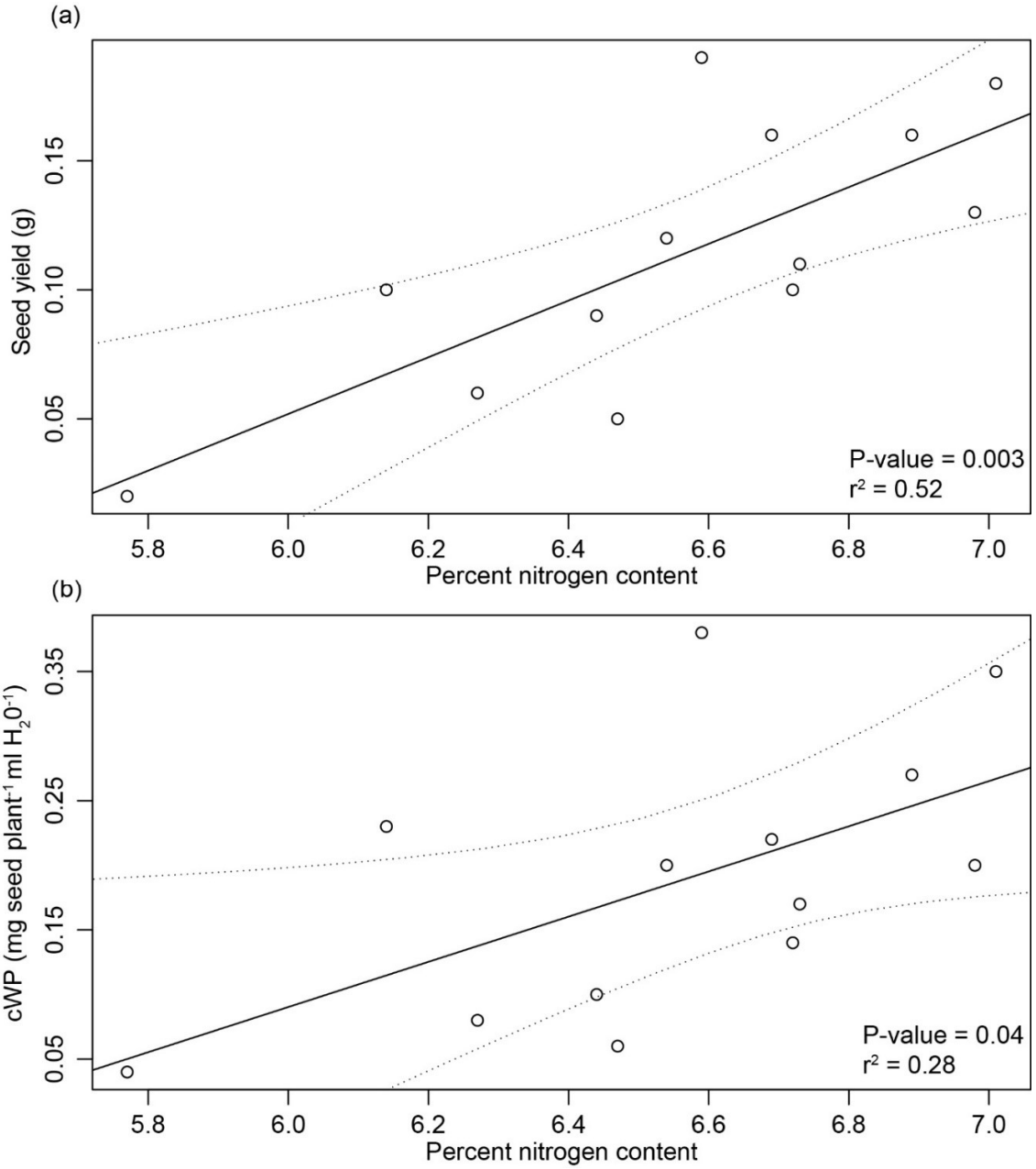

Figure 3.10. Relationship between the percent nitrogen content of leaf tissue and (a) seed yield and (b) calculated water productivity of the 13 Arabidopsis ecotypes. The bold fit line represents the equation of the linear regression model. The adjacent dashed lines represent the lower and upper 95% confidence intervals. The P-value and adjusted r^2 value associated with the linear models are provided for each association.

3.3.4 Water use and water productivity under short and long day conditions

Previous results described as part of this and the preceding Chapter describe the importance of flowering time for determining vegetative biomass accumulation and long term water use. In brief, later flowering ecotypes accumulate more rosette biomass and use more water than those that flower earlier. The standard experiments defining my research involved the growth of all ecotypes under short day conditions until either the completion of a short dehydration period or the initiation of a continuous watering regime, upon which long days persisted for both. Like all flowering plant species, Arabidopsis is very developmentally-sensitive to photoperiods (Reviewed in: Simpson & Dean, 2002), as such it could well be predicted that different experimental conditions in terms of light availability and duration could alter biomass accumulation and water use. For this reason, I tested the effect of day length on both vegetative and reproductive biomass accumulation, as well as long term water use and water productivity.

Strong positive associations were detected between the four above-mentioned parameters across the two parallel experiments conducted under short and long days (Fig. 3.11). The plants that were primarily subjected to short days used more water (Fig. 3.11b), but also accumulated more vegetative biomass (Fig. 3.11a), which is as to be expected since flowering is delayed under short days. Somewhat unexpectedly, these same plants also demonstrated improved reproductive performance (Fig. 3.11c), which compensated for the increased water use to ensure improved water productivity compared to the long day plants (Fig. 3.11d). When divided into separate day length experiments, there is no association between flowering time and reproductive performance or water productivity (Fig. 2.4). The

ecotypes that flower latest under short or long days do not demonstrate the highest reproductive performance. For this reason, it is likely that the improved productivity demonstrated by the ecotypes grown under short days is due primarily to increased partitioning of photosynthetically acquired assimilates into storage starch, which Arabidopsis is known to do to a greater degree under short days compared to long days (Gibson, 2004; Graf & Smith, 2011).

The significant positive relationship between water use, productivity, and flowering time under short and long days, suggests that my experiments performed under short day conditions hold relevance to the performance of these ecotypes under long day conditions also.

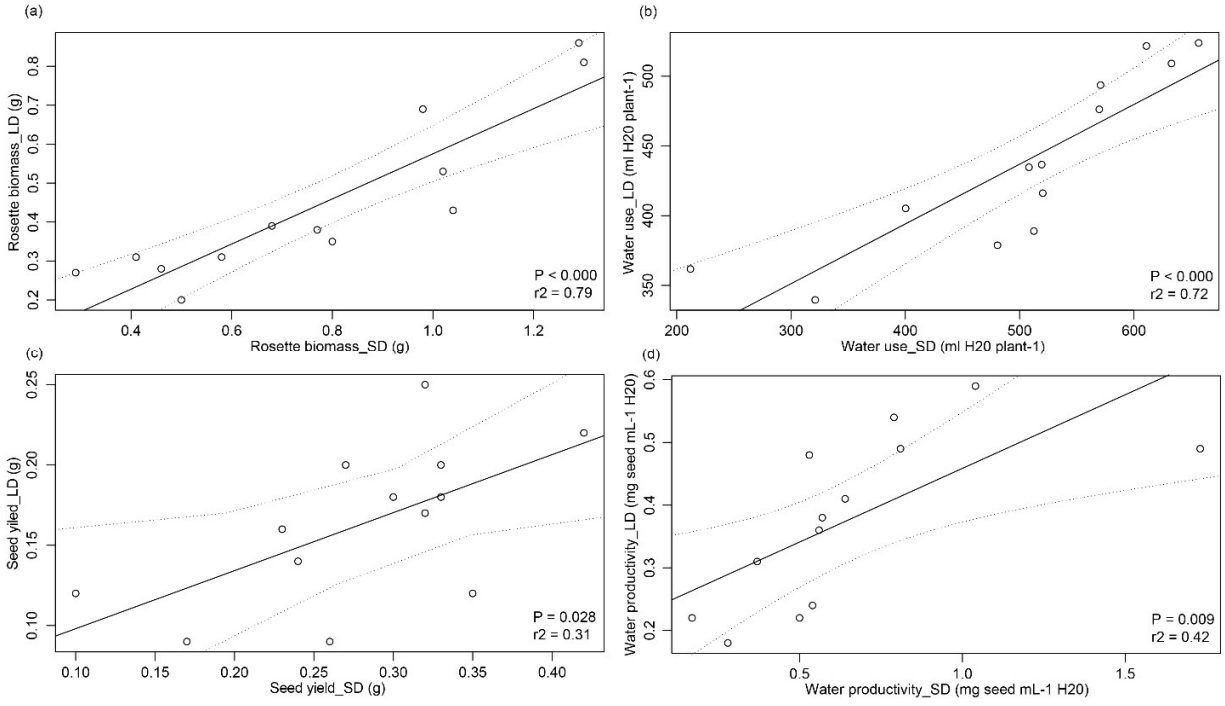

Figure 3.11. Relationship between key parameters assessed as part of the continuous watering experiments of ecotypes grown under long (LD) and short day (SD) lengths.

(a) Rosette biomass accumulation **(b)** Water use **(c)** Seed yield **(d)** Water productivity. For all plots, the bold fit line represents the equation of the linear regression model. The adjacent dashed lines represent the lower and upper 95% confidence intervals. The P-value and adjusted r2 value associated with the linear models are provided for each association.

3.3.5 Vegetative and reproductive biomass of selected ecotypes grown in field-like conditions under monocultures and polycultures and under drought stress and well-watered conditions

To test the translational significance of the key findings from the experiments performed as part of the current and preceding chapter I arbitrarily selected six ecotypes from the original 13 and tested flowering time, biomass accumulation and harvest index for all ecotypes in an outdoor garden experiment. All ecotypes were grown under covered (drought) and uncovered (well-watered) conditions. The Col-0 ecotype was also grown in polyculture and monoculture plots to assess the effect of genetic diversity on the traits of interest. The garden experiment was carried out at the University of Essex from the second week of August 2015 to the third week of October 2015 when plants were harvested. All plants were well soaked following transplanting to the garden experiment. Following this initial watering, the covered plants received no extra watering and all rainwater was blocked by the cover. As well as receiving rainwater, the uncovered plants were kept well-watered every four days. The climatic parameters assessed during the garden experimental period were typical of monthly averages during this period and at this location (Table 3.1; Historical climate data accessed at: www.metoffice.gov.uk/public/weahter/climate/).

Table 3.3. Temperature, relative and humidity, wind speed, and precipitation during the ten weeks of the garden experiment.

		Temperature (^{0}C)			Relative humidity (%)		Wind speed (mph)		
Week	**Month**	**Max.**	**Min.**	**Mean**	**AM mean**	**PM mean**	**Max.**	**Mean**	**Rainfall (mm)**
One	Aug	27.80	9.50	18.17	81.75	65.48	7.60	1.79	29.10
Two	Aug	31.50	10.80	18.07	82.18	70.92	13.60	3.15	39.90
Three	Aug/Sep	23.20	9.30	15.51	88.00	74.75	6.00	1.30	13.50
Four	Sep	22.80	7.50	14.78	79.71	61.51	6.00	1.55	0.00
Five	Sep	23.80	8.90	14.90	85.49	73.34	10.70	2.03	29.40
Six	Sep	22.40	7.30	14.01	90.71	73.51	8.30	2.00	24.00
Seven	Sep/Oct	23.20	5.60	14.17	84.00	62.77	5.40	1.19	0.00
Eight	Oct	23.90	7.80	14.36	88.98	74.35	7.60	1.73	7.80
Nine	Oct	22.20	4.70	11.57	85.42	71.12	5.40	1.05	1.80
Ten	Oct	18.60	6.50	12.07	89.73	82.73	8.30	1.57	8.40

No significant differences were detected between the covered and the uncovered treatments for flowering time for any of the ecotypes or between the Col-0 plants grown in polycultures and monocultures (Fig. 3.12, Table 3.4). This suggests firstly that the drought period the covered plants were subjected to was insufficient to initiate an early flowering drought escape mechanism. Secondly, it suggests that increased biological diversity does not alter flowering time under the particular conditions represented here.

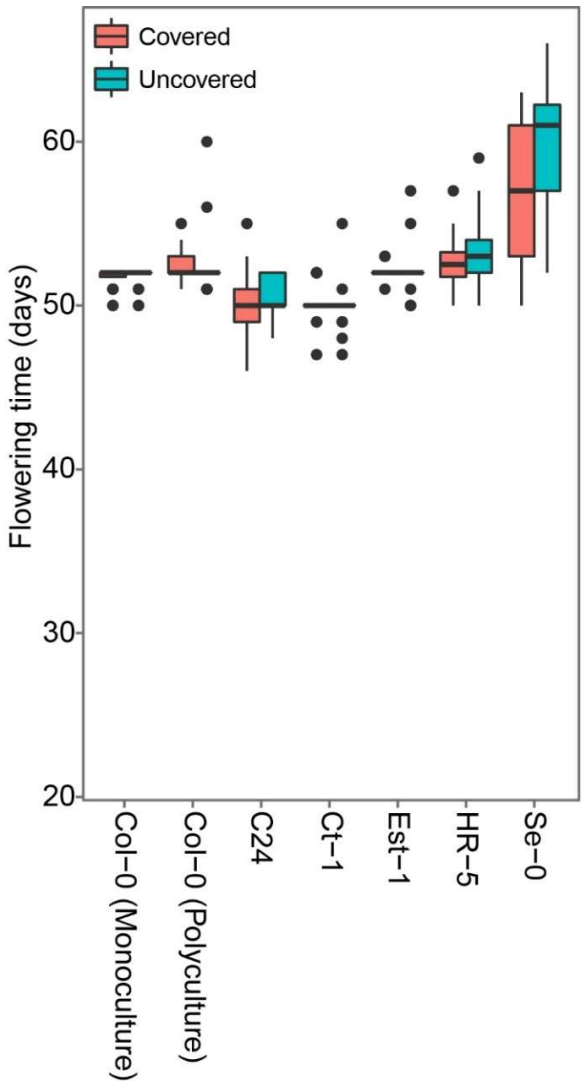

Figure 3.12. Boxplots describing the variation for flowering time demonstrated by the 6 ecotypes as part of the garden experiment. Col-0 is subdivided into those plants grown as part of the monoculture plots and the polyculture plots. All ecotypes are also subdivided into covered (drought) and uncovered (well-watered) plots. There were no significant differences of note. The bold line in the center of the boxplots represents the median, the box edges represent the 25th (lower) and 75th (upper) percentiles, the whiskers extend to the most extreme data points that are no more than 1.5x the length of the upper or lower

segment, and points away from the whiskers indicate extreme outliers. *N* = 15 (polyculture), 20 (monoculture)

Table 3.4. Results from a three-way ANOVA to test for effects of ecotype, cover treatment (covered or uncovered), and culture treatment (monoculture or polyculture) on flowering of the 6 ecotypes comprising the garden experiment.

Source	Sum of squares	Degrees of freedom	Mean square	F-value	P-value
Culture	22.4	1	22.4	3.506	0.0621
Ecotype	2061.1	5	412.2	64.495	< 0.0001
Cover	4.9	1	4.9	0.762	0.3835
Ecotype:Cover	76.2	5	15.2	2.85	<0.05
Residuals	1917.5	300	6.4		

The traits relating to biomass demonstrated significant differences between the cover treatments and between the Col-0 plants grown in mono- and polycultures (Fig. 3.12). With respect to rosette biomass accumulation, only the Col-0 ecotype demonstrated differences between the covered and the uncovered treatments. Interestingly, the difference was converse for the mono- and polyculture plants. For the monoculture Col-0 plants, the uncovered plants accumulated more rosette biomass, whereas the uncovered polyculture Col-0 plants accumulated less than their covered counterparts (Fig. 3.12a, Table 3.5). For seed yield, the only difference between cover treatments was observed between the monoculture Col-0 plants, where covered plants demonstrated improved reproductive performance (Fig. 3.12b, Table 3.6). Chaff biomass accumulation was significantly higher for covered Ct-1, Est-1, Se-0 ecotypes (Fig. 3.12c, Table 3.7). Seed yield was significantly higher for the C24 plants grown under cover and for the Col-0 monoculture plants (Fig. 3.12b, Table 3.6).

It is interesting to note the differences in biomass accumulation between the Col-0 plants grown in mono- and polycultures. Rosette biomass (Covered P-value = 0.05, Uncovered P-value = 0.05) and chaff biomass (Covered P-value = 0.05, Uncovered P-value = 0.05) accumulation were significantly higher for the monoculture treatments (Fig. 3.11). However, reproductive performance was not significantly different between the two treatments. Thus, the harvest index of the both the covered (Covered P-value = 0.05) and the uncovered (Uncovered P-value = 0.05) Col-0 polyculture plants was significantly greater than the equivalent monoculture plants (Fig. 3.11). Although the environmental conditions were substantially more heterogeneous than controlled environment conditions, all plots were subjected to similar uncontrolled conditions. This coupled with the lack of effect on flowering time by either the cover or the genetic mixture treatments suggests that the observed effect on harvest index (Table 3.8) may be a result of root and/or soil-microbiome dynamics that are adjusted through higher plant-level genetic variation.

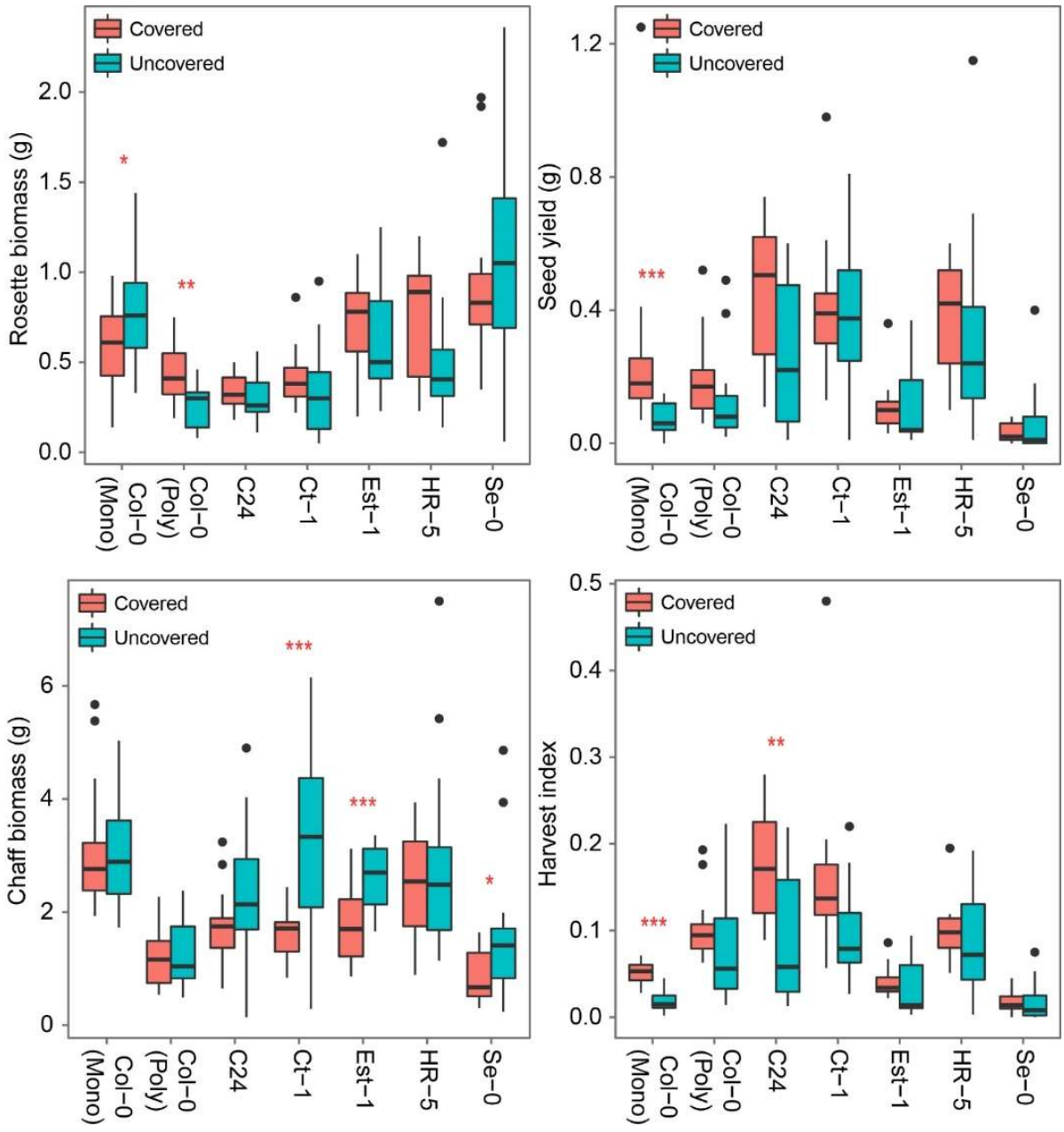

Figure 3.13. Boxplots describing the variation for (a) rosette biomass, (b) seed yield, (c) chaff biomass, and (d) harvest index demonstrated by the 6 ecotypes as part of the garden experiment. Col-0 is subdivided into those plants grown as part of the monoculture plots and the polyculture plots. All ecotypes are also subdivided into covered (drought; red) and uncovered (well-watered; blue) plots. Red asterisks denote significant

difference between the same ecotype grown under covered and uncovered conditions, where *** = P < 0.001, ** = P < 0.01, and * = P < 0.05. The bold line in the center of the boxplots represents the median, the box edges represent the 25th (lower) and 75th (upper) percentiles, the whiskers extend to the most extreme data points that are no more than 1.5x the length of the upper or lower segment, and points away from the whiskers indicate extreme outliers. *N* = 15 (monoculture), 20 (polyculture)

Table 3.5. Results from a three-way ANOVA to test for effects of ecotype, cover treatment (covered or uncovered), and culture treatment (monoculture or polyculture) on rosette biomass accumulation of the 6 ecotypes comprising the garden experiment.

Source	Sum of squares	Degrees of freedom	Mean square	F-value	P-value
Culture	0.404	1	0.4041	4.139	< 0.05
Ecotype	10.476	5	2.0951	21.46	< 0.0001
Cover	0.072	1	0.072	0.737	0.3915
Ecotype:Cover	1.357	5	0.2262	2.317	<0.05
Residuals	20.697	300	0.0976		

Table 3.6. Results from a three-way ANOVA to test for effects of ecotype, cover treatment (covered or uncovered), and culture treatment (monoculture or polyculture) on seed yield of the 6 ecotypes comprising the garden experiment.

Source	Sum of squares	Degrees of freedom	Mean square	F-value	P-value
Culture	0.387	1	0.3868	13.208	< 0.001
Ecotype	3.474	5	0.6949	23.728	< 0.0001
Cover	0.197	1	0.1975	6.743	< 0.05
Ecotype:Cover	0.306	5	0.0511	1.744	0.1123
Residuals	6.209	300	0.0293		

Table 3.7. Results from a three-way ANOVA to test for effects of ecotype, cover treatment (covered or uncovered), and culture treatment (monoculture or polyculture) on chaff biomass accumulation of the 6 ecotypes comprising the garden experiment.

Source	Sum of squares	Degrees of freedom	Mean square	F-value	P-value
Culture	40.23	1	40.23	41.029	< 0.0001
Ecotype	54.84	5	10.97	11.186	< 0.0001
Cover	17.97	1	17.97	18.33	< 0.0001
Ecotype:Cover	16.27	5	2.71	2.765	0.131
Residuals	207.87	300	0.98		

Table 3.8. Results from a three-way ANOVA to test for effects of ecotype, cover treatment (covered or uncovered), and culture treatment (monoculture or polyculture) on harvest index of the 6 ecotypes comprising the garden experiment.

Source	Sum of squares	Degrees of freedom	Mean square	F-value	P-value
Culture	0.0891	1	0.08912	13.208	< 0.0001
Ecotype	0.3668	5	0.07335	23.728	< 0.0001
Cover	0.0582	1	0.05817	6.743	< 0.0001
Ecotype:Cover	0.048	5	0.008	1.744	0.131
Residuals	0.4742	300	0.00224		

The general trends observed in the previous chapters with respect to flowering time and biomass accumulation were tested using all the ecotypes grown under the two different conditions associated with the garden experiment. As before, a highly significant positive correlation was observed between flowering time and vegetative performance, where plants that flowered later accumulated significantly more rosette biomass (Fig. 3.14a). This trend was not matched with reproductive performance, which is reflected by the significant negative relationship detected between rosette biomass and seed yield (Fig. 3.14b). These associations suggest that flowering time and vegetative performance are ineffective at predicting eventual reproductive performance.

The relationship between all the traits assessed as part of the garden experiment was further explored through PCA (Fig. 3.15). The loadings of these traits onto the first two principle components compliments the above, in the sense that yield and harvest index load in the opposite direction to flowering time and rosette biomass with respect to principle component one. Interestingly, chaff biomass did not appear to be an accurate predictor of seed yield (P-value = 0.03, r2 = 0.05), whereas it had been highly accurate predictor for this purpose as part of the controlled environment experiments (Fig 2.4). This is also reflected by its isolated nature in the biplot of the two main principle components (Fig. 3.15). It is also interesting to note that the ecotypes grown under covered treatments tend to load more strongly in a positive sense onto principle component two, which is explained primarily by variation for harvest index. Additionally, the Col-0 polyculture plants load very positively onto this same principle component, whereas the monoculture plants load in the opposite direction, which again reflects the previously described differences between these treatments.

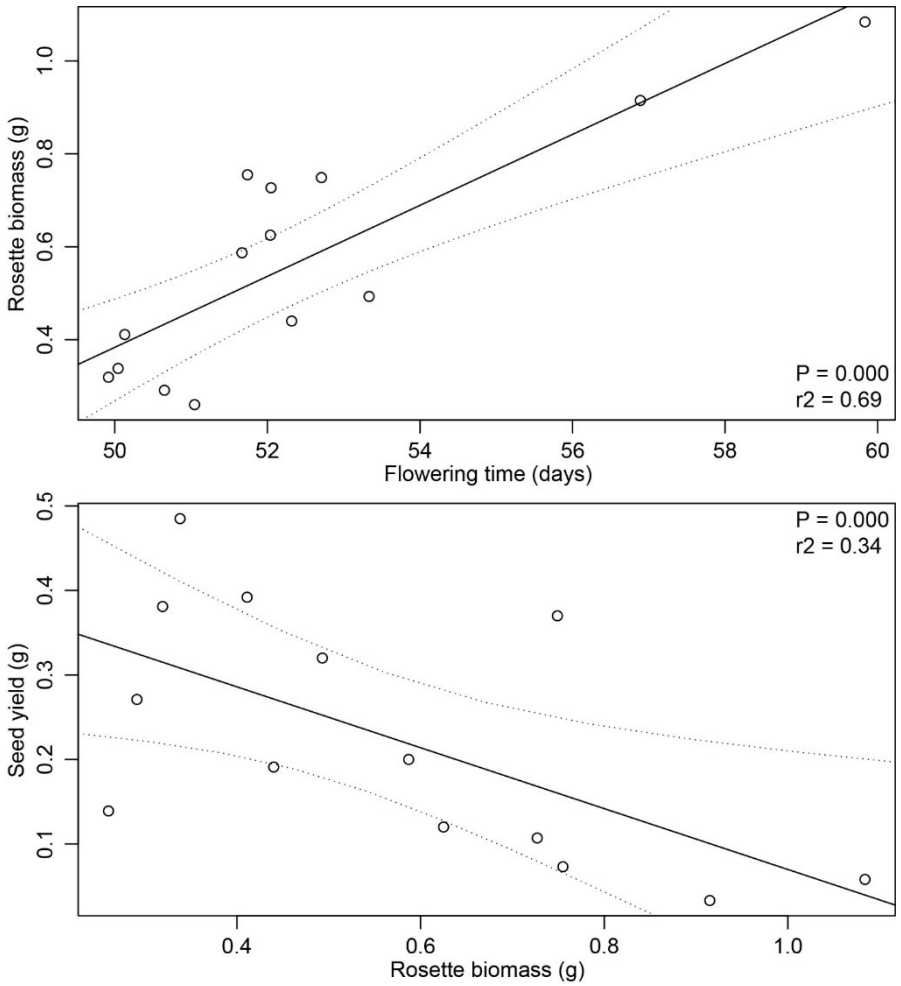

Figure 3.14. (a) Relationship between flowering time and rosette biomass for all ecotypes grown under all conditions as part of the garden experiment. (b) Relationship between rosette biomass and seed yield for all ecotypes grown under all conditions as part of the garden experiment. For both plots the bold fit line represents the equation of the linear regression model. The adjacent dashed lines represent the lower and upper 95% confidence intervals. The P-value and adjusted r^2 values associated with the linear model are provided for each association.

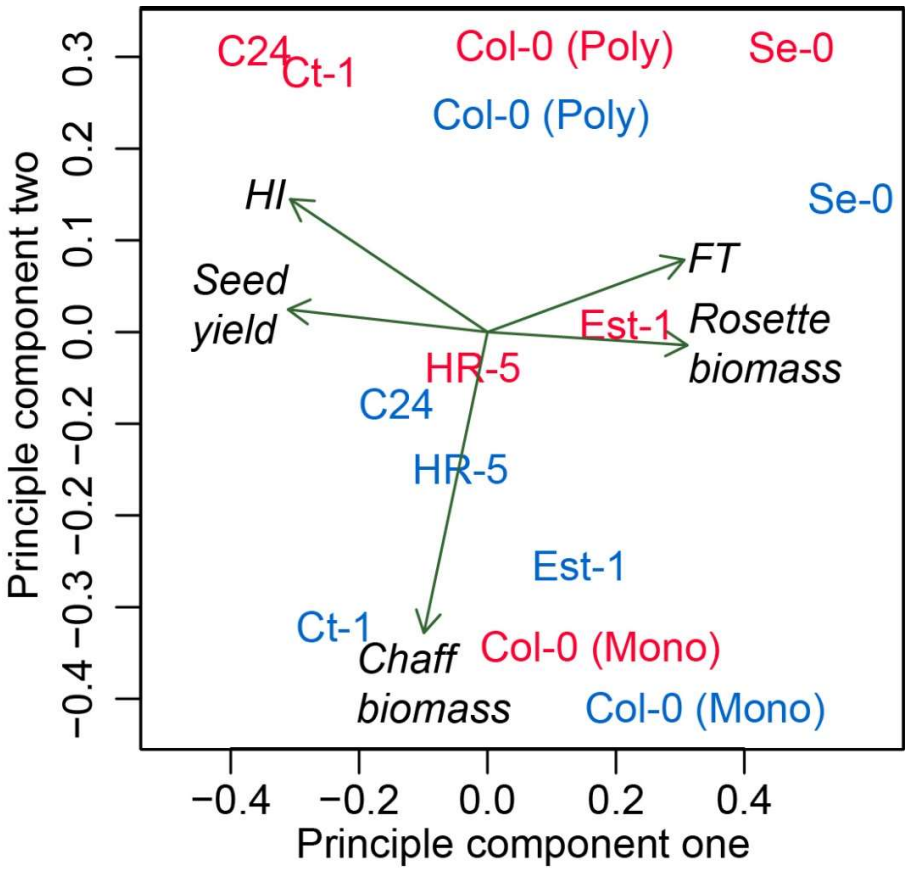

Figure 3.15. Biplot of the loadings of the five phenotypic parameters assessed as part of the garden experiment onto the first and second principle components of trait space that describe the performance of the six ecotypes. Each arrow represents a vector of loadings. The direction of each arrow represents the relationship of a variable to the two main principle components. Additionally, the loadings of the individual ecotypes onto the traits spaces are also described based on their association to the two main principle components. Each ecotype is subdivided into those grown under covered (red) and uncovered conditions (blue), additionally Col-0 is subdivided further into monoculture (mono) and polyculture (poly) plots.

To test the translational significance of the controlled environment work described in this and the preceding Chapter, I tested the relationship between flowering time and seed yield as determined in the two continuous watering experiments and the outdoor garden experiments (Fig 3.14). In all cases, the associations were highly significant and positive, thereby suggesting evaluations of biomass accumulation and calculations of water use and water productivity in controlled environmental settings are reflective of performance in outdoor and more agronomic-like environments.

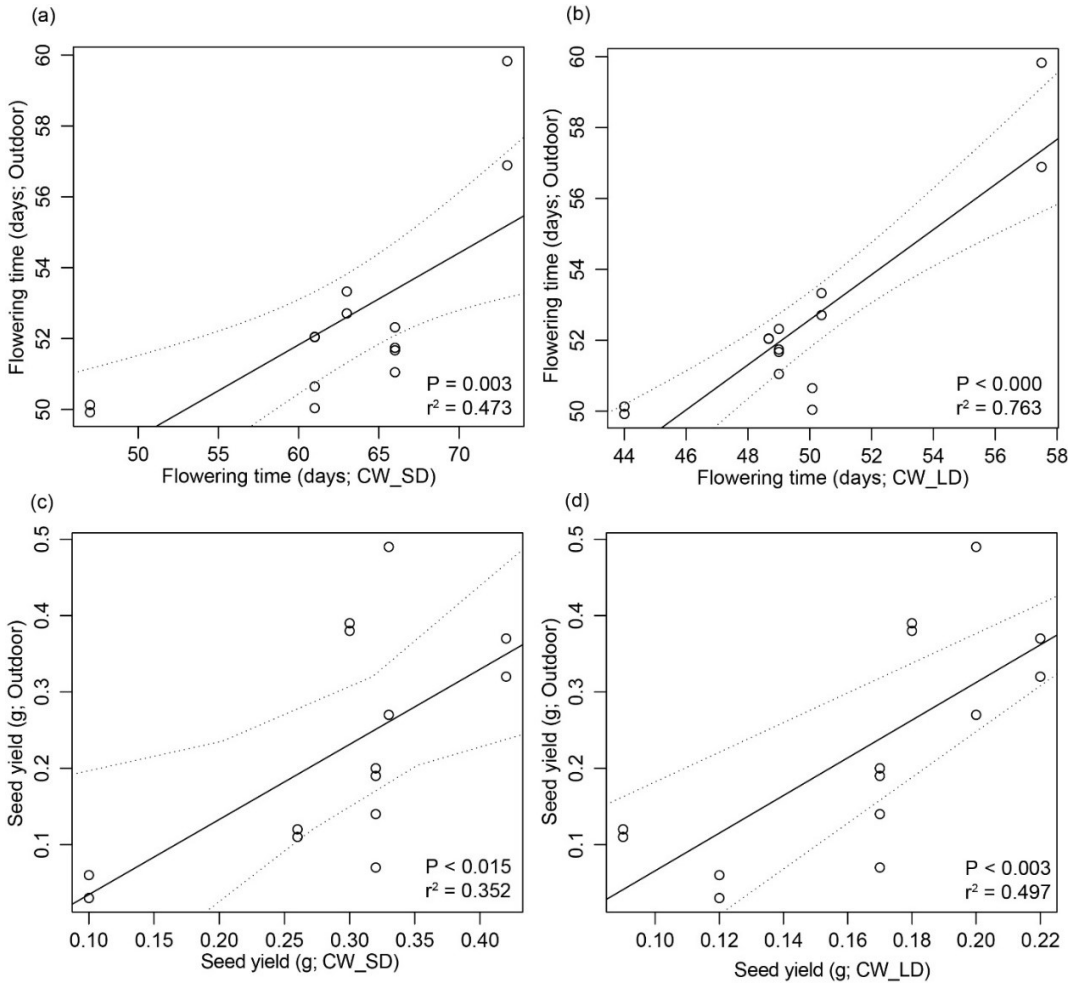

Figure 3.16. (a) Relationship between flowering of ecotypes grown as part of the short day continuous watering (CW_SD) experiment and flowering time of the same ecotypes grown

as part of the outdoor garden experiment. **(b)** Relationship between flowering of ecotypes grown as part of the long day continuous watering (CW_LD) experiment and flowering time of the same ecotypes grown as part of the outdoor garden experiment. **(c)** Relationship between seed yield of ecotypes grown as part of the CW_SD experiment and seed yield of the same ecotypes grown as part of the outdoor garden experiment. **(d)** Relationship between seed yield of ecotypes grown as part of the CW_LD experiment and seed yield of the same ecotypes grown as part of the outdoor garden experiment.

3.4 General Discussion

This study was undertaken to answer fundamental questions that presented themselves after Chapter Two. These related firstly to the usefulness of employing additional proxy parameters for predicting water use and productivity, and secondly with regards to understanding the effect of important biological and environmental variation on water use and productivity. These questions were addressed through a short dehydration experiment where potential and operational rates of photosynthesis were assessed, as well as leaf level carbon and nitrogen content, through parallel continuous watering experiments where day length was manipulated, and through an outdoor garden experiment where the effect of both biological variation and water availability were tested.

The thirteen ecotypes comprising this study represented very little natural variation for potential photosynthetic capacity, additionally this capacity did not appear a suitable proxy for either vegetative or reproductive biomass accumulation. $\delta^{13}C$ did not appear a suitable proxy of either cWUE or biomass accumulation, however variation for $\delta^{13}C$ did associate with variation for both T and g_s, but not A, providing further evidence to suggest

that changes to iWUE are primarily afforded by reductions to the denominator, i.e. water use. The percent leaf content of carbon appeared tightly linked to flowering time and rosette biomass accumulation, although this is likely an artefact of differential carbon partitioning that occurs at different development stages, which are likely correlated with flowering time variation, as opposed to directly relating to flowering time *per se*. Leaf nitrogen content was observed to be a highly accurate proxy of seed yield in Arabidopsis, which is likely to be due to the increased availability of nitrogen for mobilization during seed filling. Future research should look to further characterize the variation that exists for leaf nitrogen content using a larger number of genetically distinct ecotypes and determine with greater confidence how and whether leaf level nitrogen content associates with improved productivity and water productivity.

Comparing the flowering time and reproductive performance of the 13 ecotypes when kept well-watered and when subjected to a short dehydration period revealed that the short dehydration period is not a true drought stress in the sense that it is not fitness-limiting nor does it initiate an early flowering response which is typical of Arabidopsis when subjected to drought. This is clearly of utmost important for our estimations of water use and water productivity, since reductions in either flowering time or seed yield due to short dehydration would yield inaccurate values for cWU and cWP.

The parallel continuous watering experiments demonstrated that variation for water use and biomass accumulation is relatively stable across short and long days, advising that findings achieved through short day-based experiments are translatable across to long day conditions. To this end, it was observed that biomass accumulation is correlated across the controlled environment studies and the outdoor garden experiment, which provides incentive

to suggest that the controlled environment studies are somewhat reflective of plant performance in agricultural and/or ecological systems.

The lack of effect of short dehydration on flowering and fitness was also reflected in the garden experiment, where reduced water availability was persistent for much longer. This hints at the idea that Arabidopsis, or at least these particular genetic variants, is already reasonably drought adapted, which is somewhat irrelevant for this research where water use, as opposed to drought resistance, is of interest. Future research should attempt to quantify Arabidopsis water use in such outdoor agroecological settings through dynamic measures of soil moisture content or evapotranspiration in order to determine how water use variation changes or remains stable under such conditions.

In addition to the above, the garden experiment also demonstrated that the Col-0 ecotype appears to have a markedly elevated harvest index when grown in diverse genotypic mixtures as opposed to monocultures under both water replete and drought conditions. Improved reproductive performance of Arabidopsis grown under abiotic perturbations has been demonstrated to be achieved through biological diversity (Creissen *et al.*, 2013). The elevated performance observed in this cited study was achieved through compensation, where productivity of the total biological system, i.e. multiple ecotypes, was observed to improve as opposed to the productivity of individual ecotypes improving. The improved harvest index of Col-0 when grown in a diverse biological system as described in this study is not achieved via compensation, since biomass accumulation was determined on a genotype-by-genotype basis. Therefore, it is hard to discern the causal basis of the observed difference in harvest index between the diverse and monoculture systems. It could be envisaged that a system consisting of entirely Col-0 plants may have elevated root

competition, i.e. a reduction in the a of soil resources to roots that is caused by other roots (Schenk, 2006), compared to a system consisting of Col-0 and other ecotypes. This could be plausible if those other ecotypes had reduced exploratory root systems compared to Col-0, therefore the competition for resource availability experienced by individual Col-0 plants would be reduced. Equally, it is plausible that diverse plant systems culture diverse root microbiota communities which could function to achieve improved resource availability, thereby improving reproductive performance. Future research to this end should function to ascertain the effect of root competition in diverse Arabidopsis systems and understand how this influences biomass accumulation and water use at both the system and individual plant level. Additionally, it would be hugely interesting to understand how diversity effects the soil microbial community and how this in turn effects the transcriptional regulation of key genes of constituting microbial species and genes of interest at the plant level.

