Identifying Genes Regulating Stomatal Density in Different Carbon Dioxide Levels in *Arabidopsis thaliana*

A thesis by

Hannah Sewell

A thesis submitted to The University of Sheffield in partial fulfilment of the requirements for the degree of Doctor of Philosophy



Department of Animal and Plant Sciences

September 2018

Summary

Anthropogenic impacts on atmospheric CO2 concentration ([CO2]) are expected to cause changes to plant morphology. Specifically, alterations are anticipated in the stomatal density (SD) of *Arabidopsis thaliana*. Previous studies have suggested that in elevated [CO2], SD decreases as the plant acquires the same amount of [CO2] through fewer stomata, leading to reduced water loss and improved water use efficiency.

Whilst some genes involved in the SD response to fluctuations in [CO2­] have previously been identified, this thesis aimed to discover further genes within this pathway. Additionally, an aim was set to examine which genes had the largest control of variance in a population of recombinant inbred lines (RILs), derived from 19 parental ecotypes. This set of RILs, known as the Multi-Parent Advanced Generation Inter-Cross (MAGIC) lines, were produced for such experiments. Once sufficient diversity in SD and SD response to changes to [CO2] had been established in the 19 parental ecotypes, a Quantitative Trait Loci (QTL) analysis was carried out to identify regions of the genome associated with control of SD and SD response to [CO2].

This analysis identified 3 QTLs which controlled SD, located on chromosomes 1, 2 and 5. There were 3 QTLs present in ambient [CO2] and 1 of these QTLs was also observed in elevated [CO2]. No QTLs identified regarding the difference in SD between [CO2] treatments were found. The QTL on chromosome 2 was identified as likely being *ERECTA* (*ER*), a gene that has previously been associated with controlling stomatal density, based on chromosomal position. The gene on chromosome 5 was identified, through the phenotyping of T-DNA insertions, as likely being *ADP GLUCOSE PYROPHOSPHORYLASE 2* (*ADG2*), which is involved in starch biosynthesis.

Further analysis of both *ADG2* and *ER* were carried out to study their role in controlling SD in the MAGIC parental lines. Genomic data and expression data were analysed to try and identify the mechanism by which *ADG2* and *ER* affected SD. Results indicated that control of SD by *ADG2* was related to the starch content of the plant, however no clear mechanism was found for *ER*.

Acknowledgements

I would first like to take the opportunity to thank my supervisors Dr Stuart Casson, Professor Julie Gray and Professor David Beerling. They have provided invaluable feedback and guidance throughout my time at The University of Sheffield. In particular I would like to thank Stuart for taking the time to provide detailed guidance on experiments and writing, especially to do with molecular work, and to Julie who gave particularly useful advice while I was writing my thesis, and valuable life advice when I was feeling particularly down.

I’d like to thank my collaborator Dr Paula Kover of Bath University, who spent so much time answering my questions as I first set out on this PhD and provided the MAGIC lines seed and valuable assistance in running the QTL analysis.

I’d like to thank my partner Rich for all of the help and support he has given me throughout my PhD. Whether reading my chapters, listening to me complain about life or putting up with me sulking on the sofa, he has always been infinitely patient. I would have cried a lot more without him there.

There are people too numerous to name who have helped me in the lab but Dr Nick Zoulias and Dr Jim Rowe provided so much help while I struggled with lab work. Dr Dave Johnson gave me advice throughout my PhD while I bothered him with questions and was a lifesaver while we tried to work out how to quantify starch. Everyone else from the Casson Lab and Sorefan lab and C57 also helped in numerous small ways. I’d like to also thank Matt Parker for always taking the time to help me with any bioinformatics/coding problems I had.

I’d like to thank my family who supported me throughout, my parents particularly who showed an interest even when they didn’t understand. Friends at the university are what made doing this PhD bearable. I’d like to thank my two desk neighbours, Dave Johnson and Kirsty Elliott for listening to me grumble and all of those who spent time working in the C Mezz office. I’d like to thank Rachel Denley-Bowers, Nick Zoulias, Jim Rowe, Jordan Brown, Matt Parker, Collette Baxter and the others of my MBB office for their friendship and guidance. Finally I’d like to thank ‘Pub Club’ for providing me with friendship from the start of my PhD.

Finally, I’d like to thank my funding providers The Grantham Centre for Sustainable Futures for giving me the opportunity to do a PhD and all the training alongside. I’d especially like to thank them for providing funding for an internship at Zooniverse (who I’d also like to acknowledge as amazing) without which I may not have gotten my job!

Table of Contents

[1 General Introduction 10](#_Toc1319390)

[1.1 Global Climate Change 11](#_Toc1319391)

[1.1.1 Food Security 12](#_Toc1319392)

[1.1.2 Plant physiological Responses to [CO2] 13](#_Toc1319393)

[1.2 Stomata Physiology and Trade-offs 13](#_Toc1319394)

[1.2.1 Stomatal Density 15](#_Toc1319395)

[1.2.2 Other Stomatal traits 17](#_Toc1319396)

[1.2.3 Epidermal Cells Size 18](#_Toc1319397)

[1.3 Variation in [CO2] levels 19](#_Toc1319398)

[1.3.1 Past changes to [CO2] increases 19](#_Toc1319399)

[1.3.2 Laboratory and greenhouse and controlled environment experiments 20](#_Toc1319400)

[1.3.3 Free Air Carbon-dixide Enrichment (FACE) Experiments 21](#_Toc1319401)

[1.4 Stomatal Aperture 22](#_Toc1319402)

[1.5 Stomatal Development 23](#_Toc1319403)

[1.5.1 The Arabidopsis Core Development Pathway 23](#_Toc1319404)

[1.5.2 Integration of Signals into Stomatal Development 25](#_Toc1319405)

[1.5.3 The One Cell Spacing Rule 27](#_Toc1319406)

[1.6 The Genetic Pathway for [CO2] Response 28](#_Toc1319407)

[1.6.1 Biochemical signalling and [CO2] 30](#_Toc1319408)

[1.6.2 Stomatal Aperture response to CO2 31](#_Toc1319409)

[1.7 MAGIC Arabidopsis and QTL Models 32](#_Toc1319410)

[1.7.1 Quantitative Traits 32](#_Toc1319411)

[1.7.2 QTL Models 32](#_Toc1319412)

[1.7.3 Recombinant Inbred Lines and MAGIC Lines 34](#_Toc1319413)

[1.7.4 Previous uses of MAGIC Lines 37](#_Toc1319414)

[1.7.5 Fine Mapping 38](#_Toc1319415)

[1.8 Aims 39](#_Toc1319416)

[1.8.1 Chapter 2 39](#_Toc1319417)

[1.8.2 Chapter 3 39](#_Toc1319418)

[1.8.3 Chapter 4 40](#_Toc1319419)

[1.8.4 Chapter 5 40](#_Toc1319420)

[2 Variation in Stomatal Characteristics across 19 Arabidopsis Ecotypes 41](#_Toc1319421)

[2.1.1 Natural Accessions of Arabidopsis 41](#_Toc1319422)

[2.1.2 Response to [CO2] in Natural Arabidopsis Ecotypes 42](#_Toc1319423)

[2.1.3 When is a Cotyledon Mature? 43](#_Toc1319424)

[2.1.4 Aims 44](#_Toc1319425)

[2.2 Methods 45](#_Toc1319426)

[2.2.1 Plant Material 45](#_Toc1319427)

[2.2.2 Plant Growth 45](#_Toc1319428)

[2.2.3 Stomatal Impressions 46](#_Toc1319429)

[2.2.4 Stomatal Counting and Microscopy 46](#_Toc1319430)

[2.2.5 Measuring Cell Sizes 47](#_Toc1319431)

[2.2.6 Leaf Area 47](#_Toc1319432)

[2.2.7 Ecotype Environment Data Collection 47](#_Toc1319433)

[2.2.8 Bootstrapping 47](#_Toc1319434)

[2.2.9 Statistics 47](#_Toc1319435)

[2.3 Results 48](#_Toc1319436)

[2.3.1 Finding Ideal Sample Size 48](#_Toc1319437)

[2.3.2 Cotyledon Age and Stomatal Density 50](#_Toc1319438)

[2.3.3 Mature Leaf and Cotyledon Similarities 52](#_Toc1319439)

[2.3.4 Stomatal Density of 19 Ecotypes 55](#_Toc1319440)

[2.3.5 Stomatal Features and Environmental Conditions 60](#_Toc1319441)

[2.4 Discussion 62](#_Toc1319442)

[2.4.1 When and How Many to Phenotype 62](#_Toc1319443)

[2.4.2 Cotyledons for High-Throughput Phenotyping 62](#_Toc1319444)

[2.4.3 Response of 19 parental Ecotypes 64](#_Toc1319445)

[2.4.4 Environment and SD 65](#_Toc1319446)

[2.4.5 Conclusions 66](#_Toc1319447)

[3 Quantitative Trait Loci Analysis of Stomatal Density in Different [CO2] Treatments 68](#_Toc1319448)

[3.1.1 QTLs for [CO2] response 68](#_Toc1319449)

[3.1.2 MAGIC lines 69](#_Toc1319450)

[3.1.3 QTL and GWAS 69](#_Toc1319451)

[1.1.1 Finding the causes of QTLs 70](#_Toc1319452)

[3.1.4 Aims 71](#_Toc1319453)

[3.2 Methods 72](#_Toc1319454)

[3.2.1 Plant Material 72](#_Toc1319455)

[3.2.2 Plant Growth 72](#_Toc1319456)

[3.2.3 Phenotyping 72](#_Toc1319457)

[3.2.4 QTL analysis 72](#_Toc1319458)

[3.2.5 Haplotype analysis 73](#_Toc1319459)

[3.2.6 Gene Ontology 74](#_Toc1319460)

[3.2.7 DNA extraction 74](#_Toc1319461)

[3.2.8 PCR 74](#_Toc1319462)

[3.2.9 Gel Electrophoresis 75](#_Toc1319463)

[3.3 Results 78](#_Toc1319464)

[3.3.1 Phenotypes of the MAGIC lines 78](#_Toc1319465)

[3.3.2 Controls 82](#_Toc1319466)

[3.3.3 QTL mapping with Ler-0 haplotype on chromosome 2 85](#_Toc1319467)

[3.3.4 QTL Mapping without Ler-0 haplotype at ERECTA 88](#_Toc1319468)

[3.3.5 Haplotype Analysis of Ecotypes at each QTL with Ler-0 91](#_Toc1319469)

[3.3.6 Haplotype Analysis without Ler-0 93](#_Toc1319470)

[3.3.7 Gene Ontology analysis 97](#_Toc1319471)

[3.3.8 T-DNA Insertions 100](#_Toc1319472)

[3.4 Discussion and Conclusions 103](#_Toc1319473)

[3.4.1 Polygenetic Control of Stomatal Density 103](#_Toc1319474)

[3.4.2 Quantitative Trait Loci 103](#_Toc1319475)

[3.4.3 Evaluating the QTL analysis 106](#_Toc1319476)

[3.4.4 Identification of QTL2 chromosome 107](#_Toc1319477)

[3.4.5 Identification of QTL5 chromosome 107](#_Toc1319478)

[3.4.6 Conclusions 108](#_Toc1319479)

[4 ADG2, Starch Biosynthesis and Stomatal Density 110](#_Toc1319480)

[4.1.1 Starch, its synthesis and break down 110](#_Toc1319481)

[4.1.2 The AGP complex 110](#_Toc1319482)

[4.1.3 Regulation of ADG2 and starch 112](#_Toc1319483)

[4.1.4 Starch and regulation of stomatal function 114](#_Toc1319484)

[4.1.5 Aims 114](#_Toc1319485)

[4.2 Methods 115](#_Toc1319486)

[4.2.1 Plant Material 115](#_Toc1319487)

[4.2.2 Plant Growth 116](#_Toc1319488)

[4.2.3 DNA extractions 116](#_Toc1319489)

[4.2.4 Polymerase Chain Reaction (PCR) 116](#_Toc1319490)

[4.2.5 Agarose Gel Electrophoresis 118](#_Toc1319491)

[4.2.6 RNA Extraction 118](#_Toc1319492)

[4.2.7 cDNA Synthesis 118](#_Toc1319493)

[4.2.8 qPCR Analysis 119](#_Toc1319494)

[4.2.9 Sequencing 121](#_Toc1319495)

[4.2.10 Stomatal Impressions 121](#_Toc1319496)

[4.2.11 Starch Assay – Staining 121](#_Toc1319497)

[4.2.12 Starch Assay – Enzyme Quantification 122](#_Toc1319498)

[4.2.13 Exon Usage analysis 123](#_Toc1319499)

[4.2.14 Promotor Analysis 124](#_Toc1319500)

[4.2.15 Statistical Analysis 124](#_Toc1319501)

[4.3 Results 125](#_Toc1319502)

[4.3.1 *ADG2* Phenotyping 125](#_Toc1319503)

[4.3.2 Cotyledon Size and Stomatal Density 127](#_Toc1319504)

[4.3.3 Phenotype in Mature Leaves 129](#_Toc1319505)

[4.3.4 *ADG2* Sequence in 19 parental lines 130](#_Toc1319506)

[4.3.5 *ADG2* Sequence in Can-0 and Wu-0 134](#_Toc1319507)

[4.3.6 Promotor analysis 134](#_Toc1319508)

[4.3.7 Anti-Sense RNA 138](#_Toc1319509)

[4.3.8 The AGP Complex 138](#_Toc1319510)

[4.3.9 Starch Deficiency 140](#_Toc1319511)

[4.3.10 *ADG2* Circadian Expression Analysis 142](#_Toc1319512)

[*4.3.11* Circadian Starch Content of Ecotypes and *adg2-3* 143](#_Toc1319513)

[4.3.12 Alternative Splicing 149](#_Toc1319514)

[4.4 Discussion 152](#_Toc1319515)

[4.4.1 QTL5 and ADG2 152](#_Toc1319516)

[4.4.2 Starch and Stomatal density 153](#_Toc1319517)

[4.4.3 Conclusions 156](#_Toc1319518)

[5 The Effect of ERECTA Transcript and Splicing on the Regulation of Stomatal Density 157](#_Toc1319519)

[5.1.1 ERECTA is a pleiotropic gene 157](#_Toc1319520)

[5.1.2 QTLs and ERECTA 157](#_Toc1319521)

[5.1.3 Regulation of ERECTA through introns 158](#_Toc1319522)

[5.1.4 Regulation of ERECTA through epigenetics 159](#_Toc1319523)

[5.1.5 ERECTA and [CO2] 160](#_Toc1319524)

[5.1.6 Aims 160](#_Toc1319525)

[5.2 Methods 161](#_Toc1319526)

[5.2.1 Plant Material 161](#_Toc1319527)

[5.2.2 Plant Growth 161](#_Toc1319528)

[5.2.3 RNA Extraction 162](#_Toc1319529)

[5.2.4 cDNA synthesis 162](#_Toc1319530)

[5.2.5 qPCR 162](#_Toc1319531)

[5.2.6 Exon Usage analysis 162](#_Toc1319532)

[5.2.7 Statistics 162](#_Toc1319533)

[5.3 Results 164](#_Toc1319534)

[5.3.1 *ER* sequence in 19 parental lines 164](#_Toc1319535)

[5.3.2 Epigenetic regulation of *ERECTA* gene expression 167](#_Toc1319536)

[5.3.3 *ERECTA* gene expression levels in 19 parental ecotypes 169](#_Toc1319537)

[5.3.4 Expression of *ER* Exons 171](#_Toc1319538)

[5.3.5 Expression of ER exons between [CO2] treatments 173](#_Toc1319539)

[5.3.6 Expression of ER correlates with SD 176](#_Toc1319540)

[5.3.7 Alternative Splicing in ERECTA 177](#_Toc1319541)

[5.3.8 Changes to *ER* expression level in response to short term environmental change 179](#_Toc1319542)

[5.4 Discussion and Conclusions 181](#_Toc1319543)

[5.4.1 ER Expression and Stomatal Density 181](#_Toc1319544)

[5.4.2 Alternative Splicing 183](#_Toc1319545)

[5.4.3 Conclusion 185](#_Toc1319546)

[6 General discussion 186](#_Toc1319547)

[6.1 Introduction 186](#_Toc1319548)

[6.2 The Stomatal Density in 19 Arabidopsis ecotypes 186](#_Toc1319549)

[6.3 QTLs in MAGIC lines 187](#_Toc1319550)

[6.4 The role of starch in SD regulation 188](#_Toc1319551)

[6.5 The role of ER in the control of SD 190](#_Toc1319552)

[6.6 Future directions 191](#_Toc1319553)

[6.7 Potential applications 193](#_Toc1319554)

[6.8 Concluding Remarks 194](#_Toc1319555)

[7 References 195](#_Toc1319556)

[8 Supplementary Information 215](#_Toc1319557)

# General Introduction

## Global Climate Change

The atmospheric concentration of carbon dioxide (CO2­) has been steadily rising since the industrial revolution, from approximately 280ppm in the 1700s to a current value of 412ppm [2018] (NOAA, 2018), because of burning fossil fuels for energy (IPCC 2014). This represents a large increase in atmospheric CO2 concentration ([CO2]) in a very short space of time. It is estimated that levels could exceed 700ppm if no action is taken to curb emissions (Figure 1.1.1) with some emissions scenarios putting the predicted concentration as high as 950ppm by 2100 (IPCC 2014). The emission scenarios are called Representative Concentration Pathways (RCPs) and there are 4 suggested representing high emissions (RCP8.5), medium (RCP4.5), low (RCP4.5) and carbon negative (RCP2.6).

|  |
| --- |
|  |
| Figure 1.1.1 IPCC future emissions scenarios based on changes to behaviour with the high (RCP8.5), medium (RCP6) and low (RCP4.5) and carbon negative (RCP3) and the resulting CO2 concentration. From IPCC 2014 |

These changes to the atmospheric composition are expected to have wide reaching effects on our climate. Current climate models expect that if we stay on our business as usual path of high emissions, as shown by the RCP8.5 in Figure 1.1.1, we can expect global temperatures to rise by as much as 4°C with significant changes to precipitation and an increase frequency and intensity of weathering extremes (e.g., floods, droughts, heatwaves) (IPCC, 2014). This is likely to cause problems with agricultural yields and food security as increasing temperatures and drought affect crop production negatively (Alqudah et al, 2011).