4. Identification and characterization of QTL underlying natural variation for water use and productivity traits in Arabidopsis

4.1 Introduction

The identification of allelic variation that defines phenotypic variation has important implications for crop improvement. Additionally, the identification of genetic loci that confer ecological adaptations, which are often analogous to those that are central to crop improvement, can further our understanding of the evolutionary progression and prospects of plant populations under the context of global change. Pioneering work to elucidate plant allelic variation for continuous traits through linkage mapping, such as the studies of Sax (1923) and Thoday (1961), were constrained due to a lack of suitable molecular markers. Advances in molecular marker technology (Semagn *et al.*, 2006), have alleviated this significant restraint and expedited the construction of genetic maps with high marker densities. The increase in accessibility of the tools required to develop these markers and screen them across mapping population panels, combined with the availability of software platforms to perform requisite statistical analyses, has resulted in a huge increase in the number of studies utilizing linkage disequibrlium (LD) mapping (also referred to as quantitative trait loci (QTL) mapping). Consequently, our understanding of the genetic architecture of complex plant traits has dramatically improved over the past four decades (Lipka *et al.*, 2015).

Kowalski *et al* (1994) published the first successful application of QTL mapping in Arabidopsis, where multiple flowering time QTL were identified. Since then, QTL mapping has been used for the identification of genetic regions in Arabidopsis that control abiotic stress resistance, biotic stress resistance, viability, phenology, physiology, and molecular processes (Reviewed in: Alonso-Blanco *et al* 2009). A direct result of the surge in use of QTL mapping for gene discovery in Arabidopsis is that there are now over 100 collections of RILs, the most popular mapping population type in Arabidopsis, available to the scientific community (See: http://www7.inra.fr.vast/RILs.htm). Additionally, many of these populations have associated near isogenic lines (NILs) developed, facilitating the further analysis of particular regions and QTLs of interest (Weigel, 2012).

With specific reference to understanding the performance of Arabidopsis under the context of water limitations, QTL mapping has been successfully incorporated into previous work centered on understanding the genetic basis of drought resistance (Ghandilyan *et al.*, 2009; El-Soda *et al.*, 2014; Lovell *et al.*, 2015) and WUE (McKay *et al.*, 2003, 2008; Juenger & Mckay, 2005; Hausmann *et al.*, 2005; Masle *et al.*, 2005). The work of Masle *et al* (2005), the most cited of all of these studies, identified the well-characterized *ERECTA* gene as a key regulator of WUE, measured as $\delta^{13}C$, primarily through controlling stomatal conductance. Separate efforts to elucidate the genetic basis of WUE, again measured as $\delta^{13}C$, have built up a significant body of evidence to suggest that WUE is pleiotropically linked to flowering time (McKay *et al.*, 2003, 2008; Juenger & Mckay, 2005; Hausmann *et al.*, 2005). As discussed in Chapters one and two, these studies postulate a positive genetic correlation between flowering time and WUE. This is supported by their QTL mapping identifying flowering time genes following mapping for $\delta^{13}C$.

Consistent results demonstrated as part of Chapters Two and Three provide substantial evidence to suggest that the observed link between flowering time and WUE, as described above, is debatable. It could be argued that it is somewhat unlikely that an increase in flowering time, i.e. longer life cycle, would be concurrent with improved WUE, because a plant that lives for a longer period would likely use more water, consequently it would have reduced WUE. Therefore, I would argue that traits directly relating to water use and productivity are more agronomically relevant.

With respect to the above, the present study was undertaken to identify QTL that underlie variation for water use and water productivity. Chapter Two demonstrated that substantial variation exists for key traits relating to water use and water productivity. Crucially, this variation was observed to have a significantly heritable basis, thereby indicating the viability of dissecting out the genetic basis of these traits. The cWU and cWP of the ecotypes C24 (505.93 cWU, and 0.20 cWP) and Col-0 (625.16 cWU, and 0.09 cWP) were substantially different, as such a RIL population developed with these ecotypes as the parental lines was acquired (Törjék *et al.*, 2006; Kindly provided by Dr Rhonda Meyer) for QTL mapping for these and other related traits.

4.2 Materials and methods

4.2.1 Plant material, growth conditions, and trait parameters assessed

For this study, the RIL mapping population developed using the Arabidopsis ecotypes Col-0 and C24 as parental lines (Törjék *et al.*, 2006) was used for QTL mapping. 168 RILs, plus the two parental lines, were phenotyped for short term water use, flowering time, the

number of rosette leaves at the point of bud initiation, rosette biomass, chaff biomass, and seed yield (n = 15). Additionally, assessment of these parameters facilitated the calculation of long time water use (cWU) and water productivity (cWP), as described in chapters two and three.

All plants were grown in the controlled environment room described in Chapter Two in pots with precise volumes of soil. An SD experimental period, as described in Chapter Two, was initiated at 50 days. The SD period facilitated the assessment of short term water use as the slope of the linear regression of the rate of drying from ~100% to ~20% rSWC. Following the completion of the SD period, all plants were transferred to the glasshouse, where conditions were as described in Chapter Two. Here, flowering time and the number of rosette leaves at bud initiation were determined as plants were kept well-watered. Plants were kept well-watered in the glasshouse until the opening of the final flower, at which point the entire plant was bagged and allowed to dry down. Once completely dried down, plant biomass components were separated and measured as rosette biomass (vegetative biomass), chaff biomass (stalks and pods; reproductive biomass), and seed yield (reproductive biomass). Long time water use was calculated as a short-term water use multiplied by flowering time to give cWU. Water productivity was calculated as seed yield divided by cWU to give calculated water productivity (cWP).

4.2.2 Statistical Analysis

All statistical analyses were performed within the R software environment for statistical computing and graphics (R Development Core Team, 2008). The SD experiment to assess the above described phenotypic parameters was temporally blocked over a period

of two years. One-way ANOVA comparison of means tests were performed across all lines and all blocks to determine the existence of experimental block effects that could potentially confound further analysis and the QTL mapping. Significant block effects were detected for all traits, as such I extracted BLUPs using GLMMs, as described in Chapter Two (2.2.4). BLUPs were used to calculate estimated means, by adding them to population means. Estimated means were used for subsequent assessment of genetic correlations and for the QTL mapping. GLMMs allowed for the determination of V_P and V_G for all trait parameters. These were in turn used to obtain estimates of H^2 and C_{VG}, i.e. evolvability, as described in Chapter Two.

4.2.3 QTL mapping

The R package 'qtl' developed by Arends *et al.*, (2010) was used for all QTL analyses. The initial qtl cross object, i.e. the set of files required for QTL mapping in R-qtl, was converted from the F_2 default to a RIL object using the convert2riself() function. The checkAlleles() function, which tests for marker switching, identified no apparent problems with these data. The Lander-Green algorithm, i.e. hidden Markov model technology, was used to re-estimate the genetic map, in terms of the centimorgan (cM) distances between markers, using the est.map() function with an assumed genotyping error rate of 0.001. The re-estimated genetic map, based on the lines incorporated in this study, was preferred to the original genetic map, which was based on over 400 RILs.

Hidden Markov model technology was employed to calculate genotypes at steps of every 1cM between genotyped markers using the calc.genoprob() function. Two-dimensional QTL mapping was performed with the scantwo() function. Two-dimensional

mapping was performed using the Hayley-Knott regression method and with 10,000 permutations. Penalties were derived for the penalized logarithm (base 10) of odds (LOD) scores on the basis of the permutations performed during the two-dimensional mapping using the calc.penalties() function. These penalties were used to perform multiple QTL mapping using the stepwiseqtl() function. Multiple QTL mapping involves forward and backward selection to identify the most optimal QTL model that can include both single effect QTLs and interactions between QTLs, where there are separate penalties on main effects and interactions (Broman & Sen, 2009). This method of QTL model detection has been successfully employed in previous and similar Arabidopsis QTL mapping projects (Agren *et al.*, 2013; Oakley *et al.*, 2014).

4.2.4 Validation of *FRIGIDA* and *FLOWERING LOCUS C* as main effect QTL underlying variation for flowering time, rosette biomass, and water use

To test the hypothesis that variation for flowering and water use was determined by the *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*) flowering time genes, markers were identified that would allow for the prediction of the allelic state of these genes in all RILs. The genetic map used in this study did not have markers flanking the genomic position of *FRI*, therefore the two markers upstream of this position, MASC04123 and MASC04725, were used. For *FLC*, flanking markers were available, MASC04531 (left flanking marker) and MASC09207 (right flanking marker). These *predictor* markers are highlighted as part of Fig. 4.4. It was assumed that RILs that had the same allelic form at both *predictor* markers for each gene, i.e. both Col-0 and C24 alleles, would harbor the equivalent alleles of *FRI* or

FLC. Based on this and after determining which RILs had which alleles of these two genes, all RILs were grouped into four groups based on which alleles of *FRI* and *FLC* they were predicted to possess, i.e. *FRI*:Col-*FLC*:Col, *FRI*:Col-*FLC*:C24, *FRI*:C24-*FLC*:Col, *FRI*:C24-*FLC*:24.

The trait performances of all traits assessed as part of this study were compared between the four allelic groups. One-way ANOVA comparison of means test were performed for each trait using the aov() R base code function, where the allelic group classification was used as the predictor variable. For each trait, the appropriate one-way ANOVA model was used to perform a *post-hoc* Tukey test to determine which, if any, allelic groups were significantly different from one another. This was achieved via Tukey Honest Significant Difference (HSD) tests using the HSD.test() function from the 'agricolae' R package (de Mendiburu, 2016).

4.3 Results

4.3.1 Natural variation of water use and productivity traits within the Col-0 x C24 RIL population

A QTL mapping project was undertaken to try and identify chromosomal regions containing genes underlying the natural variation for water use and productivity that was known to exist following the study described in chapter two. A RIL population developed by Törjék *et al* (2006) which used the ecotypes Col-0 and C24 as parental lines was employed for QTL mapping. This population was selected because of the considerable phenotypic differences between the parental lines observed during the study described in Chapter Two, especially for cWU and cWP. 164 RILs (106 Col-0 x C24 and 58 C24 Col-0) were subjected

to a SD experiment in order to determine short term water use. Additionally, flowering time, the number of rosette leaves at bud initiation, biomass accumulation, cWU and cWP were also determined. The SD experiment was divided into four temporally isolated experimental blocks. Highly significant block effects were detected for all traits. Indeed, significant block effects were detected when focusing on just the check genotypes, i.e. Col-0 and C24, suggesting that the existence of the block effects were due to environmental effects as well as genotypic effects (Table 4.1.).

Table 4.1. Comparison of means testing for all traits assessed as part of the QTL mapping. Testing was performed through one-way ANOVA tests for all traits. Comparisons were made based on the means of all ecotypes, just Col-0, and just C24. Significance is indicated at *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. Non-significant differences are indicated by n.s.

Trait parameter	P-value		
	Between experimental blocks (All ecotypes)	Between experimental blocks (Col-0)	Between experimental blocks (C24)
Short term water-use (ml^{-1} H_20 $plant^{-1}$)	< .0001	< .0001	< .0001
Flowering time (days)	< .0001	n.s.	< 0.05
Rosette leaves at bud initiation	< .0001	< .001	< 0.05
Rosette biomass (g)	< .0001	< .001	< .0001
Chaff biomass (g)	< .0001	n.s.	< .0001
Seed yield (g)	< .0001	< .001	< .0001
Calculcated water-use (ml^{-1} H_20 $plant^{-1}$)	< .0001	< .0001	< 0.05
Calulcated water productivity (mg seed $plant^{-1}$ ml H_20^{-1})	< .0001	< 0.05	< .0001

To control for the detected experimental block effects, appropriate GLMMs were performed and BLUPs were extracted. To make results more understandable estimated means were obtained by adding BLUPs to population means for all traits. The variation in estimated means for all traits was not significantly different from what would be expected of a normal distribution (Fig. 4.1., $P > 0.05$). Furthermore, and with reference to all traits, this

variation was far more extreme than the difference between the parental lines, which is symptomatic of transgressive segregation (Meyer *et al.*, 2004, 2010). For example, the flowering times of Col-0 (74.86 days, estimated mean) and C24 (74.07 days, estimated mean) are not significantly different. Despite this, the variation in flowering time observed within the RIL population was much greater (32.62-111.45 days). It does appear, however, that the extreme variation here is due to outliers, with ~92% of RILs displaying flowering times between 50-90 days, which, although not quite as extreme, is still substantially more variable than the variance displayed between the two parental lines (Fig. 4.1). The pattern in variation and location of the estimated means of parental lines for short term water use, number of rosette leaves at bud initiation, chaff biomass, and cWU is similar to that described for flowering time (Fig. 4.1). The variation in estimated means of seed yield and cWP for the RILs again demonstrated extensive transgressive segregation, however in these instances the parental lines were situated toward the positive extremes of the variation of the mapping population, suggesting that the RILs selected for phenotyping and QTL mapping did not capture much allelic variation in terms of genotypes that elevated productivity (Fig. 4.1).

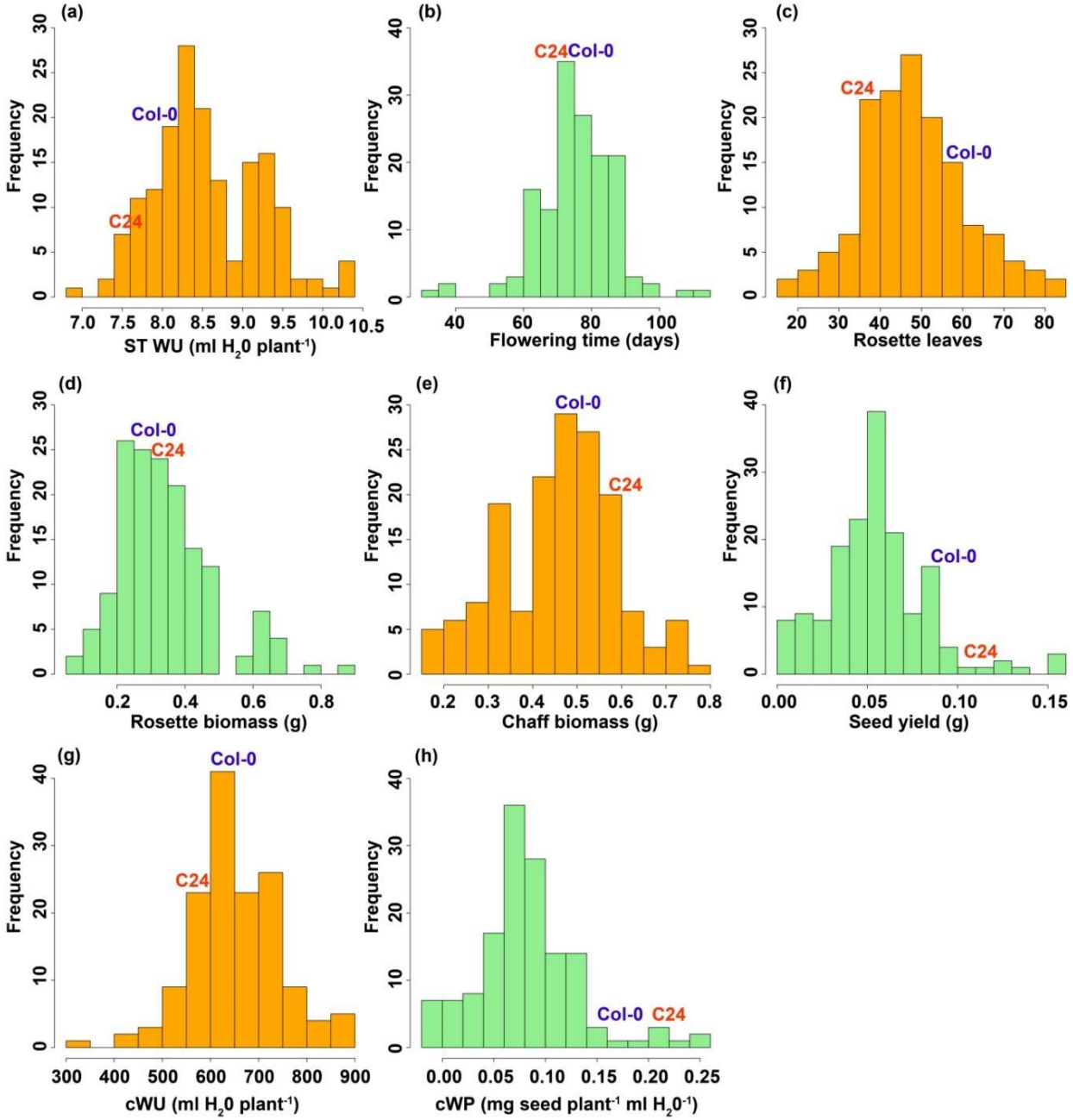

Figure 4.1 Natural variation of estimated means for all traits assessed as part of the QTL mapping. (a) Short term water use (ST WU). **(b)** Flowering time. **(c)** Number of rosette leaves at bud initiation. **(d)** Rosette biomass. **(e)** Chaff biomass. **(f)** Seed yield. **(g)** Calculated water use (cWU). **(h)** Calculated water-productivity (cWP). For all histograms the estimated means of all RILs are ploted and the estimated means of the parental lines are indicated.

For all traits, a Shaprio-Wilk test of normality was performed on the estimated means of all RILs, where all traits desmonstrated variation that is not signficantly different from a normal distribution ($P > 0.05$). $N = 15$.

The GLMMs allowed for the determination of genetic variance (V_G) and total phenotypic variance (V_P). These parameters were used in the calculation of H^2 and C_{VG} (Table 4.2). All traits assessed demonstrated a significantly heritable basis, suggesting that a signification proportion of the variation observed for all traits is determined by genetic variation as well as owing to environmental heterogeneity. Although significant, the H^2 of seed yield (0.21) and cWP (0.20) is relatively low, especially when compared to traits that are known to have a strong genetic basis, such as flowering time ($H^2 = 0.78$). Despite this, these two traits exhibited the highest evolvability, which points towards their elevated propensity to respond to environmental selection compared to other traits. This is concurrent with the observation in Chapter Two that fitness related traits show much greater variation in plasticity, i.e. they are far more sensitive to the environment (Fig. 2.6). In further support of this, traits observed to show reduced plasticity variation, such as flowering time and water use, showed much lower levels of evolvability and high levels of H^2 (Table 4.2).

Table 4.2. Genotypic and phenotypic variation of the eight traits assessed as part of the QTL mapping. The true (arithmetic) mean, standard error (SE), genetic variance (V_G), phenotypic variance (V_P), broad sense heritability (H^2), evolvability (C_{VG}), and significance of H2 (Sig.) are provided for all traits. All traits displayed highly significant heritability at the $p < 0.001$ level (***).

Trait	Mean	SE	V_G	V_P	H^2	C_{VG}	Sig.
Short term water-use	8.56	0.02	0.49	0.84	0.58	8.17	***
Flowering time (days)	74.25	0.40	132.20	170.12	0.78	15.49	***
Rosette leaves at bud initiation	46.14	0.59	186.07	360.58	0.52	29.57	***
Rosette biomass (g)	0.32	0.01	0.02	0.04	0.63	47.53	***
Chaff biomass (g)	0.51	0.01	0.02	0.06	0.36	28.60	***
Seed yield (g)	0.07	0.00	0.00	0.01	0.21	51.89	***
Calculated water-use	637.83	3.65	9454.70	13404.30	0.71	15.24	***
Calculated water-productivity	0.11	0.00	0.00	0.02	0.20	55.83	***

Genetic correlations were estimated between all traits as Pearson's product moment correlations between estimated means (Fig. 4.3). An interesting observation of note here is the significant negative genetic correlation between short term water use and flowering time. This suggests that plants that use more water in the short term, and could thus be conceived to be more drought susceptible, flower earlier. This is in contradiction to the previous two Chapters, where the contrary was observed (Fig. 2.4).

The negative genetic link between short term water use and rosette biomass is contradictory to what was detected in Chapters Two and Three (Fig. 2.4) and suggests that plants with reduced biomass use less water in the short term (Fig. 4.3). Conversely, rosette biomass is much more strongly genetically associated, in a positive sense, with cWU, which supports previous observations and conclusions. It could therefore be suggested that the link observed here between rosette biomass and short term water use is an artefact of the

minimal variation that exists for short term water use within the RILs, compared to what was observed between the different ecotypes in Chapter Two (Fig. 2.4)

Flowering time was observed to genetically correlate with rosette biomass (positive correlation) and chaff biomass (negative correlation), but not with seed yield (Fig. 4.3). Additionally, negative genetic correlations were detected between rosette biomass and the key fitness related parameters and cWP (Fig. 4.3). These genetic associations provide further support to suggest that neither flowering time nor vegetative performance are accurate proxies, or physical markers, of fitness. Moreover, the significant positive genetic association detected between rosette biomass and cWU highlights the importance of assessing seed yield either directly or as a factor of water use, i.e. cWP.

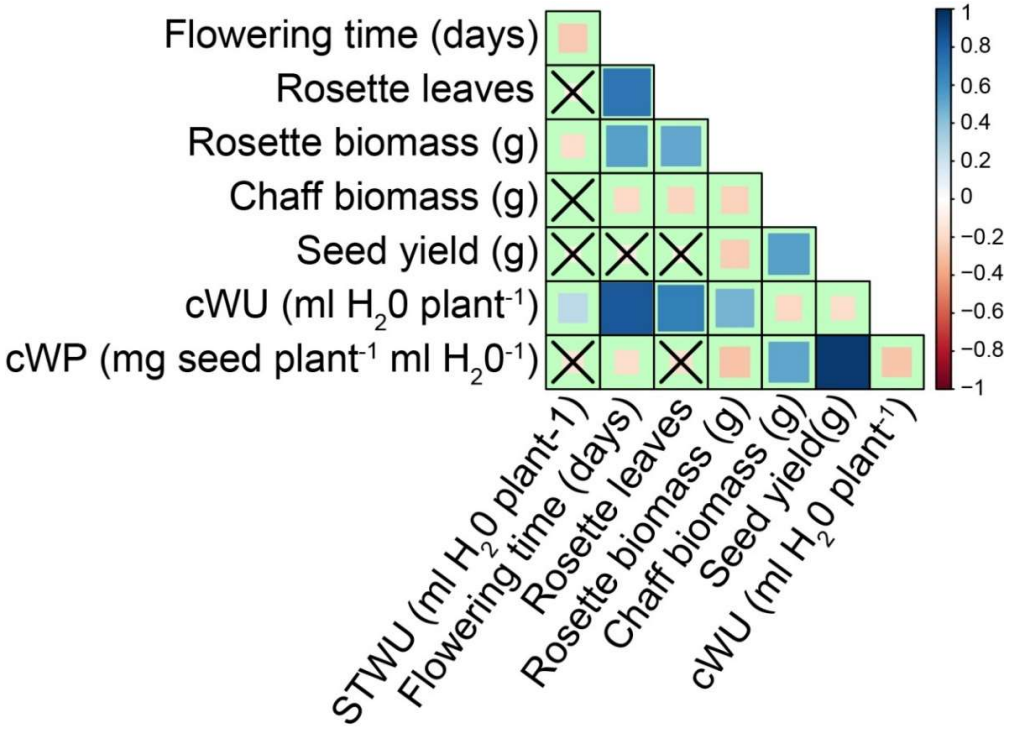

Figure 4.2. Correlation matrix plot describing the pairwise genetic correlations between all traits. The size of the square denotes the p-value, where larger squares

indicate lower p-values, i.e. greater significance. The color of the square indicates the direction of the genetic correlations, with dark red indicating a negative correlation and dark blue indicating a positive correlation (see heat bar to the right). Squares not marked with a cross denote significant genetic correlations (p-value < 0.05). STWU – Short term water use, cWU – Calculated water use, cWP – Calculated water productivity. *N* = 15

4.3.2 QTL mapping for key traits relating to water use and productivity

164 randomly selected individuals from the Col-0 x C24 RIL population were used for QTL mapping of key traits pertaining to water use and water productivity. Adequate phenotyping data were compiled for at least 84.1% of the RILs for all traits. This mapping population is supported by a high density genetic map consisting of 111 single nucleotide polymorphism (SNP) markers, with 26 SNP markers on chromosome one, 19 on chromosome two, 22 on chromosome three, 19 on chromosome four, and 25 on chromosome five (Fig. 4.4a). The Lander-Green algorithm (Lander & Green, 1987) was used to re-estimate the genetic map based on the RILs used in this study (Fig. 4.4b). Preliminary analyses indicated that 97.5% of the markers had been genotyped for all the RILs, where there was a virtually even split in the allelic form of these markers, with 50.3% coming from the Col-0 parental line and 49.7% from the C24 parental line.

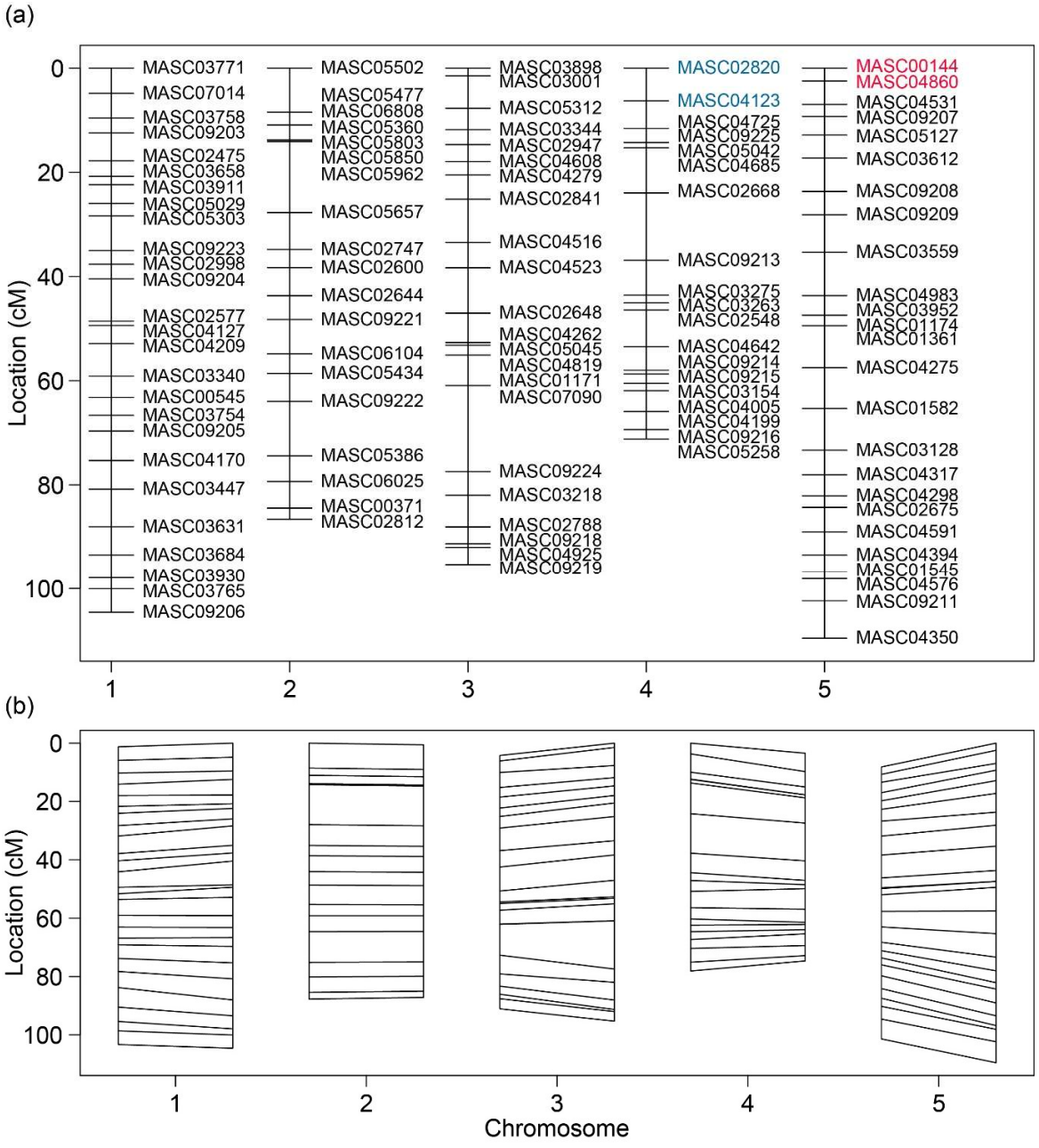

Figure 4.3. The single nucleotide polymorphism (SNP) markers used and their position on the re-estimated linkage map. (a) Position in cMs of all markers on the re-estimated genetic map. The two markers highlighted in blue on chromosome four are the *predictor* markers for *FRI* and the two highlighted in red on chromosome five are the *predictor* markers for *FLC*. **(b)** Comparison of the original genetic map (left) with the re-estimated

genetic map, where the marker distances were re-estimated using the Lander-Green algorithm.

Multiple QTL mapping was performed to identify the multiple QTL model with the maximal penalized LOD score for each trait through forward and backwards elimination of main and interaction effects. No QTL models were identified as significant for seed yield or cWP. For the remaining traits, a total of 13 main effect QTLs were detected. No interaction effects were detected as significant in any of the most significant trait models. There was a high degree of co-localization between the main effect QTLs detected on chromosomes 4 and 5 for flowering time, the number of rosette leaves at bud initiation, rosette biomass, chaff biomass (chromosome 4 only), and cWU (Fig. 4.4, Table 4.3).

A further QTL was detected for flowering time on chromosome one, FT:1, which was observed to co-localize with a QTL for short term water use, STWU:1. An additional QTL for short term water use was also detected on chromosome 3, STWU:3. Which also co-localized with a QTL for cWU, cWU:3 (Fig. 4.4, Table 4.3). Since cWU is calculated based on short term water use, it is highly likely that these are the same QTL. It is also highly likely that cWU:4 and cWU:5 are the same QTLs as FT:4 and FT:5 respectively, since cWU is calculated based on flowering time also. The strong positive genetic correlations observed between flowering time, rosette leaves at flowering, and rosette biomass suggest that the co-localizing QTL for these traits on chromosomes four and five likely represent the same genes also. Additionally, the negative association of these traits with chaff would also suggest that the Chaff:4 is the same QTL as the QTL underlying these aforementioned traits on chromosome 4, however the allelic effects is reversed, i.e. the C24 allele of the QTL on

chromosome 4 reduces chaff biomass, but increases flowering time, leaf number, rosette biomass, and cWU.

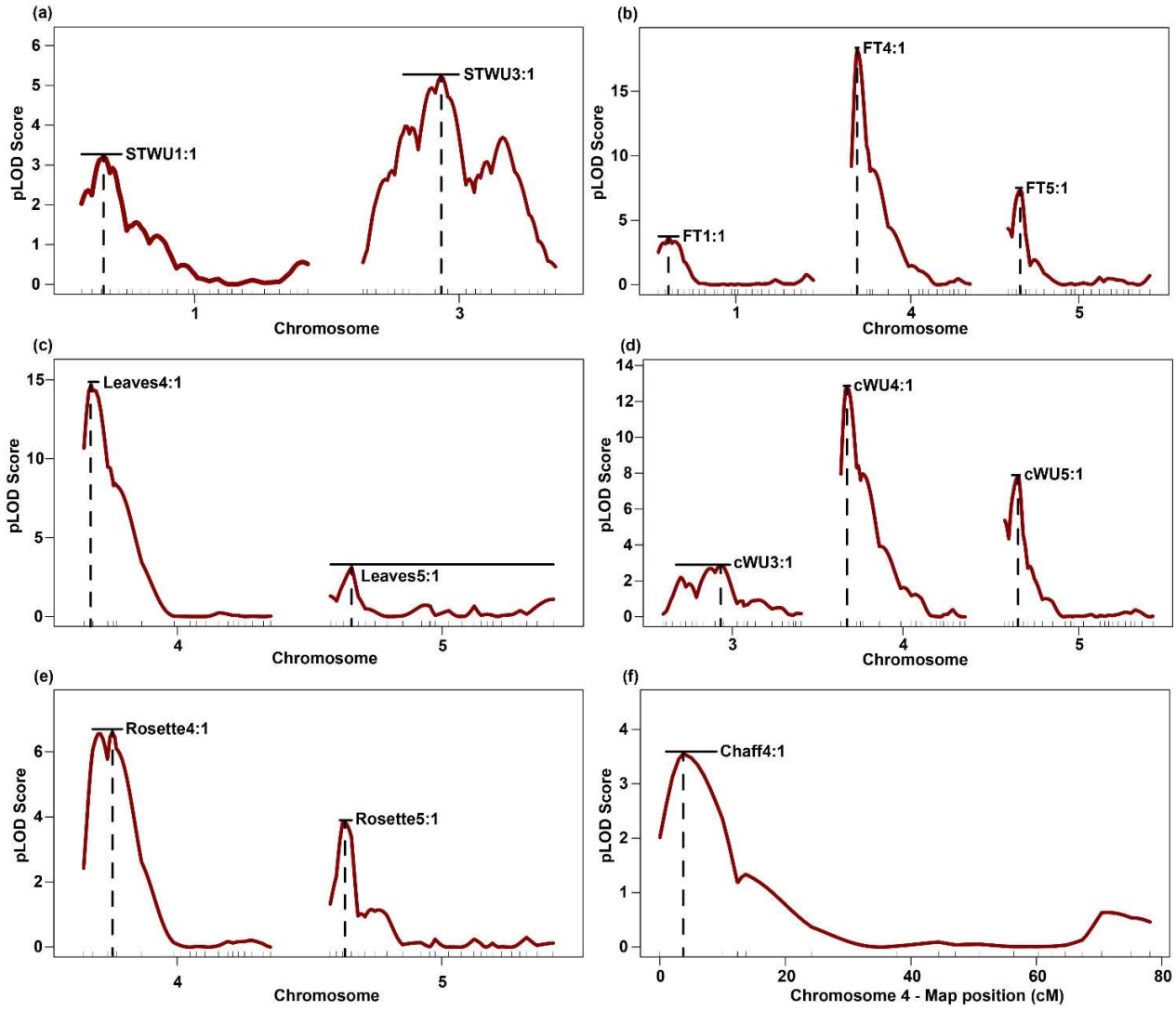

Figure 4.4. Significant QTLs detected through multiple QTL mapping for all traits assessed as part of present study (a) Short term water use (STWU), **(b)** Flowering time (FT), **(c)** Number of rosette leave at bud initiation (Leaves), **(d)** Calculated water use cWU, **(e)** Rosette biomass (Rosette), **(f)** Chaff biomass (Chaff). For all plots, the curved red lines represent the penalized LOD score (Y-axis) at multiple actual and simulated genotypic

locations across the genome (X-axis). For each QTL the 95% Bayesian confidence interval is represented by a horizontal black line and the associated vertical dashed black line points to the position of the QTL.

Table 4.3. Significant QTLs detected through multiple QTL mapping for all traits assessed as part of this study. QTL names are provided as the trait followed by the associated chromosome number. The following are provided for each QTL: position (cM), penalized LOD score, proportion of genetic variance explained by the QTL, confidence interval of the QTL (cM), p-value, and additive genetic effect (with standard error (SE)). The additive genetic effect corresponds to the direction of effect of alleles from the C24 parent, i.e. positive values indicate that the C24 allele of any particular QTL increase the trait value and vice versa.

QTL	Position (cM)	Penalised LOD score	Proportion of total genetic variation	95% Bayesian credible interval (cM)	P-value	Additive genetic effect (SE)
STWU:1	10	3.21	7.561	0-18	< 0.000	0.19 (0.05)
STWU:3	35	5.215	12.643	18-43	< 0.000	-0.25 (0.05)
FT:1	7	3.578	5.346	1-14	< 0.000	-2.86 (0.69)
FT:4	3.69	18.177	34.769	3-5	< 0.000	6.75 (0.64)
FT:5	8	7.334	11.658	5-9	< 0.000	-3.95 (0.64)
Leaves:4	3	14.68	34.409	2-6	< 0.000	7.88 (0.86)
Leaves:5	8.824	3.12	6.028	0-93.22	< 0.000	-3.16 (0.82)
Rosette:4	12	6.6	15.414	3.699-16	< 0.000	0.05 (0.01)
Rosette:5	6	3.84	8.596	4-9	< 0.000	-0.04 (0.01)
Chaff:4	3.699	3.545	10.06	1-9	< 0.000	-0.04 (0.01)
cWU:3	36	2.851	4.799	8-42	< 0.000	-21.99 (6.01)
cWU:4	4	12.785	25.455	2-6	< 0.000	48.79 (5.79)
cWU:5	8.82	7.811	14.283	5.25-10	< 0.000	-35.90 (5.69)

4.3.3 The additive genetic action of non-functional alleles of *FRI* and *FLC* reduces water use without penalizing reproductive performance

The QTLs detected on chromosomes four and five underlie a substantial proportion of the total genetic variation for numerous traits. Additionally, these QTL co-localize with two well characterized flowering time genes, namely *FRI* (Chromosome 4; Schmitz & Amasino 2007) and *FLC* (Chromosome 5; Hepworth & Dean, 2015). Given the detection of these QTL following mapping for flowering time and the strong genetic association detected between these traits in question and flowering time, it is plausible that *FRI* and *FLC* underlie the detected QTLs.

To test the above hypothesis, all phenotyped RILs were sub-divided into four groups based on their predicted allelic state at the *FRI* and *FLC* loci. Comparison of means testing and post-hoc analyses were performed in order to determine the effect of different combinations of different alleles of *FRI* and *FLC* on all of the traits assessed as part of the QTL mapping. (Fig 4.6, Table 4.4).

QTL mapping for short term water use did not detect the QTLs on chromosome 4 and 5, it was therefore expected that *FRI* and *FLC* would not influence variation for this trait. Consequently, there were no significant differences between the different allelic groups with regards to short term water use (Fig 4.6a, Table 4.4). It was observed that this was the only trait that was not affected by allelic variation at these two loci (Fig. 4.5, Table 4.5).

The parental allelic state combinations, i.e. *FRI*-Col:*FLC*-Col and *FRI*-Col:*FLC*-C24, demonstrated no difference in flowering time. This is concurrent with the observation that

the estimated means for flowering are almost identical for Col-0 (74.86 days) and C24 (74.07 days). However, possessing the C24 allele of *FRI* and the Col-0 allele of *FLC* was observed to significantly extend flowering time. Inversely, a combination of the Col-0 allele of *FRI* and the C24 allele of *FLC* significantly reduces flowering time compared against both the parental line combinations and the opposite allelic combinations. This exact trend is repeated for cWU, where *FRI*:Col-*FLC*:C24 has a significantly reduced water use, presumably reflective of its reduced life cycle span, i.e. reduced flowering time.

It is also important to note that the traits pertaining to reproductive performance, i.e. chaff biomass, seed yield, and cWP are significantly reduced for RILs harboring both the C24 allele of *FRI* and the Col-0 allele of *FLC*. However, those same RILs had the highest vegetative performance, i.e. rosette biomass accumulation. These differences support the general notion that flowering time is positively linked to water use and vegetative performance. As such, flowering time could be perceived to be an accurate proxy, or physical marker, of such traits. Importantly, these observed differences also suggest that flowering time is uncoupled from reproductive performance, reinforcing the inaccuracy of utilizing flowering time as a proxy for yield.

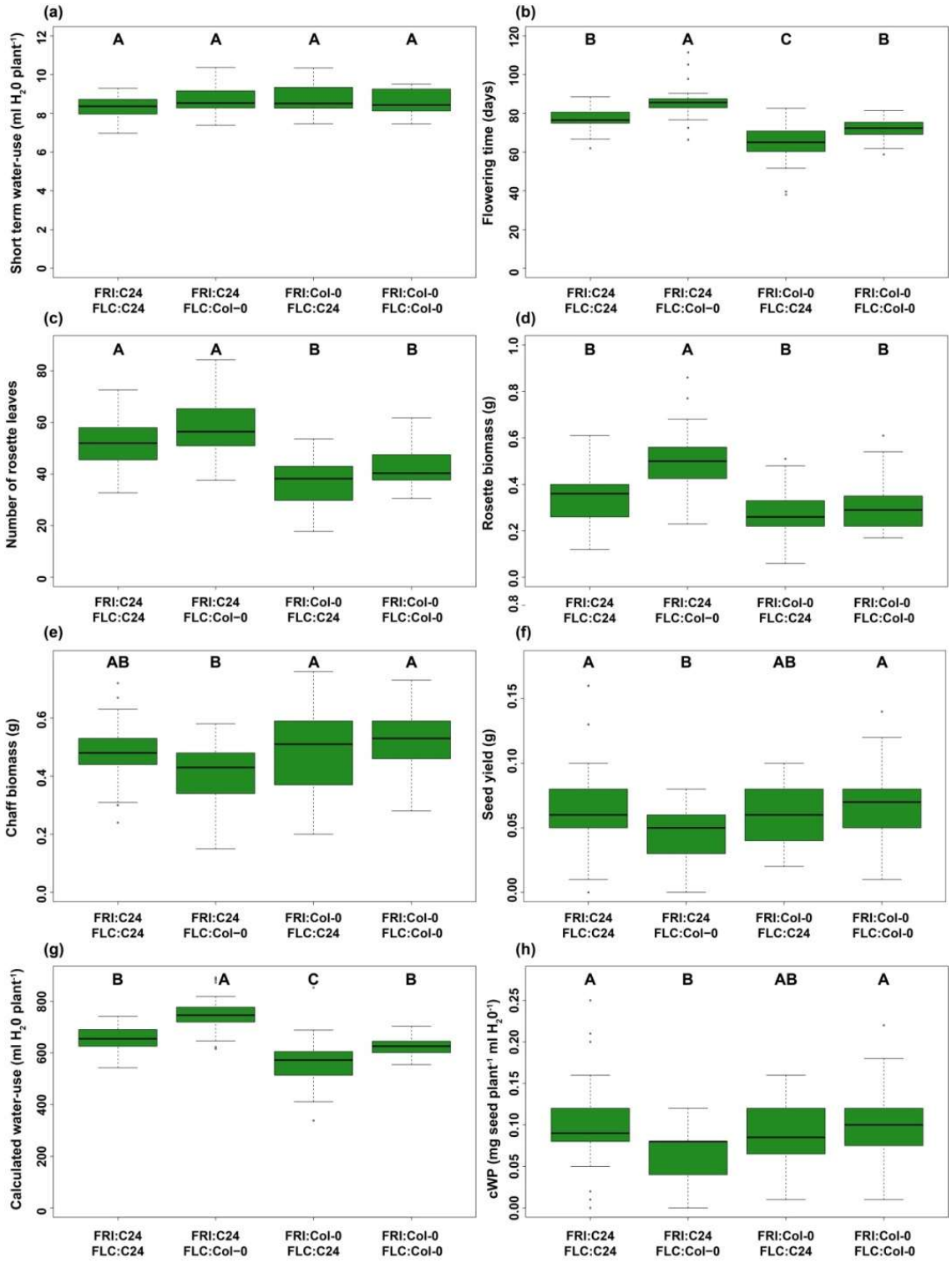

Figure 4.5. Trait performances of genotypes harboring different allelic combinations of the *FRI* (AT4G00650) and *FLC* (AT5G10140) genes. Boxplots describing the variation

for all traits assessed for the groupings of RILs based on allelic forms of both *FRI* and *FLC* that RILs were predicted to harbor based on markers used in the QTL analysis that neighbor these two genes. **(a)** Short term water use, **(b)** flowering time, **(c)** number of rosette leaves at bud initiation, **(d)** rosette biomass, **(e)** chaff biomass, **(f)** seed yield, **(g)** calculated water use, **(h)** calculated water productivity (cWP). The letters (A, B, and C) above the boxplot denote the post-hoc Tukey groups, where allelic groups whose letters are different being significantly different from one another for that particular trait (See Table 4.6. for full results).