Plants are huge drivers of the carbon cycle, removing around 25% of carbon emissions from the atmosphere every year through gas exchange (Sun et al, 2014). It is estimated that they cycle approximately 120 × 1015 g C (440 × 1015 g CO2) from the atmosphere's 730 × 1015 g C (Ciais et al, 1997) each year. It is therefore important to understand how plants will respond to changes in atmospheric make up and climatic shifts.

### Food Security

Sugarcane, maize, rice and wheat were the four crops with the highest volume of production (in tonnes) in 2013 (FAO, 2015) and all are predicted to be impacted by changes to the global climate and atmospheric composition (Zhao et al, 2017; Zhao and Li, 2015). As plants utilise CO2 during photosynthesis to produce sugars, and the amount of sugar available impacts yield, the increase in atmospheric [CO2] is predicted to have a positive effect on crop production through the process of CO2 fertilisation (Ainsworth and Long, 2005). However, the confounding factors of climate change, such as extreme heat events, are expected to have negative effects on crop yield (Zhao et al, 2017). Zhao et al (2017) estimated that, without CO2 fertilisation, for every degree increase in temperature, there would be a decrease in global yields of wheat by 6.0%, rice by 3.2%, maize by 7.4% (2017).

This expected decrease in yield is in addition to the expectation that by 2050, in comparison to 2005, we will need to double food production. If increases in production are not done in a sustainable and intensive manner, it could require up to 1 billion hectares of land to be cleared, releasing approximately 3Gt y-1 of carbon (Tilman et al, 2011). With climate change already expected to have a major impact on the global environment, including loss of rainforest from drought (Nepstad, 2008), loss of grasslands (Zhang et al, 2018) and ice sheet melt and sea-level rise (IPCC 2014), it is imperative that we minimize the effects of increased food production on our natural ecosystems. The required changes to crop production to increase food yields will need to include developing more drought tolerant and water use efficient plants to ensure that the changing water cycle does not impact yields in a negative manner.

### Plant physiological Responses to [CO2]

Land plants evolved approximately 972.4 to 669.9 million years ago (Morris et al, 2018) and have existed through CO2 as low as 180ppm in the last glacial maximum 18000-2000 years ago (Gerhart et al, 2010). The physiological responses of modern plants to changes in [CO2] have been under investigation for several decades. Changes to plant growth, physiology and reproduction can all be seen in sustained elevated or subambient [CO2] (Jauregui et al, 2015; Ainsworth and Long, 2005; Ward and Strain, 1997), particularly between very low [CO2] (180ppm) and modern concentrations (Ward and Strain, 1997).

It has been shown in a meta-analysis that numerous vascular plant species grown in elevated [CO2], increase growth and yield (Kimball, 2016). Higher [CO2] conditions increased photosynthesis and therefore increased availability of starch and sugars for growth and reproduction (Ainsworth and Long, 2004). The increased [CO2] also allows the plant to modify its development and behaviour, such as changes to flowering time (Springer and Ward, 1997) and to the number of stomata on leaves (Woodward and Kelly, 1995). Gas exchange takes place through stomata (Ainsworth and Long, 2005) and changes to stomata number and physiology in elevated [CO2], both long term and short term changes, can have an effect on plant gas exchange (Ainsworth and Long, 2005) which over large scales can result in climatic changes.

Transpiration from plants contributes as much as 70 × 1015 kg water to the atmosphere every year making it a significant contributor to the world’s water cycle (Hetherington and Woodward, 2003).A study by Rigden and Salvucci suggests that over the past 50 years there has been in a reduction in overall evapotranspiration across the United States in response to rising [CO2], due to changing plant physiology effecting gas exchange. Perturbations to evapotranspiration can cause changes to future humidity and future transpiration which can in turn feedback on whole ecosystem gas exchange (Scholze et al, 2006).

Molecular genetic controls underpinning changes to density of stomata in response to growth [CO2] levels are the subject of this thesis.

## Stomata Physiology and Trade-offs

Stomata are microscopic pores in the leaf through which gas exchange takes place. Stomata are found on both the abaxial (lower) and adaxial (upper) side of the leaf and allow CO2 in and water and oxygen and water out. In the simplest form stomata are formed of a pair of guard cells which become more turgid in order to open and more flaccid in order to close. This movement is caused by internal and external stimnuli causing changes to water pressure within the guard cells through changes to osmolites in the guard cells (Araújo et al al, 2011).

As sessile organisms, plants must be able to maintain homeostasis mostly without moving. A method the sessile plant can use to maintain homeostasis is through the manipulation of its internal environment through increased or decreased gas exchange. Stomata respond developmentally and physiologically to a range of stimuli including [CO2] (the subject of this thesis), light, humidity, drought, pathogens, and altitude (Hetherington and Woodward, 2003).

Stomata can respond to their external environment in two ways, the results of which can are pictorially represented in Figure 1.2.1. In the short term, plants respond by changing their aperture, which is how open or closed the pore is, resulting in altered gas exchange properties. Closing stomata reduces water loss which can be advantageous to a plant, particularly in drier conditions but can in reduce photosynthetic rate. In elevated [CO2] however, it has been shown that plants reduce their aperture as they are more able to absorb the same amount of CO2 through a smaller aperture, thus reducing water loss whilst maintaining photosynthetic capability (Engineer et al, 2016).

In the longer term, plants respond developmentally by changing their stomatal density (SD), which is the number of stomata per unit area. Changes to a plant’s stomatal density will alter their maximum and minimum gas exchange capabilities. A plant with more stomata will be able to carry out more gas exchange (Franks et al, 2009; Dow and Bergman, 2014), however, it also leaves it open to more transpiration loss, leading to a reduced water use efficiency (WUE) and potentially lower drought tolerance (Franks et al, 2015). In elevated [CO2], the plant is able to uptake the same amount of CO2 through fewer stomata and so around 75% of plants reduce their stomatal density when exposed to elevated [CO2] (Woodward and Kelly, 1995) potentially leading to a more WUE plant. Further discussion of the stomatal density response to [CO2] can be seen in Section 1.3.

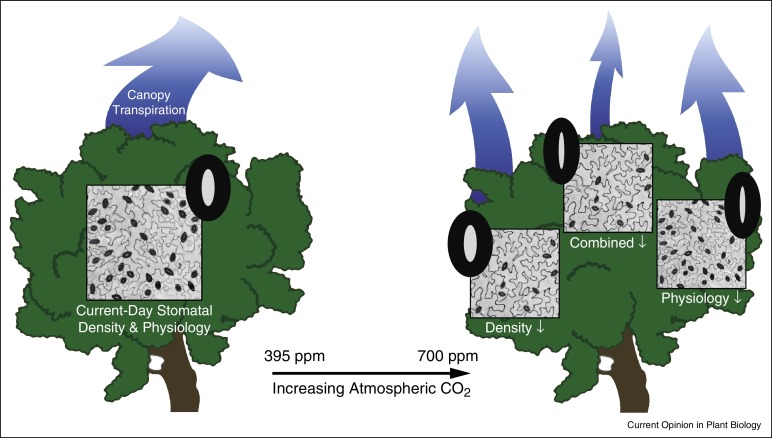


Figure 1.2.1 Pictorial representation of the expected changes to transpiration from plants in the future due to changes in both stomatal density and stomatal aperture from Dow and Bergman, 2014

### Stomatal Density

Stomatal Density of plants varies largely depending on plant species and location of collection. From the natural environment *Arabidopsis thaliana*, the model organism for dicotyledons, ecotypes’ stomatal density has been shown to vary by up to 3.5 fold (Delgado et al, 2011). Genetically manipulated Arabidopsis plants grown in controlled environment growth chambers showed a range in densities of 16 fold, showing that there is a huge potential to manipulate stomatal densities (Doheny-Adams et al, 2012), but that environmental conditions put genetic restrictions in the natural world.

The changes to SD of the genetically manipulated plants grown in controlled environment growth chambers were correlated with WUE, the amount of water used per unit of growth, with lower SD plants showing higher WUE (Doheny-Adams et al, 2012; Hepworth et al, 2015; Caine et al, 2018). However, studies investigating natural populations grown in the lab could not confirm this result. Dittberner et al found no correlation between stomatal density and WUE in ecotypes grown in controlled environments (2018), suggesting that stomatal density is not the primary method used by wild Arabidopsis to improve its WUE.

Decreased stomatal density has also been correlated with increased defence against plant pathogens (Tateda et al, 2018). Many plant pathogens have been shown to enter through the stomata, a natural gap in the plant’s defences (Underwood et al, 2007). Plants with a lower stomatal density can therefore have a lower susceptibility to pathogens as there is less opportunity for invasion (Tateda et al, 2018)

A reduction in stomatal density can have counterproductive effects on the plant as well as the beneficial increase in WUE and pathogen resistance. Transpiration is used by the plant for both heat loss and to generate mass flow throughout the plant. A reduction in stomatal density can make it difficult for a plant to cool itself and could lead to over-heating (Doheny-Adams et al, 2012). This would be a disadvantage to a plant in a non-controlled environment setting which may have to respond quickly to changes in temperature. Research by Caine et al (2018) however showed that rice with reduced stomatal densities still performed better in higher temperatures than those with reduced stomatal densities suggesting that SD may not be an issue. Caine et al (2018) suggest instead that changes in aperture are more important in high temperatures, and that plants with reduced SD have increased water availability in order to more properly respond to changes in temperature.

Changes in SD have also been suggested to lead to reduced mass flow which can result in reduced uptake of minerals leading to a nutrient starved plant (McDonald et al, 2002). McDonald et al, 2002 showed that plants grown in elevated [CO2] had a reduced nitrogen content compared to those grown at ambient [CO2] due to decreased transpiration. Contrary to this, Hepworth et al, (2015) showed that Arabidopsis with reduced stomatal density did not have reduced nutrient content. The reduction in Carbon to Nitrogen ratio (C:N) in elevated [CO2] can be for reasons other than changes in nutrient uptake. The down regulation of rubisco and other photosynthetic components can also lead to a reduction in C:N as higher [CO2] results in more efficient carbon fixation as there is less competition for Rubisco’s active site from oxygen meaning less is required (Nakano et al, 1997).

Plants must therefore balance the need to minimise water loss with the need for cooling and mass flow. The requirement for balance may be a reason for the restriction on the natural variation in stomatal density compared to that of the lab manipulated Arabidopsis lines.

### Other Stomatal traits

It is necessary to consider other traits of the epidermis in order to properly understand stomatal density. Changes in stomatal density can be achieved in a number of ways; including an increase in stomatal initiation or Stomatal Index (SI), increased epidermal cell division, or increases in overall epidermal cell expansion (see Figure 1.2.2 for examples). SI is the percentage of stomata in the total cell population of a leave’s epidermis. It is typical calculated by dividing the number of stomata by the total number of cells, including stomata. It is a good indication that stomatal density changes are a result of perturbations to stomatal initiation, rather than other changes in epidermal physiology, such as epidermal cell size.

Beyond the SD of a leaf, other stomatal properties can impact on the gas exchange. For example, Stomatal Size (SS) can also have an effect on gas exchange capacity. The equation for finding gas exchange capacity of a leaf if found using the equation of Franks and Beerling (2009) below.

where *gw* is stomatal conductance, *d* is the diffusivity of water in air (m2s−1), *D* is stomatal density (number per unit leaf area, m−2), *a* is the average stomatal pore area (m2), *v* is the molar volume of air (m3mol−1), *l* is the depth of the stomatal pore (taken as equal to guard cell width) and π = 3.142. This calculation suggests that for control of gas exchange, SD and stomatal pore area/size (SS) are the main controls of capacity. When considering methods of altering gas exchange, either SS or SD can affect stomatal conductance and both can be considered.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |
|  |  |
| Figure 1.2.2 Stomatal densities with different stomatal indexes. The stomatal density of two leaves can be similar but the index is different because of the way the plant achieved that density. The change in density can either through increased initiation of stomata, increased cell size or increased cell division. | |

SS does not cause changes in stomatal density, however it is closely related. There is a strong negative correlation between SS and SD (Franks et al, 2009). One of the dominating theories for this is that it allows proper spacing of stomata and presents stomata from taking up too much space on the epidermis (Franks et al, 2009). In Eucalyptus, it was shown that the overall percentage area stomata take up was fixed at around 5%. As the SD of Eucalyptus leaves increased, the SS decreased proportionally to ensure the stomata did not cover more than 5% of the leaf epidermis (Franks et al, 2009). This relationship between SD and SS was also found in a study that looked at 330 ecotypes of Arabidopsis (Dittberner et al, 2018). Dittberner et al (2018) went on to show that plants with smaller stomata are more water use efficient and that there is a correlation between SS and WUE.

SS is important to consider when looking at gas exchange capacity, as larger stomata have a larger gas exchange capacity (Franks et al, 2009). However, due to time constraints, density was chosen as the main topic of this thesis.

### Epidermal Cells Size

Epidermal cell size is a function of cell expansion and cell division and this can have a strong effect on stomatal density independently of SI. Carins Murphy et al (2016) propose that it is cell expansion that is the primary driver of SD in herbs and woody plants and call the phenomenon ‘passive dilution’. It is therefore important to briefly consider some of the controls of cell size and division.

Cell number and leaf size are highly correlated in some species (Granier et al, 2000). In Arabidopsis, an analysis on cell size and cell number identified the gene *ERECTA* (see Section 1.5.2 and chapter 5) as a control of both epidermal cell size and number indicating that both factors are sometimes regulated by the same gene (Tisne et al, 2008). However, it has also been demonstrated that some cell cycle mutants with increased division and cell numbers do not result in increased leaf size, suggesting that cell size and division can be regulated independently of each other (Dewitte et al, 2003).

Cell size and number are under the control of the cell cycle which decides when a cell stops expanding and divides (Sablowski, 2016). The cell cycle is controlled by homeostatic feedback loops within the cell, which take into consideration turgor pressure, phytohormones, sugar availability, and nutrient availability (Sablowski, 2016) and many of these are known to affect stomatal density too. Studying the SD solely by looking at stomatal initiation is therefore missing out a part of the puzzle. Studying cell size and index can give clues as to what is controlling SD which is the subject of this thesis.

## Variation in [CO2] levels

### Past changes to [CO2] increases

Studying past plant responses to elevated [CO2]can give us an indication as to how plants have previously adapted to high levels of [CO2]. Over geologic time, it can be seen in fossilised plant leaves that plants reduce their stomatal density in times of elevated [CO2] (Franks and Beerling, 2009). The trend in reduction in SD in elevated [CO2] can also be seen in the time between the last ice age and the present (Van de Water et al, 1994). Recent changes in SD in response to rises in [CO2] since the industrial revolution were shown by Woodward (1987) who found that some woody species have reduced their stomatal density by up to 40%. Previous plant responses to elevated [CO2] have therefore shown a consistent reduction.

Plants in the past will also have adapted to other changes in their environment that accompanied the changes in [CO2] levels. The change in temperature between 1880 and 2015 is approximately 1°C ([Dahlman](https://www.climate.gov/author/luann-dahlman), 2017) and while the change in temperature between the larger fluctuations in [CO2] concentrations cannot be completely known, Retallack, 2002 showed that increases in atmospheric [CO2] are often accompanied by increases in temperature. Importantly, Beerling and Chaloner (1993) showed that temperature is a stronger control on density than irradiance intensity and small seasonal (⩽ 10 ppmv) variations of [CO2]. The assumption that the changes in density are solely related to [CO2] level therefore cannot completely exclude the accompanying changes in climate that come with increases in [CO2]. Despite this, multiple papers have shown that predictions of atmospheric [CO2] concentration from leaves’ SD have matched modelled data from other sources suggesting that negative changes in SD are caused by increasing [CO2] (Sun et al, 2017; Sun et al, 2007), however it cannot be ruled out that this is from adaption, rather than acclimation.

In addition, past [CO2] levels are believed to have changed over much more slowly than today, giving plants time to adapt. In past ice ages, the largest change in [CO2] levels in 1000 years was an increase of 17ppm, whereas today [CO2] has increased 150ppm in 150 years (Bereiter et al, 2015). Even during periods when the [CO2] rose at a speed similar to today’s 150ppm in 150 years, plants may have adapted by the time they were fossilised. Current levels have increased 150ppm in 150 years meaning any adaption expected to occur may not have had time to take place. Instead it is necessary to study how plants not adapted to high [CO2] respond to further increases in [CO2] to understand the expected effects of our atmospheric change in the near-future.

### Laboratory and greenhouse and controlled environment experiments

To study how near-future changes to [CO2] will impact plants, many plants have been studied in labs under artificial [CO2] levels.

A meta-analysis by Woodward et al showed that around 75% of plants show a reduction in stomatal density in elevated [CO2](Woodward and Kelly, 1995) showing that not all plants have the expected physiological change. It has been shown that *Quercus-robur* (Beerling and Chaloner, 1993), *Pinus sylvestris* (Beerling, 1997) Douglas fir (Apple et al, 2000), and Maize (Driscoll et al, 2006) did not show a change in density in elevated [CO2]. However, species such as Arabidopsis (Caldera et al, 2016), Poplar (Ceulemans et al, 1995), and Olive (Tognetti et al, 2001), did show a reduction in density. A recent meta-analysis by Yan et al (2017) showed a significant decrease in stomatal density across tree species (*Betula pendula, Quercus petraea, Tsuga heterophylla, Salix herbacea*). Each species should therefore be considered seperately.

In addition, there can be variation within species as to how plants respond to changes. For example, a study of 18 ecotypes grown in ambient and elevated [CO2] showed only 5 Ecotypes has a decrease in stomatal density in elevated [CO2] and 8 showed an increase (Caldera et al, 2016). The ecotype variation of Arabidopsis seen by Caldera et al (2016), was further confirmed by the work carried out in this thesis.