Table 4.4. Trait performance differences between genotypes harboring different allelic variants of *FRI* and *FLC*. Results from one-way ANOVA comparison of means testing and post-hoc Tukey tests. All RILs were grouped according to their allelic form at both *FRI* and *FLC*. One-way ANOVAs were performed to assess differences between these allelic groups for performances of all traits. Additionally, Tukey's Honest Significant Difference method of post-hoc testing was performed to determine which groups were significantly different from one another for each trait. For all traits and allelic groups; the p-value against all other groups, minimum trait value, maximum trait value, and post-hoc Tukey group are provided.

Allelic Combination	Trait	FRI:C24 FLC:C24	FRI:C24 FLC:Col-0	FRI:Col-0 FLC:C24	FRI:Col-0 FLC:Col-0	Min	Max	Tukey post-hoc group
FRI:C24-FLC:C24	STWU	-	0.23	0.06	0.59	6.97	9.30	A
FRI:C24-FLC:Col-0	STWU	0.23	-	0.93	0.93	7.38	10.37	A
FRI:Col-0-FLC:C24	STWU	0.06	0.93	-	0.66	7.46	10.34	A
FRI:Col-0-FLC:Col-0	STWU	0.59	0.93	0.66	-	7.45	9.50	A
FRI:C24-FLC:C24	FT	-	**0.00**	**0.00**	0.05	61.95	88.47	A
FRI:C24-FLC:Col-0	FT	**0.00**	-	**0.00**	**0.00**	66.35	111.45	B
FRI:Col-0-FLC:C24	FT	**0.00**	**0.00**	-	**0.00**	38.04	82.59	B
FRI:Col-0-FLC:Col-0	FT	0.05	**0.00**	**0.00**	-	58.77	81.35	C
FRI:C24-FLC:C24	Leaves	-	0.12	**0.00**	**0.00**	32.74	72.60	A
FRI:C24-FLC:Col-0	Leaves	0.12	-	**0.00**	**0.00**	37.57	84.18	A
FRI:Col-0-FLC:C24	Leaves	**0.00**	**0.00**	-	0.19	17.82	53.60	B
FRI:Col-0-FLC:Col-0	Leaves	**0.00**	**0.00**	0.19	-	30.52	61.73	B
FRI:C24-FLC:C24	Rosette	-	**0.00**	0.12	0.73	32.74	72.60	A
FRI:C24-FLC:Col-0	Rosette	**0.00**	-	**0.00**	**0.00**	37.57	84.18	A
FRI:Col-0-FLC:C24	Rosette	0.12	**0.00**	-	0.70	17.82	53.60	B
FRI:Col-0-FLC:Col-0	Rosette	0.73	**0.00**	0.70	-	30.52	61.73	B
FRI:C24-FLC:C24	Chaff	-	0.07	0.92	0.41	0.24	0.72	A
FRI:C24-FLC:Col-0	Chaff	0.07	-	**0.02**	**0.00**	0.15	0.58	A
FRI:Col-0-FLC:C24	Chaff	0.92	**0.02**	-	0.79	0.20	0.76	AB
FRI:Col-0-FLC:Col-0	Chaff	0.41	**0.00**	0.79	-	0.28	0.73	B
FRI:C24-FLC:C24	Yield	-	**0.01**	0.66	0.99	0.00	0.16	A
FRI:C24-FLC:Col-0	Yield	**0.01**	-	0.14	**0.01**	0.00	0.08	A
FRI:Col-0-FLC:C24	Yield	0.66	0.14	-	0.69	0.02	0.10	AB
FRI:Col-0-FLC:Col-0	Yield	0.99	**0.01**	0.69	-	0.01	0.14	B
FRI:C24-FLC:C24	cWU	-	**0.00**	**0.00**	0.34	542.85	742.18	A
FRI:C24-FLC:Col-0	cWU	**0.00**	-	**0.00**	**0.00**	615.91	892.05	B
FRI:Col-0-FLC:C24	cWU	**0.00**	**0.00**	-	**0.02**	338.35	852.93	B
FRI:Col-0-FLC:Col-0	cWU	0.34	**0.00**	**0.02**	-	555.14	703.11	C
FRI:C24-FLC:C24	cWP	-	**0.02**	0.89	0.10	0.00	0.25	A
FRI:C24-FLC:Col-0	cWP	**0.02**	-	0.10	**0.02**	0.00	0.12	A
FRI:Col-0-FLC:C24	cWP	0.89	0.10	-	0.92	0.01	0.16	AB
FRI:Col-0-FLC:Col-0	cWP	0.10	**0.02**	0.92	-	0.01	0.22	B

STWU - Short term water-use, FT - Flowering time, Leaves - Number of rosette leaves at bud initiation, Rosette - Rosette biomass, Chaff - Chaff biomass, Yield - Seed yield, cWU - Calculated water-use, cWP - Calculated water productivity

4.4 Discussion

4.4.1 *FRI* and *FLC* underlie variation for flowering time and water use in the Col-0 x C24 RIL population

It is only until relatively recently that appropriate statistical genetic methods that facilitate the elucidation of genes underlying quantitative traits have been developed. Consequently, we are still only just beginning to understand the genetic basis of traits of interest and how these genes impact upon plant physiology. The work described here links extensive assessments of whole plant physiology, specifically traits relating to water use and productivity, with quantitative genetics in an effort to dissect out the genetic basis of these traits. A large set of Arabidopsis RILs derived from parental ecotype lines, Col-0 and C24, which demonstrated substantial differences in traits relating to water use and productivity were utilized for this purpose.

For all traits assessed as part of this investigation, a huge degree of genetic and phenotypic variation was detected within the RIL population (Fig. 4.1, Table 4.2). Crucially, this variation demonstrated a significantly heritable basis, which was assessed as both H^2 and C_{VG}. The H^2 of the fitness related traits were somewhat reduced compared to those of the other traits assessed. Concurrently, the fitness related traits demonstrated much higher C_{VG} estimates. This pattern of estimates of genetic variation for the fitness and non-fitness related traits is parallel to what was observed with the scan of natural variation in Chapter Two. Similarly, it is supportive of the work of (Houle, 1992), which describes the unsuitability of H^2 as an assessment of the genetic, or heritable, basis of fitness related traits, due to the high levels of residual variance these traits are typically subject too (Houle, 1992; Pigliucci,

2008). So, whilst the low-to-moderate H^2 of seed yield and cWP suggests they may not be as genetically determined as flowering time, for example, its high C_{VG} offers encouragement that it is highly selectable, therefore its genetic basis should be discernible through appropriate means.

Essentially just five QTL were mapped as part of this investigation. One of these QTL, STWU:1, was unique to short term water use. A further QTL, FT:1, was observed to control a subset of the variation observed for flowering time only. A further QTL, STWU:3, detected for short term water use, which explained slightly more variation than STWU:1, was also detected for cWU, however this is not surprising since cWU is calculated based on short term water use. However, it should be noted that mapping for cWU did not detect STWU:1, presumably because its effect on water use is somewhat reduced compared to STWU:3. In general these three QTL defined far less variation than those identified on chromosomes 4 and 5, additionally their intervals were far larger too. These QTLs do not overlap with any well-characterized genes, or previously mapped QTLs, that are known to influence flowering time, water use, drought resistance, or WUE. Future work to identify the causal genes responsible for these QTLs will require the saturation of these regions with further markers and subsequent fine mapping.

The highly overlapping QTLs detected on chromosomes four and five underlie the vast majority of the variation for flowering and water use. These two QTL colocalise with the well-characterized flowering time genes *FRI* (chromosome 4) and *FLC* (chromosome 5). It is worth noting the studies of McKay *et al* (2008) and Strange *et al* (2011) that both identified single QTLs on chromosomes 4 and 5, for WUE and vernalisation requirements respectively. The single QTLs from these studies overlapped with each other and with the corresponding

QTLs from this present study. *FRI* and *FLC* were defined as responsible for the QTLs in McKay *et al* (2008) and Strange *et al* (2011). Alongside the evidence based on colocalisation, the genetic relatedness of the traits affected by these QTLs and their corresponding allelic effects provides compelling evidence to suggest that *FRI* and *FLC* are responsible for these QTL.

Allelic variation at *FRI* has been identified as key to the well-described winter annual adaptation strategy of ecotypes persisting in particularly high latitudinal areas (Clarke & Dean, 1994; Gazzani *et al.*, 2003). In brief, ecotypes possessing functional *FRI* alleles require a period of overwintering, termed vernalisation, during which time *FRI* increases the transcript level of the *FLC* floral repressor gene. Depending on the remaining genomic background, ecotypes with functional *FRI* alleles that do not undergo a period of vernalisation will either flower late or not flower at all. *FLC* is described as a floral repressor because it functions to repress the expression of multiple genes required for the transition to flowering (Michaels & Amasino, 1999).

Despite its Portuguese origin, C24 possess a functional allele of *FRI* which would otherwise require a period of vernalisation to transition to flowering, else flowering would be very late. However, the *FLC* allele of C24 is non-functional, therefore the vernalisation requirement of C24-*FRI* is redundant (Michaels *et al.*, 2003). The *FRI* and *FLC* alleles of the Col-0 ecotype are opposite in their functionality, that is to say that Col-0 harbors non-functional and functional alleles of *FRI* and *FLC* respectively (Johanson *et al.*, 2000). Therefore, the functional Col-*FLC* allele prevents noticeably early flowering, whereas the non-functional Col-*FRI* allele forfends a vernalisation requirement. The opposing allelic forms of *FRI* and *FLC* in Col-0 and C24 explain their similar flowering times.

4.4.2 A combination of non-functional *FRI* and *FLC* alleles reduces water use without penalizing reproductive performance

Grouping the RILs comprising this present study into the various predicted allelic combinations of *FRI* and *FLC* offered the opportunity to assess the impact the functional state of these genes has on flowering time and water use. Crucially it also facilitated the assessment of how this variation impacts upon productivity as a unique trait, but also as a factor of water use. Those RILs that harbored the parental allelic combinations of *FRI* and *FLC* were not significantly different from one another in terms of flowering time or water use. This was to be as expected considering the detected importance of these genes and the lack of difference in flowering time between the parental lines. Similarly, cWU was not significantly different between these lines, which is of note considering these groups do not account for the variation at the gene, or genes, underling STWU:1/CWU:1. This, along with the lack of detection of STWU:3 following mapping for cWU, suggests that variation in long term water use is primarily a function of flowering time variation, rather than short term water use variation from which it is also calculated. This is presumably the case because variation for flowering time is far greater than variation for short term water use.

Combining functional alleles of *FRI* and *FLC*, i.e. C24-*FRI*:*FLC*-Col, resulted in an increase in flowering time. This is concurrent with what would be predicted based on the known functional effect of these genes. Plants harboring homozygous functional *FRI* alleles will flower late if they do not undergo a period of vernalisation, additionally homozygous functional *FLC* alleles act to repress flowering further. Conversely, a combination of non-functional alleles of both *FRI* and *FLC* reduced flowering time significantly. Again, this is as to be expected since the non-functional allele of *FRI* does not require a period of

vernalisation to cease positive regulation of *FLC*, and the non-functional *FLC* allele does not confer late flowering. The effects of the four different allelic combinations where mirrored for cWU, which of course is not wholly surprising since cWU is calculated based on flowering time. However, the variation in cWU that must be due to short term water use, from which cWU is also calculated, does not appear to disturb this trend. It is also worth noting that cWU was demonstrated in Chapters Two and Three to be a highly accurate proxy of measured long term water use (Fig. 2.8), so reducing flowering time is a highly efficient means of reducing water use.

Alterations in flowering time and water use must be considered alongside their impact upon biomass accumulation, since reductions in either or both would be undesirable if they resulted in reductions in harvestable yield. There were no significant differences between the parental combination of *FRI* and *FLC* for biomass accumulation (Fig. 4.5). However, RILs predicted to harbor functional alleles of both *FRI* and *FLC* demonstrated significantly improved vegetative biomass accumulation, thus reflecting the well-described positive link between flowering time and vegetative performance, i.e. rosette biomass. However, the opposite is observed for reproductive performance, where the allelic combination of functional alleles actually results in significantly reduced seed yield (Fig. 4.5). Again, this reflects what has previously been observed as part of this study, and the studies described in Chapters Two and Three, where plants that flower later do not translocate the supplemental photosynthates acquired during extended photosynthetically active vegetative periods to reproductive sinks (Fig. 2.4). Therefore, flowering time is a good proxy for water use and total biomass accumulation, but not reproductive performance. To this end, the results demonstrated here, suggest that combining allelic forms of floral repressor genes

that increase flowering time, will increase water use and biomass accumulation concurrently, but not reproductive performance.

Over the past 15 years, a significant body of work has been amassed which describes the natural variation of WUE, assessed almost exclusively via $\delta^{13}C$ as a proxy, in Arabidopsis (McKay *et al.*, 2003, 2008; Juenger & Mckay, 2005; Hausmann *et al.*, 2005; Masle *et al.*, 2005) and closely related Brassica species (Fletcher *et al.*, 2014). A recurring theme from this work is the detection of a positive genetic link between flowering time and WUE. Indeed, a major element of this work was the demonstration of *FRI* and *FLC* as key determiners of natural variation for WUE (McKay *et al.*, 2008). Based on the results described in this present study, and from a purely rationalistic standpoint, I would contend that the recurring link observed between flowering time and WUE in these studies is possibly flawed. A plant that has an extended flowering time will use more water during its lifetime. Clearly, this seems a very elementary statement to make, but this has not been experimentally demonstrated until now. I would argue for this this reason alone it seems unlikely that increases in flowering time, i.e. longer life cycles, will be met with increases in WUE, for the simple reason that plants that flower later use more water, as such they will have reduced levels of WUE. I suggest that the only scenario whereby it could be envisaged that flowering time and WUE could be positively associated, would be a situation where extended flowering times are met with increases in biomass accumulation. To this end, I have consistently demonstrated, as part of the study described in this Chapter and those described in Chapters Two and Three, that flowering time is positively associated with rosette biomass accumulation (Figs. 2.4, 3.14). Furthermore, it is predicted that combining functional alleles of *FRI* and *FLC* would increase both flowering time and rosette biomass. However, in order for WUE to be positively

associated with flowering time, simultaneous increases in biomass accumulation would have to *compensate* for the coinciding increase in water use. Further physiological analyses are required in order to determine whether this *compensation* in vegetative biomass occurs with extended flowering times to an extent that leads to improved actual WUE, not $\delta^{13}C$, compared to those plants that flower earlier and use less water.

If a link between WUE and flowering time did exist in the above described manner, it would be feasible and appropriate to employ flowering time as a proxy of WUE. However, WUE in this sense holds no agronomic relevance, since the improvement in biomass accumulation is with regards to vegetative biomass only. Accordingly, it should be noted that the combination of functional alleles of *FRI* and *FLC* had significantly reduced seed yields compared to the combinations of functional and non-functional alleles in both iterations. This is worth noting since the vast majority of commercially grown crops are harvested for their reproductive product (FAO, 2016), including most crops of the Brassica genus that are directly related to Arabidopsis (Zhu *et al.*, 2016), as such it is important for studies of this nature to consider the impact of reproductive performance.

Unfortunately, the genetic mapping described here was unable to detect genomic regions responsible for the observed variation in either seed yield or cWP. This is likely due to the observed transgressive segregation, where the parental lines were situated toward the positive extreme of the total variation, as opposed to closer to the center like the other traits assessed. This problem could potentially be addressed by incorporating further RILs into the analysis, however I would argue an association mapping approach, whereby more natural variation for reproductive performance is captured, assessed and mapped for, would be more successful in term of *de-novo* gene discovery. Despite the lack of QTL detected for

seed yield and cWP, a single QTL on chromosome 4, which is highly likely to be *FRI*, was detected following mapping for chaff biomass. Chaff biomass shares a highly significant positive genetic correlation with both seed yield and cWP, as such could be considered a proxy, or physical marker, for both. With regards to *FRI* and chaff biomass, the allelic effect is opposite to flowering time and water use, where functional alleles, which increase water use and flowering time, decrease chaff biomass. This provide further motivation to explore the use of a combination of non-functional *FRI* and *FLC* alleles as a means of reducing water use, since it will at best likely improve reproductive performance and at worst not reduce it compared to parental lines, whose water use is greater.

This investigation has identified *FRI* and *FLC* as indirect regulators of long term water use, due to their negative regulation of flowering time. A homozygous combination of non-functional alleles of both *FRI* and *FLC* is predicted to significantly reduce water use compared to wild type Col-0 and C24 ecotypes, by reducing the time to floral transitioning. Crucially, a reduction in flowering time and water use mediated in this sense is not projected to detrimentally influence reproductive or vegetative performance, in terms of biomass accumulation. As such, these findings have important agronomic inferences, because they indicate the feasibility of combining a reduction in the time to harvest with a decrease in irrigational outlays, whilst not compromising yield.

Future work, which has now been initiated, should seek to further understand the effects of functional and non-functional alleles of *FRI* and *FLC* in Col-0 and C24 genomic backgrounds. This could be achieved through the use of already-developed Near Isogenic Lines (NILs; Törjék *et al.*, 2006) or targeted genome editing. Such work would build upon the *prediction*s described here in terms of solidifying the response of flowering, water use, and

biomass accumulation to combining different alleles of these genes together. Additionally, confirmatory investigations should be initiated to ascertain that improvements to water use are being achieved through reduced flowering times, as opposed to some alternative action of *FRI* and *FLC*. It is proposed that this should be determined through assessing the long-term water use and biomass accumulation of mutant lines of Arabidopsis whose internal circadian rhythms are maintained.

5. Genome wide association mapping for key traits relating to water use and productivity in Arabidopsis

5.1 Introduction

Understanding the genetic variation that underpins phenotypic variation for complex traits has been a major goal of plant scientists since Gregor Mendel's hypothesis regarding the existence of *internal factors*, i.e. genes, that are generationally inherited (Mendel, 1865). Understanding this relationship is central to the development of elite crop varieties, through breeding or biotechnological approaches. Furthermore, it is also important for studies centered on understanding the molecular basis of ecological adaptation, which can in turn have implications for crop improvement.

As discussed and demonstrated in the previous chapter, forward genetics as achieved through QTL mapping has been instrumental for our understanding of the genetic control of complex traits in Arabidopsis (Reviewed in: Weigel (2012)), and in many other plant species (Reviewed in: Holland (2007)). Moreover, QTL analyses have been successfully employed in animal studies too (Reviewed in: Solberg Woods (2014)). Despite the demonstrated and continued success of QTL mapping for gene discovery, it is a technique that is constrained in two fundamental ways, as described below.

The first fundamental constraint of QTL mapping is that it is only possible to capture the allelic variation that segregates between the original parental lines, which may be

suitable for a few genetically related traits, but it is not guaranteed to be useful for many traits beyond this (Korte & Farlow, 2013). It is important to note that successful efforts have been made to address this by capturing increased allelic variation from more parental lines, before establishing RILs through successive rounds of selfing. This has been achieved in Arabidopsis through the MAGIC (Kover *et al.*, 2009) and AMPRIL (Huang *et al.*, 2011) populations. However, the generation of such populations requires substantial investments in time and effort. Furthermore, the populations are still limited to a few parental lines, in terms of the variation they capture.

The second constraint to QTL mapping, which is particularly pertinent when dealing with RIL populations, is the decreasing rates of recombination that occur with each round of selfing from the F_2 generation. This results in extended blocks of linkage which reduce the accuracy of linkage mapping (Korte & Farlow, 2013). Essentially, this results in QTLs with large confidence intervals which cover vast chromosomal distances, making it difficult to discern the precise gene(s) that define a QTL. What is more, these large linkage blocks give rise to the problem of linkage drag in plant breeding, where undesirable genes linked to genes of interests are transferred into elite germplasm following marker assisted breeding. For this reason, molecular markers that are based solely on QTLs can often be impractical for integration into breeding programs (Peng *et al.*, 2014).

GWAS is essentially achieved via combining thousands of one-way ANOVA comparison of means statistical tests (Hayes, 2013). The constraints that limit QTL mapping are to some extent addressed by GWAS. When working with multiple accessions of a species, as opposed to a specifically designed mapping population, the ancestral polymorphisms that transcend the genome are more informative. They allow the user to

capture far more variation, which means the same panel of multiple populations can be used to identify causal genetic variants of many more traits than is possible using a mapping population and associated QTL analysis. The obstacle here, however, is that many more lines need to be phenotyped in order to capture enough variation to identify candidate polymorphisms (Mauricio, 2001).

Arabidopsis represents the ideal model organism for GWAS. This is primarily since ecotypes can essentially be maintained as if they were RILs, due to the plant's self-pollinating nature. Additionally, the genotypic resources that are available for Arabidopsis are unrivalled in plant research. Atwell *et al* (2010) made the genotypic information (250K SNPs) for over 1000 Arabidopsis ecotypes from across the Northern Hemisphere publicly available. This has facilitated the opportunity for researchers to begin performing large scale and high resolution GWAS. Furthermore, the availability of multiple and free software platforms to carry out such analysis has made this process even more accessible (Bradbury *et al.*, 2007; Seren *et al.*, 2012). These data and resources have been successfully employed to identify polymorphisms and genes related to drought resistance (Verslues *et al.*, 2014), improved photosynthetic efficiency (van Rooijen *et al.*, 2015), germination success (Morrison & Linder, 2014), and thermotolerance (Bac-Molenaar *et al.*, 2015) in Arabidopsis.

Chapters Two and Three of this thesis highlighted the importance of water use, productivity, and water productivity as alternatives for assessing plant performance under either drought conditions or minimal water inputs. The QTL analysis described in chapter four, which utilized the RIL mapping population developed from crossing the ecotypes Col-0 and C24 (Törjék *et al.*, 2006), was ideal for discerning the genetic basis of water use. However, I had limited success identifying genetic loci contributing to variation for

productivity or cWP. This is likely due to the limitations of linkage disequilibrium mapping as previously described. Substantial variation exists between Col-0 and C24 for water use as determined by flowering time or as measured over the course of plants' life cycles (Bechtold *et al.*, 2010), consequently we were able to delineate the genetic basis of this variation through QTL mapping. The variation between these parental lines for productivity and water productivity is much smaller, as such the mapping population did not capture enough allelic variation for these traits.

To overcome the limitations of the QTL mapping and to dissect the genetic basis of productivity following an SD period and cWP, I employed a GWAS utilizing 117 ecotypes of Arabidopsis. Short term water use, cWU, flowering time, cWP, and key biomass accumulation parameters were assessed for all ecotypes. This allowed for further analysis of the relationships and heritable basis of these key traits. The MLMM approach as described by Segura *et al* (2012) was utilized for GWAS. This approach was preferred because of the increased statistical power it confers compared to alternative approaches and because of its suitability for use in Arabidopsis-based studies, as described in the materials and methods of this chapter.

5.2 Materials and methods

5.2.1 Plant material, growth conditions, and trait parameters assessed

Seed for all ecotypes comprising this study were obtained from NASC (Scholl et al 2000). 117 ecotypes that were included as part of the Atwell *et al.*, (2010) RegMap project were randomly selected and represented a wide distribution across the Northern

Hemisphere (Table 5.1.). The growth conditions of plants and SD experiment for the GWAS are described in Chapter 2 (2.2.3). Briefly, this involved the assessment of short term water use as the soil drying rate from 100% to 20% *rSWC*. Following the SD period, biomass accumulation and flowering time were assessed (n = 15). These trait parameters were used to calculate long term water use (cWU) and water productivity (cWP).

5.2.3. Statistical analyses

The phenotypic assessment of all the ecotypes had to be temporally blocked due to space constraints within the controlled growth environments. Highly significant block effects were detected between these temporal blocks based on one-way ANOVA comparison of means testing (Table 5.2). Therefore, BLUPs were extracted using appropriate GLMMs (See: 2.2.5). Estimated means were subsequently calculated by adding BLUPs to population means. Estimated means were employed to assess genetic correlations between all the traits assessed via Pearson's regression analysis and for the GWAS.

Phenotypic (V_P) and genotypic variance (V_G) was calculated from the above GLMMs (described in Chapter 2). These values were used to estimate the H^2 and C_{VG} of all traits (See: 2.2.5). Briefly, H^2 pertains to the degree of variation for a particular trait that is owing to genetic variation, whereas C_{VG} provides an estimate of the propensity of a trait to respond to selection.

5.2.4 Genome-wide association mapping

Genotypic information for all 117 ecotypes comprising this study was obtained as part of the Regional Mapping (RegMap) project (Atwell *et al.*, 2010). The RegMap project involved the genotyping of 1,307 widely distributed ecotypes of Arabidopsis using an

Affymetrix 250,000 SNP genotyping chip (Data accessed from: http://www.bergelson.uchicago.edu). The genotypic information relevant to the 117 ecotypes comprising this present study was extracted from the RegMap dataset. Missing genotype information was imputed within the TASSEL software package (Trait Analysis by aSSociation, Evolution, and Linkage; Bradbury *et al* (2007)), using the Linkage Disequibrlium K-Nearest Neighbor imputation (LD-kNNi) method. LD-kNNi imputes missing genotype information based on the *k*-nearest neighbor, i.e. most genotypically similar ecotype, but it also takes into account the extent of linkage disequibrlium of SNPs to be imputed when choosing which neighbor to base imputation on (Money *et al.*, 2015).

I employed the MLMM (Segura *et al.,* (2012) approach for GWAS. This approach is based on the popular EMMA statistical test for association mapping (Kang *et al.*, 2008). One of the fundamental issues of GWAS is the risk of false positive results arising due to population structure, i.e. relatedness between ecotypes in this case. EMMA addresses this issue by correcting for population structure by incorporating the pairwise relatedness of all ecotypes into the mixed models. EMMA has been demonstrated to eliminate the problem of false positives arising from GWAS for studies based on mice, Arabidopsis, and Maize (Kang *et al.*, 2008). MLMM builds upon EMMA by including SNPs as cofactors in the analysis, much like composite interval mapping for QTL analysis. Using flowering time and ion accumulation studies as tests, MLMM has been demonstrated to reduce the risk of detecting false positives, as well as increasing statistical power, thereby increasing the feasibility of detecting important causal loci (Kang *et al.*, 2008). MLMM has recently been successfully employed to identify key genes underlying variation for metabolites important for fruit flavor in tomato (Sauvage *et al.*, 2014).

MLMM is a step-wise approach to GWAS, where the most significant SNP is added to the mixed model at each successive step. This stepwise process continues until no additional genetic variance can be explained by adding further SNPs to the model. At this point, the stepwise backward elimination of markers is initiated and proceeds until the original model, which accounts only for population structure, is reached (Segura *et al.*, 2012). When using MLMM for GWAS, two criteria are available for the determination of the most significant model: the extended Bayesian information criterion (e-BIC; Chen & Chen (2008)) and the multiple-Bonferroni criterion (mBonf; Bonferroni (1936)). The Bonferroni criterion is typically considered to be more stringent, as such it was selected for determination of the most significant step-wise model. According to Segura *et al* (2012), the optimal MLMM GWAS model according to the Bonferroni criterion is the step where all SNPs that are incorporated as cofactors have $-\log_{10}$(p-values) above the Bonferroni significance threshold. If a situation arises where not all of the SNP co-factors from any of the stepwise models fall above the significance threshold, the original model (with no cofactors), which accounts for population structure only, is considered optimal. In this study, in situations where GWAS for particular traits resulted in no SNPs being considered significant according to the stringent Bonferroni threshold, a $-\log_{10}$(p-value) cut-off of 5 was designated as a *significance* threshold, as is typical with studies of this nature (van Rooijen *et al.*, 2015).

5.2.5 Identification of candidate genes and assessment of their response to abiotic stress using Genevestigator

Linkage disequibrlium has been demonstrated to completely decay after 10kb in Arabidopsis (Kim *et al.*, 2007). As such, I identified all the protein coding genes that where within a 20kb range (10kb upstream/downstream) of all SNPs that had $-\log_{10}$(p-values)

above the significance thresholds. This was achieved using the 'Annotations Table' section of the R-shiny application Zbrowse (Ziegler *et al.*, 2015). Adopting a 10kb threshold for candidate gene identification based on SNPs of interest has been successful in previous Arabidopsis GWA-mapping studies (Verslues *et al.*, 2014; van Rooijen *et al.*, 2015).

As a precursor to the future investigation of the characterization and function of candidate genes arising from the GWAS described here, I assessed their response to abiotic stress using the Genevestigator online platform for exploring transcriptomic data (Hruz *et al.*, 2008). For these analyses, the Affymetrix Arabidopsis ATH1 Genome Array dataset was used, with samples matching the 'Stress' condition for perturbations. A highly stringent log-2-fold change in gene expression of 3 and a significance threshold of p-value < 0.001 was set to identify significant responses of the candidate genes to abiotic stresses. Multiple significant responses were detected, with the most interesting responses reported and discussed as part of this Chapter.

5.3 Results

5.3.1 The phenotypic basis of genetic variation for key traits relating to water use and water productivity

Phenotypic variation for cWU and cWP, amongst other key water use and productivity related traits, was assessed for a randomly selected set of 117 Arabidopsis ecotypes that are part of the Horton *et al.*, (2012) RegMap project and that originate from a wide variety of environmental conditions across the Northern Hemisphere (Table 5.1.).

Table 5.1. Plant material comprising the GWAS. For each line the native name (ecotype), geographic location or origin, and the experimental block are given.

Ecotype	Location of origin	Experimental block(s)
Ag-0	Argentat, France	11
An-1	Anderlecht, Belgium	4
Ang-0	Angleur, Belgium	11
Ayu-Dag-2	Crimea, Ukraine	13
Bay-0	Bayreuth, Germany	4
Belmonte-4-94	Belmonte, Italy	13
Bil-7	Billaberget, Sweden	13
Bla-11	Blanes, Spain	9
Bor-1	Borky, Czech Republic	8
Bor-4	Borky, Czech Republic	8
Br-0	Bruno, Czech Republic	12
Bsch-2	Frankfurt, Germany	13
Buckhorn Pass	Hayden Ridge, USA	5
Bur-0	Burren, Ireland	8
C24	Coimbra, Portugal	All blocks
Cala-8	Villamayor de Calatrava, Spain	2
CIBC-17	Ascot, UK	12
CIBC-5	Ascot, UK	2
Col-0	Columbia, USA	All blocks
Ct-1	Catania, Italy	8
Cvi-0	Cape Verde Islands	2
Dem-4	Dem, USA	4
Dog-4	Dogruyol, Turkey	5
Don-0	Donana, Spain	13
DraIV 2-6	Drahonin, Czech Republic	13
DraIV 4-2	Drahonin, Czech Republic	13
Ei-2	Eifel, Germany	1
En-2	Frankfurt, Germany	6
Est-1	Vagli, Estonia	3
Fab-2	Faberget, Sweden	13
Fei-0	St. Mariad. Feiria, Portugal	7
Ga-0	Gabelstein, Germany	12
Gel-1	Geleen, Netherlands	13
Got-22	Gottingen, Germany	3
Got-7	Gottingen, Germany	4
Gu-0	Guckingen, Germany	5
Gy-0	La Miniere, France	12
Hey-1	Heythuysen, Netherlands	13
Hod	Hodja, Tadjikistand	13
Hr-10	Asoct, UK	11
HR-5	Ascot, UK	2
HSm	Horni Smrcne, Czech Republic	13
Kas-1	Kashmir, India	8
Kas-2	Kashmir, India	13
Knox-10	Knox, USA	4
Knox-18	Knox, USA	9
Knox-19	Knox, USA	9
Kondara	Kondara, Tajikistan	1
Kz-1	Karagandy, Kazakhstan	9
Kz-9	Karagandy, Kazakhstan	3
Lag2-2	Lagodechi, Georgia	5
Ler-0	Landsberg am Lech, Germany	6
LL-0	Llagostera, Spain	5
Lp2-2	Lipovec, Czech Republic	9
Lz-0	Lezoux, France	1
MOG-37	Mog, France	13
Mr-0	Mote, Italy	12
Mrk-0	Markt-Baden, Germany	5

Table 5.1. (Continued from the previous page)

Ecotype	Location of origin	Experimental block(s)
Ms-0	Ms, Russia	12
Mt-0	Martuba/Cyrenalka, Libya	6
Mz-0	Bad Camberg, Germany	10
N13	Koncehzero, Russia	13
Navajo-5	Navajo, Spain	12
Nd-1	Niederzenz, Germany	10
Nemrut-1	Nemrut Dag, Turkey	11
NFA-10	Ascot, UK	4
NFA-8	Ascot, UK	7
Nok-3	Noortlwik, Netherlands	11
Omo-23	Unknown	9
Oy-0	Oystese, Norway	11
PAR-3	Par, France	13
PHW-2	Tatsfield, UK	13
Pna-17	Benton Harbor, USA	11
Pro-0	Proaza- Spain	12
Pu2-23	Prudka, Czech Republic	10
Pu2-7	Prudka, Czech Republic	3
Ra-0	Randan, France	4
Ren-1	Rennes, France	10
Ren-11	Rennes, France	1
Rmx-A02	Michigan, USA	9
Rmx-A180	Michigan, USA	9
RRS-10	North Liberty, USA	9
RRS-7	North Liberty, USA	4
Sakhadara	Shakdara, Tajikistan	4
San Martin-1	San Martin, Spain	13
Sapporo-0	Sapporo, Japan	13
Se-0	San Eleno, Spain	1
Sf-2	San Feliu, Spain	13
Shakdara	Shakdara, Tadjikistan	10
Sn(5)-1	Unknown	13
Sorbo	Sorbo, Tadjikistan	12
Spro 2	Gotland, Sweden	9
Sq-1	Ascot, UK	5
Sq-8	Ascot, UK	11
Tamm-2	Dragsvik, Finaldn	9
Tamm-27	Dragsvik, Finaldn	9
Tascha-1	Tsagguns, Austria	13
Truk-5	Trukhaniv Island, Ukraine	13
Ts-1	Tossa del Mar, Spain	12
Ts-5	Tossa del Mar, Spain	3
Tsu-1	Tsushima, Japan	8
Ull-25	Onnestad, Sweden	12
Uod-1	Linz, Austria	12
Uod-7	Linz, Austria	5
Van-0	Vancouver, Canada	6
Var2-1	Varhallarna, Sweden	12
Wa-1	Warsaw, Poland	11
Wei-0	Weiningen, Switzerland	2
Ws-0	Wassilwsklja, Russia	3
Ws-2	Wassilwsklja, Russia	1
Wt-5	Wietze, Germany	11
Xan-1	Xanbulan, Azerbaijan	1
Yo-0	Yosemite, USA	4
Zdr-1	Zdarec, Czech Republic	9
Zdr-6	Zdarec, Czech Republic	10

An SD experiment was performed in order to determine the short term water use of all of the ecotypes comprising the GWAS. Following the SD period, all plants were kept well-watered and flowering time was recorded. Additionally, biomass accumulation was assessed and long term water use (cWU) and water productivity (cWP) estimates were calculated. As previously described in chapter one, both cWU and cWP are accurate proxies of measured water use and measured water productivity respectively. I detected significant block effects between the temporally divided experimental blocks (Table 5.2). Consequently, BLUPs were extracted as variance components and combined with the appropriate population means to give estimated means, which were used for the GWAS and all other analyses.

Table 5.2. Comparisons of means testing for all traits assessed as part of the GWAS. Testing was performed through one-way ANOVA for all traits. Comparisons were made based on all ecotypes, just Col-0, and just C24. Significance is indicated by *** $p < 0.0001$, and ** $p < 0.01$.

Trait parameter	P-value		
	Between experimental blocks (All ecotypes)	Between experimental blocks (Col-0)	Between experimental blocks (C24)
Short term water-use (ml^{-1} H_20 $plant^{-1}$)	< .0001 ***	< .0001 ***	< .0001 ***
Flowering time (days)	< .0001 ***	< .0001 ***	< .0001 ***
Leaves at bud initiation	< .0001 ***	< .0001 ***	< .0001 ***
Rosette biomass	< .0001 ***	< .0001 ***	< .0001 ***
Chaff biomass	< .0001 ***	< .0001 ***	< .0001 ***
Seed yield	< .0001 ***	< .0001 ***	< .0001 ***
Calculated water-use (ml^{-1} H_20 $plant^{-1}$)	< .0001 ***	< .0001 ***	< .0001 ***
Calculated water productivity (mg seed $plant^{-1}$ ml H_20^{-1})	< .0001 ***	< .01 **	< .0001 ***

The GLMMs facilitated the calculation of V_G, which represents the total genetic variation, and V_P, which represents the total phenotypic variation (environmental, genetic, and residual). Both variation coefficients were used to estimate H^2. All traits assessed as part of the GWAS demonstrated highly significant H^2, which varied from 0.320 for cWP to 0.545 for flowering time. As with the ecotypes assessed in Chapter Two, the fitness-related

parameters, namely seed yield, chaff, and cWP, had the lowest levels of H^2. However, the estimates were much higher in this extended study and closer to the non-fitness related parameters. This is likely because these particular estimates are based on much more genetic variation, over twice as many ecotypes were incorporated into this present study as the study described in Chapter Two (Tables 2.1, 5.1).

Interestingly, the evolvabilities of seed yield and cWP were again much higher than the other parameters (Table 5.3). This reflects the observations described for these traits in the Chapter Two (Table 2.4) and suggests that fitness related parameters are far more sensitive to environmental heterogeneity than non-fitness related traits, and as such they are more likely to respond to selection. In general, the highly significant H^2 for all parameters suggests that it is possible to dissect the genetic basis of the observed natural variation through GWAS.

Table 5.3. Genotypic and phenotypic variation of the eight traits assessed as part of the GWAS. The true (arithmetic) mean, standard error (SE), genetic variance (V_G), phenotypic variance (V_P), broad sense heritability (H^2), coefficient of genetic variance (C_{VG}), and significance of H^2 are provided for all traits. All traits displayed significant broad sense heritability indicated by *** ($p < 0.0001$).

Trait	Mean	SE	V_G	V_P	H^2	C_{VG}	Sig.
Short term water-use (ml H_20 plant$^{-1)}$	8.912	0.033	0.947	2.185	0.433	10.920	***
Flowering time (days)	75.155	0.420	162.840	298.690	0.545	16.979	***
Rosette leaves at bud initiation	51.293	0.566	199.100	530.900	0.375	27.509	***
Rosette biomass (g)	0.312	0.005	0.019	0.041	0.469	44.336	***
Chaff biomass (g)	0.431	0.006	0.034	0.090	0.374	42.522	***
Seed yield (g)	0.076	0.002	0.002	0.007	0.335	63.917	***
Calculated water-use (ml H_20 plant^{-1})	667.193	4.583	14719	33063	0.445	18.184	***
Calculated water productivity (mg seed plant^{-1} ml H_20^{-1})	0.120	0.004	0.006	0.020	0.320	66.401	***

The estimated means of all of the eight traits assessed as part of the GWAS demonstrated variation that was not significantly different from normal (Fig 5.1.). It should be noted that the estimated means with negative values are not erroneous. It is possible to obtain estimated means below zero if the BLUPs from which they are calculated are extremely negative compared to the population mean, as is the case with seed yield and cWP (Fig. 5.1). In actuality, those ecotypes with negative estimated means for these traits would have demonstrated seed yields and cWPs that were just above zero. In general, all traits displayed substantial variation, for example those ecotypes that demonstrated the highest cWU were estimated to use double the volume of water than those that demonstrated the lowest cWU. Indeed, this trend was matched by all the other traits as well (Fig. 5.1).

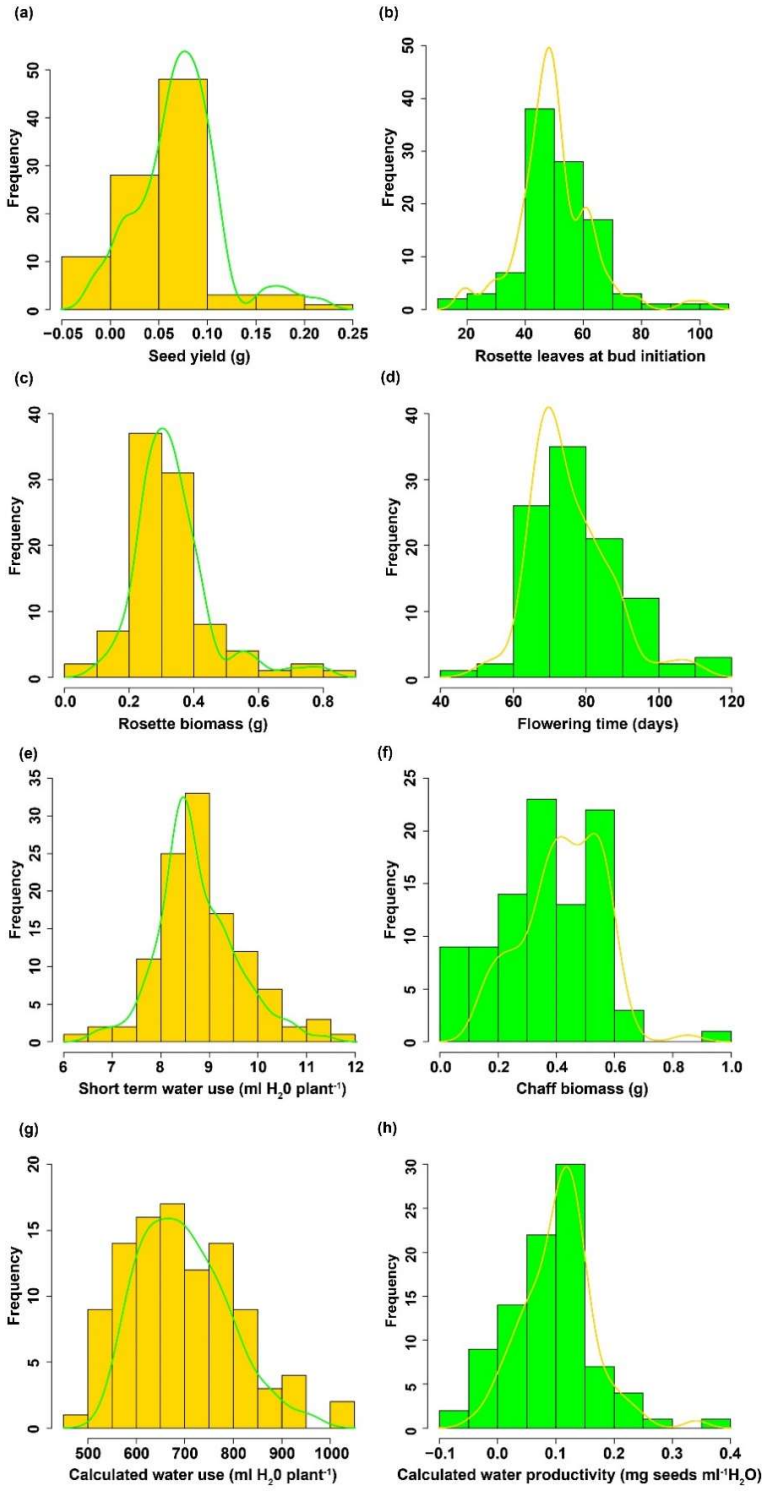

Figure 5.1. Natural variation of estimated means for all traits assessed as part of this study. (a) Histogram displaying the natural variation for seed yield of all ecotypes. **(b)**

Histogram displaying the natural variation for the number of rosette leaves at bud initiation of all ecotypes. **(c)** Histogram displaying the natural variation for rosette biomass of all ecotypes**. (d)** Histogram displaying the natural variation for flowering time of all ecotypes. **(e)** Histogram displaying the natural variation for short term water use of all ecotypes. **(f)** Histogram displaying the natural variation for chaff biomass of all ecotypes**. (g)** Histogram displaying the natural variation for calculated water use of all ecotypes. **(h)** Histogram displaying the natural variation for calculated water productivity of all ecotypes. All traits displayed distributions that were not significantly different from normal ($p > 0.05$), per Shapiro-Wilks test for normality performed on the estimated means of all ecotypes.

Genetic correlations were calculated between all traits as Pearson's product-moment correlation coefficients using the estimated means (Fig. 5.2). Flowering time displayed a significant negative genetic correlation with both chaff biomass and seed yield. Conversely, flowering time showed a significant positive genetic correlation with vegetative performance, i.e. rosette biomass accumulation, suggesting that ecotypes that flower later translocate the majority of the additional photosynthates primarily into vegetative sinks. The final genetic association of interest is that of the negative correlation shared between cWU and seed yield, where ecotypes that use more water have reduced reproductive performance compared to those that use less water (Fig 5.2).

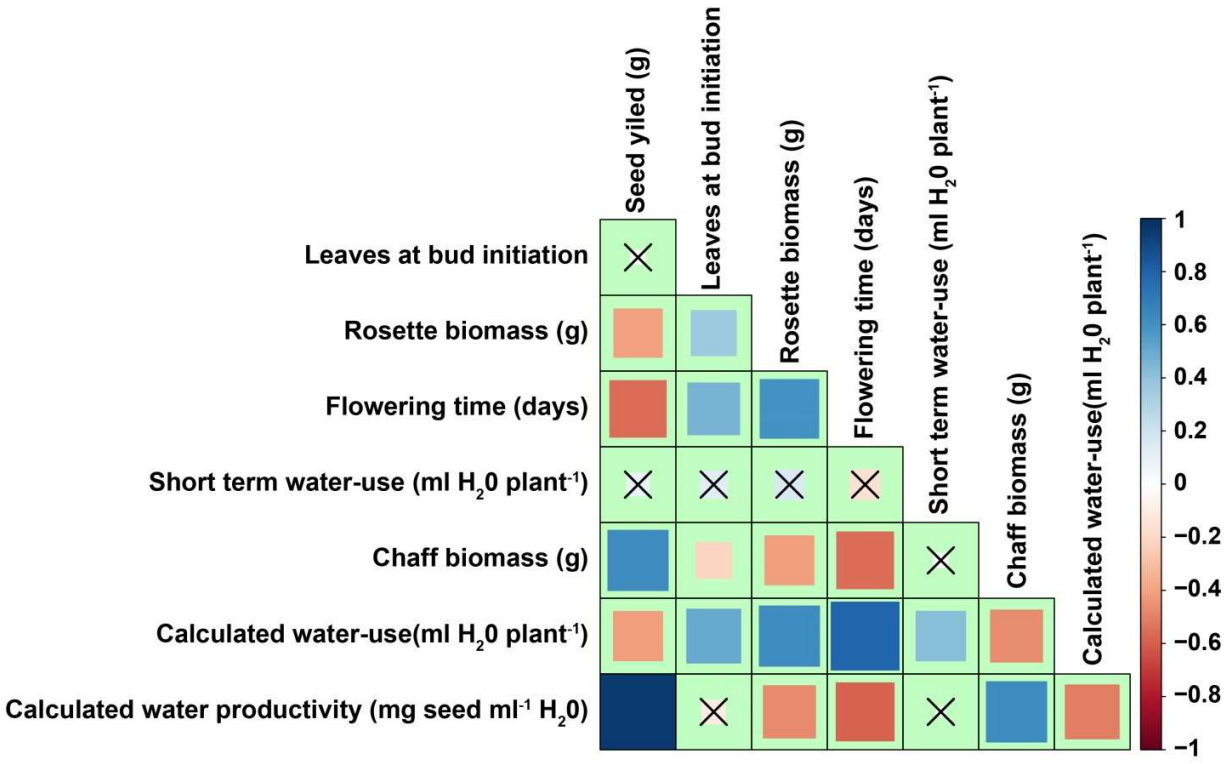

Figure 5.2. Correlation matrix plot describing the pairwise genetic correlations between all parameters assessed for all ecotypes comprising the GWAS panel. The size of the squares denotes the p-value, where larger squares indicate lower p-values, i.e. greater significance. The color of the square indicates the direction of the genetic correlations, with dark red indicating a negative correlation and dark blue indicating a positive correlation (see heat bar to the right). All squares not marked with a cross denote significant genetic correlations (p-value < 0.05). *N* = 15.

5.3.2 Genetic dissection of loci underlying the natural variation for key traits relating to water use and water productivity

To overcome the limitations of QTL mapping for identifying the genetic basis of productivity related traits, I postulated that a GWA mapping approach maybe more fruitful in terms of identifying loci important for these more agronomically relevant traits. To this end, the MLMM approach to GWAS was employed to dissect out the genetic basis of natural variation for the traits described in the previous sub-section of this chapter.

5.3.2.1 Short term water use

For the short term water use GWAS, none of the stepwise models with SNPs added as cofactors were significant according to the Bonferroni significance criteria suggested by Segura *et al* (2012; Fig. 5.3). As such, the most significant model was the first model which accounts for population structure only (Fig. 5.4). With regards to this model, two SNPs had $-\log_{10}$(p-values) above the designated threshold of five, one on chromosome three (3_14663400; Fig. 5.3) and one on chromosome five (5_1079059; Fig. 5.3). Comparison of means testing confirmed the association of these SNPs to the variation for short term water use (Fig. 5.5).

Seven protein coding genes were observed within the LD threshold of these two SNPs (Table 5.4). Analysis of the expression response of these genes to abiotic stress revealed that the *IQ-DOMAIN 2* gene (AT5G03960) was up-regulated in response to salinity stress in

root cells and an uncharacterized gene (AT3G42550) was up-regulated in response to drought stress in whole root samples (Table 5.5).

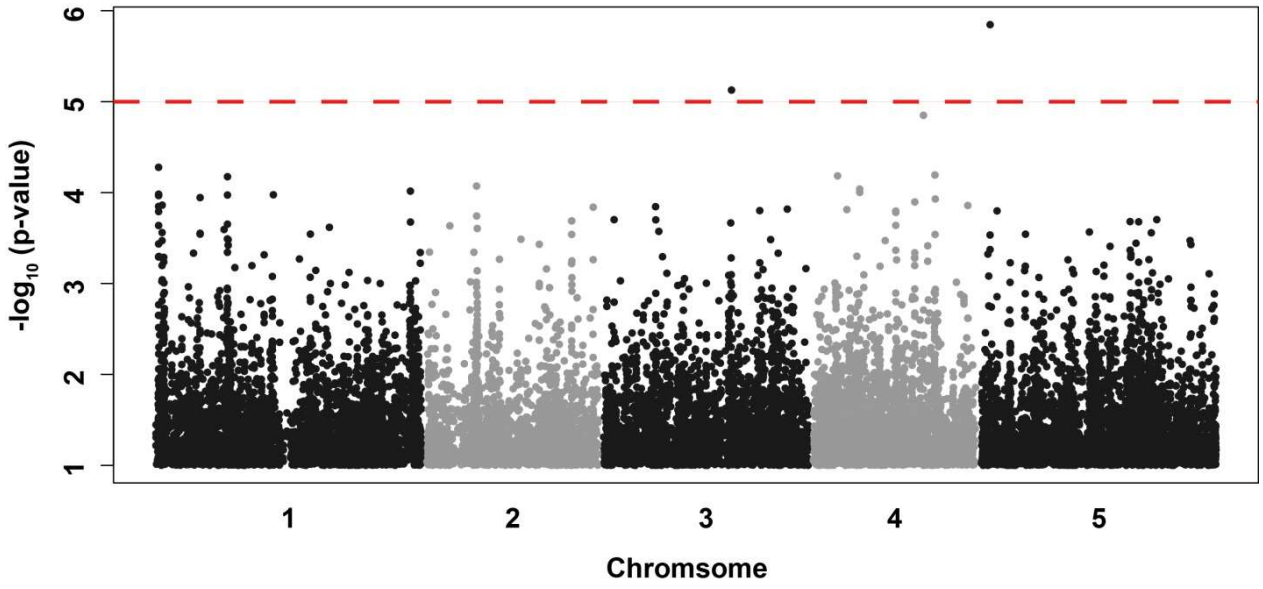

Figure 5.3. Manhattan plot of results for GWA-mapping for short term water use. Figure displays the $-\log_{10}$(p-values) (y-axis) over genomic positions (x-axis) where SNP markers included in this analysis exist. Successive chromosomes are denoted by the consecutive change in color from black to grey. The dashed red line indicates the selected threshold of 5 $-\log_{10}$(p-value).

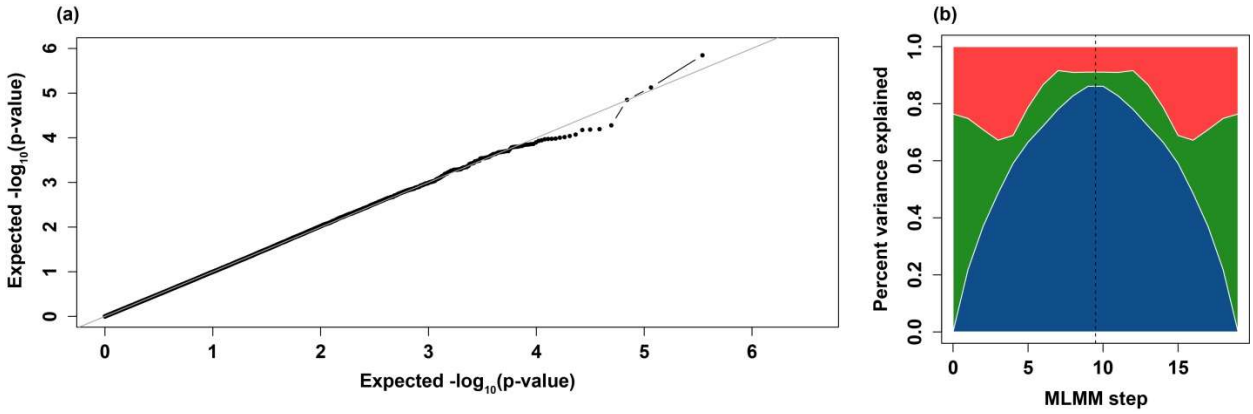

Figure 5.4. Summary statistics for GWA-mapping for short term water use (a) Quantile-quantile plot describing the observed variance in –log10(p-values) compared to the expected distribution of variance for the optimal stepwise model. **(b)** Variance plot describing the evolution in variance with each step of the multi-locus mixed model analysis (Blue: Genetic variance explained by SNP cofactors; Green: Estimated total genetic variance, Red: Error variance)

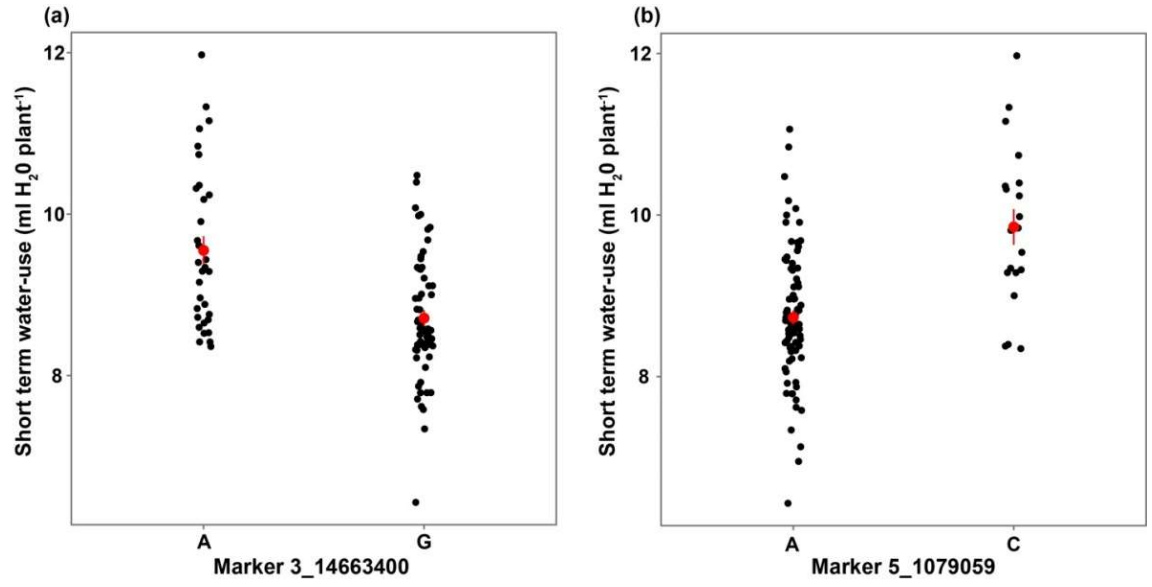

Figure 5.5. Allelic variation for short term water use of the two SNPs above the designated threshold from the associated GWAS. (a) Strip plot describing the variation

in short term water use of the A and G alleles of the SNP marker at 14663400bp on chromosome three. P-value < 0.0001 **(b)** Strip plot describing the variation in short term water use of the A and C alleles of the SNP marker at 1079059bp on chromosome five. P-value < 0.0001. For both plots the larger red dot indicates the mean value and the extend arms denote the distance of the standard errors above and below the mean. P-values were generated from a one-way ANOVA comparison of means test.

Table 5.4. Protein coding genes within a 20kb range of the two SNPs with $-\log_{10}$(p-values) above a designated threshold of five from the GWAS for short term water use. The associated SNP marker, locus code, other names, and key gene ontology (GO) terms are provided for all genes.

SNP Marker	Locus code	Other name(s)	GO Biological process(es)	GO cellular component(s)	GO molecular function(s)
5_1079059	AT5G03960	IQ-Domain 12	Unkown	Nucleus	Calmodulin binding
5_1079059	AT5G03970	Unkown function	Unkown	Nucleus	Unkown
5_1079059	AT5G03980	Unkown function	Lipid Catabolic region	Extraceullar region	Hydrolase activity, acting on ester bonds
5_1079059	AT5G03990	Unkown function	Unkown	Unkown	Unkown
5_1079059	AT5G03995	Unkown function	Unkown	Mitochondrion	Unkown
5_1079059	AT5G04000	Unkown function	Unkown	Nucleus	Unkown
3_14663400	AT3G42550	Unkown function	Proteolysis	Extraceullar region	Aspartic-type endopeptidase activity

Table 5.5. The response of genes associated with variation to short term water use to abiotic stress. The locus code, experimental details, and expression change details are provided for all experiments were the expression of genes arising from the GWAS for short term water use was significantly altered due to abiotic stress. These data were obtained using Genevestigator.

Locus code	Key experimental details	Log2-ratio	Fold-Change	P-value	References(s)
AT5603960	Salt stress (32h), root protoplast samples	-1.62	-3.09	< 0.001	Dinneny et al (2008)
AT5603960	Salt stress (8h), root protoplast samples	-1.7	-3.25	< 0.001	Dinneny et al (2008)
AT5603960	Salt stress (3h), root endodermis samples	-1.73	-3.31	< 0.001	Dinneny et al (2008)
AT5603960	Salt stress (1h), root protoplast samples	-1.75	-3.34	< 0.001	Dinneny et al (2008)
AT5603960	Salt stress (48h), root protoplast samples	-1.87	-3.67	< 0.001	Dinneny et al (2008)
AT3G42550	Drought study, root samples	2.79	6.86	<0.001	Galon et al (2010), Pandey et al (2013)
AT3G42550	Drought study, root samples	2.74	6.75	<0.001	Clauw et al (2015)

5.3.2.2 Flowering time

The optimal model for the flowering time GWAS was the second stepwise model, which included the 3_10792021 SNP as a cofactor. This model explained just over 40% of the total variation for flowering time (Figs. 5.6, 5.7). Comparisons of means testing confirmed the association of this SNP to gene(s) underlying the variation for flowering time (Fig. 5.8).

Six protein coding genes exist within the predetermined LD threshold of this SNP (Table 5.6). It is likely that the causal gene for this degree of variation for flowering time is *STRUCTURE SPECIFIC RECOGNITION PROTEIN 1* (*SSRP1*). *SSRP1* has been demonstrated as crucial for the transition from vegetative to reproductive growth, as it forms part of a transcription factor complex which binds to the promotor of *FLC*, thereby inducing its expression, which in turn delays flowering, since *FLC* suppresses *FLOWERING LOCUS T* (*FT*) and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS* (SOC1), which are genes that promote flowering (Lolas *et al.*, 2010).

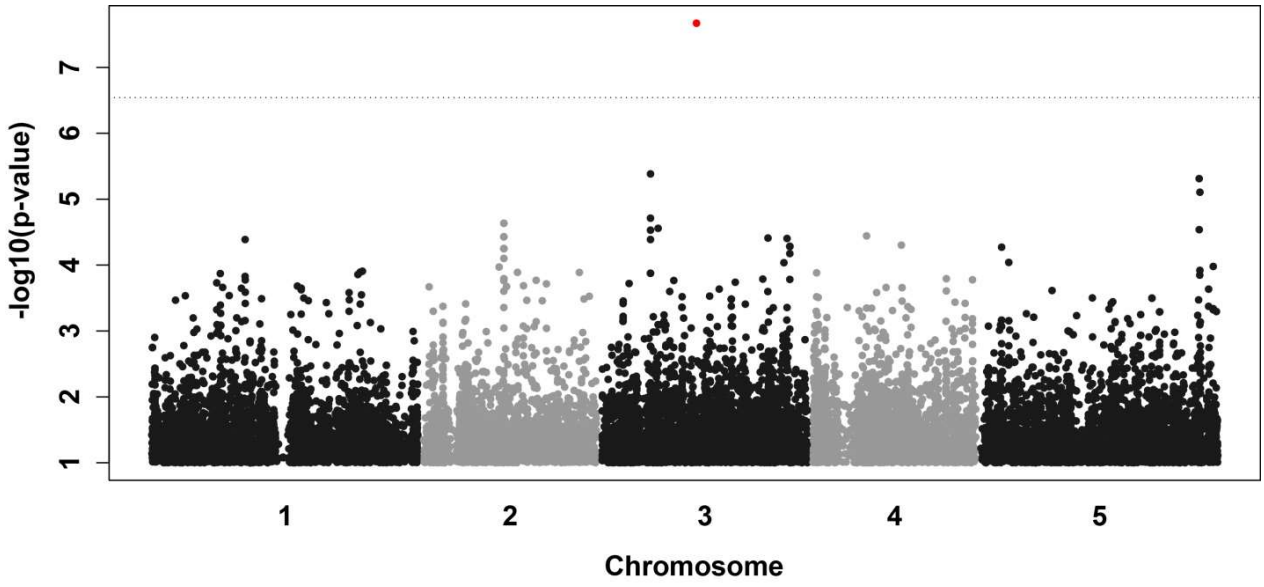

Figure 5.6. Manhattan plot of results of GWA-mapping for flowering time. Figure displays the $-\log_{10}$(p-values) (y-axis) over genomic positions (x-axis) where SNP markers included in these analyses exist. Successive chromosomes are denoted by the consecutive change in color from black to grey. SNPs used as cofactors in the multi-locus mixed model analysis are highlighted in red. The Bonferroni significance threshold is displayed as the dashed grey line.

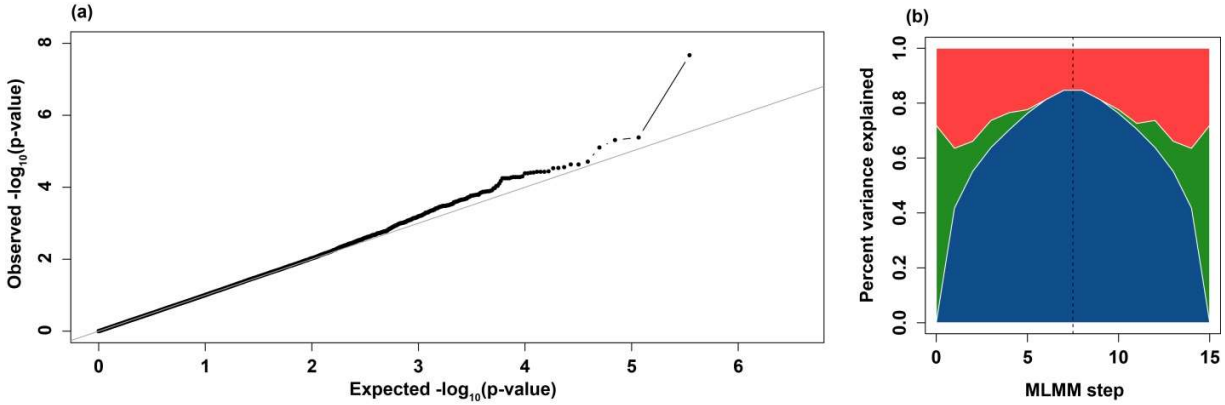

Figure 5.7. Summary statistics for GWA-mapping for flowering time (a) Quantile-quantile plot describing the observed variance in –log10(p-values) compared to the expected distribution of variance for optimal stepwise model. **(b)** Variance plot describing the evolution in variance with each step of the multi-locus mixed model analysis as part of the GWAS for flowering time (Blue: Genetic variance explained by SNP cofactors; Green: Total genetic variance; Red: Error variance).

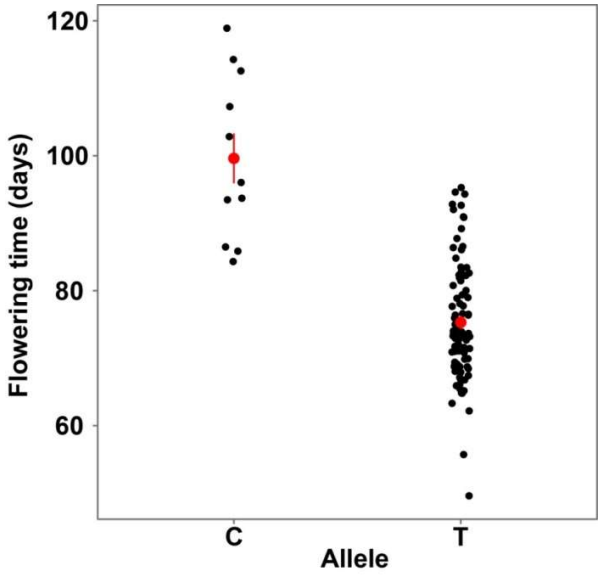

Figure 5.8. Allelic variation for flowering time of the significant SNP from the associated GWAS. Strip plot describing the variation in flowering time of the C and T alleles of the SNP marker at 1079202bp on chromosome three. P-value < 0.0001 as determined by a one-way ANOVA comparison of means test.

Table 5.6. Protein coding genes within a 20kb range of the significant SNP from the GWAS for flowering time. The associated SNP marker, locus code, other names, and key gene ontology (GO) terms are provided for all genes.

SNP Marker	Locus code	Other name(s)	GO Biological process(es)	GO cellular component(s)	GO molecular function(s)
3_10792021	AT3G28720	Unkown function	Unknown	Endoplasmic reticulum, mitochondrion	Unknown
3_10792021	AT3G28730	*Site Specific Recognition Protein 1* (*SSRP1*)	Vegetative to reprodutive phase transition of meristem	Nuclear euchromatin, nucleus	DNA binding, TF activity
3_10792021	AT3G28740	*Cytochrome P450*	Insect defense (synthesis of volatile compounds)	Membrane	Oxygen binding
3_10792021	AT3G28750	Unkown function	Unknown	Chloroplast	Unknown
3_10792021	AT3G28760	Unkown function	Aromatic amino acid biosynthetic process	Chloroplast	3-dehydroquinate synthase activity
3_10792021	AT3G28770	Unkown function	Unknown	Nucleus	Unknown

5.3.2.3 Number of rosette leaves at bud initiation

The second stepwise model was the optimal iteration for the GWAS for the number of rosette leaves at the point of bud initiation (Figs. 5.9, 5.10). Here, the cofactor SNP, 3_6586963, was the single significant SNP per the Bonferroni criterion and explained 25% of the total variation (Fig. 5.10b). Comparison of means testing confirmed the association of this SNP to variation for leaf number (Fig. 5.11). Four protein coding genes exist with the 20kb threshold of this SNP (Table 5.7). Analysis of their expression in response to abiotic stress revealed that none were significantly up or downregulated in response to stress. Furthermore, there were no obvious developmental or anatomical patterns in the expression of these genes either.

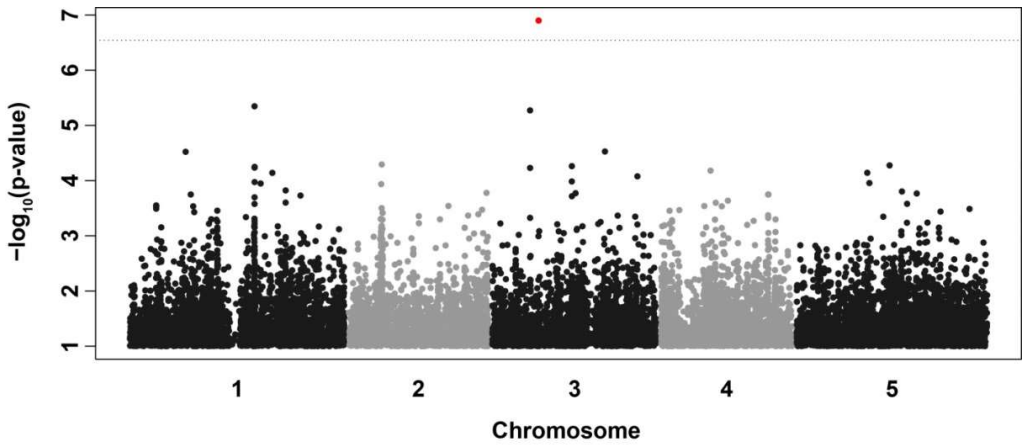

Figure 5.9. Manhattan plot of results of GWA-mapping for the number of rosette leaves at bud initiation. Figure displays the $-\log_{10}$(p-values) (y-axis) over genomic positions (x-axis) where SNP markers included in this analysis exist. Successive chromosomes are denoted by the consecutive change in color from black to grey. SNPs used as cofactors in the multi-locus mixed model analysis are highlighted in red. The Bonferroni significance threshold is displayed as the dashed grey line.

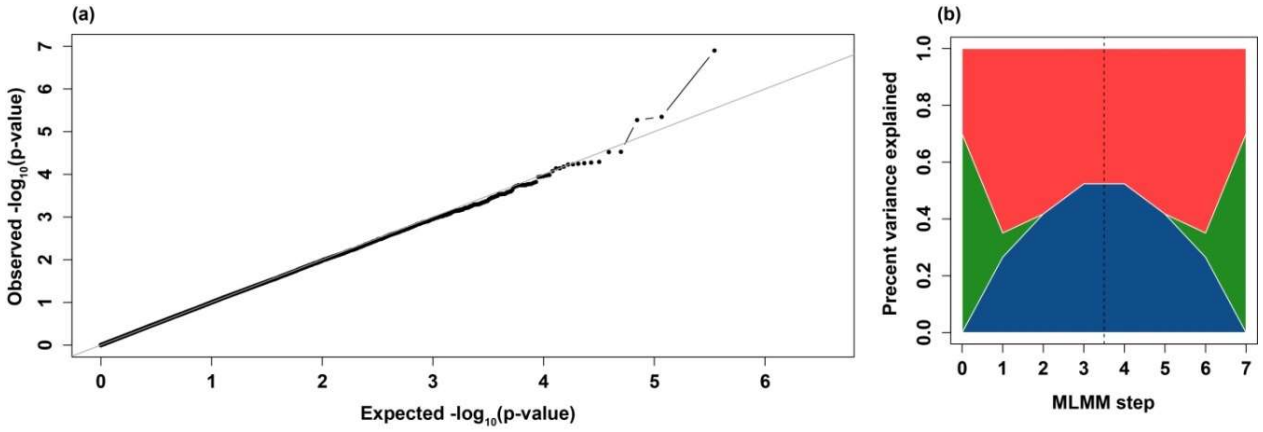

Figure 5.10. Summary statistics for GWA-mapping for the number of rosette leaves at bud initiation. (a) Quantile-quantile plot describing the observed variance in –log10(p-values) compared to the expected distribution for the optimal stepwise model. **(b)** Variance plot describing the evolution in variance with each step of the multi-locus mixed model analysis as part of the same GWAS (Blue: Genetic variance explained by SNP cofactors; Green: Total genetic variance; Red: Error variance).

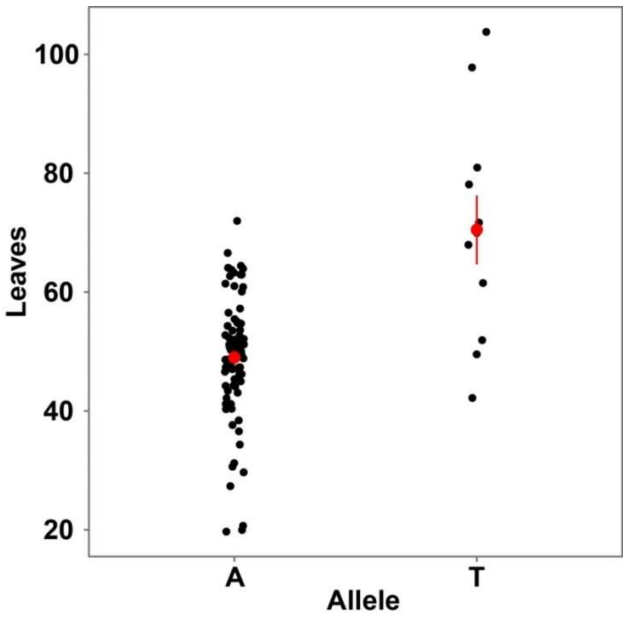

Figure 5.11. Allelic variation for the number of rosette leaves at bud initiation of the significant SNP from the associated GWAS. Strip plot describing the variation in flowering time of the A and T alleles of the SNP marker at 6586963bp on chromosome three. P-value < 0.0001 as determined by a one-way ANOVA comparison of means test.

Table 5.7. Protein coding genes within a 20kb range of the significant SNP from the GWAS for the number of rosette leaves at bud initiation. The associated SNP marker, locus code, other names, and key gene ontology (GO) terms are provided for all genes.

SNP marker	Locus code	Other name(s)	Key GO biological process(es)	GO cellular component(s)	GO molecular function(s)
3_6586963	AT3G19050	*PHRAGMOPLAST ORIENTING KINESIN 2* (*POK2*)	Cytokinesis	Cytoplasm	ATPase activity
3_6586963	AT3G19055	Unknown function	Unknown function	Mitochondrion	Unknown function
3_6586963	AT3G19070	Unknown function	Regulation of transcription	Nucleus	DNA binding
3_6586963	AT3G19080	Unknown function	Unknown function	Nucleus, plasmodesma	DNA binding

5.3.2.4 Rosette biomass

The first model was the most significant model for the GWAS for rosette biomass. Four SNPs had $-\log_{10}$(p-values) above the significance threshold for this model, two on chromosome three, one on chromosome four, and one on chromosome five (Figs. 5.12, 5.13). Comparison of means testing confirmed the associations of these SNPs to the variation for the rosette biomass accumulation (Fig. 5.14). There are 19 protein coding genes within the designated LD threshold of these four SNPs (Table 5.8). Of these 19 genes, the most responsive to abiotic stress was a previously uncharacterized gene which codes for a glycosyltransferase family protein (AT4G16710). This gene demonstrated substantial differential expression between ecotypes when subjected to a period of soil drying similar to the SD experiment employed in this study (Des Marais *et al.*, 2012; Table 5.9).

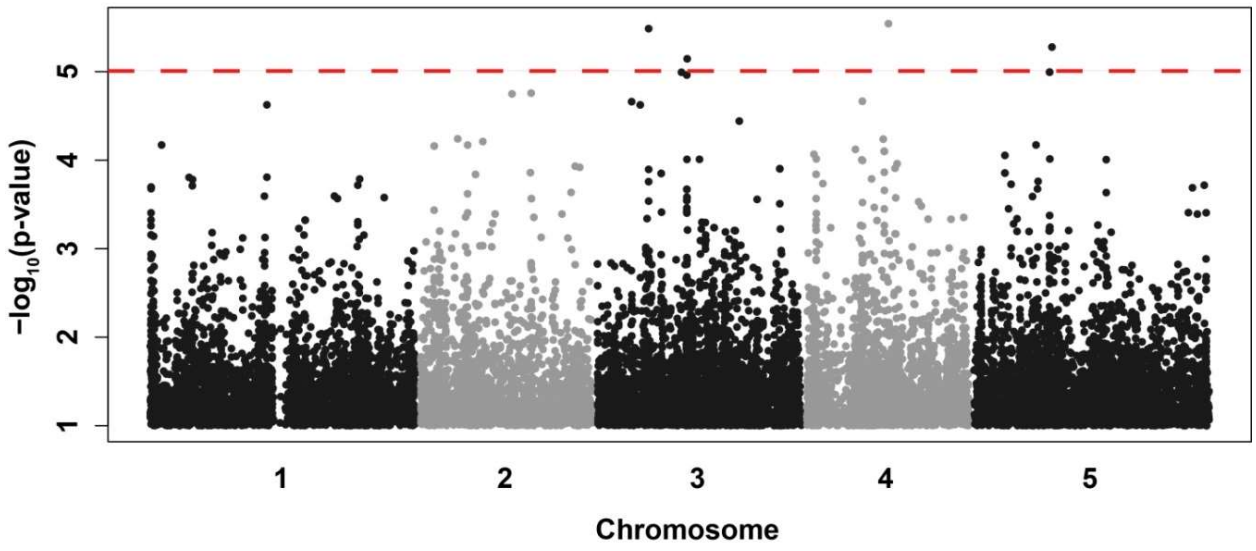

Figure 5.12. Manhattan plot of results of GWA-mapping for rosette biomass. Figure displays the $-\log_{10}$(p-values) (y-axis) over genomic positions (x-axis) for where SNP markers included in these analyses exist. Successive chromosomes are denoted by the consecutive change in color from black to grey. The dashed red line indicates the selected *significance* threshold of 5 $-\log_{10}$(p-value).

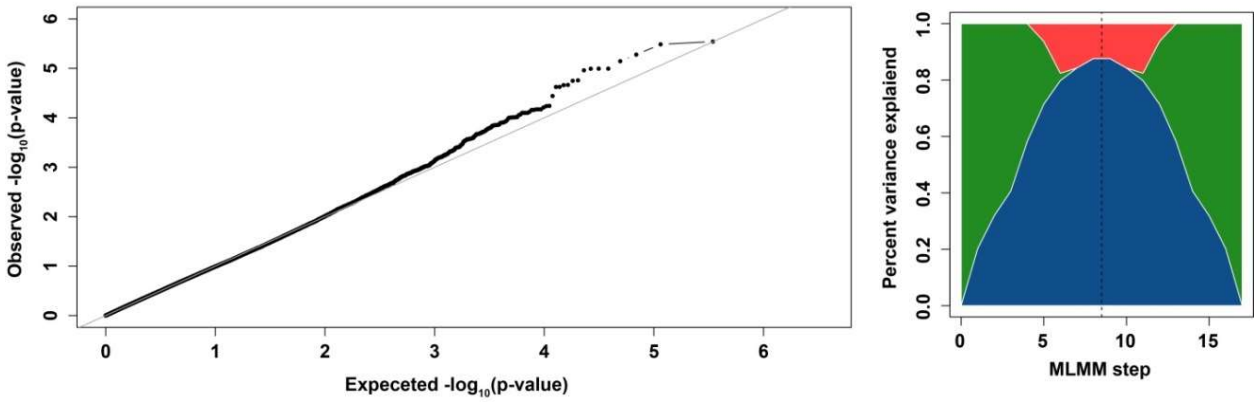

Figure 5.13. Summary statistics for GWA-mapping for rosette biomass (a) Quantile-quantile plot describing the observed variance in –log10(p-values) compared to the expected distribution of variance for the most optimal stepwise model. **(b)** Variance plot describing the

evolution in variance with each step of the multi-locus mixed model analysis as part of the same GWAS (Blue: Genetic variance explained; Green: Total genetic variance; Red: Error variance).