### Free Air Carbon-dixide Enrichment (FACE) Experiments

The above laboratory and greenhouse experiments do not account for real world interactions with the environment and make it difficult to say if their conclusions are true in real world environments. To counter this, Free Air CO2 Experiment (FACE) experiments have been set up. These experiments study the response of whole fields or whole ecosystems to increased [CO2] *in situ*, by pumping CO2 into the air into the plant canopy (Allen 1992).

Some experiments using FACE have shown reductions in stomatal density in high [CO2] for many species (Drake, 1997) despite interacting environmental effects, but the response is not universal (Reid et al, 2003). Soybean plants were shown not to change their SD in response to elevated [CO2] in FACE experiments (Levine et al, 2009) and it was demonstrated that when drought is introduced alongside elevated [CO2] the expected increases in yield due to CO2 fertilisation are not seen. Lemon trees exhibited better growth in droughted conditions in elevated [CO2] compared to ambient (Paudel et al, 2018). However elevated [CO2] only compensated for drought, droughted trees in elevated [CO2] did not have increased growth compared to ambient [CO2] grown non-droughted trees. This suggests that while changes to SD in response to elevated [CO2] may help to mitigate climate change effects, they cannot eliminate them. It is likely that the expected benefits of increased yields seen previously in FACE experiments (Van der Kooi et al, 2016) will not be as great as predicted.

The genetic “pathways” which control stomatal responses are large and there is crosstalk between signalling pathways induced by many of the environmental stimuli. There are pathways which are common to many of these stimuli and others which are specific to responses. The central pathways and the [CO2] specific response are discussed in sections 1.5 and 1.6.

## Stomatal Aperture

While stomatal aperture is not the subject of this thesis, it is important to discuss its relevance to elevated [CO2]. As the short term response to changes in the environment, changes to aperture are the first response of the plant to elevated [CO2]. The plant can only change its gas exchange rate up to the point of its gas exchange capacity (see 1.2.2), which is controlled by density (Franks and Beerling, 2009) making density and stomatal conductance intricately linked. However, aperture is the final control over evapotranspiration and gas exchange and the immediate response.

Typically, plants respond to increases in [CO2] concentration by decreasing their stomatal aperture, although exceptions to this have been found leading to a decrease in transpiration. The decrease in transpiration response has been seen in FACE experiments, laboratory experiments and in long term environmental monitoring. A meta-analysis of FACE experiments found that on average there is a reduction of 22% in stomatal conductance in elevated [CO2] (Ainsworth and Rogers, 2007) although 11.8% of species in FACE experiments were shown to have an increase in gas exchange rates in elevated [CO2] (Purcell et al, 2018). Purcell et al (2018) also found that species *in situ* exposed to an increase in [CO2] of 50ppm showed negligible change or an increase in gas exchange rate, however this was carried out on a leaf basis, not a whole plant basis.

The overall reduction in gas exchange rate can also be seen over whole continents. Rigden & Salvucci (2017) suggest that between 1961 and 2014 there was a 6% reduction in evapotranspiration over the United States which they attribute partially to a reduction in gas exchange due to changes in [CO2] and changes to atomospheric humidity. FACE experiments have also suggested that the decrease in gas exchange will be independent of the warming effect of climate change and be due to changes in [CO2] affecting SD and stomatal aperture (Kirschbaum & McMillan, 2018). Kirschbaum & McMillan (2018) found that it was likely that the effect of warming on gas exchange would not be as strong an influence as that of [CO2] and so there should still be a transpiration depression effect. Overall, the evidence suggests that there will be an overall decrease in gas exchange in elevated [CO2] and therefore decrease in transpiration from aperture changes.

## Stomatal Development

### The Arabidopsis Core Development Pathway

Cells develop into stomata through a series of asymmetric divisions with a terminal symmetric division. Development begins with protodermal cells undergoing differentiation into a meristemoid mother cell (MMC), to a meristemoid, then a guard mother cell (GMC), and then a guard cell (See Figure 1.5.1). This developmental process is known as the stomatal lineage (Zoulias et al, 2018) and is controlled by 3 transcription factors called SPEECHLESS (SPCH), MUTE and FAMA (MacAlister et al, 2007; Pillitteri et al, 2007; Ohashi-Ito and Bergmann, 2006). These are homologous basic helix-loop-helix (bHLH) transcription factors and loss of function mutants of any of them result in severely disrupted stomatal development (MacAlister et al, 2007; Pillitteri et al, 2007; Ohashi-Ito and Bergmann, 2006).

The first step in the stomatal differentiation pathway is controlled by SPCH which causes protodermal cells to undergo differentiation into MMC cells. Without active SPCH, plants do not develop stomata, as no cells can enter the stomatal lineage. Plants without active SPCH result in only pavement cells being found on the epidermis (MacAlister et al, 2007). Plants without stomata do not become mature, instead arresting as small pale seedings, unless kept in very high CO2 concentration environments (Leakey, 2016,). On the other hand plants with an over-expression of *SPCH* results in a hugely increased number of MMCs and ectopic cell divisions, however the MMCs do not all go on to differentiate into guard cells (MacAlister et al, 2007).

After the MMC cell, the cell asymmetrically divides into a daughter cell, also known as a stomatal lineage ground cell (SLGC), and a meristemoid. This division is also driven by SPCH (Lau et al, 2014). If there is a stomatal guard cell next to the developing meristemoid, it will undergo multiple rounds asymmetric (amplifying) divisions in order to properly orientate itself to keep the one cell spacing rule (see Section 1.5.3) (Lau et al, 2014). The one spacing rule ensures there is always at least one non-stomatal cell between guard cells in order to allow proper stomatal functioning (Nadeau and Sacks, 2002).

MUTE controls the differentiation of the meristemoid into a GMC, the final stage before commitment to guard cells (Pillitteri et al, 2007). Like plants lacking in SPCH, plants without MUTE do not form stomata as they are unable to finish the stomatal lineage. Instead they generate an excessive number of meristemoids which cannot progress further down the lineage (Pillitteri et al, 2007). Arabidopsis plants which over-express MUTE create an epidermis that almost entirely consists of stomata tha ignore the one cell spacing rule (Pillitteri et al, 2007).

The final fate of the GMCs is controlled by FAMA, which causes the terminal differentiation of the GMC into two guard cells following a symmetrical division. *FAMA* knock out mutants do not produce mature stomata, instead they produce clusters of small epidermal cells which are referred to as FAMA tumours. Similarly to MUTE over expressers, FAMA over expressers produce a guard cell like epidermis except that the cells are unpaired guard cells indicating FAMA is responsible for guard cell identity (Ohashi-Ito and Bergman, 2006).

SPCH, MUTE and FAMA do not act independently, instead form heterodimers with two other bHLH transcription factors, *INDUCER OF CBF EXPRESSION 1 (ICE1)/SCREAM* (*SCRM*) and *SCREAM2* (*SCRM2*) (Kanaoka et al, 2008). SCRM and SCRM2 are critical for plant development with null alleles of SCRM being plant lethal. Knock outs of *SCRM/SCRM2* phenocopy null mutants of SPCH, MUTE and FAMA (Kanaoka et al, 2008) The SCRM-D mutant leads to an over proliferation of stomata on the leaf epidermis, and many smaller epidermal cells indicating a positive drive on stomatal development. SPCH is required to induce *SCRM* expression and it is suggested that this allows transiently expressed MUTE and FAMA to be present without causing differentiation before SCRM and SCRM2s expression upon initiation by SPCH (Kanaoka et al, 2008). This simple developmental pathway, is balanced through the integration of spatial and environmental signals.

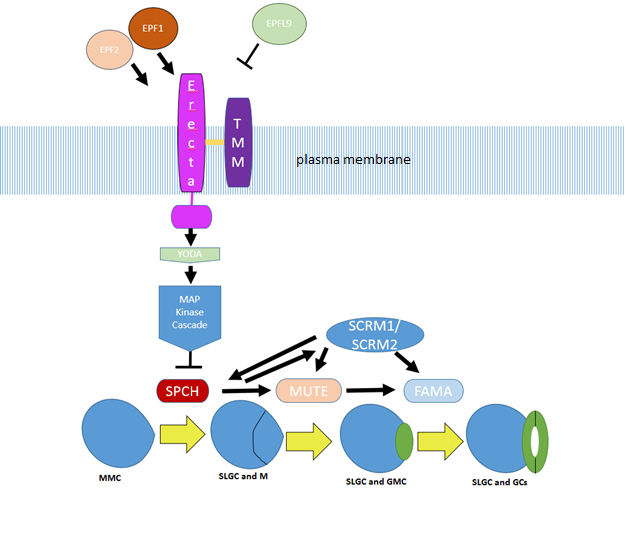


Figure 1.5.1. The stomatal development pathway and peptide signalling pathway which regulates stomatal development. EPF2, EPF1 and EPFL9/Stomagen can be seen at the top,activate and inhibit ERECTA and TMM complex which in turn regulates the MAP Kinase pathway. An activated MAP kinase pathways phosphorylates SPCH, which causes SPCH to become inacvtive and arrests stomatal development. The bHLH transcription factors associated with the stomatal lineage,SPCH, MUTE FAMA, SCRM and SCRM2 can be seen above the stages of the stomatal lineage which they regulate.. The Progression of the stomatal lineage occurs as follows; the Meristemoid Mother Cell (MMC) divides asymmetrically to become a Stomatal Lineage Ground Cell (SLGC) and a Meristemoid (M). The meristemoid becomes a Guard Mother Cell (GMC) which then divides symmetrically into a pair Guard Cells (GC).

### Integration of Signals into Stomatal Development

Many other genes are required to create proper stomatal patterning and to integrate environmental signals into the bHLH transcription factor which regulate the stomatal lineage. One of the most well researched and an important signalling pathway in the regulation of stomatal patterning and development is the peptide signalling pathway. The signalling peptides EPIDERMAL PATTERNING FACTOR 1 and 2 (EPF1, EPF2) and STOMAGEN (STOM, also known as EPIDERMAL PATTERNING FACTOR LIKE 9, EPFL9) act to negatively or positively regulate stomatal development respectively(Hara et al, 2007; Hunt and Gray, 2009; Sugano et al, 2010; Lin et al, 2017). EPF1 and EPF2 reduce stomatal density by the activation of the MAPK pathway, through binding to cell surface receptors of the ERECTA family of proteins and TOO MANY MOUTHS (TMM) (Hara et al, 2007; Hunt and Gray, 2009; Lin et al, 2017). STOM, on the other hand, upregulates stomatal initiation through its binding to the same receptors, blocking EPF1 and EPF2 from binding the receptors (Sugano et al, 2010). How [CO2] affects this pathway is discussed in section 1.6.

EPF1 and EPF2 knock-outs result in similar phenotypes however they are produced and excreted at different times in stomatal development (Hara et al, 2009). EPF2 is produced in MMCs and meristemoids while EPF1 is produced in late GMCs and late meristemoids. EPF2 is secreted into neighbouring cells by meristemoids to prevent entry into the stomatal lineage. EPF1 is secreted to properly orientate spacing (Hara et al, 2009). It has been predicted that EPF proteins need to be cleaved in order to become mature secreted peptides (Hara et al, 2007), and it has been shown that the proteases CRSP (CO2 RESPONSE SECRETED PROTEASE) cleaves EPF2.

The ERECTA family of receptors (ERf) are Leucine Rich Repeat Receptor like Kinases and consist of ERECTA (ER), ERECTA like 1(ERL1), and ERECTA like 2 (ERL2). Mutations in *ER* result in an increased stomatal density and an increase in the number of arrested SLGCs (Torii et al, 1996). Primarily, ER restricts the entry divisions into the lineage pathway, while ERL1 and ERL2 maintain the cells ability to proliferate in the earlier stages of the pathway (Shpak et al,2005). In order to properly perceive the signals from EPF1 and EPF2 and other signalling proteins, the ERf proteins must be in complex with LRR receptor-like protein TMM. When not in complex with TMM, ERf proteins cannot discriminate between ligands (Guangzhong et al, 2017). ERECTA is discussed further in chapter 5.

Similarly to ERECTA, *TMM* knock outs results in an excess of stomata that can be found in clusters (Yang and Sack, 1995). EPF2 and EPF1 bind to TMM in complex with ERf to facilitate downstream signalling (Lin et al, 2017). Without TMM, EPF2 and EPF1 do not bind to the receptor complex leading to the clustered stomata phenotype (Lin et al, 2017). TMM requires a signalling partner because it does not have an intracellular signalling domain, meaning it cannot itself signal downstream (Abrash et al, 2011) and has to form a complex with ERECTA family proteins. It is therefore suggested that TMM acts in the lineage pathway by providing specificity in the response to different ligands in complex with ERf proteins, making them essential partners in controlling stomatal density and patterning (Lin et al, 2017).

Once activated, the complex of ER, ERLs and TMM activates the downstream mitogen activated protein (MAP) kinase kinase kinase YODA, which starts the MAP kinase (MAPK) cascade that eventually inhibits SPCH through phosphorylation and stomatal initiation. YODA, which is also a controller of extra-embryonic cell fates (Lukowitz et al, 2004), is a regulator of stomatal patterning and development (Bergmann et al, 2004). In *yoda* knock-out mutants, stomatal patterning is disrupted, with excessive stomata that are also clustered. In YODA over-expressers, stomata are not present (Bergmann et al, 2004).In constitutively active YODA, which phenocopies *spch*, the downstream MAPK kinases are more highly activated, suggesting it is the start of the MAPK cascade regulating stomatal development (Wang et al, 2007).

The MAPK cascade is the final step in control of stomatal initiation before SPCH. It has been shown that MAPKs play an important role in ensuring proper spacing and stomatal initiation, particularly MITOGEN-ACTIVATED PROTEIN KINASE 3 (MPK3) and MITOGEN-ACTIVATED PROTEIN KINASE 6 (MPK6) (Wang et al, 2007). Double loss of function mutants of *mpk3mpk6*results in clusters of stomata that do not follow the one cell spacing rule (Wang et al, 2007).The MAPKs phosphorylate SPCH, downregulating it which causes reduced stomatal initiation and development (Lampard et al, 2009). Many environmental factors act to change stomatal density through this MAPK signalling pathway including light, CO2 and humidity (Pillitteri and Dung, 2012).

The full pathway mentioned here can be seen in Figure 1.5.1

### The One Cell Spacing Rule

The one cell spacing rule mentioned in section 1.5.1 is strictly controlled by genetic factors. Stomata require at least one cell space between them and another cell to ensure proper stomatal functioning. (Nadeau and Sach, 2002) This is because they need space to expand with turgor pressure for opening, and to receive sources of osmolites from their surrounding cells to control turgor pressure (Outlaw, 1983).

While WUE has been shown not to be affected by stomatal clustering, it has been shown that gas exchange is affected by improper spacing (Dow et al, 2013). This is due to impaired stomatal opening and closing as guard cells press against each other while trying to open and close (Franks and Farquhar, 2007).

Any genetic change that affects stomatal density must therefore maintain this spacing in order to maintain proper functioning.

## The Genetic Pathway for [CO2] Response

The genetic mechanisms for causing the decrease or increase in stomatal density in response to elevated [CO2] is poorly understood. In recent years, several genes have been implicated in a suggested pathway regulating a reduction in stomatal density in elevated [CO2], however there is no definitive pathway.

Carbonic Anhydrase 1 (CA1) and Carbonic Anhydrase 4 (CA4) were found to be the initial steps in the detection of [CO2] (Engineer et al, 2014). These proteins bind CO2 and carryout the reaction that combines CO2 with water to produce bicarbonate and a hydrogen ion which is used by PEPc in the citric acid cycle (Engineer et al, 2014). They were initially found to be important in the detection of [CO2] to control stomatal aperture (Hu et al, 2009). However, CA1 and CA4 were also subsequently found to be important in stomatal density (Engineer et al, 2014). Without carbon anhydrases, an increase in stomatal density in elevated [CO2] is observed in changes between 150ppm and 500ppm (Engineer et al, 2014).

Analysis of the *ca1*/*ca4* double mutant revealed that the insensitivity to elevated [CO2] was due to a lack of EPF2 upregulation compared to wildtype (Engineer et al, 2014). This indicated that the [CO2] response pathway was occurring through the known stomatal development pathway explained in Section 1.5.2. Furthermore, *epf2* knock-outs showed the same insensitivity to increased [CO2] as the *ca1ca4* mutant (Engineer et al, 2014).

It is predicted that EPF2 needs to be cleaved to become active (Hara et al. 2007), and it was using this information that Engineer et al (2014) identified the subtilisin-like serine protease CRSP (CO2 RESPONSE SECRETED PROTEASE) as a regulator of stomatal density in elevated [CO2]. They identified several candidate proteases and looked for upregulation of the genes in cotyledons in elevated [CO2]. They found that *crsp* mutants had the same SD phenotype as the *epf2* and *ca1ca4* mutants. It was also determined that CRSP could cleave EPF2 *in vitro* to become active (Engineer et al, 2014). While this paper provided some major points in the [CO2] response pathway, it does not provide a complete pathway.

Other genes in and associated with the CO2 sensing pathway have also been discovered. The first genes to be identified to be associated with [CO2] response was HIGH CARBON DIOXIDE (HIC). HIC is a putative 3-keto acyl coenzyme A synthase which synthesises very long chain fatty acids. When *HIC* is knocked out, there is an increase in stomatal density of up to 40% (Gray et al, 2000). It is suggested that this disruption in fatty acids makes the waxy cuticle of the guard cells aberrant resulting in signals being unable to move through the wax (Gray et al, 2000). This indicated that there could be a signalling mechanism involved in response to [CO2].