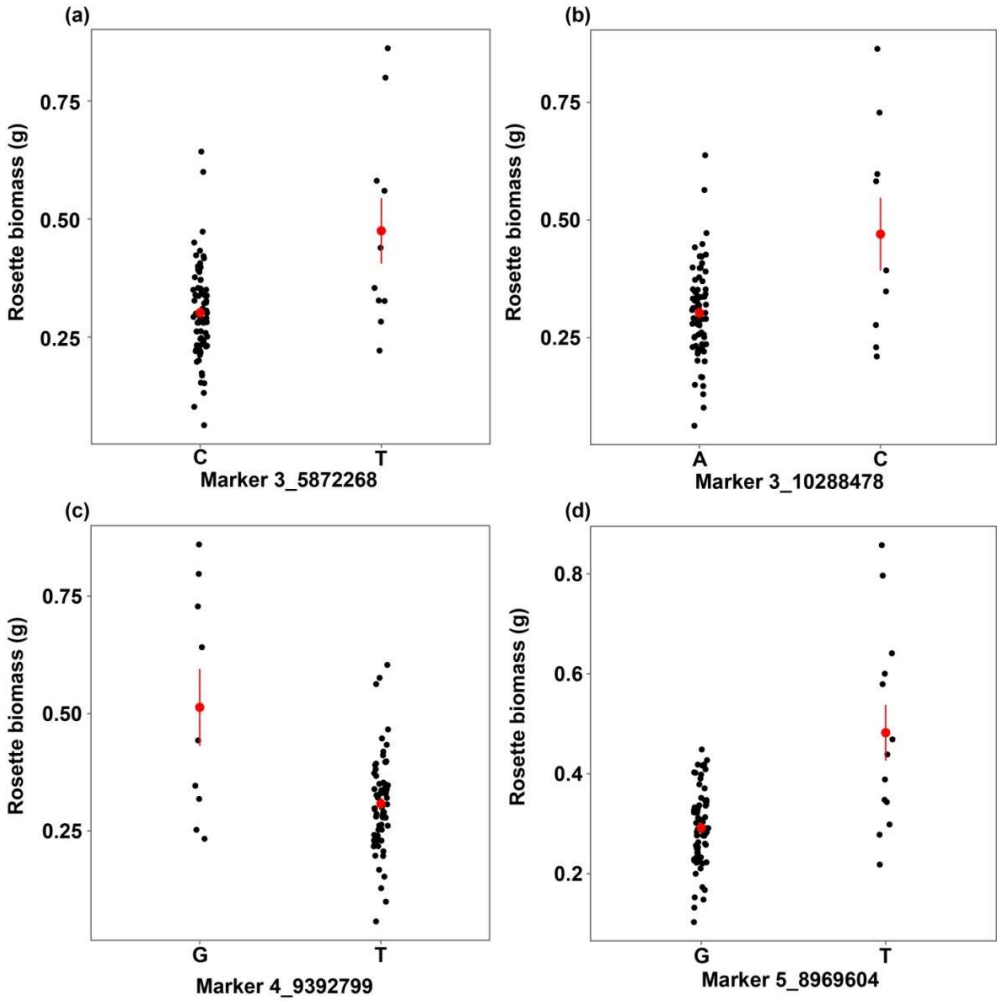

Figure 5.14. Allelic variation for rosette biomass of the four most significant SNPs from the associated GWAS. (a) Strip plot describing the variation in rosette biomass of the A and G alleles of the SNP marker at 5872268bp on chromosome three. P-value < 0.0001 **(b)** Strip plot describing the variation in rosette biomass of the A and C alleles of the SNP marker at 10288478bp on chromosome three. P-value < 0.0001 **(c)** Strip plot describing the variation in rosette biomass of the G and T alleles of the SNP marker at 9392799bp on

chromosome 4. P-value < 0.0001. **(d)** Strip plot describing the variation in rosette biomass of the G and T alleles of the SNP marker at 8969604bp on chromosome 5. For all plots the larger red dot indicates the mean value and the extend arms denote the distance of the standard errors above and below the mean. P-values were determined by a one-way ANOVA comparison of means test.

Table 5.8. Protein coding genes within a 20kb range of the two SNPs with – log_{10}(p-values) above a designated threshold of five from the GWAS for rosette biomass accumulation. The associated SNP marker, locus code, other names, and key gene ontology (GO) terms are provided for all genes.

SNP marker	Locus code	Other name(s)	Key GO biological process(es)	GO cellular component(s)	Key GO molecular function(s)
3_5872268	AT3G17190	Uknown function	Unknown	Mitochondrion	Unknown
3_5872268	AT3G17205	*UBIQUITIN PROTEIN LIGASE 6 (UPL6)*	Protein ubiquitination	Nucleus	Ubiquition-protein transferase activity
3_10288478	AT3G27730	MER3	Chiasma assembly	Nucleus	ATP binding
3_10288478	AT3G27740	*CARBAMOYL PHOSPHATE SYNTHETASE A (CARA)*	De-novo nucleotide synthesis	Chloroplast	ATP binding
3_10288478	AT3G27750	*EMBRYO DEFECTIVE 3123 (EMB3123)*	Intron splicing, chloroplast organisation	Chloroplast	RNA binding
3_10288478	AT3G27770	Uknown function	Unknown	Membrane, chloroplast	Unknown
3_10288478	AT3G27785	*MYB DOMAIN PROTEIN 118 (ATMYB118)*	Negative endosperm development regulation, glucosinate biosynthesis	Nucleus	DNA binding
4_9392799	AT4G16670	Uknown function	Signal transduction	Nucleus	Phospatidylinositol binding
4_9392799	AT4G16680	Uknown function	RNA splicing	Nucleus	ATP binding, RNA helicase activity
4_9392799	AT4G16690	*METHYL ESTERASE 16 (ATMES16)*	Carboxylesterase activity	Cytoplasm	Catalytic activity
4_9392799	AT4G16695	Uknown function	Unknown	Golgi appartus	Unknown
4_9392799	AT4G16700	*PHOSPHATIDYLSERINE DECARBOXYLASE 1 (PSD1)*	Phosophatidylethanolamine biosynthethesis	Membrane, mitochondrion	Phosphatidylserine dearboxylase activity
4_9392799	AT4G16710	Uknown function	Biosynthetic process	Plasma membrane	Transferase activity
4_9392799	AT4G16720	Uknown function	Translation	Multiple	Structural constituent of ribosome
5_8969604	AT5G25757	Uknown function	Formation of translation preinitiation complex	Multiple	Translation initiation factor activity
5_8969604	AT5G25760	*PEROXIN4 (PEX4)*	Protein ubiquitination	Multiple	Protein binding
5_8969604	AT5G25770	Uknown function	Unknown	Chloroplast	Hydrolase activity
5_8969604	AT5G25780	*EUKARYOTIC TRANSLATION INITIATION FACTOR 3B-2 (EIF3B-2)*	Formation of translation preinitiation complex	Multiple	Translation initiation factor activity
5_8969604	AT5G25790	Uknown function	Multicellular organism development	Nucleus	Metal ion binding, transcription factor activity

Table 5.9. The differential response of the AT4G16710 gene to slow soil drying in different ecotypes of Arabidopsis. Experimental details (ecotypes of interest) and expression change details are provided for all experiments. These data were obtained using Genevestigator.

Locus code	Key experimental details	Log2-ratio	Fold-Change	P-value	References(s)
AT4G16710	Slow soil drying, differential expression between CIBC-17 and Ws-2	1.91	3.76	< 0.001	Des Marais et al (2012)
AT4G16710	Slow soil drying, differential expression between HR-5 and Ws-2	1.77	3.42	< 0.001	Des Marais et al (2012)
AT4G16710	Slow soil drying, differential expression between Col-2 and Ws-2	1.71	3.28	< 0.001	Des Marais et al (2012)
AT4G16710	Slow soil drying, differential expression between NFA-10 and Ws-2	1.68	3.19	< 0.001	Des Marais et al (2012)
AT4G16710	Slow soil drying, differential expression between Col-2 and Ts-1	1.67	3.16	< 0.001	Des Marais et al (2012)
AT4G16710	Slow soil drying, differential expression between NFA-10 and Ws-2	1.67	3.17	< 0.001	Des Marais et al (2012)
AT4G16710	Slow soil drying, differential expression between SQ-8 and Ws-2	1.66	3.17	< 0.001	Des Marais et al (2012)
AT4G16710	Slow soil drying, differential expression between CIBC-17 and Ws-2	1.66	3.18	< 0.001	Des Marais et al (2012)
AT4G16710	Slow soil drying, differential expression between NFA-10 and Ts-1	1.62	3.05	< 0.001	Des Marais et al (2012)
AT4G16710	Slow soil drying, differential expression between SQ-8 and Ts-1	1.62	3.05	< 0.001	Des Marais et al (2012)

5.3.2.5 GWAS for chaff biomass

The fourth MLMM model was the optimal model for GWAS for chaff biomass accumulation. This model accounted for over 60% of the total variation and included three SNPs as cofactors in the analysis. These cofactor SNPs were the only significant SNPs and were found on chromosomes one, two, and three (Figs. 5.15, 5.16). Comparison of means testing confirmed their association to the variation for chaff biomass accumulation (Fig. 5.17). 14 protein coding genes were observed to exist within the predetermined LD threshold of these three SNPs (Table 5.10). Analysis of the expression of these 14 genes in response to abiotic stress revealed that the *BCL-2-ASSOCIATED ATHANOGENE 6* (*ATBAG6*) gene is highly upregulated in response to multiple abiotic stresses, although it appears to be primarily associated to heat stress. With respect to drought stress specifically, *ATBAG6* demonstrated substantial upregulation in roots samples in response to mild drought stress (Table 5.11).

The accumulation of chaff biomass has been demonstrated as part of this study, and in the studies comprising the preceding chapters, to be genetically linked with seed yield and water productivity. As such, the identification of *ATBAG6* through GWAS for chaff biomass and its observed importance in the response to heat and drought stress, suggests that this gene may be important for combining improved heat and drought stress resistance with stabilized reproductive performance.

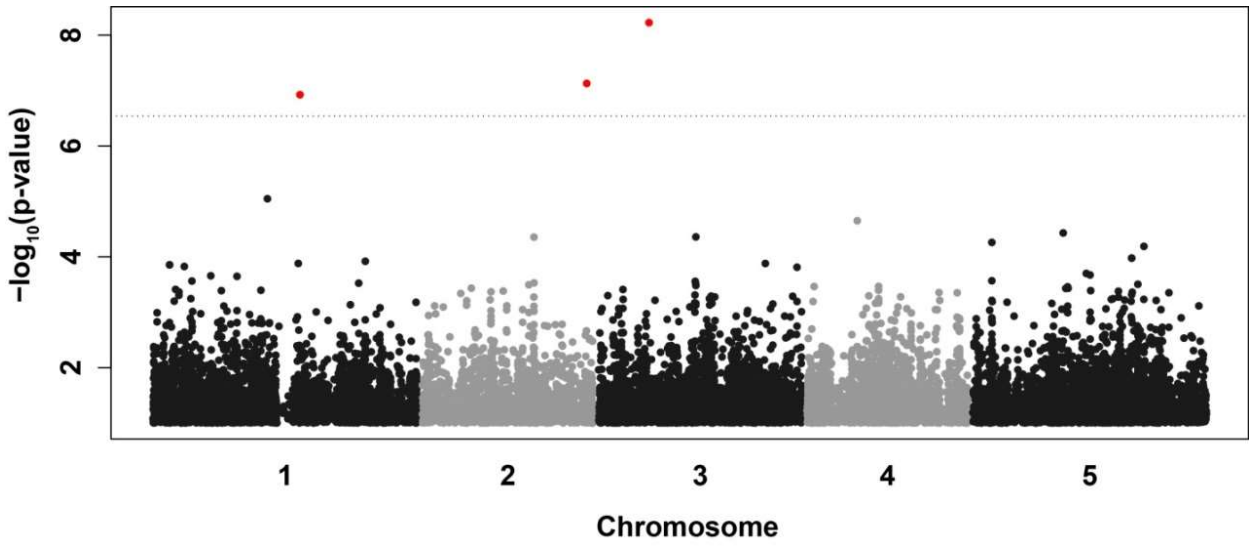

Figure 5.15. Manhattan plot of results of genome-wide association mapping for the chaff biomass. Figure displays the $-\log_{10}$(p-values) (y-axis) over genomic positions (x-axis) for where SNP markers included in these analyses exist. Successive chromosomes are denoted by the consecutive change in color from black to grey. SNPs used as cofactors in the multi-locus mixed model analysis are highlighted in red. The Bonferroni significance threshold is displayed as the dashed grey line.

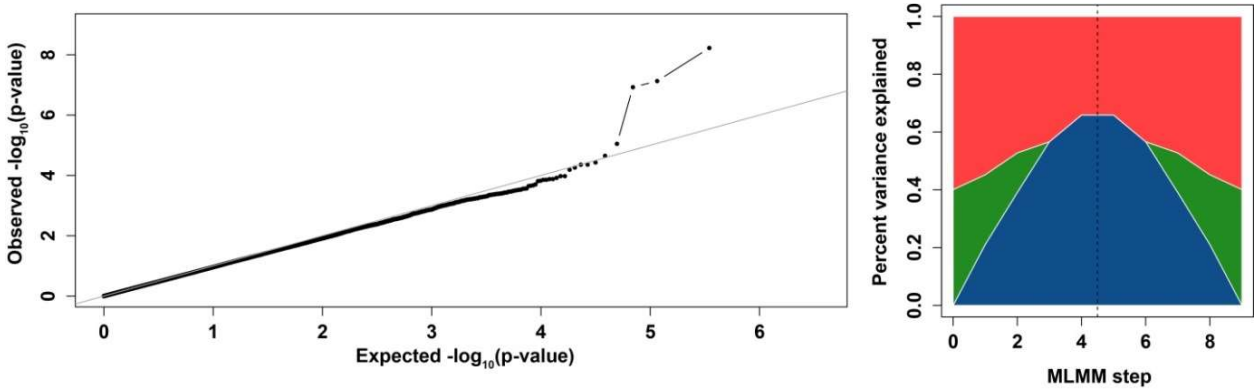

Figure 5.16. Summary statistics for GWA-mapping for chaff biomass accumulation (a) Quantile-quantile plot describing the observed variance in $-\log_{10}$(p-values) compared to the expected distribution for the optimal stepwise model. **(b)** Variance plot describing the

evolution in variance with each step of the multi-locus mixed model analysis (Blue: Genetic variance explained by SNP cofactors; Green: Estimated total genetic variance; Red: Error variance).

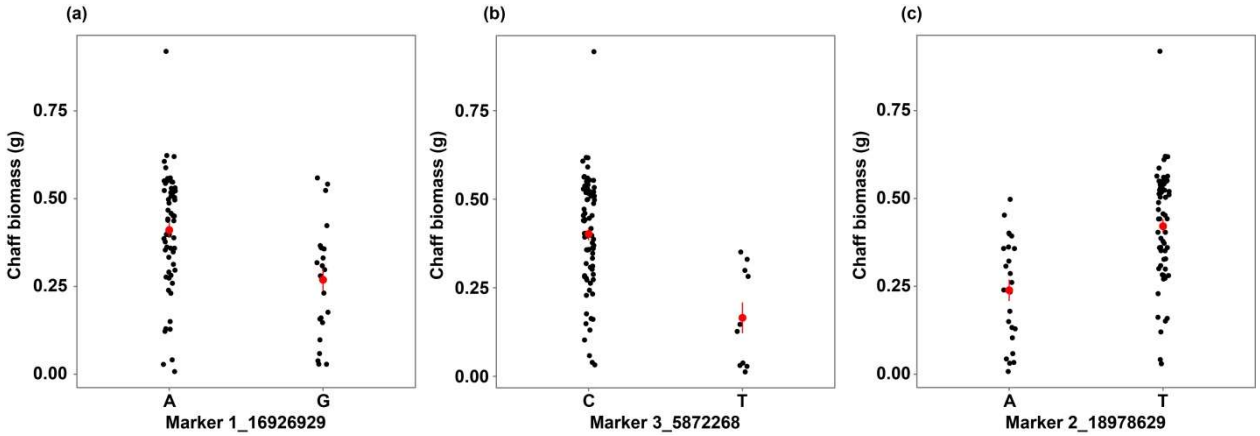

Figure 5.17. Allelic variation for chaff biomass of the three significant SNPs from the associated GWAS (a) Strip plot describing the variation in chaff biomass of the A and G alleles of the SNP marker at 16926929bp on chromosome one. P-value < 0.0001 **(b)** Strip plot describing the variation in chaff biomass of the C and T alleles of the SNP marker at 5872268bp on chromosome three. P-value < 0.0001 **(c)** Strip plot describing the variation in chaff biomass of the A and T alleles of the SNP marker at 18978629bp on chromosome 2. P-value < 0.0001. For all plots the larger red dot indicates the mean value and the extend arms denote the distance of the standard errors above and below the mean. P-values were determined by a one-way ANOVA comparison of means test.

Table 5.10. Protein coding genes within a 20kb range of the significant SNPs from the GWAS for chaff biomass. The associated SNP marker, locus code, other names, and key gene ontology (GO) terms are provided for all genes.

SNP marker	Locus code	Other name(s)	Key GO biological process(es)	GO cellular component(s)	Key GO molecular function(s)
1_16926929	AT1G44800	*SILIQUES ARE RED 1* (*SIAR1*)	Amino acid export	Membrane, mitochondrion	Amino acid transmbrane transporter activity
1_16926929	AT1G44810	Uknown function	Regulation of transcription	Nucleus	DNA binding
1_16926929	AT1G44820	Uknown function	Hydrolase activity, aminoacylase activity	Chloroplast, endoplasmic reticulum	Aminoacylase activity
1_16926929	AT1G44830	Uknown function	Multiple	Nucleus	DNA binding
2_18978629	AT2G46180	*GOLGIN CANDIDATE 4* (*GCC4*)	Golgi organisation	Golgi apparatus	Unknown
2_18978629	AT2G46190	Uknown function	Unknown	Mitochondrion	Unknown
2_18978629	AT2G46200	Uknown function	RNA splicing	Cytoplasm	Unknown
2_18978629	AT2G46210	*SPHINGOID LCB DESATURASE 2* (*ATSLD2*)	Cellular response to cold	Multiple	Metal ion binding, oxidoreductase activity
2_18978629	AT2G46220	-	Unknown	Chloroplast	Unknown
2_18978629	AT2G46225	*ABI-1-LIKE 1* (*ABIL1*)	Actin nucleation, trichome morphogenesis	Cytoplasm	Protein binding
2_18978629	AT2G46230	Uknown function	Unknown	Nucleus, small-subunit processome	Unknown
2_18978629	AT2G46240	BCL-2-ASSOCIATEED ATHANOGENE 6 (ATBAG6)	Mutiple	Nucleus, plasmodesma	Calmodulin binding, protein binding
3_5872268	AT3G17190	Uknown function	Unknown	Mitochondrion	Unknown
3_5872268	AT3G17205	*UBIQUITIN PROTEIN LIGASE 6* (*UPL6*)	Multiple	Multiple	Ligase acitivity, ubiquitin-protein transferase activity

Table 5.11 The differential response of *ATBAG6* to various forms of abiotic stress. Experimental details, expression change details, and appropriate references are provided for all experiments. These data were obtained using Genevestigator.

Locus code	Key experimental details	Log2-ratio	Fold-Change	P-value	Reference(s)
AT2G46240	Heat stress, seedling samples	8.48	389.66	< 0.001	Not available
AT2G46240	Heat stress, seedling samples	8.46	358.95	< 0.001	Suzuki et al (2011)
AT2G46240	Heat stress, seedling samples	8.36	346.51	< 0.001	Not available
AT2G46240	Heat stress, seedling samples	8.27	311.92	< 0.001	Suzuki et al (2011)
AT2G46240	Heat stress, aerial tissue samples	7.97	251.12	< 0.001	Clauw et al (2015)
AT2G46240	Heat stress, aerial tissue samples	7.40	169.01	< 0.001	Not available
AT2G46240	Heat stress, shoot apical meristem samples	7.20	147.42	< 0.001	Clauw et al (2015)
AT2G46240	Heat stress, shoot apical meristem samples	7.06	134.59	< 0.001	Clauw et al (2015)
AT2G46240	Heat stress, green tissue samples	5.94	109.56	< 0.001	Clauw et al (2015)
AT2G46240	Hypoxia stress, seedling samples	5.63	49.32	< 0.001	Not available
AT2G46240	Heat stress	5.55	44.97	< 0.001	Clauw et al (2015)
AT2G46240	Heat stress, root samples	5.43	65.00	< 0.001	Clauw et al (2015)
AT2G46240	Drought stress, root samples	5.37	41.67	< 0.001	Pandey et al (2013)
AT2G46240	Heat stress	4.63	23.84	< 0.001	Clauw et al (2015)
AT2G46240	Drought stress, root samples	4.57	23.76	< 0.001	Clauw et al (2015)
AT2G46240	Hypoxia stress, seedling samples	3.80	13.95	< 0.001	Not available
AT2G46240	Salt stress (20h), root protoplast samples	2.87	7.35	< 0.001	Dinneny et al (2008)
AT2G46240	Salt stress (48h), root protoplast samples	2.76	6.75	< 0.001	Dinneny et al (2008)
AT2G46240	Drought stress, whole plant samples	2.41	5.40	< 0.001	Not available
AT2G46240	Salt stress (8h), root protoplast samples	2.37	5.18	< 0.001	Dinneny et al (2008)
AT2G46240	Osmotic stress, root samples	2.31	5.49	< 0.001	Clauw et al (2015)
AT2G46240	Salt stress (1h), root protoplast samples	2.05	4.14	< 0.001	Dinneny et al (2008)
AT2G46240	Salt stress, root samples	2.03	4.19	< 0.001	Clauw et al (2015)
AT2G46240	Salt stress (48h), root protoplast samples	1.89	3.69	< 0.001	Dinneny et al (2008)
AT2G46240	Salt stress (32h), root protoplast samples	1.62	3.10	< 0.001	Dinneny et al (2008)

5.3.2.6 GWAS for seed yield

The optimal model for the GWAS for seed yield was the second stepwise model. Here, the associated cofactor SNP (5_7666515) was the only SNP with a –log10(p-value) above the Bonferroni significance threshold (Fig. 5.18). This second stepwise model explained 18% of the total variation for seed yield (Fig. 5.19), which is a particularly large amount of variation for a trait that is likely influenced by many genetic factors, especially since I have already demonstrated that it is strongly influenced by environmental fluctuations (2.3.1).

Comparison of means testing confirmed the association of 5_7666515 to the variation for seed yield (Fig. 5.20). Four protein coding genes were observed to be within the set 20kb range of this SNP (Table 5.12). Analysis of the response of these four genes to abiotic stress revealed that one gene in particular, *CHY ZINC-FINGER AND RING PROTEIN 1* (*CHYR1*), is differentially expressed under abiotic stress, specifically heat and cold stress. *CHYR1* is significantly down regulated in response to heat and cold stress (Table 5.13), furthermore it has recently been demonstrated to promote ABA-induced stomatal closure, thereby contributing to drought resistance and reduced water-loss through transpiration (Ding *et al.*, 2015). The downregulation of *CHYR1*, which leads to reduced stomatal conductance, in response to temperature stress is concurrent with observations that stomatal closures can be induced in response to cold (Wilkinson *et al.*, 2001) or heat stress (Song *et al.*, 2014).

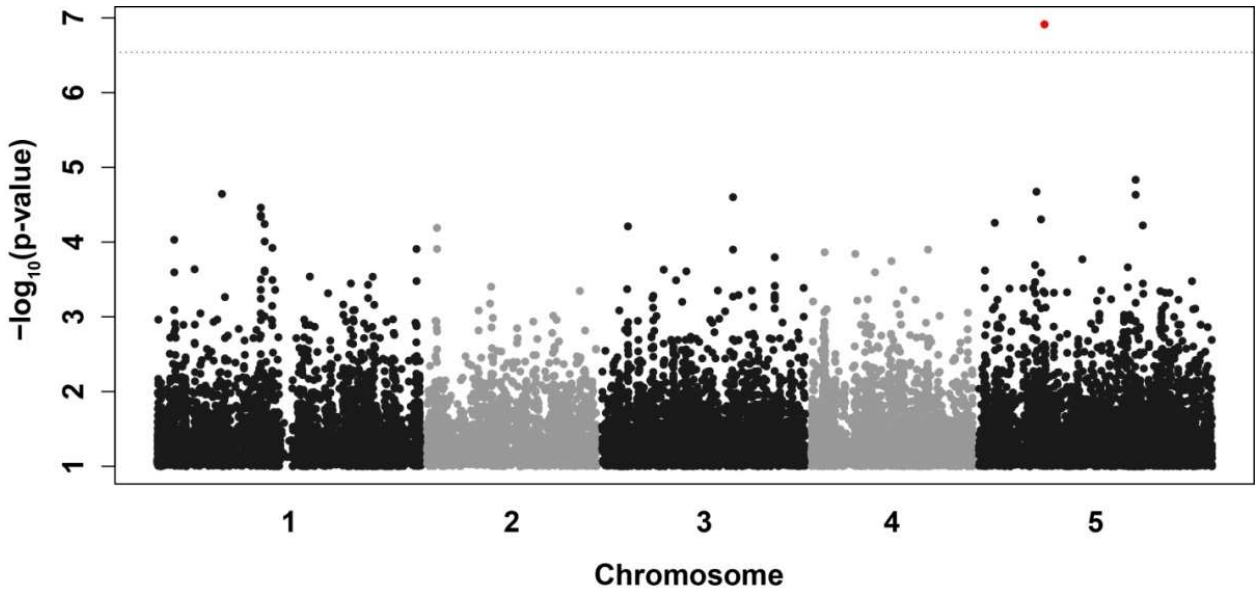

Figure 5.18. Manhattan plot of results of genome-wide association mapping for the seed yield. Figure displays the $-\log_{10}$(p-values) (y-axis) over genomic positions (x-axis) for where SNP markers included in these analyses exist. Successive chromosomes are denoted by the consecutive change in color from black to grey. SNPs used as cofactors in the multi-locus mixed model analysis are highlighted in red. The Bonferroni significance threshold is displayed as the dashed grey line.

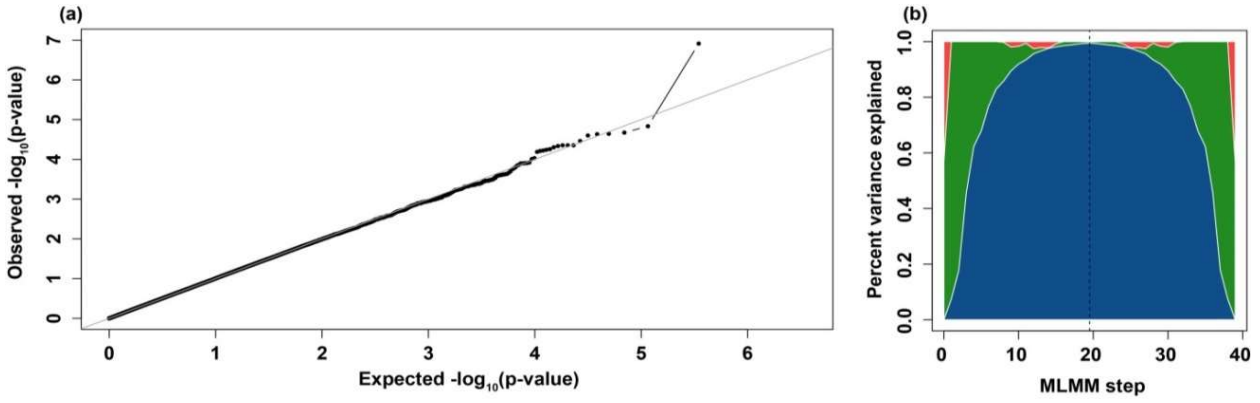

Figure 5.19. Summary statistics for GWA-mapping for seed yield (a) Quantile-quantile plot describing the observed variance in $-\log_{10}$(p-values) compared to the expected

distribution of variance for the optimal stepwise model. **(b)** Variance plot describing the evolution in variance with each step of the multi-locus mixed model analysis (Blue: Genetic variance explained by SNP cofactors; Green: Estimated total genetic variance; Red: Error variance).

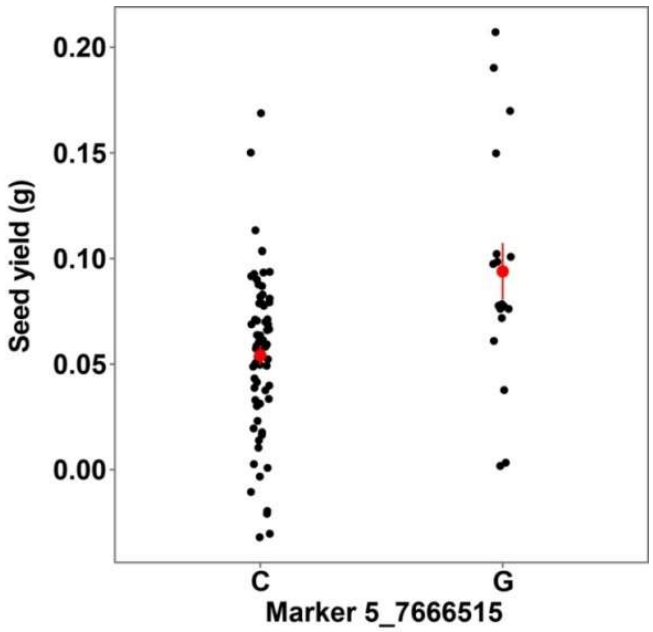

Figure 5.20. Allelic variation for the seed yield of the significant SNP from the associated genome-wide association mapping analysis. Strip plot describing the variation in seed yield of the C and G alleles of the SNP marker at 7666515bp on chromosome five. P-value < 0.0001.

Table 5.12. Protein coding genes within 10 kb up- and down-stream of the significant SNP from the GWAS for seed yield. The associated SNP marker, locus code, other names, and key gene ontology (GO) terms are provided for all genes.

SNP marker	Locus code	Other name(s)	Key GO biological process(es)	GO cellular component(s)	Key GO molecular process(es)
5_766515	AT5G22900	*ARABIDOPSIS THALINA CATION/H+ EXCAHNGER 3* (*ATCHX3*)	Cation transport	Membrane	Monovalent cation:proton anitporter activity
5_766515	AT5G22910	*ARABIDOPSIS THALINA CATION/H+ EXCAHNGER 9* (*ATCHX9*)	Cation transport	Membrane	Monovalent cation:proton anitporter activity
5_766515	AT5G22920	*CHY ZINC-FINGER AND RING PROTEIN 1* (*CHYR1*)	Regulation of stomatal opening	Nucleus	Zinc ion binding
5_766515	AT5G22930	Unknown function	Unknown	Nucleus	Unknown

Table 5.13 The differential response of *CHYR1* to temperature stress. Experimental details, expression change details, and appropriate references are provided for all experiments. These data were obtained using Genevestigator.

Locus code	Key experimental details	Log2-ratio	Fold-Change	P-value	Reference(s)
AT5622920	Heat stress, aerial tissue samples	-2.05	-4.18	< 0.001	Unavailable
AT5622920	Cold stress, rosette samples	-2.12	-4.33	< 0.001	Xu et al (2015)
AT5622920	Heat stress, aerial tissue samples	-2.32	-4.95	< 0.001	Clauw et al (2015)
AT5622920	Cold stress, rosette samples	-4	-15.96	< 0.001	Clauw et al (2015)

5.3.2.7 Calculated water use

The initial model was the optimal model for the GWAS for cWU. Two SNPs had association $-\log_{10}$(p-values) above the designated significance threshold, one on chromosome four and one on chromosome five (Figs. 5.21, 5.22). Comparison of means testing confirmed the association of these SNPs to the variation for water use (Fig. 5.23). Six proteins coding genes were observed to exist within the 20kb set LD threshold of these genes (Table 5.14). Analysis of gene expression in response to abiotic stress revealed that none of the genes showed a response to water stress. However one of the genes, a previously uncharacterized receptor like protein kinase (AT5G48540), is significantly upregulated in response to oxygen deprivation, also termed hypoxia (Table 5.15), indeed it is an ortholog of multiple cereal genes who encoded proteins are highly accumulated in coleoptiles of cereals grown under anoxia (Shingaki-Wells *et al.*, 2011). AT5G48540 is also significantly upregulated in response to drought and osmotic stress, as evident from various unique studies (Table 5.15).

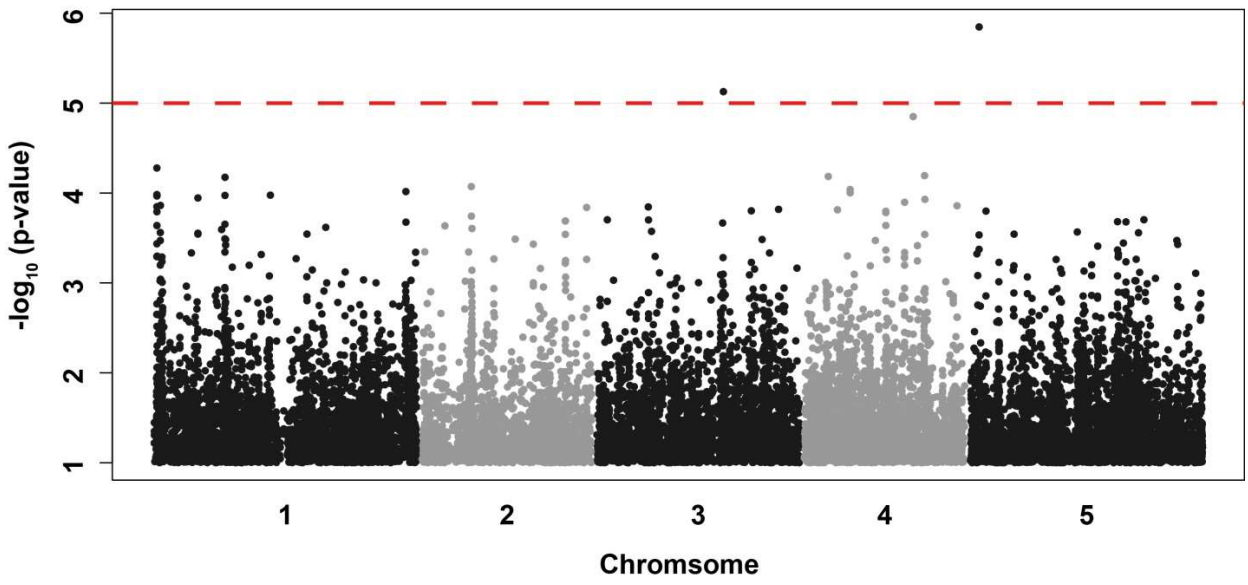

Figure 5.21. Manhattan plot of results of for GWA-mapping for calculated water use. Figure displays the $-\log_{10}$(p-values) (y-axis) over genomic positions (x-axis) for where SNP markers included in these analyses exist. Successive chromosomes are denoted by the consecutive change in color from black to grey. The dashed red line indicates the selected *significance* threshold of 5 $-\log_{10}$(p-value

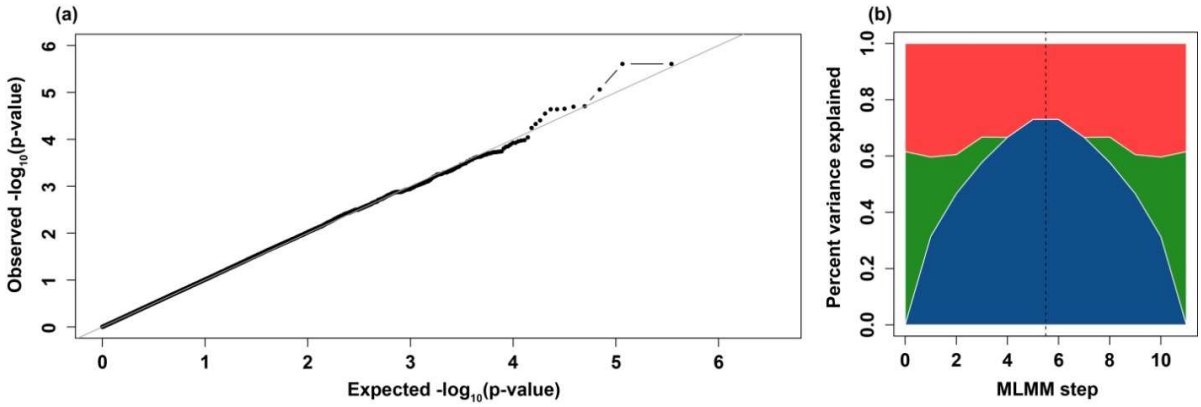

Figure 5.22. Summary statistics for GWA-mapping for calculated water use (a) Quantile-quantile plot describing the observed variance in –log10(p-values) compared to the expected distribution for the optimal stepwise model**. (b)** Variance plot describing the evolution in variance with each step of the multi-locus mixed model analysis (Blue: Genetic variance explained by SNP cofactors; Green: Estimated total genetic variance; Red: Error variance).

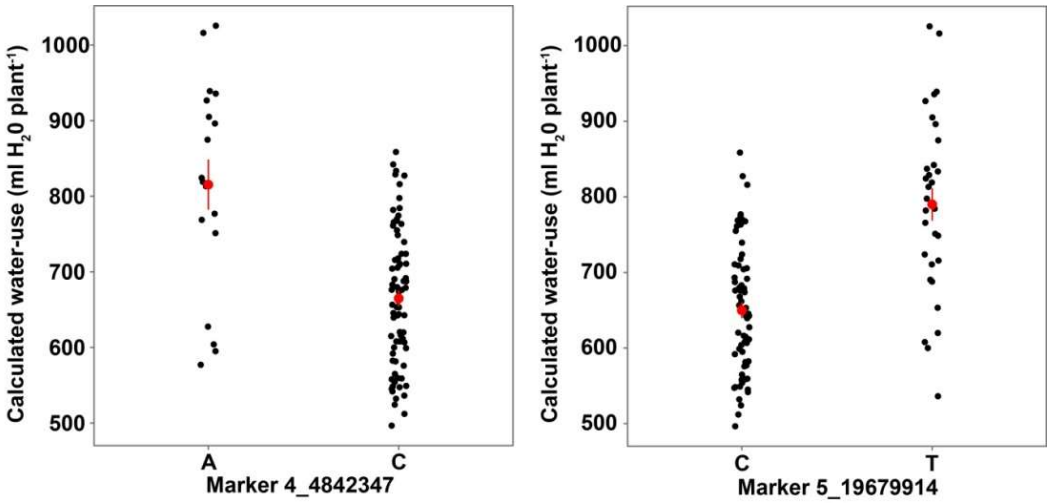

Figure 5.23. Allelic variation for calculated water use of the two most significant SNPs from the associated genome-wide association mapping analysis. Strip plot describing the variation in calculated water use of the A and C alleles of the SNP marker at 4842347

bp on chromosome four. P-value < 0.0001. **(b)** Strip plot describing the variation in calculated water use of the C and T alleles of the SNP marker at 19679914bp on chromosome five. P-value < 0.0001. For both plots the larger red dot indicates the mean value and the extended arms denote the distance of the standard errors above and below the mean. P-values were generated from one-way ANOVA comparison of means tests.

Table 5.14. Protein coding genes within a 20kb range of the two SNPS with $-\log_{10}$(p-values) above a designated threshold of 5 from the GWAS for calculated water use. The associated SNP marker, locus code, other names, and key gene ontology (GO) terms are provided for all genes.

SNP marker	Locus code	Other name(s)	Key GO biological process(es)	GO cellular component(s)	Key GO molecular process(es)
5_19679914	AT5G48540	Unknown function	Response to karrikin	Extracellular region	Unknown
5_19679914	AT5G48543	LOW-MOLECULAR-WEIGHT CYSTEINE-RICH 1 (LCR1)	Defense response to fungus	Extracellular region	Unknown
5_19679914	AT5G48545	HISTIDINE TRIAD NUCLEOTIDE-BINDING 3 (HINT3)	Purine ribonucleotid and sulfur compound metabolic processes	Nucleus, peroxisome	Adenylylsufatase activity
5_19679914	AT5G48550	Unkown function	Unknown	Extracellular region	Unknown
4_4842347	AT4G08025	Uknown function	Unknown	Extracellular region	Unknown
4_4842347	AT4G08028	Unknown function	Defense response to fungus	Extracellular region	Unknown

Table 5.15 The differential response of AT5G48540 to various abiotic stresses. Experimental details, expression change details, and appropriate references are provided for all experiments. These data were obtained using Genevestigator

Locus code	Key experimental details	Log2-ratio	Fold-Change	P-value	Reference(s)
AT5G48540	Hypoxia stress, root samples	3.74	13.34	< 0.001	Unavailable
AT5G48540	Cold stress, green tissue samples	3.25	7.87	< 0.001	Clauw et al (2015)
AT5G48540	Osmotic stress, green tissue samples	2.89	6.23	< 0.001	Clauw et al (2015)
AT5G48540	Hypoxia stress, root samples	2.87	7.21	< 0.001	Clauw et al (2015)
AT5G48540	Drought stress, whole plant samples	2.63	6.21	< 0.001	Unavailable
AT5G48540	Cold stress, rosette samples	261	6.17	< 0.001	Unavailable
AT5G48540	Hypoxia stress, rosette samples	2.46	5.5	< 0.001	Licausi et al (2011)
AT5G48540	Drought stress, whole plant samples	2.13	4.39	< 0.001	Umezawa et al (2013)
AT5G48540	Drought stress, root samples	1.97	3.92	< 0.001	Clauw et al (2015)
AT5G48540	Cold stress, root samples	1.88	3.59	< 0.001	Clauw et al (2015)
AT5G48540	Freezing stress	1.88	3.7	< 0.001	Zuther et al (2012)
AT5G48540	Drought stress, root samples	1.88	3.67	< 0.001	Pandey et al (2013)
AT5G48540	Wounding stress, green tissue samples	1.82	3.36	< 0.001	Clauw et al (2015)
AT5G48540	Cold stress, rosette samples	1.81	3.53	< 0.001	Unavailable
AT5G48540	Freezing stress	1.8	3.44	< 0.001	Xu et al (2015)

5.3.2.8 Calculated water productivity

As with cWU, the optimal GWAS model for cWP was the initial model. Here, there was a single SNP on chromosome two above the significance threshold (Figs. 5.24, 5.25), whose association to the variation for cWP was confirmed via comparison of means testing

(Fig. 5.26). There are two protein coding genes within the designated threshold of this SNP (Table 5.16), however neither are represented on the Array used on Genevestigator to evaluate gene expression. For this reason, it was not possible to assess the response of these genes to abiotic stress based on previous research. Additionally, there are no published accounts of the function of these genes. Indeed, there is a just a single Gene Ontology biological function pertaining to disease resistance being attributed to AT2G03955. Due to the lack of previous studies into the function and response of these genes to abiotic stress, they represent wholly novel candidates that may be central to elucidating the genetic and molecular basis of variation for water productivity.

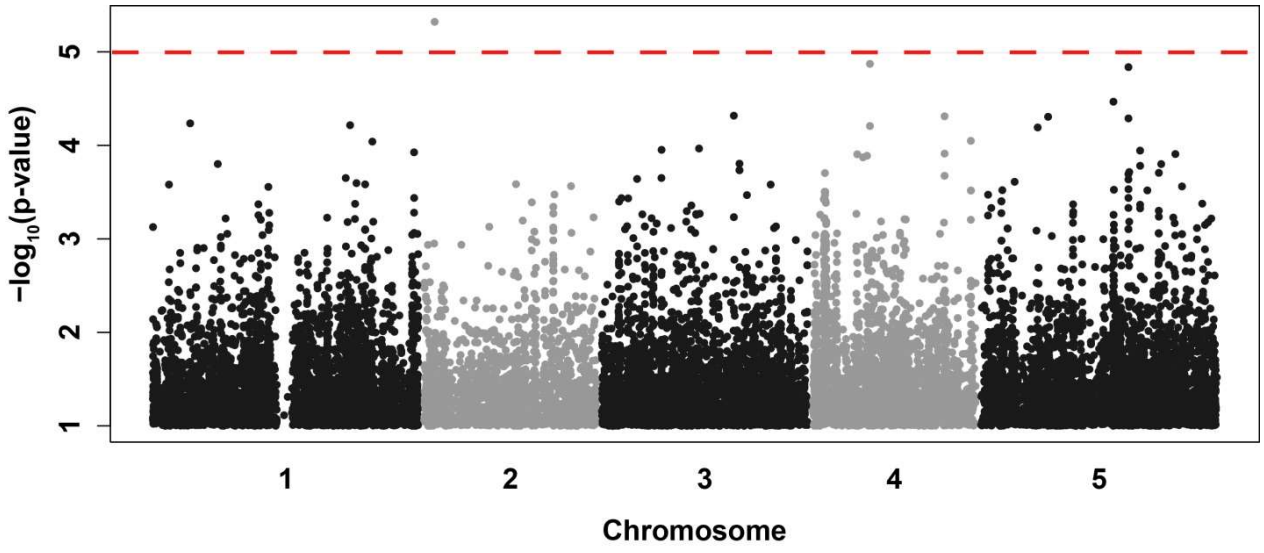

Figure 5.24. Manhattan plot of results for GWA-mapping for calculated water productivity. Figure displayed the $-\log_{10}$(p-values) (y-axis) over genomic positions (x-axis) where SNP markers included in this analysis exist. Successive chromosomes are denoted by the consecutive changes in color from black to grey. The dashed red line indicates the selected *significance* threshold of 5 $-\log_{10}$(p-value).

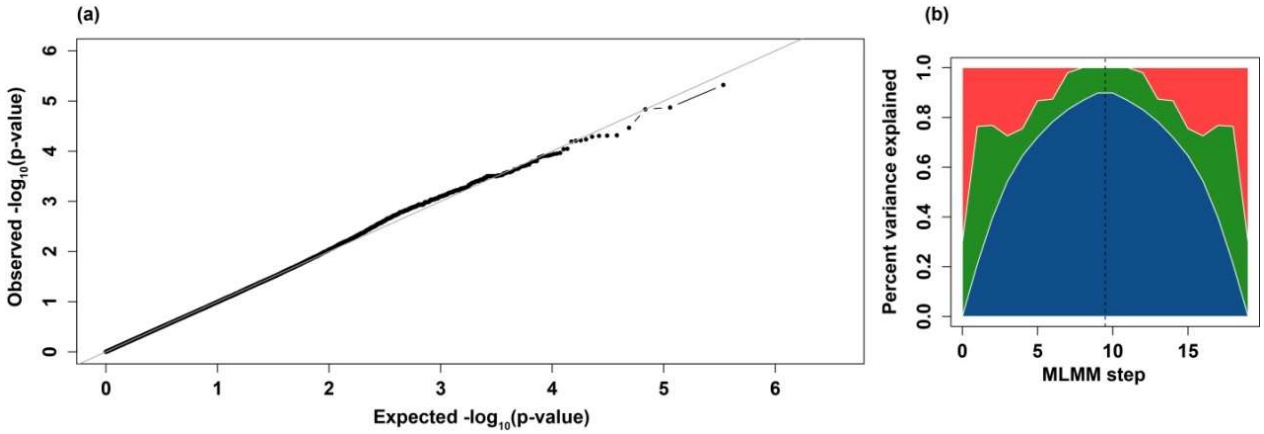

Figure 5.25 Summary statistics for GWA-mapping for calculated water productivity. **(a)** Quantile-quantile plot describing the observed variance in $-\log_{10}$(p-values) compared to the expected distribution of variance for the optimal stepwise model. **(b)** Variance plot describing the evolution in variance with each step of the multi-locus mixed model analysis (Blue: Genetic variance explained by SNP cofactors; Green: Estimated total genetic variance, Red: Error variance).

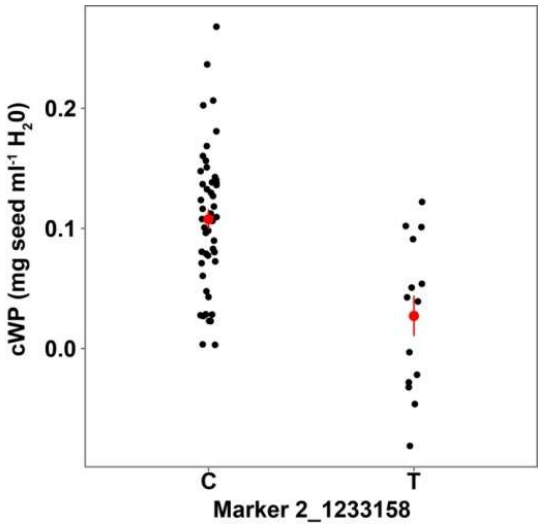

Figure 5.26. Allelic variation for calculated water productivity of the single SNP above the designated threshold from the associated GWAS. Strip plot describing the variation

in calculated water productivity of the C and T alleles of the SNP marker at 1233158bp on chromosomes 2. P-value < 0.0001 as determined by a one-way ANOVA comparison of means test. The red dot indicates the mean value and the extended arms denote the distance of the standard errors above and below the mean.

Table 5.16. Protein coding genes within a 20kb threshold of the single SNP with a $-\log_{10}$(p-value) above a designated threshold of five from the GWAS for calculated water productivity. The associated SNP marker, locus code, other names, and key gene ontology (GO) terms are provided for both genes.

SNP marker	Locus code	Other name(s)	Key GO biological process(es)	GO cellular component(s)	Key GO molecular process(es)
2_1233158	AT2G03937	Unknown function	Unknown	Extracellular region	Unknown
2_1233158	AT2G03955	Unknown function	Defense response to fungus	Extracellular region	Unknown

5.4 Discussion

5.4.1 Phenotypic basis of genetic variation for traits pertaining to water use and reproductive fitness

The genetic variation present between the 117 ecotypes translated into substantial variation for water use, flowering time, and both vegetative and reproductive performance (Fig. 5.1). Arabidopsis has been employed as a model for assessing natural variation for flowering time (Méndez-Vigo *et al.*, 2011; Sasaki *et al.*, 2015; Sanchez-Bermejo & Balasubramanian, 2016) and biomass accumulation (Meyer *et al.*, 2007; Atwell *et al.*, 2010; Sulpice *et al.*, 2013). Indeed this natural variation has been employed to dissect the genetic basis of these traits in Arabidopsis via traditional linkage mapping (Flowering time: Clarke &

Dean (1994) Clarke *et al* (1995); Biomass: Meyer *et al* (2004, 2010)) and also GWAS (Flowering time: The 1001 Genomes Consortium (2016)).

Despite the above, there have been no published accounts of efforts to study the natural variation of long term measures of water use. However, it should be noted that there have been numerous studies centered on understanding the genetic and phenotypic variation for traits such as WUE and drought resistance (Juenger & Mckay, 2005; Hausmann *et al.*, 2005; Masle *et al.*, 2005; Mojica *et al.*, 2016). However, as demonstrated as part of Chapters two (2.3.1) and three (3.3.1), these parameters are not accurate proxies of reproductive performance or long term water use, as such it is important that efforts are made that allow us to begin to understand the variation that exists here and how it relates to reproductive performance.

The variation observed for flowering, biomass accumulation, and water use demonstrated a highly significant genetic basis within this GWAS mapping panel, both in terms of heritability and evolvability (Table 5.3). In general, studies of this nature do not tend to address the idea of evolvability, instead focusing on the idea of heritability (McKay *et al.*, 2003; Lau *et al.*, 2007; Kenney *et al.*, 2014; Easlon *et al.*, 2014). Evolvability is arguably more relevant for traits pertaining to fitness and provides a measure of their selective potential (Pigliucci, 2008). However, the values of H^2 in this study are very similar to values obtained for traits whose genetic basis has been successfully elucidated in previous GWA studies, such as hypocotyl phenotypes important for shade avoidance (Filiault & Maloof, 2012) and germination-related traits (Morrison & Linder, 2014). As such, the moderate heritabilities demonstrated by the traits of interest suggest that a significant proportion of the total variation observed is due to genetic variation and that this variation should be

discernible. The detection of significant block effects between the Col-0 and C24 ecotypes included in every experimental block suggests that there is also substantial variation due to environmental heterogeneity, however this is not uncommon for studies of this nature and was appropriately accounted for via the use of BLUPs and estimated means.

The opposing genetic associations flowering time displays with reproductive and vegetative performance points toward the existence of a resource allocation trade off. It has long been understood that the differential allocation of resources toward reproductive performance as opposed to continued growth and survival is key to the evolution of life history strategies in all organisms, but especially plant species due to their sessile life cycles (Roff & Fairbairn, 2007). Flowering is often considered a defense mechanism, e.g. drought escape (Reviewed in: Kooyers (2015)), by those concerned with crop improvement. However, when considering wild plant species in their natural environments, I would argue that it is probably more pertinent to consider this phenomenon as more of a *last resort* in order to ensure some reproductive success. Indeed ill-timed diversion of resources into reproductive transitioning and fitness has long been suggested as costly from an ecological perspective, given the likelihood of encountering biological or environmental perturbations in the wild (Bell, 1979; Reznik, 1985; Lovett Doust, 1989).

Despite the long standing recognition of the above described trade off, little is actually known with respect to its functional or genetic basis. A relatively recent reciprocal transplant QTL study from Remington *et al* (2013) demonstrated that the allelic variants of QTLs that improve vegetative growth and extend flowering time in *Arabidopsis lyrata* actually reduce reproductive output and vice versa, which is in concurrence with the negative genetic correlation observed between reproductive performance and both flowering time and

vegetative biomass accumulation here (Fig. 5.2). With respect to this point, it is worth remembering that both *Arabidopsis thaliana* and *Arabidopsis lyrata* are undomesticated species. Therefore, in order to persist in the wild, where environmental conditions are heterogeneous and very rarely continually optimal, they have to balance tradeoffs between investment in vegetative growth and defense against environmental perturbations with reproductive transitioning (Lovett Doust, 1989). Domesticated crop species do not have to adopt similar reproductive and resource allocation strategies, since they are almost exclusively bred and cultivated under optimal conditions without concern for economy of water use. As such, they have been bred to delay flowering and senescence in order to boost yields (Blum, 2009). This is unsustainable from a water availability standpoint, but also with respect to nutrient availability, as such it is important that concerted efforts are made toward understanding resource allocation strategies in food and fuel crops.

5.4.2 Candidate genes underlying key water use and productivity related traits

5.4.2.1 Flowering time

The *SSRP1* gene, which is linked to the significant SNP from the flowering time GWAS, is the only candidate gene for which a defined role in floral transitioning has been demonstrated. *SSRP1* has been well characterized as central to the timing of flowering, because it induces the expression of the floral repressor gene *FLC*, thereby negatively regulating flowering time. Expression of *FLC* in mutant *ssrp1* lines has been demonstrated to result in early flowering (Lolas *et al.*, 2010). Five further genes were identified in close proximity to this significant SNP, and of those only *CYP81D11* has previously been

characterized. *CYP81D11* is significantly upregulated in response to jasmonic acid accumulation and is known to be important for defense against herbivory, however neither mutant or overexpressor lines of *CYP81D11* appear to elicit changes in flowering time (Matthes *et al.*, 2010, 2011). The SD period does not elicit an early flowering response, i.e. drought escape, in Arabidopsis (3.3.1). For this reason, it is unlikely that the identification of *SSRP1* is indicative of its importance for drought resistance. However, the significant negative genetic correlation between flowering and seed yield, could suggest that those ecotypes with weaker or non-functional alleles of *SSRP1* could perhaps be higher yielding.

5.4.2.2 Short term water use

As discussed previously, the SD period employed as part of this study does not represent a typical drought stress, as neither flowering time nor reproductive performance appear to be detrimentally affected (3.3.1). However, the mild soil drying that occurs during this period elicits a response in terms of transcriptional changes and the induction of stress signaling (Bechtold *et al.*, 2016), as evident from the physiological responses described in chapter two (2.3.1). To this end, reduced short term water use can be perceived to be associated with drought resistance, specifically drought avoidance, because those ecotypes who use less water in the short term will not succumb to severe drought as quickly as those who use more. It is therefore conceivable that genes implicated in the response to drought stress may also correspond to genes underlying variation for both drought resistance and reduced short term water use.

With respect to the above, two genes identified from the GWAS for short term water use demonstrated highly differential expression in response to abiotic stress. The significant

SNP on chromosome five is in close proximity to *IQ-DOMAIN 12* which encodes a calmodulin binding nuclear protein (Table 5.4). All the genes known to belong to the IQ-Domain family in Maize have recently been demonstrated to be highly responsive to drought (Cai *et al.*, 2015). In Arabidopsis, however, no previous studies have specifically addressed the effect of abiotic stress on genes belonging to this family. However, analysis of *IQ-DOMAIN 12* with Genevestigator revealed that it is substantially downregulated in response to salt stress in root protoplast cells, which suggests that it is central to the response of high levels of salinity, especially since calmodulin proteins are archetypal sensors of Ca^{2+} signatures, which are elicited in response to salt stress. It is well understood that salinity and drought stresses generate similar Ca^{2+} signatures, termed calcium spikes (Reviewed in: Knight & Knight (2001)), which when detected by appropriate sensors initiate conformational changes that facilitate appropriate responses (Yang & Poovaiah, 2003). Salinity and drought initiate common plant defense responses, primarily reductions in the rate of photosynthetic activity (Chaves *et al.*, 2009). To this end, it is reasonable to suggest that *IQ-DOMAIN 12* may be important for detecting calcium spikes that are associated with both mild soil drying and salinity stress. Signaling in this sense may initiate reductions in photosynthetic activities that are related to reduced water use, drought avoidance, and salinity tolerance. *IQ-DOMAIN 12* may not be as differentially expressed under water stress as it is under salt stress, because salinity stress is known to affect gene expression more intensely as it elicits the combined effect of dehydration and osmotic stress also (Chaves *et al.*, 2009). For this reason, its response to drought stress may not have been detected using the most stringent terms on Genevestigator, even though its response to salinity was (Table 5.5).

The previously uncharacterized aspartyl protease gene which is linked to the significant SNP for short term water use on chromosome 3 was observed to demonstrate significant upregulation in root samples in response to drought stress in three separate studies (Tables. 5.5). Aspartic proteases are a subfamily of proteolytic enzymes, of which there are at least 51 within the Arabidopsis genome (Faro & Gal, 2005). However, there have been very few studies into the physiological and biochemical functions they afford in Arabidopsis. With respect to drought stress, aspartic proteases have been demonstrated to be important in the response to drought in buckwheat (Timotijevic *et al.*, 2010), cowpea (de Carvalho *et al.*, 2001), and common bean (Contour-Ansel *et al.*, 2010). Additionally, the study of (Yao *et al.*, 2012) demonstrated that the *ASPARTIC PROTEASE IN GUARD CELL 1* (*ASPG1*) gene in Arabidopsis functions in drought avoidance through ABA signaling in guard cells, where its overexpression enhanced ABA sensitivity in guard cells, thereby reducing water loss. Based on its upregulation in root tissues in response to drought and the apparent importance of the aspartic protease gene family in the response to drought in numerous plant species, the uncharacterized aspartyl protease gene identified here represents an exciting candidate gene offering the potential for reducing short term water use.

5.4.2.3 Rosette biomass accumulation

Four SNPs were significantly associated with the variation for vegetative biomass and 19 protein coding genes were observed to exist within the designated LD threshold. I previously demonstrated that the accumulation of vegetative biomass is not as sensitive to environmental heterogeneity as reproductive performance, i.e. it is a less plastic trait (2.3.1). However, like reproductive performance, growth and the accumulation of vegetative biomass

is highly polygenic (Meyer *et al.*, 2010), with certain genes being of greater importance under certain conditions (Bac-Molenaar *et al.*, 2015; Miller *et al.*, 2015). Consequently, it is difficult to discern which of the 19 gene(s) identified through GWAS for rosette biomass underlie the observed variation. Despite this, the previously uncharacterized glycosyltransferase family gene located on chromosome four does stand out, because of the known influence of plasma membrane located glycosyltransferases for growth, development, and environmental responses (Reviewed in: Perrin (2008). Membrane-bound glycosyltransferases are critical for growth, as they are the predominant synthesizers of cellulose and callose, which are the main cell wall components (Reviewed in: Williams & Davies (2001)). In addition, it was observed that differences in expression between ecotypes of the glycosyltransferase gene during soil drying are largely between Ws-0 and other ecotypes, where Ws-2 demonstrates markedly reduced expression (Table 5.9; Des Marais *et al* (2012)). Interestingly, Ws-2 also demonstrated a much reduced level of biomass accumulation. It is worth noting that Ws-2 also demonstrated reduced cWU and relatively high seed yield, as such the above suggests that manipulation of the previously uncharacterized glycosyltransferase gene identified through GWAS for rosette biomass may facilitate the opportunity to reduce resource allocation to vegetative sinks, whilst concurrently reducing water use and maintaining reproductive performance.

5.4.2.4 Chaff biomass accumulation

The GWAS for chaff biomass identified 14 candidate genes. Chaff biomass is genetically linked to reproductive performance (Fig. 5.3), and many studies have used counts of inflorescence stems or siliques as physical markers of fitness (E.g. Wagner *et al.*, 2011; Akiyama & Ågren, 2012; Dittmar *et al.*, 2014). As a major component of above ground

biomass and a driver of reproductive fitness, chaff biomass accumulation is certain to have a highly polygenic basis, however the genetic basis of chaff biomass *per se* has not been explored to date.

Due to its polygenic nature, it is again somewhat difficult to discern exactly which of the 14 associated protein coding gene(s) underlies the genetic variation explained by the associated GWAS model. One of these genes, *SILIQUES ARE RED* 1 (*SIAR1*), is central to the allocation of organic nitrogen and amino acid biosynthesis in developing siliques. However, its manipulation does not appear to affect the dry weight or number of siliques (Ladwig *et al.*, 2012). For this reason, it is unlikely that allelic variation at *SIAR1* is responsible for the observed variation for chaff biomass.

Of the remaining genes, *ATBAG6* stands out because of its known role in heat tolerance (Nawkar *et al.*, 2016; Table 5.11) and its upregulation in response to drought and various other abiotic stresses (Table 5.11). *ATBAG6* is a member of the BCL-2-ASSOCIATED ATHANOGENE (BAG) gene family. For Arabidopsis, there is limited information available concerning the function of BAG genes, however a number have been investigated with respect to their role in controlling programmed cell death (PCD) in response to cold, heat, UV, and biotic stresses (Williams *et al.*, 2010). Indeed, a recent study focusing especially on *ATBAG6* confirmed this upregulation in relation to heat stress and demonstrated that loss of *ATBAG6* resulted in impairment of basal thermotolerance and increased cell death upon multiple abiotic stresses (Nawkar *et al.*, 2016). Since the availability of water and atmospheric temperature are typically parallel (Lobell & Gourdji, 2012), allelic variation that can facilitate resistance to drought and heat stress is clearly of huge interest. *ATBAG6* facilitates tolerance of multiple stresses through the regulation of

PCD. Furthermore, its potential association to reproductive biomass accumulation offers the potential to combine improved abiotic stress tolerance with improved fitness. Additional incentive is provided to this hypothesis because of the known importance of PCD in facilitating appropriate plant development and growth (Reviewed in: Reape *et al.*, 2008; Reape & McCabe, 2010) which potentially underlies the discovery of *ATBAG6* through GWAS for chaff biomass.

5.4.2.5 Seed yield

Four proteins coding genes were identified arising from the GWAS for seed yield. Two of these genes, *ARABIDOPSIS THALIANA CATION/H+ EXCHANGER 3* (*ATCHX3*) and *ARABIDOPSIS THALIANA CATION/H+ EXCHANGER 9* (*ATCHX9*), belong to a putative Na+/H+ antiporter family and are likely involved in the modulation of intracellular and/or intercellular pH regulation. However, their precise functions have not been explored to date. Of the two remaining genes, one encodes for a protein domain of unknown and function and the other encodes for a ring zinc-finger protein, *CHYR1*, which has been well demonstrated to contribute to drought tolerance (Hsu *et al.*, 2014; Ding *et al.*, 2015).

Hsu *et al* (2014) demonstrated that loss-of-function of *CHYR1* reduced the relative stomatal aperture of Arabidopsis under non-stressful conditions, and that this loss-of-function phenotype could be rescued via the expression of a rice homolog of *CHYR1, OsRZFP34*. This same study also demonstrated that expression of *OsRZFP34* in Arabidopsis increased evaporative cooling under heat stress; however elevated levels of water loss were also observed with ABA treatment, suggesting *OsRZFP34* expression is ABA-insensitive in Arabidopsis. Ding *et al* (2015) built upon this preceding study by

demonstrating that the expression of *CHYR1* specifically is in fact highly responsive and positively regulated by both ABA and drought stress. *CHYR1* was observed to promote ABA-induced stomatal closure, ROS production, and drought tolerance (Ding *et al.*, 2015).

The above describes how *CHYR1* is hugely important for regulating water loss; as such manipulation of *CHYR1* offers the potential to reduce long term water use. At this point, it is difficult to discern whether *CHYR1* is responsible for the variation for seed yield explained by the significant GWAS model – this would require appropriate validation testing of all four genes linked to the associated SNP. However, its association with the regulation of stomatal conductance and therefore photosynthetic activity in general suggests that it may well be important for reproductive biomass accumulation as well as for water use, which is an exciting proposition.

5.4.2.6 Calculated water use and calculated water productivity

The previously uncharacterized receptor like kinase (AT5G48540) stands out as a particularly probable candidate gene underlying the associated variation for cWU. This gene was observed to be significantly upregulated in response to oxygen deprivation of the roots and drought in multiple studies (Table 5.15). A known response of root hypoxia is the increased concentration of ABA in leaf tissues (Bradford & Yang, 1980; Kozlowski, 1997; Gil *et al.*, 2009), which plays a role in root-to-shoot signaling inducing stomatal closure in waterlogged plants (Else *et al.*, 1995; Gil *et al.*, 2009), similar to the well characterized drought response mechanism (Reviewed in: Jia & Zhang, 2008). Consequently, the protein kinase in question may be implicated in the response to both drought and hypoxia, where it mediates a reduction in the rate of stomatal conductance via a common root-to-shoot ABA

signaling pathway. This candidate gene may well be important for drought tolerance, however in situations where water availability is reduced, it may also be important for reducing long term water use. For this reason, it is plausible that this gene is directly responsible for a substantial proportion of the variation observed for cWU.

Both of the candidate genes associated with cWP are uncharacterized DEFENSIN-LIKE (*DEFL*) family protein coding genes. There is some evidence that DEFL genes are important in the response to drought and other abiotic stresses (Yamada *et al.*, 1997; Maitra & Cushman, 1998), thus it is reasonable to predict that these genes could play a role in reducing water use through drought tolerance mechanisms. The exact function of these genes in response to drought or other stresses is not presently understood. Furthermore, there are no published accounts of the role of DEFL genes with respect to reproductive performance, or biomass accumulation. As such, it is difficult to predict whether these two genes contribute to the natural variation observed for water productivity without performing reverse genetics-based validation studies.

6. General discussion

6.1 Project background, aims, and approaches

Traditionally, efforts to understand the impact of water availability and water use on plant performance have focused on the concepts of drought resistance and WUE respectively (Blum, 2005; Morison *et al.*, 2008). During the past two decades, the natural variation present within Arabidopsis has been harnessed to further our understanding of the physiological and genetic basis of both drought resistance and WUE (Reviewed in: Juenger, 2013; Kooyers, 2015). During this time period, substantial criticism of drought resistance and WUE has arisen because of the lack of evidence supporting their association to reproductive performance (Passioura, 2004; Condon *et al.*, 2004; Blum, 2005, 2009; Sinclair & Purcell, 2005; Morison *et al.*, 2008). This present study was performed in order to elucidate the physiological and genetic basis of water use and water productivity as alternatives to drought resistance and WUE. Dissecting the bases of these traits is of considerable interest since it will facilitate a systematic understanding of the mechanisms through which water use can be reduced without inferring a reduction in harvestable yield.

An assessment of the natural variation of multiple phenological and physiological traits relating to water use and fitness was performed as a precursor to genetic mapping. A maximum of 46 ecotypes were subjected to specifically designed SD and CW experiments that enabled both the calculation and measurement of water use and water productivity. Water use and productivity were compared to important life history traits as well as measures of WUE, namely iWUE and $\delta^{13}C$. This allowed for the determination of traits that are

genetically linked to water use and productivity, as well as those that are in evolutionary constraint of these target traits.

The ecotypes Col-0 and C24 have previously been observed to demonstrate substantial difference in water use and water productivity (Bechtold *et al.*, 2010). For this reason, the RIL population derived from these ecotypes (Törjék *et al.*, 2006) was employed for QTL mapping of these traits. 146 RILs were subjected to an SD experimental period and phenotypically characterized for flowering, biomass accumulation, and water use. To complement the QTL mapping, a GWAS was performed based on 117 distinct ecotypes for the same traits, again using an SD experimental period. Both the QTL and GWAS projects allowed for further exploration of the general trends observed as part of the original scan of Arabidopsis natural variation and for the elucidation of the genetic factors underlying water use and productivity.

6.2 Key Findings

- Seed yield in Arabidopsis is more sensitive to environmental heterogeneity than water use, vegetative biomass accumulation, and flowering time. Elevated sensitivity is evident from the greater variation in plasticity for seed yield and from the estimates of heritability (reduced) and evolvability (elevated).
- Flowering time is genetically coupled to vegetative biomass accumulation and water use, where plants that flower later use more water in both the short and the long term and also accumulate far more vegetative biomass.
- Seed yield is in evolutionary constraint of flowering time and vegetative biomass accumulation, in a relationship that is symptomatic of a resource allocation trade-off.

Plants that flower earlier appear to have improved reproductive productivity and use less water.

- It is possible to accurately determine long term water use and water productivity of Arabidopsis by multiplying short term water use, i.e. the rate of soil drying, by flowering time.
- QTL mapping determined that allelic variation at *FRI* and *FLC* is responsible for the majority of the variation that exists for water use in the Col-0 x C24 RIL population. A combination of non-functional alleles of both genes significantly reduces water use by reducing the flowering time of Arabidopsis. Crucially, a reduction in water use achieved in this manner does not impair reproductive performance.
- GWAS identified multiple candidate genes that potentially underlie the variation observed between 117 ecotypes for multiple traits. The identification of the known floral transitioning gene *SSRP1* provides validation that pertinent phenotyping and analyses were performed. It also represents an avenue through which it may be possible to reduce flowering time, and consequently water use. Further promising candidate genes include *IQ-DOMAIN 12* (short term water use), *ATBAG6* (reproductive biomass accumulation), and *CHYR1* (reproductive biomass accumulation).

6.3 Findings in relation to previous work

6.3.1 The physiological basis of improved plant performance under the context of water use

6.3.1.1 Drought Resistance

Pioneering efforts to understand plant performance under reduced water availability entailed studying species or genotypes that persist in habitually dry environments. This persistence is a result of natural selection driving the evolution of drought resistance mechanisms inherent to said species or genotypes (Levit, 1972). Studies of this nature have resulted in a very broad understanding of the physiological and phenological mechanisms that confer drought resistance (Reviewed in: Kooyers, 2015).

The assessment of short term water use as described in this present study is to some extent parallel to the idea of drought avoidance, which is the predominant component of drought resistance (Levit, 1972; Blum, 2005). It is analogous in the sense that a plant that uses less water in the short term will avoid succumbing to severe drought stress as quickly as a plant that uses more water. With respect to the ecotype-based studies described here, short term water use appeared to be predominantly determined by the size of the plant, where larger plants demonstrated increased short term water use (Fig. 2.4). Plants with larger overall leaf surface areas tend to have a greater total number of stomata (Carins Murphy *et al.*, 2014), as such it is likely that the association between size and short term water use is a result of the increased opportunity for water loss in larger plants. Within the Arabidopsis community there has been a surprising lack of research into the effect of plant

size on water use or drought resistance, however a relatively recent *Arabidopsis lyrata* based study from (Paccard *et al.*, 2014) demonstrated a very similar trend to that described here between plant size and short term water use.

6.3.1.2 Water use efficiency

Today there is substantial evidence to suggest that studies aimed at interpreting drought resistance and drought escape mechanisms are impractical because they are often linked to low productivity (Condon *et al.*, 2004; Sinclair & Purcell, 2005; Morison *et al.*, 2008; Bechtold *et al.*, 2010). For this reason, many researchers now focus on the idea of WUE in order to assess plant performance under drought stress or to minimize water inputs (Reviewed in: Lawson & Blatt, 2014; Vadez *et al.*, 2014). WUE is a preferential assessment here because it provides a measure of the maintenance of photosynthesis as a ratio of the stomatal regulation of transpiration (water loss).

The assessments of WUE, measured as both iWUE and $\delta^{13}C$, used in this study provide substantial evidence to suggest that alterations to WUE are predominantly achieved via changes to stomatal conductance and transpiration, as opposed to variation in the rate of photosynthetic assimilation (Figs. 2.5, 2.6, 2.7, 3.7). This is in concurrence with previous work in Arabidopsis (Easlon *et al.*, 2014) and also in cereal species (Reviewed in: (Condon *et al.*, 2004; Blum, 2009). In general, I observed that iWUE and $\delta^{13}C$ shared neutral relationships with vegetative biomass accumulation (Fig. 2.4, Section 3.3.3). Additionally, there was no link between these proxies and either measured or calculated WUE, i.e. biomass accumulated as a ratio of water used over the course of the lifetime of the plant (Figs. 2.4, 4.3, 5.2). To the best of my knowledge, no published work has demonstrated a

significant link between either iWUE or $\delta^{13}C$ and biomass in Arabidopsis. Indeed, Kenney *et al* (2014) describes how the relationship between $\delta^{13}C$ and biomass production is verging on negative, suggesting that ecotypes that demonstrate lower $\delta^{13}C$ (higher WUE) under drought stress actually have reduced biomass production. As such, my observations with respect to WUE and biomass accumulation are in line with previous work (Kenney *et al.*, 2014) and point toward the ineffective nature of the different ways employing and interpreting WUE in Arabidopsis. Despite this, the link between these physiological parameters is somewhat different in many cereal species, where numerous studies have in fact detected a strong positive relationship between WUE and total above-ground biomass, however the link with yield is much more commonly neutral (Reviewed in: Condon *et al* 2004).

The lack of an observed relationship between $\delta^{13}C$ and biomass production in Arabidopsis is likely due its non-domesticated nature. Domesticated crops have consistently been bred to improve yield without consideration for water lost through transpiration, as such they have been bred to divert all photosynthetically acquired resources toward growth and productivity (Lawson *et al.*, 2012). It is therefore unsurprising that a link between $\delta^{13}C$ and biomass production does occasionally exist in some crop species (Condon *et al.*, 2004; Morison *et al.*, 2008). Conversely, Arabidopsis ecotypes have not undergone selection that mirrors this artificial breeding. As such, it is unlikely that a link exists between $\delta^{13}C$ and biomass production in Arabidopsis, since those ecotypes that have reduced $\delta^{13}C$ (improved WUE) are not certain to divert the additional photosynthates they acquire toward growth, unlike crops that have been bred to do so. These particular ecotypes are perhaps more likely to divert these photosynthates toward abiotic stress defense mechanisms that are associated with reduced water availability and drought.

6.3.1.3 Water use and productivity

Very few studies have previously attempted to measure lifetime water use in Arabidopsis or any plant species (Bechtold *et al.*, 2010, 2013). Furthermore, these studies have assessed water use and water productivity in only a few ecotypes or transgenics. As such, this present study allows for informed assertions toward the physiological and genetic basis of improved water use and water productivity in Arabidopsis.

One of the fundamental findings of this study is the development of proxy parameters that accurately predict water use and water productivity. This was achieved through multiplying short term water use by the point of bud initiation, i.e. flowering time. The accurate nature of the calculated measure of long term water use was demonstrated by comparing cWU and mWU of the ecotypes that constituted both the CW experiment (Fig. 2.10a), but also based on these same measures of the ecotypes that constituted the parallel SD experiment (Fig. 2.10b). The ability to calculate long term water use and water productivity was essential for this type of study, since manually measuring these parameters for the entire QTL and GWAS mapping panels would have been unfeasible in the given time.

Despite the observation that both iWUE and $\delta^{13}C$ appeared to be primarily determined by stomatal conductance and thus water lost through transpiration (Figs. 2.5, 2.6, 3.7), there was no association between iWUE or $\delta^{13}C$ and mWU or cWU. Furthermore, stomatal conductance and transpiration did not appear to be accurate predictors of water use either (Fig. 2.4). This suggests that instantaneous measures of stomatal conductance, transpiration, and iWUE are neither representative of the whole plant nor do they provide an accurate measure of these parameters over the lifetime of the plant. This points towards the

need for the adoption of whole plant gas exchange, as demonstrated by Easlon *et al* (2014), performed on a continuous time scale. This would allow for the determination as to whether lifetime rates of conductance, transpiration, or iWUE do in fact determine long term water use. Likewise, this could also be less intrusively achieved via imaging based assessments of WUE which allow for dynamic assessments of these parameters (McAusland *et al.*, 2013).

There are no previous examples relating to the assessment of $\delta^{13}C$ as a physical marker of water use. The neutral association between $\delta^{13}C$ and water use is initially harder to discern than those described above, however it is possible that it is related to the time of harvest of the tissue samples. $\delta^{13}C$ is often described as an integrated measure of WUE, providing a measure of WUE that is more representative of actual WUE, i.e. unit of biomass gained per unit of water lost, than iWUE (Araus *et al.*, 2002). However, the tissue harvested to assess $\delta^{13}C$ is often harvested before flowering and seed setting (Condon *et al.*, 1993, 2004; Rebetzke *et al.*, 2002; Hausmann *et al.*, 2005; Masle *et al.*, 2005; Kenney *et al.*, 2014), as is the case with this study. For this reason, if $\delta^{13}C$ does indeed provide an integrated measure of WUE, it will only be a measure that is relevant up to the point of tissue harvesting, furthermore it will likely only be relevant for the tissue type analyzed. Water use is hugely dependent on flowering time (Fig. 2.14); as such it is not wholly surprising that $\delta^{13}C$ here doesn't directly relate to water use, since the associated tissues samples were harvested within one-to-three days of each other and before flowering, which was vastly different between these ecotypes.

With respect to the above, cereal based studies have previously described how the relationship between leaf-$\delta^{13}C$ and grain-$\delta^{13}C$ is very weak (Condon *et al.*, 1993; Merah *et al.*, 2001), which suggests that $\delta^{13}C$ measures are only relevant for the tissue they are

harvested from. This proposition can only really be supported by comparing leaf-$\delta^{13}C$ measures attained from tissue harvested along a temporal scale. As an aside, it is also worth noting that $\delta^{13}C$ estimated from grain is often far more tightly linked to yield in a positive sense than estimates from leaf tissue (Voltas *et al.*, 1999; Merah *et al.*, 2001; Royo *et al.*, 2002; Araus *et al.*, 2002), which may also go some way to explaining the lack of association between $\delta^{13}C$ and yield observed here in Arabidopsis (Section 3.3.3).

In general, water use appeared to be predominantly determined by flowering time and vegetative biomass (Figs 2.4, 2.14). Also, it is worth noting that lines that flowered later accumulated significantly more vegetative biomass (Figs. 2.4, 4.3, 5.2), but less (Ecotypes; 2.4, 5.2) or insignificantly different (RILs; Fig. 4.3) reproductive biomass. It is assumed that a plant that flowers later and uses more water has a reduced WUE, unless the excess biomass it accumulates overcompensates for the increased volume of water used. This *overcompensation* scenario did not hold true for this study (Section 3.3.3), as such it is reasonably safe to make this assumption. Herein lies the fundamental discrepancy that exists between this present study and the extensive body of work that has centered on understanding the natural $\delta^{13}C$/WUE variation in Arabidopsis (E.g. (McKay *et al.*, 2003, 2008; Juenger & Mckay, 2005; Hausmann *et al.*, 2005; Kenney *et al.*, 2014). These studies consistently report a positive genetic correlation between flowering time and WUE, measured as $\delta^{13}C$, under drought stress. This would suggest that plants with a greater lifespan have a higher WUE. Based on my consistent observation that water use is driven by flowering time (Figs 2.4, 2.14), I would argue that a positive link between flowering time and WUE is unlikely. I believe this discrepancy is a fundamentally due to $\delta^{13}C$ not providing a truly integrated or predictive measure of WUE in Arabidopsis.

6.3.2 Genetic basis of water use and productivity related traits

In Arabidopsis, multiple genetic factors have been identified that contribute to improved drought resistance through multiple means. In general, these genetic loci are implicated in stomatal signaling (Vahisalu *et al.*, 2008; Brandt *et al.*, 2012; Kanno *et al.*, 2012; Behnam *et al.*, 2013), early stress response and transduction pathways (Yamaguchi-Shinozaki & Shinozaki, 2006; Utsumi *et al.*, 2012), and in protecting photosynthesis, i.e. improving WUE (Rossel *et al.*, 2006, 2007; Moulin *et al.*, 2008; Lee *et al.*, 2009). Manipulation of these genes may well serve to improve water use and/or water productivity, especially in the case of genes that reduce water loss or enhance the capture of soil moisture (Blum, 2005).

The numerous studies that have explored the natural variation for $\delta^{13}C$ in Arabidopsis have demonstrated trends with flowering (McKay *et al.*, 2003, 2008; Juenger & Mckay, 2005; Hausmann *et al.*, 2005; Kenney *et al.*, 2014). Consequently, when these studies have also employed linkage disequilibrium mapping, QTLs that colocalise with flowering time genes have often been elucidated. The most pertinent example of this is that of McKay *et al* (2003) who demonstrated that *FRI* and *FLC* are tightly linked to both flowering time and $\delta^{13}C$. Using NILs, McKay *et al* (2003) demonstrated that the introgression of a functional allele of *FRI* into a genotype with a functional *FLC* allele delays flowering time. This is entirely as expected, and is mirrored by my observations based on grouping Col-0 x C24 RILs into allelic forms of these flowering time genes (Fig. 4.5). Concurrently with their observation regarding the positive association between flowering time and $\delta^{13}C$, McKay *et al* (2003) also

observed that those NILs with functional *FRI* and *FLC* alleles also demonstrated elevated $\delta^{13}C$, suggesting that they have improved WUE. This is contrary to my observation that combining functional alleles of these genes increases water use, due to an increased lifespan, i.e. delayed flowering. Furthermore, functional alleles of these genes have reduced reproductive biomass, as they appear to allocate additionally acquired photosynthetic resources to vegetative biomass sinks (Fig. 4.3, 4.5).