A signalling mechanism is further supported by the evidence that mature leaves exposed to elevated [CO2] could signal to leaves developing under ambient conditions to alter their SD. This was shown by encasing Arabidopsis leaves in cuvettes in elevated [CO2] while new leaves were free to develop in ambient [CO2]. There was still a reduction in stomatal density in few day old leaves compared to plants with no exposure (Lake et al, 2001). Unfortunately, a microarray experiment to look at the genes driving this response did not find anything that led to identification of the signal. There were 915 significantly different genes in the expression of immature leaves from elevated [CO2] exposed plants and ambient exposed plants (Coupe et al, 2005) but no definitive signal was found. This means this long distance signal is still unknown.

More recently Watson Lazowski et al (2016) looked at the differencial transcriptomes of *Plantago lanceolata* L. in different [CO2] concentrations though RNA seq to look at differential expression. They found 147 differentially expressed genes including the stomatal genes *ER*, *YODA*, *MYB88* and *BCA1.* *ER* is the subject of chapter 5 of this thesis and its significance in the stomatal development pathway is discussed in section 1.5.2. *YODA* is similarly discussed in section 1.5.2. *MYB88* is a transcription factor which restricts mitosis in guard cells to prevent clusters of stomata forming. When *MYB88* is knocked out it results in clusters of stomata (Lee et al, 2013). BCA1 is the *Plantago lanceolata* L. homologue of CA1 mentioned above (Watson Lazerwoski et al, 2016). The mRNA of *ER* and *YODA* are found in the cells of developing arabidopsis guard cells (Wang et al, 2007) and as Wang et al (2007) found an increase in stomata in elevated [CO2], it is possible that these increases in ER expression are a result of changes in stomatal developement, rather than the cause of the changes in stomatal development. *CA1* upregulation is consistent with the work on gene expression in elevated [CO2] in Arabidopsis and is therefore supporting evidence for the pathway discovered by Engineer et al (2014).

### Biochemical signalling and [CO2]

Long distance signalling in Arabidopsis has been shown to frequently be non-protein dependant. The phytohormones ethylene, abscisic acid (ABA), gibberellins and cytokinins often act as long distance signals (Woodward and Bartell, 2005), as well as sugars (Smeekens et al, 2000), and reactive oxygen species (ROS) (Apel and Hirt, 2004) suggesting the long range [CO2] signal is not necessarily a protein one.

Chater et al found that both ROS and ABA are necessary for both proper stomatal closure and changes to stomatal density. ABA signalling mutants had previously been shown to be defective in stomatal opening (Webb and Hetherington, 1997) and Chater et al found that ABA synthesis and perception genes are necessary for SD reduction in elevated [CO2] or for closure. The responsiveness to [CO2] could specifically be rescued through expression of ABA synthesis genes in the guard cells (Chater et al, 2016).

ROS had previously been shown to increase in the presence of bicarbonate ions (Kolla et al, 2007), which are the result of carbon fixation by carbonic anhydrases which are upregulated in elevated [CO2] (Engineer et al, 2014). Chater et al (2016) showed that ROS deficient plants showed an increase in stomatal density in elevated [CO2]. Miller et al (2009) showed that accumulation of ROS travelled at a rate of 8.4cm per minute to rapidly signal wounding, heat, light and salinity stress. This suggests that ROS could be the long distance signal associated with [CO2] signalling, or ABA.

The role of sugars in signalling has been previously associated with stomatal density (Akita et al, 2013) and the soluble sugar content of plants in elevated [CO2] has been shown to be increased (Teng et al, 2006). Teng et al also showed an increase in starch content, starch grain number and size and an increase in overall carbohydrate content in elevated [CO2] which could all increase the availability of sugars to the plant. Sugar signalling and carbohydrate content is therefore a strong possibility for biochemical signalling in [CO2] sensing.

The identification of key parts of the signalling pathway of the SD response to elevated [CO2] was the aim of this thesis and attempts to identify key genes are discussed in chapters 3-5.

### Stomatal Aperture response to CO2

Although responses to elevate [CO2] are usually defined as short or long term responses, there is evidence of crosstalk between the two pathways.CA1 and CA4 are both required for the proper closure of stomata in elevated [CO2] (Hu et al, 2010) and are also required in Col-0 for a decrease in SD at elevated [CO2] (Engineer et al, 2014). Other genes and hormones that are responsible for the proper function of stomatal aperture to the environment could therefore be genes of interest when looking for genes that control SD and will be briefly discussed.

HT1 (Hasimoto, 2006) was one of the first genes identified as being required for the proper response of stomatal aperture to elevated [CO2]. Plants with ethyl methanesulfonate (EMS) induced mutations in the *HT1* gene showed abnormal leaf temperature changes in response to elevated [CO2], but still responded as expected to light intensity changes. HT1 was therefore identified as being specifically required for [CO2] responsive aperture change.

Further genes implicated in response to [CO2] include OPEN STOMATA 1 (OST1) and SLOW ANION CHANNEL-ASSOCIATED 1 (SLAC1). SLAC1 is an anion channel required for opening. SLAC1 allows anions into the guard cell to cause the increase in turgor pressure (Vahisalu et al, 2008). OST1 is an activator of SLAC1 and without OST1 there is improper activation of SLAC1 (Xue et al, 2011). OST1 and SLAC1 have been shown to respond to internal stimuli of bicarbonate ions and internal [CO2] to cause changes in aperture in response to available CO2 for photosynthesis (Xue et al, 2011).

Many other genes are involved in the integration of environmental signals into stomatal aperture response, and these could therefore be considered when looking for candidate genes responsible for stomatal density changes.

## MAGIC Arabidopsis and QTL Models

### Quantitative Traits

A quantitative trait is one that is measured on a scale rather than falling into distinctive categories. Most traits of interest in crop breeding are quantitative traits, including SD. A quantitative trait is often controlled by more than one gene and many quantitative traits are controlled by genes which interact together to produce phenotypes which are more than additive (Brem & Kruglyak, 2005). Genes that interact to produce phenotypes that are more than additive are called epistatic. Traits such as seed number and germination rate have been shown to be under the control of epistatic loci (Malmberg et al, 2005).

In some quantitative traits, the trait is largely controlled by a few genes of large effect alongside multiple genes of small effect (Brem & Kruglyak, 2005). This, for example, is seen in flowering time, where one gene explains 40% of variance while many smaller genes control the remaining 40% (Kover et al, 2009). Other traits have been shown to be controlled by many genes of small effects. For example, in Maize, it was predicted that oil concentration in the seed kernel is controlled by over 50 loci (Lauri et al, 2004).

### QTL Models

QTL models identify regions of the genome (loci) associated with a quantitative trait, such as stomatal density. QTL modelling is a form of genetic mapping. The idea of QTL mapping has been around since 1923 when Sax used beans to estimate the number of genes involved in the quantitative trait of bean weight. Sax showed that bean weight was a quantitatively controlled trait that was linked to bean colour thus showing that bean colour genes could be a marker for bean weight genes (Sax, 1923; Lander and Botstein, 1989). This process used single marker mapping (The bean colour), which was the only method possible at the time without detailed genetic maps (Lander and Botstein, 1989). More accurate and powerful interval mapping is now more commonly used as described by Doerge (2002)

The first step in mapping QTL traits requires the construction of a population to carry out phenotyping. These are usually large numbers of recombinant inbred lines (RILs). RILs are lines that genetically differ from each other marginally having been bred from two parents of opposing ends of the desired trait. RILs show a range of phenotypes due to the quantitative, multi-allelic nature of the phenotype (Doerge, 2002). The two lines are crossed and then their offspring crossed multiple times to produce a mosaic of the two genomes. They are then inbred for several generations to stabilise the genome.

Alternatively, Standard natural populations of inbred ecotypes can be used. These are created by collecting specimens from the wild and inbreeding them until a stable genome has been created. In Arabidopsis, over 1135 Arabidopsis ecotypes have been fully sequenced and are available for phenotyping (Alonso-Blanco, 2016)

A QTL model works by correlating changes in the phenotype between lines with the accompanying changes in the genotype. Traditional QTL models use single marker analysis. Single marker analysis works on the null hypothesis that the phenotype is independent of a genotype at a particular marker. The marker is often a single nucleotide polymorphism (SNP), a single base pair change in the genome that varies between individuals (Single Nucleotide Polymorphisms…, 2019). They take the marker and assess whether its presence or absence affects the phenotype (Doerge et al, 2002). The position in the chromosome is not taken into account.

Interval mapping is the most common form of QTL mapping today. It uses ordered pairs of markers to estimate if there is likely to be QTL within an interval of the markers or not (Lander and Botstein, 1989). It uses maximum likelihood models to test for the likelihood that the null hypothesis is true, that there is no QTL, compared with the hypothesis that there is a QTL at that position. It results in a LOD (logarithm of the odds) score which are compared to a threshold (Lander and Botstein, 1989). This method however does not take into account linked QTLs or those that interact to create epistasis, which occurs in many quantitative traits.

Instead, a fixed effects QTL model can be used, as has been used in this thesis. This method evaluates the presence of the QTLs and takes into account the presence of other QTLs which could be affecting the current one. This can be achieved by running QTL model multiple times (In this thesis 500 times) and evaluating the QTLs each time in a multiple QTL model. In these models, 80% of the lines are selected each time and the QTLs are added to the model one at a time until the fit cannot be improved. The location of these loci are recorded and the number of models that contain each QTL gives the probability the QTL is real (Kover et al, 2009).

In addition to QTLs, the falling price of RNA sequencing (RNA-seq) has made it possible to carry out expression QTLs (eQTLs) in which as well as QTL mapping. RNA-seq is the process of sequencing all RNA extracted and comparing the presence and quantity of each gene transcript to create a transcriptome (Wang et al, 2009). In eQTL analysis, the transcriptome of the plant is used to infer differences in expression in different conditions, and map these differences to the QTLs. This gives a clearer idea of which genes could be having an effect within a QTL region, especially as many differences are not the result of Single Nucleotide Polymorphisms (SNPs) in exons, but rather the result of differences in expression which can be difficult to find when looking for gene candidates (Yang et al, 2007). Exons are the parts of the gene which encode the protein, with introns found between them (Gilbert, 1978).

Some QTLs are only of large effect in certain environmental conditions. This is because of gene x environment interactions (GXE). GXE can be seen when a trait is affected by the environment, such as SD varying in different [CO2] levels to a different degree in different genotypes (El-Soda et al, 2014). In QTL analysis, GXE can be seen by variation in QTLs in different environments, with some QTLs not being of significance in some environments, or by some QTLs explaining different amounts of variance (El-Soda et al, 2014).

In crop breeding, QTLs that are consistent across different environments are preferred (El-Soda et al, 2014). In this analysis however, variation in QTL would give clues as to which genes are affected by changes in [CO­2]. A QTL X environment interaction (QXE) can be detected in several way, however the simplest of them is to run a QTL on the differences in the trait between environments (El-Soda et al, 2014). Other methods require more complex multi-environment mixed model QTL analysis (van Eeuwijk et al, 2010) which are available for populations from 2 parent RILs, and have more recently become available for multiparent RILs (Verbyla et al, 2014).

### Recombinant Inbred Lines and MAGIC Lines

Multi-Parental Advanced Generational Inter-Cross (MAGIC) Arabidopsis lines are a series of Arabidopsis lines produced from 19 parents from across the northern hemisphere (Kover et al, 2009). The MAGIC lines were created by intermating the 19 founder Arabidopsis lines for 4 generations to create 342 F4 outcross families. From these families, 3 inbred lines were chosen and inbred by back-crossing for 6 generations (See Figure 1.7.1). Each MAGIC Arabidopsis line incorporates approximately 9.97 founder lines. Using these MAGIC Arabidopsis lines in a QTL analysis has many advantages over both standard 2 parent RILs and the use of natural ecotypes.

|  |
| --- |
|  |
| Figure 1.7.1 Methods of creating different types of recombinant inbred lines. Pictured here are traditional RILs (a), more advanced RILs with increased crossing to create more linkage disequilibrium(b), and Multi-parent Advanced Genetic Inter-Cross lines(c). It can be seen that the extra crossing of b and c results in increased fragmentation and greater linkage disequilibrium for finer mapping (Kover and Mott, 2012) |

Standard QTL with natural ecotypes contain more minor alleles, although at very low frequency (Clark et al, 2007). This allows for rarer phenotypes to exist in the population, but because of their very low frequency in the population, also means they are less likely to be picked up by the QTL model. Use of ecotypes however does produce much finer mapping, as the higher linkage disequilibrium creates higher resolution mapping. The linkage disequilibrium creates a more varied genome to carry out interval mapping thereby making it possible to map to a finer level (Kover et al, 2009). The use of multiple parental lines also helps to identify which genes are important on a population level. Many 2 parent RIL crosses result in QTLs which cannot be found again in other mapping populations (Holland, 2007).

However, traditional RILs have an advantage over natural accessions in that there is a high power for detection of QTLs. The power of detection is calculated by the amount of phenotypic variation the QTL explains. If there are only two inbred parents, then this means there will only be two genes, meaning the QTL is diallelic and the power is proportional to the formula *p(1-p)* where *p* is the minor allele frequency (Kover at al, 2009). In two parent synthetic populations this is usually around 0.5*.* There is also no population structure that can result in ghost QTLs. Ghost QTLs sometimes arise when linked parts of the genome are present in natural populations, often at great distances (Kover et al, 2009). These can be controlled for in the model used, for example by efficient mixed-model association (EMMA) which also controls for the population structure of a sample (Kang et al, 2008)

Traditional RILs however do have the disadvantage of poor mapping resolution, as there is much lower linkage disequilibrium between the RILs than a natural population, although this is what makes QTL detection easier (Kover et al, 2009). The resultant QTLs of 2 parent RILs can have confidence intervals of up 40cM (Bandaranayake et al, 2004) which translates to more than 4.8Mb. This means that further fine mapping is then required to more thoroughly locate the gene of interest within a QTL among hundreds of candidate genes (Kover at al, 2009).

MAGIC line populations take the best of both ecotypes and RILS. They are a set of recombinant inbred lines which are derived from 19 parental lines. This gives them the advantages of containing rare alleles found in the parental ecotypes at a higher frequency than found in wild populations as there is no selection pressure. As the QTLs come from 19 parents, they are also of significance outside of2 lines. In traditional QTLs populations, many QTLs are only significant in those QTL populations and are not significant QTLs in populations with different parents (Holland, 2007). Thus, using 19 parents to create a population gives the QTL wider significance across the species. Ensuring the QTL is relevant outside of two parents is possible by using multiple populations of RILs, however MAGIC lines can ensure relevance too without the need for extra phenotyping.

The multiple parents also give better mapping ability due to better linkage disequilibrium which allows for finer mapping than RILs. The MAGIC lines in the Kover, 2009 paper showed an average of 300kB confidence interval around the QTL marker and showed that genes of effect size of 10% or above should always be detected with 460 or more lines used, while those of 5-10% had around 52% chance of detection. If the full 1026 cohort of plants were used then 79% of 5% effect QTLs would be detected. Detection of QTLs is therefore dependent on number of lines used and increasing the number of lines in an experiment should be a priority when planning.

MAGIC lines have now been created in many other plant species including wheat (Gardner et al, 2016), rice (Bandillo et al, 2013; Meng et al, 2016), maize (Dell’Acqua, 2015), sorghum (Ongom and Ejeta, 2018) and Cowpea (Huynh et al, 2018).

Carrying out a QTL analysis using the MAGIC lines has an additional advantage of a haplotype analysis that gives predicted phenotypic results for all 19 parental ecotypes at that locus (Kover et al, 2009). This can help with deciding upon candidate genes after QTL analysis.

### Previous uses of MAGIC Lines

The first use of MAGIC lines in Arabidopsis was that by Kover et al, to check the efficacy of the MAGIC population. They used the MAGIC lines to study flowering time in short and long days, days to germination, growth rate, days to bolting and days from bolt to flower (Kover et al, 2009). For flowering time, they were able to map 3 of 4 QTLs to known flowering genes and the final gene to a suspected flowering gene showing the efficacy of the system (Kover et al, 2009).

Further to this the lines have been used to study the relationship between seed number and seed size and find QTLs to break the known relationship between them (Gnan et al, 2014). Seed number and seed size are linked because there is a limited amount of resources available to the plant, and a plant will either make larger, more fecund seeds, or produce more seeds (Venable, 1992). Breaking this relationship would be important in plant breeding as it would be advantageous to the plant breeder to able to increase seed size without decreasing seed number, or to increase both. MAGIC lines are ideal for researching this because there has been no selection pressure on them like natural populations to develop trade-offs, but they still maintain the diversity to be able to find novel low frequency alleles that would have been missed in a two-parent population (Gnan et al, 2014).

MAGIC lines have also been used to study adaption to new conditions (Fournier-Level et al, 2016). MAGIC lines were used to study the adaption of plants to changing climates to look at the changes in fitness in simulated environments. The MAGIC lines make a good model for study here as they incorporate genetic material from a wide geographical area and from a range of climates (Kover et al, 2009). The unlinked nature of traits in the MAGIC lines again makes them a good model for studying advantageous traits without the problem of linkage.

### Fine Mapping

Fine mapping is the process of refining QTL results to get a more accurate QTL region to discover the gene, and sometimes SNP, causing the QTL. In some association mapping, this is unnecessary as the mapping can give confidence intervals small enough that it is easy to identify the gene in question. In fine mapping, molecular markers around the QTL for up to 1000 lines are genotyped to a very fine degree. After genotyping, phenotyping of the lines is carried out and another QTL model is run (Price, 2006). This gives a location of under 1cM which is equivalent of around 50 genes.

Fine mapping was successfully used to identify the BREVIS RADIX gene in Arabidopsis. This gene controls root length and was identified in a screen between Uk-1 and Sav-0. After an initial QTL was identified through a screen of 200 RILs, recombination mapping was used to further narrow down the region, successfully isolating *BRX* (Mouchel et al, 2004). The Hardtke lab again successfully carried out a QTL that returned a gene of interest on a screen of seed dormancy in 2014 when they used Eil‐0 and Lc‐0 to isolate the gene down to 180kb using QTL models. After this they used genome sequences and SNP analysis to identify the QTL as a gene they called *IBO* which was later identified by others to be *REDUCED DORMANCY 5 (RDO5)* (Amiguet‐Vercher et al, 2014).