Further contradiction is found with respect to the above in the study of Hausmann *et al* (2005) who used a RIL mapping population derived from the ecotypes Col and Ler to identify QTL related to $\delta^{13}C$. This particular study also observed a positive genetic correlation between flowering time and $\delta^{13}C$. Additionally, Hausmann *et al* (2005) described how those plants with the greatest WUE ($\delta^{13}C$) not only flowered later, but they were also smaller and produced fewer branches and siliques. This is again in complete contradiction to my observations in the sense that flowering time was also tightly linked to vegetative biomass accumulation. Therefore, plants that flowered later produced more total above ground biomass, so were larger than lines that flowered earlier (Figs. 4.3, 5.2.). Additionally, it is worth remembering that flowering was also negatively linked to reproductive performance in the ecotypes (Figs. 5.2).

Based on these findings, it can be argued that it is important to measure actual water use and productivity, since proxies do not appear to be accurate predictors of these traits. As a follow on to this, GWAS was employed in order to identify wholly novel genes that may contribute to the variation observed for these traits in Arabidopsis. The promising candidate genes illuminated through GWAS are reviewed in brief below.

The known flowering time gene, *SSRP1,* was identified to be responsible for a substantial degree of the variation for flowering time between the GWAS ecotypes (Fig. 5.9., 5.10.). *SSRP1* is a key regulator of the expression of *FLC*, as it is a component of the FACT transcription factor complex (Lolas *et al.*, 2010; Van Lijsebettens & Grasser, 2010). Identification of *SSRP1* through GWAS for flowering time offered encouragement that apposite phenotyping and statistical mapping had been performed. Additionally, *SSRP1* may offer an avenue through which time to flowering may be reduced, thereby potentially minimizing water use.

Three promising candidate genes were identified following GWAS for water use. Two of these genes were associated to variation for short term water use, *IQ DOMAIN 12* and a previously uncharacterized aspartic protease gene (Table 5.4). Based on their response to water availability and because of the known function of their respective gene families (Table 5.5, Yao *et al.*, 2012; Cai *et al.*, 2015), both represent targets through which short term water use and drought resistance may be improved. Additionally, the uncharacterized receptor like kinase protein identified through GWAS for cWU (Table 5.14) also offers encouragement as a candidate through which water use may be reduced due to its upregulation in response to both oxygen deprivation and drought stress in root samples. As such, it may be important for conferring root-to-shoot ABA signaling. Improving ABA signaling may represent an additional means through which water use could be reduced given appropriate conditions, since this typically induces stomatal closure (Osakabe *et al.*, 2014), thereby reducing water use.

GWAS for cWP revealed two associated genes, both of which were DEFL-like proteins according to gene ontology (Table 5.16). There is no previously documented

account of DEFL-like proteins contributing to drought resistance, water use, or productivity. Additionally, neither of these genes appear to be differentially expressed in response to abiotic stress. Despite this, two promising candidate genes arose from the GWAS for reproductive biomass accumulation as a unique trait, i.e. without consideration for water use. The *ATBAG6* gene is potentially the causal gene underlying a significant proportion of the variation for chaff biomass (Fig. 5.16, Table 5.10). It is perhaps more likely to be the casual gene associated to the SNP of interest than the other nearby genes due its role in programmed cell death (Williams *et al.*, 2010), which is known to be crucial for plant growth (Reviewed in: Reape & McCabe, 2010). Moreover, *ATBAG6* is also involved in resistance to drought stress (Nawkar *et al.*, 2016), so may offer the potential to combine elevated reproductive performance with improved drought resistance and/or reduced water use. The *CHYR1* gene was identified following GWAS for seed yield (Table 5.13). It has been implicated in the control of water use through regulating ABA-induced stomatal closure in both rice (Hsu *et al.*, 2014) and Arabidopsis (Ding *et al.*, 2015). As such, its identification through GWAS for seed yield and its known biological functionality points toward a route through which reduced water use and improved productivity may be combined to increase water productivity.

6.4 Future work

In context of the extensive assessment of natural variation for $\delta^{13}C$ in Arabidopsis, the general trends reported by my research suggest that commonly utilized proxies for WUE and productivity are inefficient for predicting the amount of reproductive biomass produced, either as a unique trait or as a factor of the volume of water used by the plant, i.e. water productivity. I would argue that this is especially true of $\delta^{13}C$ and iWUE, of which neither

appeared related to water use or reproductive performance. The garden experiment performed as part of Chapter three confirmed that those ecotypes that perform best in terms of reproductive performance in controlled conditions, also do so in outdoor conditions (Fig. 3.16). Additionally, it provided further support to suggest that flowering is linked to vegetative biomass accumulation and that larger plants produce less reproductive product (Figs. 3.14, 3.15). However, neither $\delta^{13}C$ nor iWUE were assessed as part of this study. As such, it would be interesting to repeat such an experiment whilst also assessing these proxy measures to understand how they relate to biomass accumulation in a more *natural/agronomic* situation.

It is possible that one of the fundamental reasons why $\delta^{13}C$ does not appear to be an accurate proxy of WUE or biomass-accumulation in Arabidopsis is because it only provides an integrated measure up to the point at which tissue is harvested. For this reason and because of its extensive application, I think it would be appropriate to conduct a proof-of-concept study, where $\delta^{13}C$ is assessed along a temporal scale at various different developmental stages and based on vegetative and reproductive tissue. This would allow for a true representation of how $\delta^{13}C$ relates to water use, actual WUE, and water productivity.

The QTL detected on chromosomes four and five underlying flowering time, water use, and vegetative biomass accumulation were validated as *FRI* and *FLC* respectively (Section 4.5). Two additional QTL were detected on chromosomes one and two, both underlying variation for short term water use (Fig. 4.4). The first of these, STWU:1, colocalises with *GI*, as such before proceeding with fine mapping, it would be prudent to initially test whether *GI* is indeed the gene underlying STWU. Since there are no obvious genes colocalising with the second short term water use QTL, STWU:3, fine mapping is

required to delineate the precise genetic basis of this QTL. This would be most efficiently achieved through saturating these regions with further SNP markers and re-performing QTL analysis. The required molecular markers would be easy to design, since the genomic sequences of Col-0 and C24 are publicly available (The 1001 Genomes Consortium, 2016). Identification of genes underlying these short-term water use QTLs would represent additional targets through which water use may be manipulated in Arabidopsis.

Since *FRI* and *FLC* have been confirmed to control water use in the mapping population, it is necessary to obtain NILs to confirm their importance. All RILs were grouped according to whether they harbored Col-0 or C24 alleles of both *FRI* and *FLC*. Those RILs possessing functional alleles of both demonstrated the greatest water use and reduced reproductive performance (Fig. 4.5). This is in accordance with my hypothesis regarding flowering time, water use, and reproductive fitness. Despite this, testing NILs which have the different alleles of both genes introgressed into both genomic backgrounds is required to fully support this hypothesis. Indeed, these NILs have already been developed (Törjék *et al.*, 2006).

A further natural progression of the QTL mapping-based research is to translate these findings into a crop species. To this end it would be interesting, and potentially of agronomic significance, to investigate the effect of homologs of *FRI* and *FLC* in *Brassica napus*, rapeseed, which is closely related to Arabidopsis. *B. napus* is of great economic importance to the global agricultural sector, as such improving harvestable yield whilst reducing irrigational outlays is of considerable interest (White *et al.*, 2015; Zhu *et al.*, 2016). Multiple homologous of *FRI* and *FLC* have previously been identified in *B. napus* and the functionality of many of them is conserved. Moreover, associated allelic variation has been demonstrated

to control flowering time variation (Wang *et al.*, 2011; Zou *et al.*, 2012). Initially, it would be interesting to test the water use and productivity of multiple accessions of *B. napus* that harbor different alleles of *BnFRIs* and *BnFLCs*. Additionally, it is worth noting that mapping populations have been developed and successfully employed to map flowering time QTL, including *FLC* (Long *et al.*, 2007) and *FRI* (Wang *et al.*, 2011) homologs. As such it may be possible to take a similar NIL-based approach to that previously described for Arabidopsis in order to understand how combining *BnFRI and BnFLC* alleles of different functionality effects water use and productivity in *B. napus*.

Using the RIL population developed from the Col-0 and C24 ecotypes, QTL mapping was inefficient at delineating the heritable genetic basis of seed yield or water productivity. This could potentially be addressed through incorporating more RILs into the study to try and capture more allelic variation. Alternatively, it may be more efficaciously achieved through the development of a more appropriate mapping population. For example, a population developed using the very high yielding ecotype DraIV 2-6 with the moderately yielding variety DraIV 4-2 may be more fruitful in terms of identify QTL associated with yield. Furthermore, these two ecotypes would be particularly interesting to use as parental lines since they are from the same area of Czech Republic and have thus likely been subjected to very similar environmental perturbations during their evolutionary history. For this reason, the elevated productivity demonstrated by DraIV 2-6 is likely to have a distinct genetic basis, as opposed to owing to allelic variation at loci that confer tolerance to abiotic stress. Equally, the ecotypes C24 (high yield/cWP) and Se-0 (low yield/cWP) could be appropriate for this purpose since they both arise from arid Iberian environments and have similar flowering times.

GWAS identified 62 protein coding genes associated to flowering time, water use, and biomass accumulation. Apart from *SSRP1*, which is highly likely to be responsible for the associated significant SNP for flowering time, the presented elucidation of which genes are responsible for the associated trait variation is based on their response to abiotic stress and on previously characterized gene-specific and/or family-specific functionality. In order to truly validate which genes are responsible for underlying trait variation, it is necessary to test mutant and overexpressor lines to validate the phenotypic consequence of repressing and enhancing the expression of these genes respectively. As with the *FRI* and *FLC* NILs, it would be sensible to test the effect of such lines via continuous watering experiments, to precisely measure water use and water productivity.

Comparing the expression of all of the candidate genes under drought and control conditions between multiple ecotypes would be an extremely useful precursor to the above-outlined gene validation studies. Differential expression between the ecotypes that mirrors the performance of the same ecotypes with respect to the trait of interest, for example water use, would provide substantial evidence that certain genes are involved in variation of that trait. This would be of great biological interest in of itself, but it would also help to narrow down the candidate gene list. Additionally, comparison of expression differences between these ecotypes in controlled environments and outdoor conditions would help to illuminate whether genes that are apparently important in controlled conditions maintain their effect and importance in field-like conditions.

6.5 Summary

In summary, this present research has described novel means through which to assess plant performance under the context of water availability and has confirmed the usefulness of these means in relation to more commonly employed assessments of performance, primarily WUE. Furthermore, genetic mapping was employed in order to understand the genetic basis of traits relating to water use and productivity. Arabidopsis represents the ideal study system for research of this nature. The ready availability of plant material that represents genetic variation in the form of ecotypes and mapping populations facilitates the opportunity to perform phenotypic scans of natural variation in order to address questions relating to the phenological and developmental basis of improved performance and how proxies do, or do not, relate to performance. Additionally, public access to appropriate genotypic information allows for follow up genetic dissection efforts as demonstrated in Chapters four and five. The short generation time, small size, and manipulability of Arabidopsis also contributes to the feasibility of studies of this nature, especially given time and space constraints. As previously described, the natural progression of this present study will involve reverse genetics to understand the importance of candidate genes for water use and productivity and population genetic analyses in order to ascertain the ecological and evolutionary significance of allelic variation at these genes. Again, Arabidopsis is perfectly suited for research to this end thanks the public availability of mutant lines and due to the resources and tools tied to the 10001 genomes project that allow for population genetic analyses (The 1001 Genomes Consortium, 2016).

7. References

Adiredjo AL, Navaud O, Munos S, Langlade NB, Lamaze T, Grieu P. **2014**. Genetic control of water use efficiency and leaf carbon isotope discrimination in sunflower (Helianthus annuus L.) subjected to two drought scenarios. *PLoS ONE* **9**.

Agren J, Oakley CG, McKay JK, Lovell JT, Schemske DW. **2013**. Genetic mapping of adaptation reveals fitness tradeoffs in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **110**: 21077–21082.

Agren J, Schemske DW. **2012**. Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. *New Phytologist* **194**: 1112–1122.

Akiyama R, Ågren J. **2012**. Magnitude and timing of leaf damage affect seed production in a natural population of Arabidopsis thaliana (Brassicaceae). *PLoS ONE* **7**.

Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M. **2009**. What has natural variation taught us about plant development, physiology, and adaptation? *The Plant Cell* **21**: 1877–96.

Alsdurf JD, Ripley TJ, Matzner SL, Siemens DH. **2013**. Drought-induced trans-generational tradeoff between stress tolerance and defence: Consequences for range limits? *AoB PLANTS* **5**: 1–12.

Anithakumari AM, Nataraja KN, Visser RGF, van der Linden CG. **2012**. Genetic dissection of drought tolerance and recovery potential by quantitative trait locus mapping of a diploid potato population. *Molecular Breeding* **30**: 1413–1429.

Aprile A, Havlickova L, Panna R, Marè C, Borrelli GM, Marone D, Perrotta C, Rampino P, De Bellis L, Curn V, *et al.* 2013. Different stress responsive strategies to drought and heat in two durum wheat cultivars with contrasting water use efficiency. *BMC Genomics* **14**: 1–18.

Araus JL, Slafer GA, Reynolds MP, Royo C. **2002**. Plant Breeding and Drought in C3 Cereals: What Should We Breed For? *Annals of Botany* **89**: 925–940.

Arends D, Prins P, Jansen RC, Broman KW. **2010**. R/qtl: High-throughput multiple QTL mapping. *Bioinformatics* **26**: 2990–2992.

Arms EM, Bloom AJ, St Clair DA. **2015**. High-resolution mapping of a major effect QTL from wild tomato Solanum habrochaites that influences water relations under root chilling. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **128**: 1713–1724.

Asseng S, Ritchie JT, Smucker AJM, Robertson MJ. **1998**. Root growth and water uptake during water deficit and recovering in wheat. *Plant and Soil* **201**: 265–273.

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, *et al.* 2010. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* **465**: 627–631.

Bac-Molenaar J a., Granier C, Keurentjes JJB, Vreugdenhil D. **2015**. Genome wide association mapping of time-dependent growth responses to moderate drought stress in Arabidopsis. *Plant, Cell & Environment*: 88–102.

Banta JA, Ehrenreich IM, Gerard S, Chou L, Wilczek A, Schmitt J, Kover PX, Purugganan MD. **2012**. Climate envelope modelling reveals intraspecific relationships among flowering phenology, niche breadth and potential range size in Arabidopsis thaliana. *Ecology Letters* **15**: 769–777.

Bates D, Maechler M. **2009**. lme4: Linear mixed-effects models using s4.

Bechtold U, Albihlal WS, Lawson T, Fryer MJ, Sparrow PAC, Richard F, Persad R, Bowden L, Hickman R, Martin C, *et al.* 2013. Arabidopsis HEAT SHOCK TRANSCRIPTION FACTORA1b overexpression enhances water productivity, resistance to drought, and infection. *Journal of Experimental Botany* **64**: 3467–3481.

Bechtold U, Lawson T, Mejia-Carranza J, Meyer RC, Brown IR, Altmann T, Ton J, Mullineaux PM. **2010**. Constitutive salicylic acid defences do not compromise seed yield, drought tolerance and water productivity in the Arabidopsis accession C24. *Plant, cell & environment* **33**: 1959–73.

Bechtold U, Penfold CA, Jenkins DJ, Legaie R, Moore JD, Lawson T, Matthews JSA, Vialet-Chabrand SRM, Baxter L, Subramaniam S, *et al.* 2016. Time-Series Transcriptomics Reveals That AGAMOUS-LIKE22 Affects Primary Metabolism and Developmental Processes in Drought-Stressed Arabidopsis. *The Plant cell* **28**: 345–66.

Behnam B, Iuchi S, Fujita M, Fujita Y, Takasaki H, Osakabe Y, Yamaguchi-Shinozaki K, Kobayashi M, Shinozaki K. **2013**. Characterization of the promoter region of an arabidopsis gene for 9-cis-epoxycarotenoid dioxygenase involved in dehydration-inducible transcription. *DNA Research* **20**: 315–324.

Bell G. **1979**. The Costs of Reproduction and their Consequences. *The American Naturalist* **116**: 45–76.

Bennett D, Reynolds M, Mullan D, Izanloo A, Kuchel H, Langridge P, Schnurbusch T. **2012**. Detection of two major grain yield QTL in bread wheat (Triticum aestivum L.) under heat, drought and high yield potential environments. *Theoretical and Applied Genetics* **125**: 1473–1485.

Bernacchi CJ, Kimball B a, Quarles DR, Long SP, Ort DR. **2007**. Decreases in Stomatal Conductance of Soybean Under Open-Air Elevation of [Co2] Are Closely Coupled With Decreases in Ecosystem Evapotranspiration. *Plant Physiology* **143**: 134–144.

Bevan M, Walsh S. **2004**. Positioning Arabidopsis in plant biology. A key step toward unification of plant research. *Plant physiology* **135**: 602–606.

Bishopp A, Lynch JP. **2015**. The hidden half of crop yields. *Nature Plants* **1**: 15117.

Blum A. **2005**. Drought resistance, water use efficiency, and yield potential—are they compatible, dissonant, or mutually exclusive? *Crop and Pasture Science*: 1159–1168.

Blum A. **2009**. Effective use of water (EUW) and not water use efficiency (WUE) is the target of crop yield improvement under drought stress. *Field Crops Research* **112**: 119–123.

Bonferroni CE. **1936**. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* **8**: 3–62.

Borevitz JO, Nordborg M. **2003**. Update on genomics and natural variation in Arabidopsis: The impact of genomics on the study of natural variation in Arabidopsis. *Plant Physiology* **132**: 718–725.

Borrell AK, Hammer GL, Douglas ACL. **2000**. Does maintaining green leaf area in sorghum improve yield under drought? I. Leaf growth and senescence. *Crop Science* **40**: 1026–1037.

Bouchabke O, Chang F, Simon M, Voisin R, Pelletier G, Durand-Tardif M. **2008**. Natural variation in Arabidopsis thaliana as a tool for highlighting differential drought responses. *PloS one* **3**: e1705.

Brachi B, Morris GP, Borevitz JO. **2011**. Genome-wide association studies in plants: the missing heritability is in the field. *Genome biology* **12**: 232.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. **2007**. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–2635.

Bradford KJ, Yang SF. **1980**. Xylem Transport of 1-Aminocyclopropane-1-carboxylic Acid, an Ethylene Precursor, in Waterlogged Tomato Plants. *Plant physiology* **65**: 322–326.

Brandt B, Brodsky DE, Xue S, Negi J, Iba K, Kangasjarvi J, Ghassemian M, Stephan AB, Hu H, Schroeder JI, *et al.* 2012. Reconstitution of abscisic acid activation of SLAC1 anion channel by CPK6 and OST1 kinases and branched ABI1 PP2C phosphatase action. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 10593–8.

Brennan AC, Méndez-Vigo B, Haddioui A, Martínez-Zapater JM, Picó FX, Alonso-Blanco C. **2014**. The genetic structure of Arabidopsis thaliana in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe. *BMC plant biology* **14**: 17.

Broman KW, Sen S. **2009**. *A Guide to QTL Mapping with R/qtl*. New York: Springer.

Buckley TN. **2005**. The control of stomata by water balance. *New Phytologist* **168**: 275–292.

Cai R, Zhang C, Zhao Y, Zhu K, Wang Y, Jiang H, Xiang Y, Cheng B. **2015**. Genome-wide analysis of the IQD gene family in maize. *Molecular Genetics and Genomics* **291**: 543–558.

Campitelli BE, Des Marais DL, Juenger TE. **2016**. Ecological interactions and the fitness effect of water use efficiency: Competition and drought alter the impact of natural MPK12 alleles in Arabidopsis. *Ecology Letters* **19**: 424–434.

Carins Murphy MR, Jordan GJ, Brodribb TJ. **2014**. Acclimation to humidity modifies the link between leaf size and the density of veins and stomata. *Plant, Cell and Environment* **37**: 124–131.

de Carvalho MHC, Lameta D, Macauley R, Gareil M, Maarouf E, Thi P, Fodil Z, de Carvalho MHC, d'Arcy Lameta A, Macauley HR, *et al.* 2001. Aspartic protease in leaves of common bean (Phaseolus vulgaris L.) and cowpea (Vigna unguiculata L. Walp): enzymatic activity, gene expression and relation to drought susceptibility. *Febs Letters* **492**: 242–246.

Chardon F, Jasinski S, Durandet M, Lecureuil A, Soulay F, Bedu M, Guerche P, Masclaux-Daubresse C. **2014**. QTL meta-analysis in Arabidopsis reveals an interaction between leaf senescence and

resource allocation to seeds. *Journal of experimental botany* **65**: 3949–3962.

Chaves M, Flexas J, Pinheiro C. **2009**. Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of botany*: 551–560.

Chen J, Chang SX, Anyia AO. **2012**. Quantitative trait loci for water use efficiency in barley (Hordeum vulgare L.) measured by carbon isotope discrimination under rain-fed conditions on the Canadian Prairies. *Theoretical and Applied Genetics* **125**: 71–90.

Chen J, Chen Z. **2008**. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95**: 759–771.

Chimenti CA, Pearson J, Hall AJ. **2002**. Osmotic adjustment and yield maintenance under drought in sunflower. *Field Crops Research* **75**: 235–246.

Christman MA, Donovan LA, Richards JH. **2009**. Magnitude of nighttime transpiration does not affect plant growth or nutrition in well-watered Arabidopsis. *Physiologia Plantarum* **136**: 264–273.

Christopher J, Christopher M, Jennings R, Jones S, Fletcher S, Borrell A, Manschadi AM, Jordan D, Mace E, Hammer G. **2013**. QTL for root angle and number in a population developed from bread wheats (Triticum aestivum) with contrasting adaptation to water-limited environments. *Theoretical and Applied Genetics* **126**: 1563–1574.

Clarke JH, Dean C. **1994**. Mapping FRI, a locus controlling flowering time and vernalization response in Arabidopsis thaliana. *MGG Molecular & General Genetics* **242**: 81–89.

Clarke JH, Mithen R, Brown JK, Dean C. **1995**. QTL analysis of flowering time in Arabidopsis thaliana. *Molecular & general genetics : MGG* **248**: 278–86.

Cohen JE. **2003**. Human population: the next half century. *Science* **302**: 1172–5.

Condon a. G, Richards R a., Farquhar GD. **1987**. Carbon Isotope Discrimination is Positively Correlated with Grain Yield and Dry Matter Production in Field-Grown Wheat1. *Crop Science* **27**: 996.

Condon AG, Richards RA, Farquhar GD. **1993**. Relationships Between Carbon Isotope Discrimination , Water Use Efficiency and Transpiration Efficiency for Dryland Wheat. *Australian Journal of Agricultural Research* **44**: 1693–1711.

Condon AG, Richards RA, Rebetzke GJ, Farquhar GD. **2004**. Breeding for high water use efficiency. *Journal of experimental botany* **55**: 2447–60.

Cong Z, Yang D, Gao B, Yang H, Hu H. **2009**. Hydrological trend analysis in the Yellow River basin using a distributed hydrological model. *Water Resources Research* **45**: 1–13.

Contour-Ansel D, Torres-Franklin ML, Zuily-Fodil Y, de Carvalho MHC. **2010**. An aspartic acid protease from common bean is expressed 'on call' during water stress and early recovery. *Journal of Plant Physiology* **167**: 1606–1612.

Cooper M, van Eeuwijk FA, Hammer GL, Podlich DW, Messina C. **2009**. Modeling QTL for complex traits: detection and context for plant breeding. *Current Opinion in Plant Biology* **12**: 231–240.

Courtois B, Audebert A, Dardou A, Roques S, Ghneim-Herrera T, Droc G, Frouin J, Rouan L, Goze E, Kilian A, *et al.* 2013. Genome-wide association mapping of root traits in a japonica rice panel. *PLoS ONE* **8**: 1–18.

Craig H. **1957**. Isotopic standards for carbon and oxygen and correction factors for mass-spectrometric analysis of carbon dioxide. *Geochimica et Cosmochimica Acta* **12**: 133–149.

Creissen HE, Jorgensen TH, Brown JKM. **2013**. Stabilization of yield in plant genotype mixtures through compensation rather than complementation. *Annals of Botany* **112**: 1439–1447.

Dekkers BJW, Costa MCD, Maia J, Bentsink L, Ligterink W, Hilhorst HWM. **2015**. Acquisition and loss of desiccation tolerance in seeds: from experimental model to biological relevance. *Planta* **241**: 563–577.

Demmig-Adams B, Adams WW. **2006**. Photoprotection in an ecological context: The remarkable complexity of thermal energy dissipation. *New Phytologist* **172**: 11–21.

Dermody BJ, Van Beek RPH, Meeks E, Klein Goldewijk K, Scheidel W, Van Der Velde Y, Bierkens MFP, Wassen MJ, Dekker SC. **2014**. A virtual water network of the Roman world. *Hydrology and Earth System Sciences* **18**: 5025–5040.

Dhanapal AP, Ray JD, Singh SK, Hoyos-Villegas V, Smith JR, Purcell LC, King CA, Fritschi FB. **2015**. Association mapping of total carotenoids in diverse soybean genotypes based on leaf extracts and high-throughput canopy spectral reflectance measurements. *PLoS ONE* **10**: 1–15.

Ding S, Zhang B, Qin F. **2015**. Arabidopsis RZFP34/CHYR1, a Ubiquitin E3 Ligase, Regulates Stomatal Movement and Drought Tolerance via SnRK2.6-Mediated Phosphorylation. *The Plant Cell* **27**: tpc.15.00321.

Dittmar EL, Oakley CG, Agren J, Schemske DW. **2014**. Flowering time QTL in natural populations of Arabidopsis thaliana and implications for their adaptive value. *Molecular Ecology* **23**: 4291–4303.

Dodd IC, Ryan AC. **2016**. Whole-Plant Physiological Responses to Water-Deficit Stress. *Encyclopedia of Life Sciences*: 1–9.

Dong X, Yi H, Han C-T, Nou I-S, Hur Y. **2016**. GDSL esterase/lipase genes in Brassica rapa L.: genome-wide identification and expression analysis. *Molecular genetics and genomics : MGG* **291**: 531–542.

Driever SM, Lawson T, Andralojc PJ, Raines C a, Parry M a J. **2014**. Natural variation in photosynthetic capacity, growth, and yield in 64 field-grown wheat genotypes. *Journal of experimental botany*: 1–15.

Easlon HM, Nemali KS, Richards JH, Hanson DT, Juenger TE, McKay JK. **2014**. The physiological basis for genetic variation in water use efficiency and carbon isotope composition in Arabidopsis thaliana. *Photosynthesis research* **119**: 119–29.

Else M a., Hall KC, Arnold GM, Davies WJ, Jackson MB. **1995**. Export of Abscisic Acid, 1-Aminocyclopropane-1-Carboxylic Acid, Phosphate, and Nitrate from Roots to Shoots of Flooded Tomato Plants (Accounting for Effects of Xylem Sap Flow Rate on Concentration and Delivery). *Plant physiology* **107**: 377–384.

El-Soda M, Malosetti M, Zwaan BJ, Koornneef M, Aarts MGM. **2014**. Genotype x environment interaction QTL mapping in plants: Lessons from Arabidopsis. *Trends in Plant Science* **19**: 390–398.

Von Euler T, Ågren J, Ehrlén J. **2014**. Environmental context influences both the intensity of seed predation and plant demographic sensitivity to attack. *Ecology* **95**: 495–504.

FAO. **2016**. FAO Statistical Data. *FAOSTAT*.

Faro C, Gal S. **2005**. Aspartic proteinase content of the Arabidopsis genome. *Current protein & peptide science* **6**: 493–500.

Farquhar G, O'Leary M, Berry J. **1982**. On the Relationship Between Carbon Isotope Discrimination and the Intercellular Carbon Dioxide Concentration in Leaves. *Australian Journal of Plant Physiology* **9**: 121.

Farquhar GD, Richards RA. **1984**. Isotopic composition of plant carbon correlates with water use efficiency of wheat genotypes. *Australian Journal of Plant Physiology* **11**: 539–552.

Farrant JM, Cooper K, Hilgart A, Abdalla KO, Bentley J, Thomson JA, Dace HJW, Peton N, Mundree SG, Rafudeen MS. **2015**. A molecular physiological review of vegetative desiccation tolerance in the resurrection plant Xerophyta viscosa (Baker). *Planta* **242**: 407–426.

Fess TL, Kotcon JB, Benedito VA. **2011**. Crop breeding for low input agriculture: A sustainable response to feed a growing world population. *Sustainability* **3**: 1742–1772.

Filiault DL, Maloof JN. **2012**. A genome-wide association study identifies variants underlying the Arabidopsis thaliana shade avoidance response. *PLoS genetics* **8**: e1002589.

Fischer R a, Rees D, Saire KD, Lu ZM, Condon AG, Larque Saavedra A. **1998**. Wheat Yield Associated with Higher Stomatal Conductance and Photosynthetic Rate , and Cooler Canopies. *Crop Science* **38**: 1467–1475.

Fishcer RA, Rees D, Sayre KD, Lu M, Condon AG, Larque Saavedra A. **1998**. Wheat Yield Progress Associated with Higher Stomatal Conductance and Photosynthetic Rate, and Cooler Canopies. *Crop Science* **38**: 1467–1475.

Fletcher RS, Herrmann D, Mullen JL, Li Q, Schrider DR, Price N, Lin J, Grogan K, Kern A, McKay JK. **2016**. Identification of Polymorphisms Associated with Drought Adaptation QTL in Brassica napus by Resequencing. *G3: Genes|Genomes|Genetics* **6**: 1–39.

Fletcher RS, Mullen JL, Heiliger a., McKay JK. **2014**. QTL analysis of root morphology, flowering time, and yield reveals trade-offs in response to drought in Brassica napus. *Journal of Experimental Botany* **66**: 245–256.

Foolad MR, Zhang LP, Subbiah P. **2003**. Genetics of drought tolerance during seed germination in tomato: inheritance and QTL mapping. *Genome* **46**: 536–545.

Gazzani S, Gendall AR, Lister C, Dean C. **2003**. Analysis of the Molecular Basis of Flowering Time Variation in Arabidopsis Accessions. *Plant Physiology* **132**: 1107–1114.

Ge Z, Rubio G, Lynch JP. **2000**. The importance of root gravitropism for inter-root competition and

phosphorus acquisition efficiency: results from a geometric simulation model. *Plant and soil* **218**: 159–171.

Gechev TS, Dinakar C, Benina M, Toneva V, Bartels D. **2012**. Molecular mechanisms of desiccation tolerance in resurrection plants. *Cellular and Molecular Life Sciences* **69**: 3175–3186.

Gery C, Zuther E, Schulz E, Legoupi J, Chauveau A, McKhann H, Hincha DK, Teoule E. **2011**. Natural variation in the freezing tolerance of Arabidopsis thaliana: Effects of RNAi-induced CBF depletion and QTL localisation vary among accessions. *Plant Science* **180**: 12–23.

Ghalambor CK, McKay JK, Carroll SP, Reznick DN. **2007**. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology* **21**: 394–407.

Ghandilyan A, Barboza L, Tisné S, Granier C, Reymond M, Koornneef M, Schat H, Aarts MGM. **2009**. Genetic analysis identifies quantitative trait loci controlling rosette mineral concentrations in Arabidopsis thaliana under drought. *New Phytologist* **184**: 180–192.

Gibson SI. **2004**. Sugar and phytohormone response pathways: Navigating a signalling network. *Journal of Experimental Botany* **55**: 253–264.

Gil PM, Gurovich L, Schaffer B, García N, Iturriaga R. **2009**. Electrical signaling, stomatal conductance, ABA and ethylene content in avocado trees in response to root hypoxia. *Plant signaling & behavior* **4**: 100–108.

Gnan S, Priest a., Kover PX. **2014**. The Genetic Basis of Natural Variation in Seed Size and Seed Number and Their Trade-Off Using Arabidopsis thaliana MAGIC Lines. *Genetics* **198**: 1751–1758.

Graf A, Smith AM. **2011**. Starch and the clock: The dark side of plant productivity. *Trends in Plant Science* **16**: 169–175.

Greco M, Chiappetta A, Bruno L, Bitonti MB. **2012**. In Posidonia oceanica cadmium induces changes in DNA methylation and chromatin patterning. *Journal of Experimental Botany* **63**: 695–709.

Gregersen PL, Culetic A, Boschian L, Krupinska K. **2013**. Plant senescence and crop productivity. *Plant Molecular Biology* **82**: 603–622.

Guan M, Moller IS, Schjoerring JK. **2015**. Two cytosolic glutamine synthetase isoforms play specific roles for seed germination and seed yield structure in Arabidopsis. *Journal of Experimental Botany* **66**: 203–212.

Gupta PK, Kulwal PL, Jaiswal V. **2014**. Association mapping in crop plants: Opportunities and challenges. *Advances in Genetics* **85**: 109–147.

Gutiérrez RA. **2012**. Systems biology for enhanced plant nitrogen nutrition. *Science* **336**: 1673–1675.

Haley CS, Knott SA. **1992**. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–24.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. **2011**. Arabidopsis thaliana Genome. *Science* **334**: 83–86.

Hao Z, Li X, Xie C, Weng J, Li M, Zhang D, Liang X, Liu L, Liu S, Zhang S. **2011**. Identification of functional genetic variations underlying drought tolerance in maize using snp markers. *Journal of Integrative Plant Biology* **53**: 641–652.

Harrell FE. **2006**. Hmisc: Harrell Miscellaneous.

Hausmann NJ, Juenger TE, Sen S, Stowe K a, Dawson TE, Simms EL. **2005**. Quantitative trait loci affecting delta13C and response to differential water availibility in Arabidopsis thaliana. *Evolution; international journal of organic evolution* **59**: 81–96.

Hayes B. **2013**. Overview of Statistical Methods for Genome-Wide Studies (GWAS). Genome-Wide Association Studies and Gene Prediction.149–69.

He J, Chee CW, Goh CJ. **1996**. Photoinhibition of Heliconia under natural tropical conditions: the importance of leaf orientation for light interception and leaf temperature. *Plant, Cell and Environment* **19**: 1238–1248.

He J, Zhao X, Laroche A, Lu Z-X, Liu H, Li Z. **2014**. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science* **5**: 484.

Hector A, Schmid C, Beierkuhnlein C, Caldeira MC, Diemer M, Dimitrakopoulos PG, Finn J a, Freitas H, Giller PS, Good J, *et al.* 1999. Plant diversity and productivity experiments en European grasslands. *Science* **286**: 1123–1127.

Hegazy AK, Barakat HN, Kabiel HF. **2006**. Anatomical significance of the hygrochastic movement in Anastatica hierochuntica. *Annals of Botany* **97**: 47–55.

Hepworth J, Dean C. **2015**. Flowering Locus C's Lessons: Conserved Chromatin Switches Underpinning Developmental Timing and Adaptation. *Plant physiology* **168**: 1237–45.

Hess T, Aldaya M, Fawell J, Franceschini H, Ober E, Schaub R, Schulze-Aurich J. **2014**. Understanding the impact of crop and food production on the water environment-using sugar as a model. *Journal of the Science of Food and Agriculture* **94**: 2–8.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. **2005**. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965–1978.

Hijmans RJ, van Etten J. **2012**. raster: Geographic analysis and modeling with raster data.

Hiwasa-Tanase K, Ezura H. **2016**. Molecular Breeding to Create Optimized Crops: From Genetic Manipulation to Potential Applications in Plant Factories. *Frontiers in Plant Science* **7**: 1–7.

Hoffmann MH. **2002**. Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). *Journal of Biogeography* **29**: 125–134.

Hoffmann MH. **2005**. Evolution of the Realized Climatic Niche in the Genus Arabidopsis (Brassicaceae). *Evolution* **59**: 1425–1436.

Holland JB. **2007**. Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology* **10**: 156–161.

Holmes MG, Keiller DR. **2002**. Effects of pubescence and waxes on the reflectance of leaves in the ultraviolet and photosynthetic wavebands: A comparison of a range of species. *Plant, Cell and Environment* **25**: 85–93.

Honsdorf N, March TJ, Berger B, Tester M, Pillen K. **2014**. High-throughput phenotyping to detect drought tolerance QTL in wild barley introgression lines. *PLoS ONE* **9**.

Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A, Sperone FG, Vilhjálmsson BJ, *et al.* 2012. Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. *Nature Genetics* **44**: 212–216.

Hosy E, Vavasseur A, Mouline K, Dreyer I, Gaymard F, Porée F, Boucherez J, Lebaudy A, Bouchez D, Very A-A, *et al.* 2003. The Arabidopsis outward K+ channel GORK is involved in regulation of stomatal movements and plant transpiration. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 5549–5554.

Houle D. **1992**. Comparing evolvability and variability of quantitative traits. *Genetics* **130**: 195–204.

Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P. **2008**. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Advances in bioinformatics* **2008**: 420747.

Hsu KH, Liu CC, Wu SJ, Kuo YY, Lu CA, Wu CR, Lian PJ, Hong CY, Ke YT, Huang JH, *et al.* 2014. Expression of a gene encoding a rice RING zinc-finger protein, OsRZFP34, enhances stomata opening. *Plant Molecular Biology* **86**: 125–137.

Huang X, Han B. **2014**. Natural variations and genome-wide association studies in crop plants. *Annual review of plant biology* **65**: 531–51.

Huang X, Paulo M-J, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk F a. **2011**. Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 4488–4493.

Jensen MK, Lindemose S, Masi F de, Reimer JJ, Nielsen M, Perera V, Workman CT, Turck F, Grant MR, Mundy J, *et al.* 2013. ATAF1 transcription factor directly regulates abscisic acid biosynthetic gene NCED3 in Arabidopsis thaliana. *FEBS Open Bio* **3**: 321–327.

Jia W, Zhang J. **2008**. Stomatal movements and long-distance signaling in plants. *Plant signaling & behavior* **3**: 772–777.

Jiaqin S, Ruiyuan L, Dan Q, Congcong J, Yan L, Morgan C, Bancroft I, Jianyi Z, Jinling M. **2009**. Unraveling the complex trait of crop yield with quantitative trait loci mapping in Brassica napus. *Genetics* **182**: 851–861.

Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C. **2000**. Molecular Analysis of FRIGIDA , a Major Determinant of Natural Variation in Arabidopsis Flowering Time. **290**.

Juenger TE. **2013**. Natural variation and genetic constraints on drought tolerance. *Current opinion in plant biology* **16**: 274–281.

Juenger T, Mckay J. **2005**. Identification and characterization of QTL underlying whole- plant physiology in Arabidopsis thaliana: d13C, stomatal conductance and transpiration efficiency. *Plant, Cell and Environment*: 697–708.

Juenger TE, Sen S, Bray E, Stahl E, Wayne T, McKay J, Richards JH. **2010**. Exploring genetic and expression differences between physiologically extreme ecotypes: comparative genomic hybridization and gene expression studies of Kas-1 and Tsu-1 accessions of Arabidopsis thaliana. *Plant, cell & environment* **33**: 1268–84.

Kaiser HF. **1960**. The Application of Electronic Computers to Factor Analysis. *Educational and Pschological Measurement* **XX**.

Kaminski KP, Krup K, Andersen MN, Sonderkaer M, Andersen MS, Kirk HG, Nielsen KL. **2015**. Cytosolic glutamine synthetase is important for photosynthetic efficiency and water use efficiency in potato as revealed by high-throughput sequencing QTL analysis. *Theoretical and Applied Genetics* **128**: 2143–2153.

Kang J, Park J, Choi H, Burla B, Kretzschmar T, Lee Y, Martinoia E. **2011**. Plant ABC Transporters. *The Arabidopsis book / American Society of Plant Biologists* **9**: e0153.

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. **2008**. Efficient Control of Population Stucuture in Model Organism Association Mapping. *Genetics* **178**: 1709–1723.

Kanno Y, Hanada A, Chiba Y, Ichikawa T, Nakazawa M, Matsui M, Koshiba T, Kamiya Y, Seo M. **2012**. Identification of an abscisic acid transporter by functional screening using the receptor complex as a sensor. *Proceedings of the National Academy of Sciences* **109**: 9653–9658.

Kenney AM, McKay JK, Richards JH, Juenger TE. **2014**. Direct and indirect selection on flowering time, water use efficiency (WUE, δ13C), and WUE plasticity to drought in Arabidopsis thaliana. *Ecology and Evolution* **4**: 4505–4521.

Kim HK, Park J, Hwang I. **2014**. Investigating water transport through the xylem network in vascular plants. *Journal of Experimental Botany* **65**: 1895–1904.

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M. **2007**. Recombination and linkage disequilibrium in Arabidopsis thaliana. *Nature genetics* **39**: 1151–1155.

Kim SH, Sicher RC, Bae H, Gitz DC, Baker JT, Timlin DJ, Reddy VR. **2006**. Canopy photosynthesis, evapotranspiration, leaf nitrogen, and transcription profiles of maize in response to CO2 enrichment. *Global Change Biology* **12**: 588–600.

Knight H, Knight MR. **2001**. Abiotic stress signalling pathways: Specificity and cross-talk. *Trends in Plant Science* **6**: 262–267.

Kobata T, Okuno T, Yamamoto T. **1996**. Contributions of capacity for soil water extraction and water use efficiency to maintenance of dry matter production in rice subjected to drought. *Japanese Journal of Crop Science* **65**: 652–662.