MAGIC lines give an average region return of 300kB up or downstream which can equate to as few as 200 genes. While fine mapping is possible, it is therefore not always necessary and can be eliminated through larger numbers of lines in the original phenotyping to get smaller confidence intervals (Kover et al, 2009).

An alternative method of locating genes under QTLs is the use of T-DNA insertion lines. These are Arabidopsis lines in which random insertion of T-DNAs by Agrobacterium tumefaciens has been used to disrupt genes throughout the genome (Valvekens et al, 1988). The National Arabidopsis Stock Centre stocks T-DNA insertion lines for Arabidopsis. This includes Gabi-Kat lines, Flag lines and Salk lines. These cover approximately 80% of the Arabidopsis protein coding genome (Li et al, 2006). This allows the phenotyping of most genes found in a small region and is the method chosen in this thesis. With 200 genes to study, it provides for greater for speed and cost efficiency than fine mapping and was therefore used in this study.

## Aims

The aim of this study was to establish further the genetic mechanisms by which stomatal densities respond to changes in [CO2]. By investigating the genetic mechanisms it was hoped to understand why plants have a variable response in SD to elevated [CO2]. The aims of each chapter as listed below.

### Chapter 2

This chapter aimed to first establish stomatal response across 19 Arabidopsis ecotypes (the MAGIC parental lines) to changes in elevated [CO2]. Secondly it aimed to establish the best methods for high through-put phenotyping of Arabidopsis for stomatal density response with minimal confounding factors. Based on the data by Caldera et al (2016), it was hypothesised that ecotypes would show a range of SD responses to [CO2] with Col-0 and others showing a decrease, some showing an increase and some showing no response.

For methods for high-throughput phenotyping, it was hypothesised that cotyledons would make a suitable organ of study for stomatal density and that there would be a point where SD becomes fixed. It was hypothesised that previous data on SD in cotyledons could be analysed to find the minimum number of measurements of SD required to get a good representation of SD for an ecotype.

Finally it was hypothesised that the parents of the MAGIC lines would show enough diversity that the MAGIC lines would be a suitable population for a QTL analysis.

### Chapter 3

Chapter 3 aimed to exploit the MAGIC lines in a QTL analysis to find regions of the genome associated with stomatal density response to elevated [CO2] and to find regions of the genome responsible for control of SD between ecotypes. The MAGIC lines would be phenotyped in the best practice manner reported in chapter 2 as it was hypothesised that this would result in data of a high enough quality to run an effective QTL analysis.

It was hypothesised that the MAGIC lines would show normal distribution of phenotypes, as it was hypothesised that the trait of SD is under polygenetic control. It was hypothesised that the MAGIC lines would be a suitable set of RILs to carry out a QTL analysis on the trait of SD.

Based on the work of Engineer et al (2014), it was hypothesised that CRSP, EPF2 and CA1/CA4 would be identified by QTL as controls of SD response to [CO2]. It was further hypothesised that genes in the stomatal development pathway would be identified as QTLs for stomatal density in both [CO2] treatments.

Following on from identification of regions of the genome found by the QTL analysis, genes of interests were hypothesised to be found within 300kB up or downstream of the QTL. It was further hypothesised that the gene causing the QTL could be found using T-DNA insertions if they were not known genes of interest.

### Chapter 4

This chapter aimed to further phenotype and characterise a T-DNA insertion line identified as affecting SD under the QTL on chromosome 5 found in Chapter 3. It was hypothesised that it would be possible to relate the function of the gene identified by the T-DNA insertion to stomatal density and explain why it has an effect on SD.

The chapter further aimed to relate the specific alleles in MAGIC parental ecotypes to the SD phenotype found by the haplotype analysis of the QTL produced in Chapter 3. It was hypothesised that differences in the sequence or the expression of the gene would explain the haplotype analysis pattern of the parental ecotypes produced by the QTL analysis. It was hypothesised that study of the parental lines would lead to understanding of why this gene affects SD.

### Chapter 5

This chapter aimed to study how variation in ER, identified by the QTL analysis as a control of SD, controlled stomatal density in different ecotypes. It was hypothesised that changes to the sequence, expression or splicing of the gene could be having an effect on stomatal density. As knock-outs of ER have an increased stomatal density, it was hypothesised that down-regulation of ER would result in an increase in stomatal density.

# Variation in Stomatal Characteristics across 19 Arabidopsis Ecotypes

### Natural Accessions of Arabidopsis

Arabidopsis is found all over the world growing in many different environmental conditions (Al-Shehbaz et al, 2002). Thousands of individual Arabidopsis plants have been collected from different environments and selfed to create ecotype stocks. Each ecotype has unique characteristics related to the environment they have adapted to grow in, creating a valuable resource to investigate biological variation across natural systems without each laboratory having to collect plant resources (Koornneef et al, 2004).

This variation in phenotypic characteristics can be linked back to variation in the genome. In *Arabidopsis thaliana*, across 1,135 ecotypes sequenced, on average about 1 in 20bp is polymorphic showing huge genomic variation (Alonso-Blanco et al, 2016). This variation can be used to elucidate genetic mechanisms for specific traits and responses. For example, Johanson et al, (2000) used studied Arabidopsis natural ecotypes to show how variation associated with the gene FRIGIDA is responsible for whether ecotypes are early or late flowering.

A level of genetic variation similar to that observed between natural populations (Alonso-Blanco et al, 2016) is present in the genomes of the 19 parental lines that were used to create the Multi-Parental Advanced Generation Inter-Cross (MAGIC) lines (Gan et al, 2011). The 19 MAGIC parental lines not only show a similar level of nucleotide diversity, but also similar variation in genomic features and structural variants to that seen across diverse global samples. This makes them a good population to use as a representation of how the larger population varies.

The 19 parental ecotype lines have been used several times to study how particular Arabidopsis traits are associated with specific genetic variation. They have been used to study defence metabolism (Hanschen et al, 2018), transposable elements (Guo, 2017), omega-3 desaturation of fatty acids (Menard et al, 2017), root metabolites (Monchgesang et al, 2016) and seed size and number (Gnan et al, 2014). They are ideal for exploring variation in traits in natural accessions as they have transcriptome, genome (Gan et al, 2011) and epigenome (Kawakatsu et al, 2016) sequence data available. When variation is shown, QTL studies can be carried out with MAGIC lines to identify underlying genetic mechanisms.

The MAGIC lines derived from the 19 parent lines are a set of recombinant inbred lines (RILs) produced for carrying out QTL analysis. QTL analysis is used to find regions of the genome associated with interesting traits. For a set of RILs to be suitable for QTL analysis, their parents must differ in their genotype at the regions associated with the trait to give different phenotypes (Huang et al, 2011). This can be established by studying the differences in the phenotype of parental lines. If a particular phenotype differs between parental lines, then there should be genomic diversity in the genotypes. Here, the Arabidopsis MAGIC parental ecotypes were studied to see if SD response to [CO2] showed sufficient phenotypic diversity for the MAGIC lines to be used for a QTL analysis.

### Response to [CO2] in Natural Arabidopsis Ecotypes

In previous studies it has been have shown that in elevated [CO2], Arabidopsis exhibits alterations in the level of stomatal development. 75% of plant species are believed to show a decrease in stomatal density in elevated [CO2] (Woodward and Kelly, 1995). Caldera et al (2017) studied 18 ecotypes from around Europe to look at variation in SD. Despite most plant species so far studied showing a decrease in SD in increased [CO2], Caldera et al found significant evidence of increases in SD in response to elevated [CO2] in 8 Arabidopsis ecotypes, as well as 4 that did not change their SD, and 5 that decrease their SD showing a wide diversity of responses. Some of these changes contradict previously reported values for changes, for example Col-0 showed an increase in density in response to rising [CO2] in Caldera et al (2017) despite being found to show a decrease in other experiments (Chater et al, 2015; Engineer et al, 2014). This suggests that stomatal development in response to [CO2] is variable and affected not only be ecotype, but also other experimental conditions perhaps such as light, water availability, or relative humidity.

Much of the work looking at SD in elevated [CO2] has been carried out in mature leaves of Arabidopsis, however cotyledons also show a change in stomatal density in elevated [CO2]. Engineer et al (2014) showed that 10-day old cotyledons of Col-0 exposed to elevated [CO2] showed a decrease in SD, similar to that of the mature leaves. Further to this, Engineer et al showed that the SD response of *crsp, ca1 and ca4* knock-out plant cotyledons are altered SD in elevated [CO2] suggesting the changes in SD in cotyledons involve these same genetic components. This suggests cotyledons use similar genetic pathways to true leaves to control SD and can be used to study [CO2] responses. SD response to [CO2] was only been shown to be similar in cotyledons and mature leaves in Col-0 however, and further research is required to see if cotyledons are truly representative of mature leaves in terms of SD response to elevated [CO2] in other ecotypes.

Delgado et al (2011) studied how stomatal traits in cotyledons and true leaves in 62 wild accessions of Arabidopsis correlate. The study found that there is strong, positive, statistically significant correlation between SD in cotyledons and true leaves suggesting that cotyledons are a good representation of true leaves for SD. Correlation between SI in cotyledons and true leaves was less strong than between SD in cotyledons and mature leaves. Delgado et al (2011) also found good correlation between epidermal cell size between true and cotyledons leaves. This suggests that cotyledons are a good representation of some stomatal and leaf traits. Whether cotyledons are a good representation of mature leaves for SD response to [CO2] is explored in this chapter.

### When is a Cotyledon Mature?

For a measurement to be representative of a cotyledon and to prevent different growth development stages influencing results, harvesting of cotyledons should occur when SD becomes fixed in the cotyledon or cotyledons are at a fixed growth stage. Tsukaya et al (1994) showed that within 9 days after germination the cotyledon of Arabidopsis has grown to 100X the size it was in the seed showing huge expansion which could impact measurements of density. Stoynova-Bakalova et al (2003) measured cotyledon size in Col-0 and found that after 5 days cotyledon area is fixed suggesting that cotyledons reach their mature size relatively quickly.

Some studies, eg Delgado et al (2011), harvest cotyledons at the same time as harvesting mature leaves before bolting, however this method requires much longer growth times . It is therefore important to establish when SD becomes fixed or when cotyledons reach maturity under the particular growth conditions used.

Similar to this, the sufficient number of samples should be taken to ensure that the calculated SD is representative. Within leaves of the *Alnus glutinosa* (L.) tree, there was a 2.5-fold difference in SD of leaves in just one tree showing large variation within one plant (Poole et al, 1996). It is therefore necessary to establish how many measurements are required to account for variation within the plant, and create consistent, representative data.

### Aims

1. To establish a rapid but accurate method of high-throughput phenotyping of large numbers of lines. The optimal growth period before plant harvest; the optimal number of samples needed; and ensuring that cotyledons stomatal responses are representative of mature plants needs to be established.
2. To examine the phenotypic diversity in SD response among the MAGIC parental lines to establish if there is sufficient genotypic diversity to use the MAGIC lines in a QTL analysis on SD.
3. To explore whether stomatal phenotypes of the 19 parental ecotypes can be related to the environmental conditions they were collected from, thereby giving suggestion as to why SD is different between the 19 parental lines.

## Methods

### Plant Material

Seed of Col-0, Rsch-4 and Can-0 came from the Nottingham Arabidopsis Stock Centre (NASC). Seed of 19 parental ecotypes of MAGIC lines was kindly donated by Dr Paula Kover (Bath University). Full information about where the 19 ecotypes were collected can be seen in Table 2.2-1

Table 2.2‑1 Table of 19 parental ecotypes of Arabidopsis. Contains the location of their collection. Data from Arabidopsis.org

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Line | Origin country | Origin area | Climate | latitude | longitude |
| Can-0 | Canary Islands | Las Palmas | Sub-Tropical Desert Climate | 29.214 | -13.481 |
| Bur-0 | Ireland | Burren | Temperate | 54.1 | -6.2 |
| Col-0 | Usa | Columbia | Humid-subtropical | 38.3 | -92.3 |
| Ct-1 | Italy | Catania | Mediterranean hot | 37.3 | 15 |
| Edi-0 | UK | Edinburgh | Temperate maritime | 55.9494 | -3.16028 |
| Hi-0 | Netherlands | Hilversum | Temperate Marine | 52 | 5 |
| Kn-0 | Lithuania | Kaunas | Humid Continental | 54 | 23 |
| Ler-0 | Germany/Lab | - | - | - | - |
| Mt-0 | Libya | Martuba | Hot semi-arid | 32.34 | 22.46 |
| No-0 | Germany | Halle | Warm and Temperate | 51.0581 | 13.2995 |
| Oy-0 | Norway | Oystese | Warm and Temperate | 60.38554 | 6.193019 |
| Po-0 | Germany | Poppelsdorf | Oceanic | 50 | 7 |
| Rsch-4 | Russia | Rschew | Humid Continental | 56.3 | 34 |
| Sf-2 | Spain | San Feliu | Warm and Temperate | 41.7833 | 3.03333 |
| Tsu-0 | Japan | Tsushima | Humid sub-tropical | 34.43 | 136.31 |
| Wil-2 | Lithuania | Wilna | Humid Continental | 55 | 25 |
| Ws-0 | Berlarus | Wassilewskija | Moderate Continental | 52.3 | 30 |
| Wu-0 | Germany | Wurzburg | Marina West Coast | 49.7878 | 9.9361 |
| Zu-0 | Switzerland | Zurich | Marina West Coast | 47.3667 | 8.55 |

### Plant Growth

Plants were grown in F2+sand Levington, Everris Professional compost in Sanyo-Gallenkamp SGC970/C/HQI or Conviron BDR 16 cabinet cabinets in either 180ppm, 450ppm or 1000ppm [CO2]. Before sowing, plants were stratified for 72 hours in 4°C in water in Eppendorf tubes before being spread using a pipette on compost. Plants were grown in constant 20°C in 10 hr days at 200 µmol m2 s-1 PAR. Plants were watered every 3 days. Plants were grown in 10X6 trays.

Treatment conditions of 180ppm were achieved in a Conviron BDR 16 cabinet with a soda-lime [CO2] scrub and an additive CO2 injection to achieve 200ppm. Elevated [CO2] conditions were achieved by additive CO2 injection to the Sanyo-Gallenkamp SGC970/C/HQI. 450ppm was achieved by use of ambient CO2 with CO2 injection to achieve 450ppm in Sanyo-Gallenkamp SGC970/C/HQI. CO2 was supplied from CO2 cylinders (BOC, UK).

Plants for mature leaf work were grown for 6 weeks. Cotyledons were grown for 18 days unless otherwise specified.

### Stomatal Impressions

Impressions of the abaxial side of leaves and cotyledons were made using dental resin (coltene, PRESIDENT, light body dental resin). For mature leaves 3 leaves from 3 plants were taken. For cotyledons 8-12 cotyledons were taken. Leaves were placed abaxial side down in resin and left to dry for 10 minutes. Leaves were then removed from resin and clear nail varnish was spread over the mould. Once dry, nail varnish was removed using clear tape and placed on a glass microscope slide.

### Stomatal Counting and Microscopy

Stomata were counted on a Leica DM IRBE Inverted Microscope with Planachromat 20x/ 0.4∞/ 0.17-A lens. Micro-Manager 1.4 software was used to view an image of the abaxial surface of the microscope slide and ImageJ was used to draw a box 400µm X 400µm to give an area of 1600µm or 0.16mm. For some experiments, Z stack images of the abaxial surface were taken using micro-manager. For each cotyledon, 2-3 areas, away from the edge and the base, had the number of stomata inside the box counted for density and, if index was required, number of epidermal cells inside the box was counted. For mature leaves, 3 areas from the base, middle and tip of the leaf were studied. Data was inputted into ms excel and the following equations were used to calculate SD and stomatal index

### Measuring Cell Sizes

Cell sizes of ecotypes were measured using imageJ. A Z-stack image of a 600X400µm area of the cotyledon abaxial surface was opened with imageJ. The scale of the image was appropriately set and the freehand tool was used to draw around cells. All cells that were completely in the image were included. Cell area was then taken by the measure tool. 5 images per ecotype were analysed.

### Leaf Area

Leaf area was measured using ImageJ. Distal images of cotyledons were captured using Panasonic Lumix DMC-G7 Digital Camera with a ruler alongside for scale. Scale was set in image J using the ruler for scale. Cotyledons were then drawn around and size was calculated with measure.

### Ecotype Environment Data Collection

For the analysis of environmental conditions, environmental data was collected from Bioclim (Fick and Hijmans, 2017) using the raster package in R, except for days of rain which was downloaded from NOAA website in October 2015 and PAR which was collected from <http://globalsolaratlas.info/> and converted to PAR using the approximation 1 W/m2 ≈ 4.6 µmol m2 s-1 (Based on Sager and McFarlane (1997)). Information on ecotype collection location was taken from the Arabidopsis information resource (Arabidopsis.org).

### Bootstrapping

Bootstrapping was carried out using R programming language. A data set of 30 cotyledon SD measurements taken by Jordan Brown (University of Sheffield) were randomly sampled 100 times to find the mean SD measurement when n was equal to 5, 10, 12, 13, 14, 15, 16, 17, 20, 25, 30 or 45.

Code found at <https://github.com/CrookedY/bootstrapping>

### Statistics

Statistics were carried out using R statistical programming language. Pearson’s correlation coefficient or Spearman’s correlation was used to look at the correlation between factors. ANOVA was used to investigate the difference between samples and treatments and Tukey’s post hoc analysis was used to look at the difference between groups. Graphics were produced in R using ggplot2

## Results

### Finding Ideal Sample Size

Finding the optimal sampling number is necessary when carrying out large scale phenotyping in order to lower costs and reduce experimental time. Finding the optimal number can be found by bootstrapping, which uses random number sampling to find the number of samples at which the mean and variation become similar enough that further sampling is unnecessary.

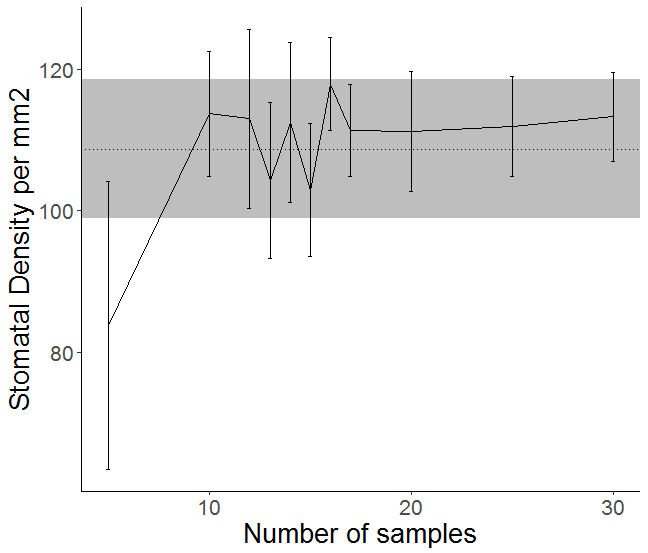


Figure 2.3.1 Variation in SD mean from random sampling. Results of mean SD of different N from one iteration of random sampling of stomatal density of cotyledons is shown. Points show mean stomatal density of different N (eg 5 samples, 10 samples, 12 samples) sampled from dataset of 30 stomatal densities. Error bars show standard error

Figure 2.3.1 illustrates how variation decreases with more samples and how the mean varying with each random sample number. The data randomly sampled at set of 30 stomatal densities from 15 cotyledons of 15 plants, with 2 measurements taken from each plant. This experiment suggested that approximately 15-20 samples are need to reach a mean that is representative of the SD for the line.

|  |
| --- |
|  |
| Figure 2.3.2 The spread of means across 100 iterations of random sampling. Out of a dataset of 30, data set of N (X axis) were randomly samples 100 times and the mean density each time recorded and shown in box plots. Data suggests sampling 15-20 measurements gives steady mean. Dots show outliers (1.5X IQR outside of the IQR) |

Sampling in bootstrapping should be carried out multiple times in order to get an accurate idea of the number of samples needed, as outliers can be omitted in initial sampling disguising the true number of samples required for a representative mean. By carrying out multiple iterations, it is possible to see how adding extra samples affects the resulting mean. In this sample of 30, it can be seen in Figure 2.3.2 that variation in mean SD decreases as N increases. There is a limit however to how informative this can be as towards N=30 there is no variation as all samples are included. Instead, it can be seen that between N=15 and N=20 there is little variation suggesting that any number of samples between 15 and 20 would be a good estimate of SD. For this reason, 16 was chosen. The method chosen to obtain 16 samples was to analyse 8 cotyledons from 8 plants with 2 measurements of 0.16mm2 each.

### Cotyledon Age and Stomatal Density

|  |
| --- |
|  |
| Figure 2.3.3 The stomatal density of cotyledons of Col-0 and Can-0 as they age. Mean stomatal density shown for ecotypes Col-0 and Can-0 at age 7, 10, 14 and 21 days in 450ppm and 1000ppm [CO2]. Data suggests stomatal densities become fixed around 14 days. Point represents mean with standard error shown by error bars. N=16 |

The age at which the cotyledon density becomes fixed, or a good representation of its final density, is important in deciding the most appropriate time to take measurements. Figure 2.3.3 shows the SD through the growth of two ecotypes of Arabidopsis, Can-0 and Col-0. As the cotyledon develops, SD changes, with Can-0 developing a fixed SD much quicker than Col-0. The plants were grown in two [CO2] conditions and it can be seen that the differences may not be apparent until past 10 days of growth. Similar to the results shown here, Rui et al, 2017 showed that the SD of true leaves becomes fixed after two weeks of growth. This suggests that after an initial period of growth SD becomes fixed. As SD appears to become fixed after 10 days it would suggest that two weeks is long enough to be sure of seeing a representative SD.

In addition, in order to ensure that cotyledons were harvested at a time when they had stopped expanding in case expansion affected SD, the expansion of ecotypes was studied. Cotyledons of plants at 7,10,14 and 21 days were harvested and leaf area was examined. Based on leaf area, cotyledon expansion had largely finished by day 14 with no significant difference between 14 days and 21 days in any ecotype (ANOVA, P<0.05). This suggested that after 14 days is a suitable time for harvest.

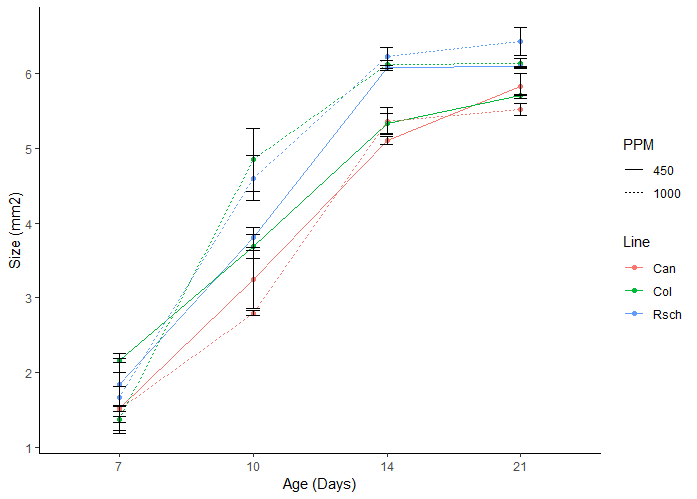


Figure 2.3.4 Cotyledon leaf expansion over time. Mean size of cotyledons of plants grown in 450ppm [CO2] and 1000ppm [CO2] harvested age 7, 10, 14 and 21 days. Points represent mean cotyledon size and error bars standard error (N=4-8)

To study further the effect of cotyledon size on SD, the size and SD of the cotyledon was studied. If there was a strong correlation between cotyledon size and SD, measuring SD could be more of a measure of cotyledon size than changes in stomatal development. This may mean that cotyledon size should be included in the analysis when looking at SD.

The measurement of 469 SD across 219 cotyledons showed that there is a relationship between SD and size however, it is not one that necessarily needs to be taken into account. There is a significant relationship between cotyledon size and SD (P=0.0001), however the relationship is small. The linear model gives an output of 0.080255x-0.000166 meaning that for every 1mm2 increase in cotyledon area, SD decreases by 0.080255 per mm2. Considering the small change in SD with changing cotyledon size, while it may be advantageous and provide extra information to explain changes in SD, it was decided that in this project it was not necessary to take into account cotyledon size when measuring SD.

|  |
| --- |
|  |
| Figure 2.3.5 Comparison of leaf size and SD in cotyledons. The size of the cotyledons (mm2) and the SD per mm2 of 19 lines of arabidopsis grown in 450ppm and 1000ppm [CO2] is shown. Blue line shows line of regression (0.080255x-0.000166 ; P=0.0001) and shading 95% confidence intervals. N=219 |

### Mature Leaf and Cotyledon Similarities

Cotyledons make an ideal organ to study SD because they are present from germination reducing growth time before harvest. They must also, however, be representative of true, mature leaves in order to properly help us understand how the whole plant responds to [CO2]. Stomatal characteristics of cotyledons and mature leaves were therefore compared.

|  |
| --- |
| A) |
| B) |
| Figure 2.3.6 Stomatal characteristics of mature leaves and cotyledons of Arabidopsis. Data shows Stomatal Density (A) and Stomatal Index (B) of 4 lines (Col-0, Can-0, Wil-2 and Wu-0) grown in 450ppm and 1000ppm [CO2]. Data suggests correlation of stomatal density in mature and cotyledon leaves (R= 0.66, P = 0.07), but weaker correlation in SI (R=0.58, P=0.13). N=8 |

|  |
| --- |
|  |
| Figure 2.3.7 The effect of [CO2] on stomatal density in mature leaves and cotyledons. Points show the percentage difference in SD of plants grown in 450ppm and 1000ppm [CO2] in mature leaves and cotyledons suggesting good, though not significant correlation (Pearson’s correlation, R=0.90, P=0.09. N=4) |

There was no significant correlation between mature leaf SD and cotyledon SD, however this is likely to be due to the low number of lines compared (R2= 0.66, P = 0.07 N=8). It can be seen that there is relatively good correlation between mature leaf and cotyledons with the exception of Can-0 in 1000ppm. Delgado et al (2011) found significant correlation in SD between first leaves and cotyledons of R2=0.71 which is similar to what was found here. Cotyledon SD would therefore appear to be a reasonable indicator of mature SD.

Stomatal index between cotyledons and mature leaves was found to have less strong correlation with a correlation of R2=0.58 and was also not significant (P=0.13). As N was again small (N=8) it is possible that this would change with the addition of SD measurements from further ecotypes. The lower correlation in SI between mature leaves and cotyledons is similar to what was found by Delgado et al (2011). Delgado et al found a significant correlation of R2=0.43. This suggests that while both SD and SI are correlated, SD correlates more consistently between mature leaves and that cotyledons are less good a proxy for SI than SD.

It is also important that cotyledon SD response to [CO2] is representative of mature leaves. It can be seen in Figure 2.3.7 that there is correlation between the differences in SD between treatments of mature leaves and cotyledons with an R3 of 0.90. This correlation is not significant (P=0.090) however, this analysis has a low N and the addition of further data points on SD difference would make it more robust. The strong R2 value, however, suggests that cotyledons are a good representation of how mature leaves SD responds to elevated [CO2] and they have therefore been used from this point on.

### Stomatal Density of 19 Ecotypes

For the MAGIC lines to be suitable for QTL analysis, it must be shown that there is phenotypic diversity among the parental lines, thus showing that there is genetic diversity at the loci that control the quantitative trait. Stomatal density in the MAGIC parental lines in 180, 450 and 1000ppm [CO2] was therefore analysed using the method developed in the above experiments i.e 16 samples from 8 cotyledons from 8 18 day old cotyledons.

|  |
| --- |
| A)  B) |
|  |
| Figure 2.3.8 The stomatal characteristics of 19 Ecotypes of arabidopsis in 3 [CO2] treatments. The stomatal density (A) and stomatal index (B) of 19 ecotypes of Arabidopsis grown in 180ppm, 450ppm and 1000ppm [CO2] is shown. Error bars represent standard error and \* indicate significant difference between treatments within line (ANOVA, P<0.05) n=8 |

SD of 19 parental ecotypes can clearly be seen to vary widely. The lowest densities can be seen in Can-0 in all treatments and the highest in Ler-0 in all treatments. For stomatal index,

|  |
| --- |
|  |
| Figure 2.3.9 The percentage change in stomatal density across 19 lines. The percentage change show is SD relative to each line at 450ppm [CO2] for 19 ecotypes of Arabidopsis. Percentage change in mean stomatal density of 19 lines in 180ppm and 1000ppm compared to 450ppm [CO2] shown. Light blue indicated a reduction in SD and dark blue an increase in SD (N=16) |

it can be seen in Figure 2.3.8 that the lowest and highest parental SI is different in the different [CO2] treatments. Can-0, Wil-2 and Sf-2 all show low indexes overall. In contrast, Wu-0, Mt-0 and Hi-0 show high indexes overall. The range of stomatal densities within a single treatment is 2.5X in 450ppm, 2.8 in 1000ppm and 2.5x in 180ppm.

Previous experiments by others have shown that Col-0 shows a decrease in SD as [CO2] increases (Engineer et al, 2014, Chater et al, 2015). In this analysis, while there was a decrease in SD with increasing [CO2], it was not significant (P=0.22). In addition to this, most lines did not show a decrease in SD as [CO2] increased but showed a range of responses to [CO2].

It can be seen in Figure 2.3.9 that most parental ecotypes show a decrease in SD between sub-ambient and ambient [CO2], but an increase between ambient and elevated [CO2]. Col-0 and Rsch-4 stand out as the only lines that consistently decrease their SD when in increasing [CO2] and Wu-0 and Zu-0 appear to show no change in density across [CO2] concentrations. As in the same lines, there were different changes in SD between different [CO2] concentrations, it could be interpreted that there may be two different SD responses to [CO2] occurring in some lines. It suggests there might be one pathway responsible for SD changes between 180ppm and 450ppm [CO2], and another between 450ppm and 1000ppm [CO2].

Some of these changes in index and density can be explained by epidermal cell size rather than alterations to stomatal development. Can-0, for example, shows an increase in SD in elevated [CO2] but a decrease in SI. This is because there are more, smaller pavement epidermal cells in elevated [CO2] as epidermal pavement cell size is significantly smaller in elevated [CO2] (P<0.001) (Figure 2.3.10). Wil-2 also shows a decrease in epidermal cell size in elevated [CO2] (P=0.01) but this is not accompanied by a significant change in SD and SI shows no significant change between [CO2] treatments. Cell size changes could therefore be an important measurement when looking at what causes changes in SD between [CO2] treatments.

|  |
| --- |
|  |
| Figure 2.3.10 Violin plots of the size of the epidermal pavement cells of cotyledons in 2 [CO2] treatments. Stomatal density in ambient 450ppm (in cream) and elevated 1000ppm (in grey) [CO2] of Can-0, Hi-0, Wil-0 and Wu-0 is shown. \* indicates significant difference between [CO2] treatments within line (T-test, P<0.05). White diamond represents mean, box plot shows IQR and density plot the spread of values. |

### Stomatal Features and Environmental Conditions

The variable phenotypic traits exhibited by Arabidopsis ecotypes are often adaptions to the environment they were originally collected from. investigating how traits and environment correlate can provide information as to why plants have acquired particular traits. Correlations between SD and source environmental conditions were therefore examined.

For the analysis of environmental conditions and SD, the ecotypes Col-0 and Ler-0 were removed from the analysis. Ler-0 was removed because it has been mutated in the lab via radiation meaning many of its phenotypes no longer represent adaptions to its environment. Col-0 was removed because its original collection place is unknown. It was first characterised in Columbia, USA however genetic analysis suggests that it was originally collected somewhere in West Germany (Nordborg et al, 2005).

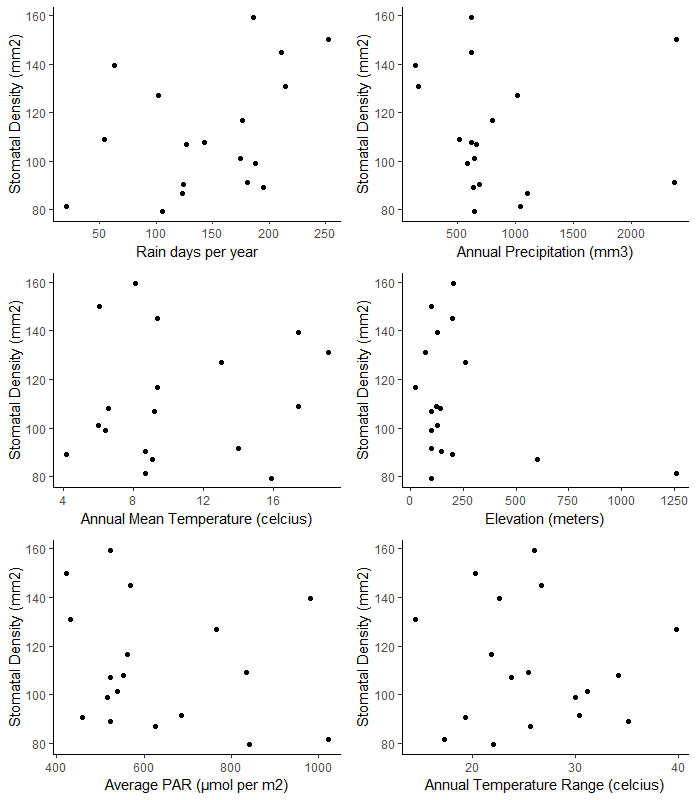


Figure 2.3.11 Scatter plots showing the relationship between stomatal density and the environmental conditions of their collected environment. Plots show stomatal density and A) Days of Rain, B) Annual Precipitation, C) Annual mean temperature, D) Elevation (m), E) Average PAR, F) Annual temperature range (°C). Data for environmental conditions from bioclim except Days of Rain which is from NOAA and PAR which is converted from World Bank dataset. Stomatal densities for ecotypes are density in 450ppm [CO2] (See Figure 2.3.8). N=17

Due to the small sample size (N=17), it is hard to draw firm correlations between climatic data and SD. However, ‘Days of Rain’ showed a Spearman’s correlation coefficient of R2=0.4779412 and P=0.054 suggesting a weak positive correlation that is non-significant. No other climatic variable showed correlation close to significance including Elevation (R=0.19, P=0.46), Annual Mean Temperature (R = 0.1264, p = 0.62), Mean Annual PAR (R = -0.324, p = 0.20), Annual Precipitation (R = -0.373, P= 0.14), or Annual Temperature Range (R=-0.7, P=0.78). This suggests that there is little effect of single environmental factors on stomatal density. There are however few data points (N=17) making conclusions difficult to reach.

## Discussion

### When and How Many to Phenotype

While ideally as many individual plants of each line as possible would be used in an analysis, when looking at large numbers of lines it is not always possible to look at many individuals, as is the case in a QTL analysis. Bootstrapping was therefore used to give a suggestion of the minimum number of measurements required. Here, 16 measurements, which equates to 2 measurements from 8 individuals, gave a good representation of SD for the line Col-0. 16 measurements for each line has therefore been used as sample number throughout this study. Ideally, further ecotypes would have been studied to ensure that 16 measurements is sufficient to be representative of other lines, however when starting this project only data for Col-0 was available.

Further to knowing the ideal number of measurements to take, the ideal time for harvesting must be considered. Cotyledons expand rapidly and it is not known when or if SD becomes fixed. Size of the cotyledon has not been measured here to see when a cotyledon is fully expanded, however, the data suggests that SD is fixed by two weeks after germination in the same manner that true leaves show fixed SD two weeks after initiation (Rui et al, 2017). This analysis showed that in Can-0 SD was fixed after 10 days and in Col-0 changed little after 14 days. 18 days was therefore decided upon as the ideal time to harvest plants for SD measurements. The extra time above 14 days was included to allow for any slower to germinate lines.