Kobayashi Y, Koyama H. **2002**. QTL analysis of Al tolerance in recombinant inbred lines of *Arabidopsis thaliana*. *Plant and Cell Physiology* **43**: 1526–1533.

Koester RP, Nohl BM, Diers BW, Ainsworth EA. **2016**. Has photosynthetic capacity increased with 80 years of soybean breeding? An examination of historical soybean cultivars. *Plant, cell & environment* **39**: 1058–67.

Kolling K, Thalmann M, Muller A, Jenny C, Zeeman SC. **2015**. Carbon partitioning in Arabidopsis thaliana is a dynamic process controlled by the plants metabolic status and its circadian clock. *Plant, Cell and Environment* **38**: 1965–1979.

Koornneef M. **2013**. My favourite flowering image. *Journal of Experimental Botany* **64**: 5791–5793.

Kooyers NJ. **2015**. The evolution of drought escape and avoidance in natural herbaceous populations. *Plant Science* **234**: 155–162.

Korte A, Farlow A. **2013**. The advantages and limitations of trait analysis with GWAS: a review. *Plant methods* **9**: 29.

Korves TM, Schmid KJ, Caicedo AL, Mays C, Stinchcombe JR, Purugganan MD, Schmitt J, Tonia M. Korves, Karl J. Schmid, Ana L. Caicedo, Charlotte Mays, John R. Stinchcombe, Michael D. Purugganan JS. **2007**. Fitness effects associated with the major flowering time gene FRIGIDA in Arabidopsis thaliana in the field. *The American naturalist* **169**: E141–E157.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R. **2009**. A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. *PLoS Genetics* **5**.

Kowalski SP, Lan TH, Feldmann KA, Paterson AH. **1994**. QTL mapping of naturally-occurring variation in flowering time of Arabidopsis thaliana. *MGG Molecular & General Genetics* **245**: 548–555.

Kozlowski TT. **1997**. Responses of woody plants to flooding and salinity. Tree Physiology Monograph No. 1. Victora, Canada: Heron Publishing, 1–29.

Kramer PJ, Boyer JS. **1995**. *Water Relations of Plants and Soil*. Cambridge, Massachusetts: Acamdeic Press, Inc.

Kruger EL, Volin JC. **2006**. Reexamining the empirical relation between plant growth and leaf photosynthesis. *Functional Plant Biology* **33**: 421–429.

Kwak JM, Mori IC, Pei ZM, Leonhardt N, Angel Torres M, Dangl JL, Bloom RE, Bodde S, Jones JDG, Schroeder JI. **2003**. NADPH oxidase *AtrbohD* and *AtrbohF* genes function in ROS-dependent ABA signaling in *Arabidopsis*. *EMBO Journal* **22**: 2623–2633.

Ladwig F, Stahl M, Ludewig U, Hirner A a, Hammes UZ, Stadler R, Harter K, Koch W. **2012**. Siliques are Red1 from Arabidopsis acts as a bidirectional amino acid transporter that is crucial for the amino acid homeostasis of siliques. *Plant Physiol* **158**: 1643–1655.

Lamb HF, Gasse F, Benkaddour A, El Hamouti N, van der Kaars S, Perkins WT, Pearce NJ, Roberts CN. **1995**. Relation between century-scale Holocene arid intervals in tropical and temperate zones. *Nature* **373**: 134–137.

Lander ES, Green P. **1987**. Construction of multilocus genetic linkage maps in humans. *Proceedings*

of the National Academy of Sciences of the United States of America **84**: 2363–7.

Lannenpaa M. **2014**. Heterologous expression of AtMYB12 in kale (Brassica oleracea var. acephala) leads to high flavonol accumulation. *Plant Cell Reports* **33**: 1377–1388.

Lau J a, Shaw RG, Reich PB, Shaw FH, Tiffin P. **2007**. Strong ecological but weak evolutionary effects of elevated CO2 on a recombinant inbred population of Arabidopsis thaliana. *New phytologist* **175**: 351–62.

Lawson T, Blatt MR. **2014**. Stomatal size, speed, and responsiveness impact on photosynthesis and water use efficiency. *Plant Physiol.* **164**: 1556–1570.

Lawson T, Kramer DM, Raines C a. **2012**. Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. *Current opinion in biotechnology* **23**: 215–20.

Lee SC, Lan W, Buchanan BB, Luan S. **2009**. A protein kinase-phosphatase pair interacts with an ion channel to regulate ABA signaling in plant guard cells. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 21419–24.

Lefebvre V, Kiani SP, Durand-Tardif M. **2009**. A focus on natural variation for abiotic constraints response in the model species Arabidopsis thaliana. *International Journal of Molecular Sciences* **10**: 3547–3582.

Levit J. **1972**. *Response of Plants to Environmental Stresses*. New York: Acamdeic Press, Inc.

Li C, He X, Zhu S, Zhou H, Wang Y, Li Y, Yang J, Fan J, Yang J, Wang G, *et al.* 2009. Crop Diversity for Yield Increase. *PLoS ONE* **4**: 0–5.

Van Lijsebettens M, Grasser KD. **2010**. The role of the transcript elongation factors FACT and HUB1 in leaf growth and the induction of flowering. *Plant signaling & behavior* **5**: 715–7.

Lipka AE, Kandianis CB, Hudson ME, Yu J, Drnevich J, Bradbury PJ, Gore MA. **2015**. From association to prediction: Statistical methods for the dissection and selection of complex traits in plants. *Current Opinion in Plant Biology* **24**: 110–118.

Lobell DB, Gourdji SM. **2012**. The Influence of Climate Change on Global Crop Productivity. *Plant Physiology* **160**: 1686–1697.

Lolas IB, Himanen K, Grønlund JT, Lynggaard C, Houben A, Melzer M, Van Lijsebettens M, Grasser KD. **2010**. The transcript elongation factor FACT affects Arabidopsis vegetative and reproductive development and genetically interacts with HUB1/2. *Plant Journal* **61**: 686–697.

Long SP, Ainsworth EA, Leakey ADB, Ort DR, No J. **2006**. Food for Thought : Lower-Than-Expected Crop Yield Stimulation with Rising CO2 Concentrations. *Science* **312**: 1918–1921.

Long SP, Marshall-Colon A, Zhu XG. **2015**. Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* **161**: 56–66.

Long Y, Shi J, Qiu D, Li R, Zhang C, Wang J, Hou J, Zhao J, Shi L, Park BS, *et al.* 2007. Flowering time quantitative trait loci analysis of oilseed Brassica in multiple environments and genomewide alignment with Arabidopsis. *Genetics* **177**: 2433–2444.

Lovell JT, Mullen JL, Lowry DB, Awole K, Richards JH, Sen S, Verslues PE, Juenger TE, McKay JK. **2015**. Exploiting Differential Gene Expression and Epistasis to Discover Candidate Genes for Drought-Associated QTLs in Arabidopsis thaliana. *The Plant Cell*: tpc.15.00122.

Lovett Doust J. **1989**. Plant reproductive strategies and resource allocation. *Trends in ecology & evolution* **4**: 230–4.

Lynch JP. **2015**. Root phenes that reduce the metabolic costs of soil exploration: Opportunities for 21st century agriculture. *Plant, Cell and Environment* **38**: 1775–1784.

Lynch M, Walsh B. **1998**. *Genetics and Analysis of Quantiative Traits*. Sunderland, MA: Sinauer Associates.

Maire V, Gross N, Da Silveira Pontes L, Picon-Cochard C, Soussana JF. **2009**. Trade-off between root nitrogen acquisition and shoot nitrogen utilization across 13 co-occurring pasture grass species. *Functional Ecology* **23**: 668–679.

Maitra N, Cushman J. **1998**. The electronic plant gene register. *Plant physiology* **118**: 1533–1536.

Des Marais DL, McKay JK, Richards JH, Sen S, Wayne T, Juenger TE. **2012**. Physiological Genomics of Response to Soil Drying in Diverse Arabidopsis Accessions. *The Plant Cell* **24**: 893–914.

Marguerit E, Bouffier L, Chancerel E, Costa P, Lagane F, -Marc Guehl J, Plomion C, Brendel O. **2014**. The genetics of water use efficiency and its relation to growth in maritime pine. *Journal of Experimental Botany* **65**: 4757–4768.

Martin B, Tauer CG, Lin RK. **1999**. Carbon Isotope Discrimination as a Tool to Improve Water use Efficiency in Tomato. : 732–735.

Masle J, Gilmore SR, Farquhar GD. **2005**. The ERECTA gene regulates plant transpiration efficiency in Arabidopsis. *Nature* **436**: 866–70.

Matthes M, Bruce T, Chamberlain K, Pickett J, Napier J. **2011**. Emerging roles in plant defense for cis-jasmone-induced cytochrome P450 CYP81D11. *Plant signaling & behavior* **6**: 563–565.

Matthes MC, Bruce TJA, Ton J, Verrier PJ, Pickett JA, Napier JA. **2010**. The transcriptome of cis-jasmone-induced resistance in Arabidopsis thaliana and its role in indirect defence. *Planta* **232**: 1163–1180.

Mauricio R. **2001**. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature reviews. Genetics* **2**: 370–81.

McAusland L, Davey PA, Kanwal N, Baker NR, Lawson T. **2013**. A novel system for spatial and temporal imaging of intrinsic plant water use efficiency. *Journal of Experimental Botany* **64**: 4993–5007.

McKay JK, Richards JH, Mitchell-Olds T. **2003**. Genetics of drought adaptation in Arabidopsis thaliana: I. Pleiotropy contributes to genetic correlations among ecological traits. *Molecular Ecology* **12**: 1137–1151.

McKay JK, Richards JH, Nemali KS, Sen S, Mitchell-Olds T, Boles S, Stahl E a, Wayne T, Juenger TE.

2008. Genetics of drought adaptation in Arabidopsis thaliana II. QTL analysis of a new mapping population, KAS-1 x TSU-1. *Evolution; international journal of organic evolution* **62**: 3014–26.

Mendel G. **1865**. Experiments of Plant Hybridization. Natural History Society of Brno.3–47.

Méndez-Vigo B, Picó FX, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C. **2011**. Altitudinal and climatic adaptation is mediated by flowering traits and FRI, FLC, and PHYC genes in Arabidopsis. *Plant physiology* **157**: 1942–55.

de Mendiburu F. **2016**. agricolae: Statistical Procedures for Agricultural Research.

Menendez CM, Hall AE. **1995**. Heritability of Carbon Isotope Discrimination and Correlations with Earliness in Cowpea. *Crop Science* **35**.

Merah O, Deleens E, Souyris I, Nachit M, Monneveux P. **2001**. Stability of carbon isotope discrimination and grain yield in durum wheat. *Crop Science* **41**: 677–681.

Merk HL, Yarnes SC, Deynze A Van, Tong N, Menda N, Mueller L a, Mutschler M a, Loewen S a, Myers JR, Francis DM. **2012**. Trait Diversity and Potential for Selection Indices Based on Variation Among Regionally Adapted Processing Tomato Germplasm. *Journal of the Amercian Society for Horticultural Scienceam* **137**: 427–437.

Meyer RC, Kusterer B, Lisec J, Steinfath M, Becher M, Scharr H, Melchinger AE, Selbig J, Schurr U, Willmitzer L, *et al.* 2010. QTL analysis of early stage heterosis for biomass in Arabidopsis. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **120**: 227–37.

Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Törjék O, Fiehn O, Eckardt A, Willmitzer L, Selbig J, *et al.* 2007. The metabolic signature related to high plant growth rate in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 4759–64.

Meyer R, Törjék O, Becher M, Altmann T. **2004**. Heterosis of biomass production in Arabidopsis. Establishment during early development. *Plant physiology* **134**: 1813–1823.

Meyerowitz EM. **2001**. Prehistory and history of Arabidopsis research. *Plant physiology* **125**: 15–19.

Michaels SD, Amasino RM. **1999**. FLOWERING LOCUS C Encodes a Novel MADS Domain Protein That Acts as a Repressor of Flowering. **11**: 949–956.

Michaels SD, He Y, Scortecci KC, Amasino RM. **2003**. Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 10102–7.

Miller M, Song Q, Shi X, Juenger TE, Chen ZJ. **2015**. Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of Arabidopsis. *Nature Communications* **6**: 7453.

Milthorpe FL. **1960**. The income and loss of water in arid and semi-arid zone. Plant-Water Relations in Arid and Semi-Arid Conditions. Paris: UNESCO, 9.

Mitchell JH, Siamhan D, Wamala MH, Risimeri JB, Chinyamakobvu E, Henderson SA, Fukai S. **1998**.

The use of seedling leaf death score for evaluation of drought resistance of rice. *Field Crops Research* **55**: 129–139.

Mitchell-Olds T, Schmitt J. **2006**. Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature* **441**: 947–52.

Moinuddin, Khanna-Chopra R. **2004**. Osmotic Adjustment in Chickpea in Relation to Seed Yield and Yield Parameters. *Crop Physiology & Metabolism* **44**: 449–455.

Mojica JP, Mullen J, Lovell JT, Monroe JG, Paul JR, Oakley CG, McKay JK. **2016**. Genetics of water use physiology in locally adapted Arabidopsis thaliana. *Plant Science* **251**: 12–22.

Monclus R, Dreyer E, Villa M, Delmotte FM, Delay D, Petit J, Barbaroux C, Thiec D Le, Brechet C, Brignolas F. **2006**. Impact of drought on productivity and water use effeciency in 29 genotypes of Populus deltoides \times Populus nigra. *New Phytol.* **169**: 765–777.

Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong G-Y, Myles S. **2015**. LinkImpute: Fast and Accurate Genotype Imputation for Non-Model Organisms. *G3: Genes|Genomes|Genetics* **5**: g3.115.021667.

Monneveux P, Sánchez C, Beck D, Edmeades GO. **2006**. Drought tolerance improvement in tropical maize source populations: Evidence of progress. *Crop Science* **46**: 180–191.

Morgan JM. **2000**. Australian Journal of Agricultural Research. *Australian Journal of Agricultural Research* **51**: 971–978.

Morison JIL, Baker NR, Mullineaux PM, Davies WJ. **2008**. Improving water use in crop production. *Philosophical transactions of the Royal Society of Biological sciences* **363**: 639–58.

Morrison GD, Linder CR. **2014**. Association mapping of germination traits in Arabidopsis thaliana under light and nutrient treatments: searching for G×E effects. *G3 (Bethesda, Md.)* **4**: 1465–78.

Moulin M, Mccormac AC, Terry MJ, Smith AG. **2008**. Tetrapyrrole profiling in Arabidopsis seedlings reveals that retrograde plastid nuclear signaling is not due to Mg-protoporphyrin IX accumulation. *Proceedings of the National Academy of Sciences* **105**: 15178–15183.

Nardini A, Gortan E, Ramani M, Salleo S. **2008**. Heterogeneity of gas exchange rates over the leaf surface in tobacco: An effect of hydraulic architecture? *Plant, Cell and Environment* **31**: 804–812.

Narusaka M, Hatakeyama K, Shirasu K, Narusaka Y. **2014**. Arabidopsis dual resistance proteins, both RPS4 and RRS1, are required for resistance to bacterial wilt in transgenic Brassica crops. *Plant signaling & behavior* **9**: e29130.

Nawkar GM, Maibam P, Park JH, Woo SG, Kim CY, Lee SY, Kang CH. **2016**. In silico study on Arabidopsis BAG gene expression in response to environmental stresses. *Protoplasma*: 1–13.

Ngeow J, Eng C. **2015**. New Genetic and Genomic Approaches After the Genome-wide Association Study Era--Back to the Future. *Gastroenterology* **149**: 1138–41.

Ngugi ECK, Galwey NW, Austin RB. **1994**. Genotype × environment interaction in carbon isotope discrimination and seed yield in cowpea (Vigna unguiculata L. Walp.). *Euphytica* **73**: 213–224.

Nord EA, Lynch JP. **2008**. Delayed reproduction in Arabidopsis thaliana improves fitness in soil with suboptimal phosphorus availability. *Plant, Cell and Environment* **31**: 1432–1441.

Nunes-Nesi A, Nascimento V de L, de Oliveira Silva FM, Zsogon A, Araujo WL, Sulpice R. **2016**. Natural genetic variation for morphological and molecular determinants of plant growth and yield. *J. Exp. Bot.* **67**.

Oakley CG, Ågren J, Atchison R a., Schemske DW. **2014**. QTL mapping of freezing tolerance: Links to fitness and adaptive trade-offs. *Molecular Ecology* **23**: 4304–4315.

Ogura T, Busch W. **2015**. From phenotypes to causal sequences: Using genome wide association studies to dissect the sequence basis for variation of plant development. *Current Opinion in Plant Biology* **23**: 98–108.

Ollas C, Dodd IC. **2016**. Physiological impacts of ABA – JA interactions under water-limitation. *Plant Molecular Biology* **91**: 641–650.

Osakabe Y, Arinaga N, Umezawa T, Katsura S, Nagamachi K, Tanaka H, Ohiraki H, Yamada K, Seo S-U, Abo M, *et al.* 2013. Osmotic stress responses and plant growth controlled by potassium transporters in Arabidopsis. *The Plant cell* **25**: 609–24.

Osakabe Y, Osakabe K, Shinozaki K, Tran L-SP. **2014**. Response of plants to water stress. *Frontiers in plant science* **5**: 86.

Paccard A, Fruleux A, Willi Y. **2014**. Latitudinal trait variation and responses to drought in Arabidopsis lyrata. *Oecologia* **175**: 577–587.

Parry MAJ, Flexas J, Medrano H. **2005**. Prospects for crop production under drought: Research priorities and future directions. *Annals of Applied Biology* **147**: 211–226.

Parry MAJ, Reynolds M, Salvucci ME, Raines C, Andralojc PJ, Zhu XG, Price GD, Condon AG, Furbank RT. **2011**. Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. *Journal of Experimental Botany* **62**: 453–467.

Passioura J. **2004**. Increasing Crop Productivity When Water is Scarce – From Breeding to Field Management. Proceeding of the 4th International Crop Science Congress. Brisbane, Australia, 1–17.

Passioura JB. **2010**. Plant-Water Relations. *Encyclopedia of Life Sciences*: 1–7.

Paterson AH, DeVerna JW, Lanini B, Tanksley SD. **1990**. Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. *Genetics* **124**: 735–742.

Patz JA, Grabow ML, Limaye VS. **2014**. When it rains, it pours: Future climate extremes and health. *Annals of Global Health* **80**: 332–344.

Peng T, Sun X, Mumm RH. **2014**. Optimized breeding strategies for multiple trait integration: I. Minimizing linkage drag in single event introgression. *Molecular Breeding* **33**: 89–104.

Penman H, Schofield R. **1951**. Some physical aspects of assimilation and transpiration. *Symp Soc Exp Biol* **5**: 115–129.

Perrin RM. **2008**. Glycosyltransferases in Plant Cell Wall Synthesis. *Encyclopedia of Life Sciences*.

Picó FX. **2012**. Demographic fate of Arabidopsis thaliana cohorts of autumn- and spring-germinated plants along an altitudinal gradient. *Journal of Ecology* **100**: 1009–1018.

Pigliucci M. **2008**. Is evolvability evolvable? *Nature Reviews Genetics* **9**: 75–82.

Pinheiro C, Chaves MM. **2011**. Photosynthesis and drought: can we make metabolic connections from available data? *Journal of experimental botany* **62**: 869–82.

Platt A, Vilhjálmsson BJ, Nordborg M. **2010**. Conditions under which genome-wide association studies will be positively misleading. *Genetics* **186**: 1045–1052.

Pritchard J. **2001**. Turgor Pressure. *Encyclopedia of Life Sciences*: 1–3.

Provart NJ, Alonso J, Assmann SM, Bergmann D, Brady SM, Brkljacic J, Browse J, Chapple C, Colot V, Cutler S, *et al.* 2016. 50 years of Arabidopsis research: Highlights and future directions. *New Phytologist* **209**: 921–944.

Qiao Y, Liu H, Kellomaki S, Peltola H, Liu Y, Dong B, Shi C, Zhang H, Zhang C, Gong J, *et al.* 2014. Comparison of the effects of symmetric and asymmetric temperature elevation and CO2 enrichment on yield and evapotranspiration of winter wheat (Triticum aestivum L.). *Ecology and Evolution* **4**: 1994–2003.

Quarrie SA, Laurie DA, Zhu J, Lebreton C, Semikhodskii A, Steed A, Witsenboer H, Calestani C. **1997**. Increased information leakage from text. *Plant Molecular Biology* **35**: 155–165.

R Development Core Team. **2008**. R: A language and environment for statistical computing.

Ray DK, Gerber JS, Macdonald GK, West PC. **2015**. Climate variation explains a third of global crop yield variability. *Nature Communications* **6**: 1–9.

Reape TJ, McCabe PF. **2010**. Apoptotic-like regulation of programmed cell death in plants. *Apoptosis* **15**: 249–256.

Reape TJ, Molony EM, McCabe PF. **2008**. Programmed cell death in plants: Distinguishing between different modes. *Journal of Experimental Botany* **59**: 435–444.

Rebetzke GJ, Condon AG, Richards RA, Farquhar GD. **2002**. Selection for reduced carbon isotope discrimination increases aerial biomass and grain yield of rainfed bread wheat. *Crop Science* **42**: 739–745.

Remington DL, Leinonen PH, Leppala J, Savolainen O. **2013**. Complex genetic effects on early vegetative development shape resource allocation differences between Arabidopsis lyrata populations. *Genetics* **195**: 1087–1102.

Reymond M, Svistoonoff S, Loudet O, Nussaume L, Desnos T. **2006**. Identification of QTL controlling root growth response to phosphate starvation in *Arabidopsis thaliana*. *Plant, Cell & Environment* **29**: 115–125.

Reznik D. **1985**. Costs of reproduction: and evaluation of the empirical evidence. *OIKOS* **44**: 257–

267.

Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, *et al.* 2003. The Arabidopsis Information Resource (TAIR): A model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Research* **31**: 224–228.

Riboni M, Galbiati M, Tonelli C, Conti L. 2013. GIGANTEA enables drought escape response via abscisic acid-dependent activation of the florigens and SUPPRESSOR OF OVEREXPRESSION OF CONSTANS. *Plant physiology* **162**: 1706–19.

Richards RA. 2000. Selectable traits to increase crop photosynthesis and yield of grain crops. *Journal of experimental botany* **51**: 447–58.

Rijsberman F, Mohammed A. 2003. Water, food and environment: Conflict or dialogue? *Water Science and Technology* **47**: 53–62.

Roff DA, Fairbairn DJ. 2007. The evolution of trade-offs: Where are we? *Journal of Evolutionary Biology* **20**: 433–447.

van Rooijen R, Aarts MGM, Harbinson J. 2015. Natural genetic variation for acclimation of photosynthetic light use efficiency to growth irradiance in Arabidopsis thaliana. *Plant Physiology* **167**: 1412-1429

Rosas U, Mei Y, Xie Q, Banta J a, Zhou RW, Seufferheld G, Gerard S, Chou L, Bhambhra N, Parks JD, *et al.* 2014. Variation in Arabidopsis flowering time associated with cis-regulatory variation in CONSTANS. *Nature communications* **5**: 1-8.

Roscher C, Scherer-Lorenzen M, Schumacher J, Temperton VM, Buchmann N, Schulze ED. 2011. Plant resource-use characteristics as predictors for species contribution to community biomass in experimental grasslands. *Perspectives in Plant Ecology, Evolution and Systematics* **13**: 1–13.

Rossel JB, Walter PB, Hendrickson L, Chow WS, Poole A, Mullineaux PM, Pogson BJ. 2006. A mutation affecting ASCORBATE PEROXIDASE 2 gene expression reveals a link between responses to high light and drought tolerance. *Plant, Cell and Environment* **29**: 269–281.

Rossel JB, Wilson PB, Hussain D, Woo NS, Gordon MJ, Mewett OP, Howell K a, Whelan J, Kazan K, Pogson BJ. 2007. Systemic and intracellular responses to photooxidative stress in Arabidopsis. *The Plant cell* **19**: 4091–4110.

Roussel M, Dreyer E, Montpied P, Le-Provost G, Guehl JM, Brendel O. 2009. The diversity of 13C isotope discrimination in a Quercus robur full-sib family is associated with differences in intrinsic water use efficiency, transpiration efficiency, and stomatal conductance. *Journal of Experimental Botany* **60**: 2419–2431.

Royo C, Villegas D, García del Moral LF, Elhani S, Aparicio N, Rharrabti Y, Araus JL. 2002. Comparative performance of carbon isotope discrimination and canopy temperature depression as predictors of genotype differences in durum wheat yield in Spain. *Australian Journal of Agricultural Research* **53**: 561–569.

Van Ruijven J, Berendse F. **2005**. Diversity-productivity relationships: Initial effects, long-term patterns, and underlying mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 695–700.

Ruts T, Matsubara S, Wiese-Klinkenberg A, Walter A. **2012**. Aberrant temporal growth pattern and morphology of root and shoot caused by a defective circadian clock in Arabidopsis thaliana. *Plant Journal* **72**: 154–161.

Samis KE, Murren CJ, Bossdorf O, Donohue K, Fenster CB, Malmberg RL, Purugganan MD, Stinchcombe JR. **2012**. Longitudinal trends in climate drive flowering time clines in north american arabidopsis thaliana. *Ecology and Evolution* **2**: 1162–1180.

Sanchez-Bermejo E, Balasubramanian S. **2016**. Natural variation involving deletion alleles of FRIGIDA modulate temperature-sensitive flowering responses in Arabidopsis thaliana. *Plant, Cell and Environment*: 1353–1365.

Sasaki E, Zhang P, Atwell S, Meng D, Nordborg M. **2015**. 'Missing' G x E Variation Controls Flowering Time in Arabidopsis thaliana. *PLoS Genetics* **11**: 1–18.

Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, Fernie AR, Causse M. **2014**. Genome-Wide Association in Tomato Reveals 44 Candidate Loci for Fruit Metabolic Traits. *Plant physiology* **165**: 1120–1132.

Sax K. **1923**. The Association of Size Differences with Seed-Coat Pattern and Pigmentation in Phaseolus Vulgaris. *Genetics* **8**: 552–560.

Sayre KD, Acevedo E, Austin RB. **1995**. Carbon-Isotope Discrimination and Grain-Yield for 3 Bread Wheat Germplasm Groups Grown at Different Levels of Water-Stress. *Field Crops Research* **41**: 45–54.

Schenk HJ. **2006**. Root competition: Beyond resource depletion. *Journal of Ecology* **94**: 725–739.

Schmitz RJ, Amasino RM. **2007**. Vernalization: A model for investigating epigenetics and eukaryotic gene regulation in plants. *Biochimica et Biophysica Acta - Gene Structure and Expression* **1769**: 269–275.

Scholl RL, May ST, Ware DH. **2000**. Seed and molecular resources for Arabidopsis. *Plant physiology* **124**: 1477–80.

Seemann JR, Sharkey TD, Wang J, Osmond CB. **1987**. Environmental effects on photosynthesis, nitrogen-use efficiency, and metabolite pools in leaves of sun and shade plants. *Plant physiology* **84**: 796–802.

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M. **2012**. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44**: 825–830.

Semagn K, Bjørnstad Å, Ndjiondjop M. **2006**. Principles, requirements and prospects of genetic mapping in plants. *African Journal of Biotechnology* **5**: 2569–2587.

Seren U, Vihjalmsson BJ, Horton MW, Meng D, Forai P, Huang YS, Long Q, Segura V, Nordborg M. **2012**. GWAPP : A Web Application for Genome-Wide Association Mapping in Arabidopsis. *The Plant cell* **24**: 4793–4805.

Sharkey TD, Bernacchi CJ, Farquhar GD, Singsaas EL. **2007**. Fitting photosynthetic carbon dioxide response curves for C3 leaves. *Plant, Cell and Environment* **30**: 1035–1040.

Sharma A, Li X, Lim YP. **2014**. Comparative genomics of Brassicaceae crops. *Breeding Science* **64**: 3–13.

Shibles R, Sundberg DN. **1998**. Relation of Leaf Nitrogen Content and Other Traits with Seed Yield of Soybean. *Plant Production Science* **1**: 3–7.

Shingaki-Wells RN, Huang S, Taylor NL, Carroll AJ, Zhou W, Millar a H. **2011**. Differential molecular responses of rice and wheat coleoptiles to anoxia reveal novel metabolic adaptations in amino Acid metabolism for tissue tolerance. *Plant physiology* **156**: 1706–1724.

Simpson GG, Dean C. **2002**. the Rosetta Stone of Arabidopsis. *Science* **296**: 285–289.

Sinclair TR, Purcell LC. **2005**. Is a physiological perspective relevant in a 'genocentric' age? *Journal of experimental botany* **56**: 2777–82.

Singarayer JS, Ridgwell A, Irvine P. **2009**. Assessing the benefits of crop albedo bio-geoengineering. *Environmental Research Letters* **4**: 45110.

Skirycz A, Vandenbroucke K, Clauw P. **2011**. Survival and growth of Arabidopsis plants given limited water are not equal. *Nature Biotechnology* **29**: 1–4.

Sletvold N, Agren J. **2015**. Climate-dependent costs of reproduction: Survival and fecundity costs decline with length of the growing season and summer temperature. *Ecology Letters* **18**: 357–364.

Solberg Woods LC. **2014**. QTL mapping in outbred populations: successes and challenges. *Physiological genomics* **46**: 81–90.

Song Y, Chen Q, Ci D, Shao X, Zhang D. **2014**. Effects of high temperature on photosynthesis and related gene expression in poplar. *BMC Plant Biol* **14**: 111.

Steduto P, Hsiao TC, Fereres E. **2007**. On the conservative behavior of biomass water productivity. *Irrigation Science* **25**: 189–207.

Steudle E. **2000**. Water uptake by roots: effects of water deficit. *Journal of experimental botany* **51**: 1531–1542.

Strange A, Li P, Lister C, Anderson J, Warthmann N, Shindo C, Irwin J, Nordborg M, Dean C. **2011**. Major-Effect Alleles at Relatively Few Loci Underlie Distinct Vernalization and Flowering Variation in Arabidopsis Accessions. *PLoS ONE* **6**.

Sulpice R, Nikoloski Z, Tschoep H, Antonio C, Kleessen S, Larhlimi A, Selbig J, Ishihara H, Gibon Y, Fernie AR, *et al.* 2013. Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of Arabidopsis accessions. *Plant physiology* **162**: 347–63.

Suter L, Widmer A. **2013**. Phenotypic effects of salt and heat stress over three generations in Arabidopsis thaliana. *PLoS ONE* **8**: 1–12.

The 1001 Genomes Consortium. **2016**. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell* **166**: 481–491.

Thoday JM. **1961**. Location of polygenes. *Nature* **191**: 368–370.

Thumma BR, Naidu BP, Chandra a, Cameron DF, Bahnisch LM, Liu C. **2001**. Identification of causal relationships among traits related to drought resistance in Stylosanthes scabra using QTL analysis. *Journal of experimental botany* **52**: 203–214.

Tilman D, Wedin D, Knops J. **1996**. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Proceedings. Biological sciences / The Royal Society* **379**: 718–720.

Timotijevic GS, Milisavljevic MD, Radovic SR, Konstantinovic MM, Maksimovic VR. **2010**. Ubiquitous aspartic proteinase as an actor in the stress response in buckwheat. *Journal of Plant Physiology* **167**: 61–68.

Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, Epple P, Kuhns C, Sureshkumar S, Schwartz C, Lanz C, *et al.* 2010. Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. *Nature* **465**: 632–6.

Tolk JA, Howell TA. **2003**. Water use efficiencies of grain sorghum grown in three USA southern Great Plains soils. *Agricultural Water Management* **59**: 97–111.

Tomfohrde J, Silverman A, Barnes R, Fernandez-Vina M a, Young M, Lory D, Morris L, Wuepper KD, Stastny P, Menter A. **1994**. Gene for familial psoriasis susceptibility mapped to the distal end of human chromosome 17q. *Science* **264**: 1141–5.

Tonsor SJ, Alonso-Blanco C, Koornneef M. **2005**. Gene function beyond the single trait: Natural variation, gene effects, and evolutionary ecology in Arabidopsis thaliana. *Plant, Cell and Environment* **28**: 2–20.

Törjék O, Witucka-Wall H, Meyer RC, von Korff M, Kusterer B, Rautengarten C, Altmann T. **2006**. Segregation distortion in Arabidopsis C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **113**: 1551–61.

Trontin C, Tisné S, Bach L, Loudet O. **2011**. What does Arabidopsis natural variation teach us (and does not teach us) about adaptation in plants? *Current Opinion in Plant Biology* **14**: 225–231.

Utsumi YO, Tanaka MAHO, Morosawa TA, Kurotani AT, Yoshida TA, Mochida KE, Matsui AK, Umemura YO, Ishitani MA, Shinozaki KA, *et al.* 2012. Transcriptome Analysis Using a High-Density Oligomicroarray under Drought Stress in Various Genotypes of Cassava : An Important Tropical Crop. *DNA Research* **19**: 335–345.

Vadez V, Kholova J, Medina S, Kakkera A, Anderberg H. **2014**. Transpiration efficiency: New insights into an old story. *Journal of Experimental Botany* **65**: 6141–6153.

Vahisalu T, Kollist H, Wang Y-F, Nishimura N, Chan W-Y, Valerio G, Lamminmäki A, Brosché M, Moldau H, Desikan R, *et al.* 2008. SLAC1 is required for plant guard cell S-type anion channel function in stomatal signalling. *Nature Letters* **452**: 487–91.

Valladares F, Sanchez-Gomez D, Zavala M a. 2006. Quantitative estimation of phenotypic plasticity: Bridging the gap between the evolutionary concept and its ecological applications. *Journal of Ecology* **94**: 1103–1116.

Vasseur F, Bontpart T, Dauzat M, Granier C, Vile D. **2014**. Multivariate genetic analysis of plant responses to water deficit and high temperature revealed contrasting adaptive strategies. *Journal of Experimental Botany* **65**: 6457–6469.

Vatén A, Bergmann DC. **2012**. Mechanisms of stomatal development: an evolutionary view. *EvoDevo* **3**: 11.

Verslues PE, Juenger TE. **2011**. Drought, metabolites, and Arabidopsis natural variation: A promising combination for understanding adaptation to water-limited environments. *Current Opinion in Plant Biology* **14**: 240–245.

Verslues PE, Lasky JR, Juenger TE, Liu T-W, Kumar MN. **2014**. Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in Arabidopsis. *Plant physiology* **164**: 144–59.

Voltas J, Romagosa I, Lafarga A, Armesto AP, Sombrero A, Araus JL. **1999**. Genotype by environment interaction for grain yield and carbon isotope discrimination of barley in Mediterranean Spain. *Australian Journal of Agricultural Research* **50**: 1263–1271.

Wagner R, Aigner H, Pružinská A, Jänkänpää HJ, Jansson S, Funk C. **2011**. Fitness analyses of arabidopsis thaliana mutants depleted of ftsh metalloproteases and characterization of three ftsh6 deletion mutants exposed to high light stress, senescence and chilling. *New Phytologist* **191**: 449–458.

Wagner GP, Altenberg L. **1996**. Perspective : Complex Adaptations and the Evolution of Evolvability Published by. *Evolution* **50**: 967–976.

Wang N, Chen B, Xu K, Gao G, Li F, Qiao J, Yan G, Li J, Li H, Wu X. **2016**. Association Mapping of Flowering Time QTLs and Insight into Their Contributions to Rapeseed Growth Habits. *Frontiers in Plant Science* **7**: 1–11.

Wang N, Qian W, Suppanz I, Wei L, Mao B, Long Y, Meng J, Müller AE, Jung C. **2011**. Flowering time variation in oilseed rape (Brassica napus L.) is associated with allelic variation in the FRIGIDA homologue BnaA.FRI.a. *Journal of Experimental Botany* **62**: 5641–5658.

Wehner GG, Balko CC, Enders MM, Humbeck KK, Ordon FF. **2015**. Identification of genomic regions involved in tolerance to drought stress and drought stress induced leaf senescence in juvenile barley. *BMC plant biology* **15**: 125.

Weigel D. **2012**. Natural variation in Arabidopsis: from molecular genetics to ecological genomics. *Plant physiology* **158**: 2–22.

White J, Castillo J, Ehleringer J. **1990**. Associations Between Productivity, Root Growth and Carbon Isotope Discrimination in *Phaseolus vulgaris* Under Water Deficit. *Australian Journal of Plant Physiology* **17**: 189.

White CA, Sylvester-Bradley R, Berry PM. **2015**. Root length densities of UK wheat and oilseed rape crops with implications for water capture and yield. *Journal of Experimental Botany* **66**: 2293–2303.

Wijnen CL, Keurentjes JJB. **2014**. Genetic resources for quantitative trait analysis: novelty and efficiency in design from an Arabidopsis perspective. *Current opinion in plant biology* **18**: 103–9.

Wilkinson S, Clephan a L, Davies WJ. **2001**. Rapid low temperature-induced stomatal closure occurs in cold-tolerant Commelina communis leaves but not in cold-sensitive tobacco leaves, via a mechanism that involves apoplastic calcium but not abscisic acid. *Plant physiology* **126**: 1566–1578.

Williams SJ, Davies GJ. **2001**. Protein-carbohydrate interactions: Learning lessons from nature. *Trends in Biotechnology* **19**: 356–362.

Williams B, Kabbage M, Britt R, Dickman MB. **2010**. AtBAG7, an Arabidopsis Bcl-2-associated athanogene, resides in the endoplasmic reticulum and is involved in the unfolded protein response. *Proc Natl Acad Sci U S A* **107**: 6088–6093.

Wingler A, Purdy SJ, Edwards SA, Chardon F, Masclaux-Daubresse C. **2010**. QTL analysis for sugar-regulated leaf senescence supports flowering-dependent and -independent senescence pathways. *New Phytologist* **185**: 420–433.

Wolfe MD, Tonsor SJ. **2014**. Adaptation to spring heat and drought in northeastern Spanish Arabidopsis thaliana. *New Phytologist* **201**: 323–334.

Würschum T. **2012**. Mapping QTL for agronomic traits in breeding populations. *Theoretical and Applied Genetics* **125**: 201–210.

Xu Y, Iacuone S, Li SF, Parish RW. **2014**. MYB80 homologues in Arabidopsis, cotton and Brassica: regulation and functional conservation in tapetal and pollen development. *BMC plant biology* **14**: 278.

Yamada S, Komori T, Imaseki H. **1997**. The electronic plant gene register. *Plant physiology* **115**: 313–315.

Yamaguchi-Shinozaki K, Shinozaki K. **2006**. Transcriptional Regulatory Networks in Cellular Responses and Tolerance To Dehydration and Cold Stresses. *Annual Review of Plant Biology* **57**: 781–803.

Yang D, Liu Y, Cheng H, Chang L, Chen J, Chai S, Li M. **2016**. Genetic dissection of flag leaf morphology in wheat (Triticum aestivum L .) under diverse water regimes. *BMC Genetics*: 1–15.

Yang T, Poovaiah BW. **2003**. Calcium/calmodulin-mediated signal network in plants. *Trends in Plant Science* **8**: 505–512.

Yang J, Zhang J. **2006**. Grain filling of cereals under soil drying. *New Phytologist* **169**: 223–236.

Yao X, Xiong W, Ye T, Wu Y. **2012**. Overexpression of the aspartic protease ASPG1 gene confers

drought avoidance in Arabidopsis. *Journal of Experimental Botany* **63**: 2579–2593.

Zhan A, Schneider H, Lynch JP. **2015**. Reduced Lateral Root Branching Density Improves Drought Tolerance in Maize. *Plant physiology* **168**: 1603–15.

Zhang H, Cui F, Wang L, Li J, Ding A, Zhao C, Bao Y, Yang Q, Wang H. **2013**. Conditional and unconditional QTL mapping of drought-tolerance-related traits of wheat seedling using two related RIL populations. *Journal of Genetics* **92**: 213–231.

Zhang JW, Feng Z, Cregg BM, Schumann CM. **1997**. Carbon isotopic composition, gas exchange, and growth of three populations of ponderosa pine differing in drought tolerance. *Tree physiology* **17**: 461–6.

Zhang YM, Gai J. **2009**. Methodologies for segregation analysis and QTL mapping in plants. *Genetica* **136**: 311–318.

Zhu M, Monroe JG, Suhail Y, Villiers F, Mullen J, Pater D, Hauser F, Jeon BW, Bader JS, Kwak JM, *et al.* 2016. Molecular and systems approaches towards drought-tolerant canola crops. *New Phytologist* **210**: 1169–1189.

Ziegler GR, Hartsock RH, Baxter I. **2015**. Zbrowse: an interactive GWAS results browser. *PeerJ Computer Science* **1**: e3.

Zou F. **2009**. QTL Mapping in Intercross and Backcross Populations. Cardiovascular Genomics. New York: Humana Press, 157–171.

Zou X, Suppanz I, Raman H, Hou J, Wang J, Long Y, Jung C, Meng J. **2012**. Comparative Analysis of FLC Homologues in Brassicaceae Provides Insight into Their Role in the Evolution of Oilseed Rape. *PLoS ONE* **7**.