### Cotyledons for High-Throughput Phenotyping

While it would have been preferred to use true leaves for high-throughput phenotyping of SD in Arabidopsis, the practicalities involved required the assessment of alternative methods. Phenotyping cotyledons requires minimal space as 24 lines can be assessed in a single tray in around 3 weeks. For mature leaves, which would require a minimum of 3 pots and 6 weeks of growth, considerably more space would be required even if growth time were reduced. The space and/or time required to carryout a QTL with over 400 lines would make this impractical. In addition, 3 plants would be the minimum required and bootstrapping would need to be carried out to ensure that this was an appropriate number. More plants may be required making the experiment even less practical.

The use of cotyledons for high-throughput phenotyping would save resources and allow for increased numbers of lines to be phenotyped. The cotyledons in this analysis showed similar densities and indexes, and SD responses to elevated [CO2], to those observed in mature leaves mature leaves. This suggested that cotyledons are representative of mature true leaves and are therefore suitable for high-throughput analysis.

The similarities in SD and SI between cotyledon and mature leaves is in line with Delgado et al (2011) who also found similar correlations between cotyledons and true leaves suggesting further that cotyledon responses are a good representation of whole plant responses. SI did not correlate as well as SD between cotyledons and true leaves in both this analysis and that reported by Delgado et al, suggesting that cotyledons are less suited for studying index. Index is a measure of stomatal initiation and, while it is correlated with density, is not used to estimate, for example, gas exchange capability. SD can therefore be considered to give better information about how plants will be affected by our future environment and it is SD that will be focused on in further experiments.

The difference between the SD in ambient and elevated [CO2] shows a strong but not significant correlation between cotyledons and mature leaves. Engineer et al (2014), found that Col-0 cotyledons show similar responses to [CO2] compared to mature leaves and also that cotyledons show changes to their response when key [CO2] response genes are knocked out. This suggests that it is the same set of genes causing response to [CO2] in both cotyledons and mature leaves.

The use of cotyledons may also have advantage over mature leaves of a reduction in confounding effects from systemic signals. The SD of Arabidopsis developing leaves has been shown to be affected by the conditions of other parts of the plant through systemic signals (Lake et al, 2002). While this should not be a problem in a controlled growth chamber, variation in conditions does exist. Conditions earlier in the plant’s growth could affect the plant later which could be a particular problem in a population which grows at different rates. Cotyledons prevent this as they are the first leaves.

The seed to be used in the QTL analysis was provided by Dr Paula Kover (University of Bath) who grew all lines at the same time in greenhouse conditions. While some variation may exist in growth conditions, it is likely that this method of growth will have reduced maternal and epigenetic effects. Cotyledons SD response to carbon dioxide is strongly impacted by the conditions the maternal plant was grown in, with plants grown in elevated co2 shown to have offspring with reduced stomatal density (Haus et al, 2018). Maternal effects can therefore influence SD of cotyledons.

The evidence of Engineer et al (2014) showing cotyledons show the same changes in density and gene expression in different [CO2] treatments, and the correlation shown here between cotyledons and mature leaves suggests that it is suitable to use cotyledons as a proxy to look at mature plant SD response to elevated [CO2].

### Response of 19 parental Ecotypes

The 19 parental ecotypes that were used to create the MAGIC lines show a wide range of SD, SI and SD responses to elevated [CO­2] (Figure 2.3.8). Can-0 consistently showed low SD while Ler-0, the *er* mutant created in the lab by irradiation (Redei et al, 1993), showed the highest SD. The wide range of SDs between the parental ecotypes suggests that there is enough diversity to use them for a QTL analysis on SD. That there are changes in SD within lines between [CO2] treatments suggests that there is Gene X Environment (GXE) interaction (as outlined in section 1.7.2) making the MAGIC lines suitable for use in a QTL to find genes associated with response to [CO2] as well. The changes in SD between 180ppm to 450ppm and 450ppm to 1000ppm suggest that it may not be a single set of genes controlling response to [CO2]. For example, in this study, many ecotypes showed a decrease in SD between subambient and elevated [CO2] but an increase in SD between ambient and elevated [CO2]. Engineer et al suggest a pathway for the change in SD between 150ppm and 500ppm [CO2] and it may be that there are further differences in SD response to [CO2] and different genes responsible for SD changes between 450ppm to 1000ppm [CO2].

As future changes to atmospheric [CO2] are expected to be increases in [CO2], the difference between 450ppm and 1000ppm is the main focus of this thesis. The range in SD and response to [CO2] in the parental lines of the MAGIC lines makes them suitable to be used in a QTL analysis as, while many lines are similar in the direction of change in SD, there are differences in scale and some changes in direction. The cotyledons of the MAGIC lines were therefore used for QTL analysis to identify genetic loci in SD responses as described in chapter 3.

### Environment and SD

It was investigated whether the SD of the parental lines could be related to the environment that they originated from. There was limited correlation between environmental factors studied and SD in this study. This is likely to be because there was limited data available for the study with only 17 lines used. There are 1000s of Arabidopsis lines available as inbred ecotypes and a more robust study would include many more of them than included here. However insights can be gleaned from this limited dataset.

The correlation shown between days of rain and SD, and the lack of correlation between annual mean rainfall and SD suggests that for SD, it is the frequency of rain that is more important for adaption in stomatal numbers rather than the quantity of rain. Plants from areas with high rainfall that fell in a short period of time, such as Can-0 from the Canary Islands, have to survive through significant periods of drought. Having fewer stomata has been shown to be an advantage in surviving drought conditions (Doheny-Adams et al, 2012) and days of rain could therefore be a better predictor of SD than total annual rainfall. A better prediction of SD may be water vapour pressure deficit (VPD), the measure of water vapour in the air compared to what it can hold before water starts to precipitate (Prenger and Lin, 2001). VPD has previously been shown to affect SD in plants as an increased VPD, indicating drier air, results in increased transpiration (Marchin et al, 2016). Ideally, VPD values for ecotypes would therefore also be examined and may be a better measure for water stress than days of rain.

Light intensity has been shown to affect SD in many plants including *Sinapis Alba,* Oak, tomato and Arabidopsis (Wild and Wolf, 1979; Furukawa, 1997, Gay and Hurd, 1974, Kang et al, 2009). However, there is no current evidence of adaption to light levels from plants originating from different environments, only plasticity when exposed to different laboratory conditions. In high light, plants develop more stomata which is believed to allow for better control of cooling under highly photosynthetically active periods and to allow adequate [CO2] uptake to meet photosynthetic demands (Boccalandro et al, 2009). In this study, there was no correlation between average PAR of the source environment and SD among the 17 studied ecotypes, suggesting that either there are other factors more important to the plant than adaption to light or that plants have high plasticity to light anyway, and there is no need for light adaption.

Elevation has previously been shown to effect SD in Arabis alpine however it was shown not to be affected in Arabidopsis (Kammer et al, 2015). The rise in elevation results in less available [CO2] due to changes in air partial pressure which can result in plants increasing SD in order to uptake the same amount of CO2. In this study, there was no correlation between SD and source elevation, however the small number of lines originating from a limited elevation range may be limiting the analysis.

Arabidopsis plants grown in elevated temperature have a reduced SD (Crawford et al, 2012). Crawford et al (2012) theorise that the decreased SD in elevated temperatures allows for increased space for diffusion of vapour. There was no evidence of ecotypes originating from higher temperature environments having a lower SD in this sample of 17 ecotypes. The source temperature range of 14.9°C shows a large variation between samples however this did not lead to a clear trend. This suggests that temperature is not a major factor in adaption of *SD* to environmental conditions

Overall this brief analysis of SD and environment suggests that it is water availability in the form of frequency of rain that is the major driver that affects adaption of SD to environment. Other factors may not have as much of an effect because a lot of plasticity is available already allowing for acclimation without need for adaption. It may also be that it is the adaption to a combination of factors that controls SD and there is no one overarching regulator. The limited sample size prevents clear conclusions from being drawn. It is likely that with data from more ecotypes, this analysis would result in clearer trends in correlation between climatic variables and SD.

### Conclusions

In conclusion, this analysis has shown that fully expanded cotyledons at least 18 days after germination are a suitable organ for the investigation of SD response to [CO2]. The parental ecotypes of the MAGIC Arabidopsis lines show a wide range of cotyledon SD and SD response to elevated [CO2] phenotypes suggesting that they have enough variety in their genetics at the loci that control SD traits that the MAGIC lines will be suitable for use in a QTL analysis. This analysis also suggests that the environment that the ecotype was collected from does not have a clear correlation with cotyledon SD, but suggests that frequency of rainfall maybe a factor in the adaption of cotyledon stomatal densities to source environment.

# Quantitative Trait Loci Analysis of Stomatal Density in Different [CO2] Treatments

### QTLs for [CO2] response

In the previous chapter (Chapter 2.3.4), the SD response of 19 Arabidopsis ecotypes to elevated [CO2] was examined and it was determined that there was significant variation between these ecotypes in their responses to [CO2]. Previously a transcriptomic approach was used by Engineer et al (2014) to identify genes involved in plant stomatal responses to elevated [CO2] however, this approach was dependent on the trait being regulated at the transcriptional level. Any alternative approach, which does not make *a priori* assumptions regarding the regulatory mechanism (e.g. transcriptional, post-transcriptional, epigenetic etc), is to perform a Quantitative Trait Loci analysis. A QTL analysis utilises measured trait data and correlates it to genomic data to statistically determine which regions of the genome have the most significant control over a trait.

There are currently no known QTLs for stomatal density (SD) responses to [CO2], only for SD in a particular [CO2]. A previous QTL analysis was carried out on SD response to [CO2] on a poplar hybrid grown for a full growth season in either ambient (400ppm) or elevated (600ppm) [CO2] by Ferris et al (2002). This analysis found 2 QTL for abaxial SD and 3 for abaxial stomatal index in ambient [CO2] but no QTLs were found for abaxial SD or abaxial stomatal index in elevated [CO2] suggesting these were not detected. This could indicate that under these conditions a large number of genes make small contributions to the trait and that this was below the QTL detection limit. It also did not show any QTLs for the change in SD between the two [CO2] treatments, although it is also not clear this was analysed (Ferris et al, 2002). The areas of the genome that control stomatal density responses to [CO2] are therefore currently unknown, however the discovery of QTLs in ambient but not elevated [CO2] suggests that there are differences in genetic controls between the treatments.

QTLs have been identified for the response of stem and root phenotypes to elevated [CO2] (Rae et al, 2002). Therefore, QTLs for responsiveness of other tissues to [CO2] exist, but they have not been found for SD. In this chapter, the aim was identify the genes that regulate the SD response to elevated [CO2].

### MAGIC lines

MAGIC lines were chosen to look at the response of Arabidopsis to elevated CO2 because of the reasons outlines in Chapter 1.7.3 but the main reasons will be briefly described here. The MAGIC lines are derived from 19 parental ecotypes (studied in Chapter 2) and are a set of recombinant inbred lines (RILS) designed for use in QTL studies. Multi-parent RILs show advantages over biparental RILs as they contain more minor alleles allowing for detection of less common phenotypes to make them more representative of natural populations. They also show advantages over natural accession populations as they are less prone to ghost QTLs which arise due to unknown population structures (Kover et al, 2009).

There are 1039 MAGIC lines available for use. Of these lines, 703 have had improved sequencing from the initial 1260 SNPs resolution to 500K SNP resolution (http://mtweb.cs.ucl.ac.uk/mus/www/magic/). This improved resolution allows for better MAGIC genome assembly by the hidden Markov model so that the mosaic genomes of the MAGIC lines have resolution that is correct to 1kb. This in turn allows for better QTL mapping and haplotype phenotype analysis. More recently, these lines have been renamed as High-resolution Sequenced Recombinant Inbred Lines (HSRIL) and were the lines used in this study.

### QTL and GWAS

The choice between using QTL analysis and Genome Wide Association Study (GWAS) is usually due to the set of lines that have been chosen for phenotyping. Traditionally, QTL analysis has been used in 2 parent RILs and GWAS is used when natural ecotype populations are used. As the MAGIC lines are a combination of these two types of lines, both GWAS and QTL analysis can be run. QTL analysis can find genes with effects as small as 5% (Kover et al, 2009). GWAS are less likely to detect these small effect QTLs however, they provide greater mapping resolution (Wang et al, 2005). A combination of QTL analysis and GWAS can be used to identify first QTLs, and second SNPs of interest and this has been successfully used in Arabidopsis to identify flowering genes (Brachi et al, 2010).

In addition to traditional methods of mapping, gene interactions in different environments can be analysed using RILs. If different genotypes show different slopes of reaction to the environmental change, then there is an environmental effect on the trait caused by a genetic factor, known as a G X E interaction. There are many methods of mapping the genes or QTLs involved in these G X E effects, however the simplest is to run a QTL analysis on the difference between the environments (El-Soda et al, 2014).

### Finding the causes of QTLs

Once a QTL has been identified, traditionally the method of finding the genes underlying the QTL is by fine mapping. In this method, the two most dissimilar lines are crossed and inbred multiple times to create a new set of RILs which are then sequenced. The rationale for this, is the assumption that these lines carry different alleles of the underlying gene/s. A QTL on the region of interest is then run, mapping the QTL to a much finer degree and can create resolution down to the Single Nucleotide Polymorphisms (SNPs) within the gene associated with the QTL (Price, 2006).

Alternatively, given that T-DNA insertions cover approximately 80% of the Arabidopsis genome (Li et al, 2006), it is possible to use a candidate gene approach. Candidate gene approaches for gene discovery have successfully worked for identifying the roles of *CRSP*, *CA1* and *CA4* in the [CO2] pathway (Engineer et al, 2014). An advantage of the QTL approach in these cases is that it can significantly reduce the number of genes to choose candidates from. There are several datasets which can be consulted to identify genes of interest in QTL regions associated with SD and SD response to [CO2] to begin phenotyping. The data sets considered were:

1. The RNA-Seq data set by Gan et al (2011) of the 19 parental lines transcriptomes can be used to look for genes which are differentially expressed between the most different lines according to the haplotype model. This paper also provides the genomic sequences of the 19 lines and can be used to look for non-synonymous mutations which could be affecting protein function.
2. The gene expression profiles of different stomatal lineage cell types during key stomatal developmental time by Adrian et al (2015) shows which genes are upregulated upon stomatal initiation. Differences in these genes have the potential to interrupt stomatal development and increase or reduce SD. In particular, the gene *SPCH* is a key initiator of stomatal development, and looking for regulation by *SPCH* could be strong indicator that a gene is involved in stomatal development.
3. The paper from Watson-Lazowski et al, 2016, discusses the transcriptional changes of the plant *Plantago lanceolata L.* in ambient and elevated [CO2]. They also looked at the effects of long term adaption of the plant to elevated [CO2] by collecting plants from a natural CO2 spring and carrying out RNA-seq to compare to those from non-elevated [CO2] areas. The RNA-seq is mapped to the Arabidopsis genome giving an easy way of looking at genes that are differentially expressed in different [CO2] concentrations.

There are many genes and processes which have previously been shown to affect the SD of Arabidopsis as well. While stomatal initiation can change density, so can changes to hormones, changes to the epidermal cells, changes to metabolites, or to changes in the ability to signal (See chapter 1). Looking for genes related to these changes can therefore provide candidate genes for further analysis.

The final list of criteria for genes to be looked at around QTLs was therefore based on the following criteria:

1. Had significantly different expression between highest and lowest haplotype in the RNA-Seq data set (Gan et al, 2011)
2. Had a non-synonymous mutation in an exon that was different between highest and lowest haplotype (Gan et al, 2011)
3. Has a significant change in expression in the *SPCH* dataset (Adrian et al, 2015)
4. Showed a change of expression in the RNA-Seq CO2 spring plants dataset (Watson-Lazowski et al, 2016)
5. Related to ABA (Tanaka et al, 2013) or Auxin (Le et al, 2014)
6. Related to cell wall synthesis or modification (Gray et al, 2000)
7. Related to sugar/starch metabolism (Akita et al, 2013)

### Aims

The aims of this chapter were to:

1. Quantify the variation of SD in the MAGIC lines to see if they are a suitable choice of RILs for QTL analysis.
2. Carry out QTL analysis to identify regions of the genome associated with SD in 450ppm and 1000ppm [CO2] and the difference in SD between [CO2] treatments.
3. Identify the genes underlying the QTLs and analyse their role in regulating SD.

## Methods

### Plant Material

MAGIC line seed was kindly donated by Dr Paula Kover (University of Bath). T-DNA insertion lines (See Table 3.2‑3 for full list) were ordered from the Nottingham Arabidopsis Stock Centre (NASC)

### Plant Growth

Plants were grown in F2+sand Levington, Everris Professional compost in Sanyo-Gallenkamp SGC970/C/HQI cabinets for 18 days in either 450ppm or 1000ppm [CO2]. CO2 was supplied from CO2 cylinders (BOC, UK). Before sowing, plants were stratified for 72 hours in the dark at 4°C in water before being spread using a pipette on compost. Plants were grown at 20°C in 10 hr days at 200 µmol m2 s-1 light. Plants were grown between April 2015 and May 2016 in 31 batches. Some batches were disregarded because of poor growth or cabinet failures.

Plants were grown in batches of up to 24 MAGIC lines at a time in 6X10 trays with the outside rows being unused. Lines were grown in a random pattern in one of two cabinets. Interspersed with the MAGIC lines, at least 2 pots each of Col-0 and Rsch-4 were grown in each batch as controls to check for cabinet effect, batch effect and positional effect.

### Phenotyping

5-8 cotyledons from 5-8 plants of each MAGIC line with harvested and moulds made in the manner described in chapter 2.2.3. For stomatal density counts, two regions of area 0.16mm2 where chosen in the same manner as described by chapter 2.2.4. A total of 10-16 areas were therefore counted for each line. Lines where there were fewer than 5 cotyledons available for counting were disregarded or grown again in another batch to reach the required number of 5-8 cotyledons. The number of images per line measured was decided using the bootstrapping in Chapter 2.3.1

### QTL analysis

The QTL analysis was carried out in the manner described by Kover et al, 2009 with the updates and instructions described by <http://mtweb.cs.ucl.ac.uk/mus/www/19genomes/magic.html>. The three step process carried out in the r package Happy is described below.

First, the approximate genomes of the MAGIC lines are created. This is done by a hidden Markov model which uses the 19 parental genomes and the SNP markers for the MAGIC lines. The model creates a probability that based on the current SNP, and the surrounding SNPs, what is the most probable parental line at that part of the genome. This creates a mosaic genome with approximately 72% accuracy at each loci (Kover et al, 2009).

Next the program carries out genome scanning. It does this using a fixed effects model in each SNP interval and at first ignores the effects of the other QTLs. It looks at the difference in phenotype against the difference in genotype and performs ANOVA at each position along the genome to evaluate if there is a QTL. The original SNP resolution of Kover et al, was 96Kb, however with the 500k SNP resequencing the SNP resolution is now much improved.

The final step is to rerun the model 500 times, fitting the QTL model to a resampling of the data, taking 80% of the lines each time. This checks the robustness of the QTL, and the number of times the QTL is detected creates support for the QTL being real and not a ‘ghost’ QTL. It is also at this point that the QTLs are evaluated in context of each other. Each time 80% of the data is taken, QTLs from the models are added to see at what point the model best fits the data. The location of these QTLs is then recorded. This part of the model is run using bagphenotype R software.

QTL effect size is estimated by the fraction of variance that the QTL accounts for in each QTL model and the final value is the average of all the 500 resampling models run.

### Haplotype analysis

This is the estimation of the founder strain effect at the QTL. This model uses the results of the fixed effects multiple regression model run for the QTL. It takes the least squares estimates (which show how far the data is from the predicted data) from the model runs and uses one way ANOVA to estimate the average distribution of phenotypes of each genotype which are then outputted as a mean for each run of the model. The mean of each model is then used to produce the box plots of founder strain effect to give the estimated haplotype for each parental line.

### Gene Ontology

Gene Ontology for the each gene was found using Panther DB gene ontology tool (http://pantherdb.org/). Full names of gene lists were found using arabidopsis.org bulk data downloader (https://www.arabidopsis.org/tools/bulk/index.jsp).

### DNA extraction

For genotyping, genomic DNA was extracted from leaf or seedlings using the Edwards prep method (Edwards et al. 1991). A leaf disc, of approximately 1cm in diameter, or several 7-14-day old seedlings were ground in 400 μl Edwards Solution [200mM Tris-HCl (pH7.5), 250mM NaCl, 25mM EDTA, 0.5% SDS] using a vortex and ball bearing and then centrifuged at 13,500rpm for 5-10mins. The supernatant was transferred to a new 1.5ml Eppendorf and 400μl isopropanol (Fisher Scientific Laboratory Grade Propan-2-ol, 1067432) was added. The sample was mixed and centrifuged at 13,500rpm for 5-10mins to pellet the DNA. The supernatant was aspirated using a pipette without disturbing the pellet and the pellet was air dried for 5mins. The pellet was then reconstituted in 100μl sterile H2O and vortexed to mix. DNA was then stored at -20°C.

### PCR

The following method has been adapted from Sigma Aldrich JumpStart™ RedTaq® ReadyMix™ PCR Reaction protocol. PCR components, including RNA free, sterile H2O, forward and reverse primers (100μM stocks), DNA template and RedTaq (Sigma-Aldrich JumpStart™ RedTaq® ReadyMix™ PCR Reaction Mix with MgCl2) were first melted at room temperature. The forward and reverse primers for PCR can be seen in Table 3.2‑1.

Table 3.2‑1 Primers used in PCR

|  |  |  |
| --- | --- | --- |
| Primer Name | Direction | Sequence |
| LB SAIL | R | GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC |
| ADG2 F2 | F | GTGCCTCAGATTTTGGGCTT |
| ADG2R2 | R | ACTCCCATTGAAGCTATGTATGGT |

2X master mix was prepared according to the number of samples. 22.5μl of master mix was pipetted in to each PCR tube (0.2ml). 2.5μl of template DNA was in to each tube, mixed and briefly centrifuged to spin down the contents and eliminate air bubbles.

For example, the following set up was used to make master mix for one sample:

12.5ul RedTaq 2x mix

2.5ul gDNA

0.5ul For primer (10pmole/ul)

0.5ul Rev primer (10pmol/ul)

9ul H2O.

Samples were placed in the thermal cycler following the program shown in Table **3.2**‑**2**. Reaction volume was set at 25μl and run for 35 cycles. Results were analysed by agarose gel electrophoresis with PCR products resolved on 1% agarose gels.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3.2‑2 The thermal profile, incubation temperature and times periods for PCR. | | | | | | |  |
|  | Step 1 | Step 2 | Step 3 | Step 4 | Step 5 | Step 6 | Step 7 |
| Temperature(°C) | 95 | 95 | 55 | 72 | 35 repeats of steps 2-4 | 72 | 12 |
| Time (minutes) | 3 | 0.5 | 0.5 | 1 | 5 minutes | end |

### Gel Electrophoresis

DNA separation via agarose gel electrophoresis was used to visualise the products of PCR. Gels were 1% agarose. Agarose powder (Sigma Agarose Gel powder) was mixed with 1 X TAE buffer and dissolved by boiling. The fluorescent indicator Ethidium Bromide (stock 10mg/ml) was added to the liquid gel to a final concentration of 1μg/ml. The gel was then poured into a gel tray fitted with a comb.

The gel was placed in a tank filled with 1X TAE buffer and 10µl of samples loaded in to each well and a DNA ladder (2.5μl GeneRuler DNA LadderMix; Fisher Scientific 10181070) loaded alongside to allow for fragment size comparisons. Gels were run at 100-120V for 20-40 minutes using BioRad mini sub-cell and power supply. Gels were visualised using GelDoc-It™ system (UVP LLC) and images were taken using VisionWorks® LS analysis software (UVP LLC).

Table 3.2‑3 Gene and their respective T-DNA insertion lines. Line abbreviations refer to names given in text and figures

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gene | Gene name | T-DNA line | Background | Line abbreviation |
| AT5G18450 |  | SAIL\_1160\_E12 | Col-3 | Sail 116 |
| AT5G18620 | CHR17; CHROMATIN REMODELING FACTOR17 | SALK\_080144C | Col-0 | SALK\_080144 |
| AT5G18670 | BAM9; BETA-AMYLASE 3; BETA-AMYLASE 9; BMY3 | SALK\_021635C | Col-0 | SALK\_021635 |
| AT5G18840 |  | SALK\_079666C | Col-0 | SALK\_079666 |
| AT5G18930 | BUD2; BUSHY AND DWARF 2; SAMDC4 | SALK\_007279 | Col-0 |  |
| AT5G18960 | FAR1-RELATED SEQUENCE 12; FRS12 | SALK\_030182C | Col-0 | SALK\_030182 |
| AT5G19095 |  | SALK\_135427 | Col-0 |  |
| AT5G19330 | ARIA; ARM REPEAT PROTEIN INTERACTING WITH ABF2 | SALK\_024530 | Col-0 | SALK24330 |
| AT5G19530 | ACAULIS 5; ACL5 | SALK\_028736 | Col-0 |  |
| AT5G19540 |  | SALK\_015620C | Col-0 | SALK\_015620 |
| AT5G19550 | AAT2; ASP2; ASPARTATE AMINOTRANSFERASE 2 | SALK\_055785C | Col-0 | SALK\_055785 |
| AT5G19720 |  | SAIL\_406\_D11 | Col-3 | SAIL\_406 |
| AT5G19760 |  | SALK\_017483C | Col-0 | SALK\_017483 |
| AT5G19810 |  | SAIL\_358\_C03 | Col-3 | SAIL\_358 |
| AT5G19170 |  | SALK\_145593 | Col-0 | SALK\_14559 |
| AT5G19220 | ADG2; ADP GLUCOSE PYROPHOSPHORYLASE 2; ADP GLUCOSE PYROPHOSPHORYLASE LARGE SUBUNIT 1; APL1 | SAIL\_108\_D03C1 | Col-3 | Sail 108 |
| AT5G18550 |  | GK-115C04-012503  GK-738B11-023864 | Col-0 | GK11  GK73 |
| AT5G19450 | CALCIUM-DEPENDENT PROTEIN KINASE 19 ; CDPK19 | SALK\_036581.51.25.x | Col-0 | SALK\_03658 |
| AT5G19770 | TUBULIN ALPHA-3 ; TUA3 | CS68878 | Col-0 | Tua3 |
| AT5G19010 | PROTEIN KINASE 16 | SALK\_059737.52.15.x | Col-0 | SALK0597 |
| AT5G19260 | FANTASTIC FOUR 3 FF3 | SM\_3\_40331 | Col-0 | FF3 |
| AT5G19740 | Peptidase M28 family protein/ LAMP1 | SALK\_117012C | Col-0 | SALK\_117012 |
| AT5G19240 |  | SALK\_147600.20.20.x | Col-0 | SALK\_147600 |
| AT5G19300 |  | SALK\_202658 | Col-0 | SALK\_2026 |

## Results

### Phenotypes of the MAGIC lines

Before the MAGIC lines can be used in a QTL analysis, it must be established whether the phenotypes of the lines are normally distributed and if the phenotypes suggest the trait of SD and difference in SD between [CO2] treatments is polygenic.

|  |
| --- |
| A) |
| B) |
| Figure 3.3.1 The frequency of stomatal density phenotypes in the MAGIC lines. Distribution plots of MAGIC lines phenotype including those with Ler-0 haplotype at ERECTA shown in (A) (ambient N=393; elevated N=386; difference N=333) and not including Ler-0 haplotype at ERECTA (B) (ambient N=370; elevated N=359; difference N=310). A shift towards higher stomatal densities can be seen in the distribution of elevated [CO2] compared to ambient |

The MAGIC lines show a wide range of SD phenotypes (Figure 3.3.1) with the lowest SD in 450ppm being 50 per mm2 in HRIL489 and the highest SD in 450ppm is HSRIL170 with a SD of 230 per mm2. This is a 4.6X difference in SD between the largest and smallest SD in 450ppm. The MAGIC lines in 1000ppm show a smaller range in SD than that seen in 450ppm [CO2]. The largest SD in 1000ppm is 226 per mm2 in HSRIL 215 and the lowest SD is 55 in HSRIL229. The overall range in SD in 1000ppm is 4.1X (

Table 3.3‑1).

The range in MAGIC lines SD is larger than the natural variation in SD found by Delgado et al (2011) of 3.5X in wild ecotypes. However, it is still smaller than the laboratory manipulated range of SDs of 16X created by Doheny-Adams et al (2012). The SD range is greater than the range shown by the MAGIC parental lines of 2.5X (Figure 3.3.2), which suggests that SD is under polygenic control.

Ler-0 has an exaggerated phenotype because of a SNP in *ER* which causes loss of function and is entirely consistent with the role of ER as a major regulator of stomatal development (Torii et al, 1996). The Ler-0 phenotype is a mutagenised phenotype that appeared to dominate the analysis with its strong phenotype. Due to these problems with Ler-0 *er* haplotype, analysis was run again without the Ler-0 haplotype at *ER* to look at naturally available variation in SD. Removal of Ler-0 haplotype at *ER* reduces the range of SD. The highest SD in 450ppm [CO2] becomes 160 per mm2 in HSRIL196 and the range in SD is 3.2X which is smaller than the range of SDs seen by Delgado et al (2015). In 1000ppm [CO2] with Ler-0 haplotype at *ER* removed the highest SD is 174 and the SD range is 3.1X (See

Table 3.3‑1).

The range of SD responses to elevated [CO2] in individual HSRIL lines ranges from a 40% reduction in SD (HSRIL 229) to a 70% increase (HSRIL 684) in SD with the Ler-0 *ER* haplotype included. With Ler-0 *ER* haplotype removed the largest increase is reduced to 60% (HSRIL 39). This is again much wider than seen in the parental lines where the largest SD increase is 26% in Sf-2 and the largest SD decrease is 25% in No-0. The larger range of differences in SD in individual MAGIC lines suggests that the trait of SD response to [CO2] is polygenic in Arabidopsis, similar to the trait of SD.

Table 3.3‑1 Phenotypes of the MAGIC lines. The table shows the highest and lowest phenotypes of the MAGIC lines including and excluding lines with the Ler-0 er haplotype

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Phenotype** | **PPM** | **Including Ler-0** | | **Excluding Ler-0** | |
| **Phenotype** | **HSRIL** | **Phenotype** | **HSRIL** |
| **Lowest SD** | **450ppm** | 50 per mm2 | 489 | 50 per mm2 | 489 |
| **1000ppm** | 55 per mm2 | 229 | 55 per mm2 | 229 |
| **Highest SD** | **450ppm** | 230 per mm2 | 170 | 159 per mm2 | 196 |
| **1000ppm** | 226 per mm2 | 215 | 174 per mm2 | 736 |
| **Range SD** | **450ppm** | 4.6X | - | 3.2X | - |
| **1000ppm** | 4.1X | - | 3.1X | - |
| **Highest positive SD change** | **-** | 70% | 684 | 60% | 39 |
| **Highest negative SD change** | **-** | 40% | 229 | 40% | 229 |

There is little difference in SD overall between the two [CO2] treatments. There is a small overall increase in average SD from 95.7 per mm2 in 450ppm to 101.4 per mm2 in 1000ppm [CO2] which is significant (T Test, P=0.001), however this is a very small shift in SD. As the shift is small and the lines show strong correlation with each other, it is not suggestive of a G X E interaction, however it is possible this is because the lines are shifting in different directions due to haplotype and thus cancelling out an observable change.

The MAGIC line SD distributions suggest that the trait of SD is ideal for QTL analysis showing both normal distribution, and polygenic control. However, the SD difference between [CO2] treatments suggests that there is no obvious GXE meaning the QTL on SD differences between [CO2] treatments was unlikely to be effective. This analysis was carried out in cotyledons, and while there is good correlation shown between mature leaves and cotyledons in Chapter 2.3.3, it is not a perfect correlation (R2 = 0.66) . It is possible that analysis in mature leaves would find stronger shifts in SD between treatments from a stronger GXE effect.

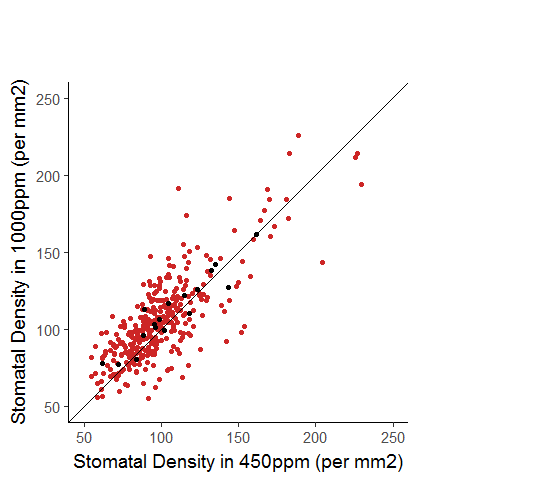


Figure 3.3.2 The range of stomatal density phenotypes of MAGIC lines in ambient and elevated [CO2]. Red indicates MAGIC line SD and black shows the MAGIC parental lines SD. Line indicates equal SD in both ambient (450ppm) and elevated (1000ppm) [CO2]. There is significant difference between SDs in ambient and elevated CO2 (T-Test, P=0.001)

### Controls

In order to ensure the phenotyping of the MAGIC lines was consistent across batches and positions, Col-0 and Rsch-4 were grown in every batch. The variation between batches was then analysed to test the hypothesis that batch and position would not affect SD.

|  |
| --- |
| A) |
| B) |
| Figure 3.3.3 The variation in Stomatal Density of controls between batches. SD of Col-0 (A) and Rsch-4 (B) grown in batches can be seen to vary between batches, as can response to [CO2] Error bars indicate standard error. Dark grey indicates 450ppm and light grey 1000ppm [CO2] N=16-32 |

There was a significant effect of both batch and position on SD (P<0.001) in Col-0 (Figure 3.3.3A). However, Tukeys test showed no significantly different batches, and only 5 significantly different positions within the tray (P<0.05). This suggests that either the ANOVA or Tukey’s is giving a type I or Type II statistical error, however it is not clear which. There is a significant amount of data in the ANOVA (N=1469) contributing to its robustness but that does not preclude error. The mean SD of Col-0 in each batch show recognisable differences, ranging from 105 per mm2 to 120 per mm2 in 450ppm [CO2] and 88 to 108 per mm2 in 1000ppm [CO2] which would support the original ANOVA result of significant differences and suggest Tukey’s is giving a type II error.

Rsch-4 shows similar variation in mean SD between growth batches with a SD range of 69 – 82 per mm2 in 450ppm [CO2] and 70-85per mm2 in 1000ppm [CO2]. However, Rsch-4 does not show significant difference in SD for batch and position (ANOVA; P>0.05), although post hoc Tukey’s did suggest there are some differences in SD between batches (P<0.05).

Looking at the data and its variability, batch and position are likely to be having an effect on SD. However, this variability has not been taken into account in the MAGIC lines SD data put into the QTL model. While stomatal densities could have been adjusted based on position and batch, it is not possible to know how much adjustment would be needed and how much would be over-fitting. Within batch 3, for example, there is no significant effect of position on Col-0 cotyledon SD (P=0.39) and it would be inappropriate to adjust based on position. Similarly it can be seen that in batch 24, the batch with the highest average SD for Col-0 in 450ppm [CO2], Rsch-4 has its third lowest average SD in 450ppm [CO2]. This implies variation within cabinets affects different lines in different ways and it would not necessarily be possible to control for batch and positional effects on SD based on this data alone.

It can also be seen in Rsch-4 that the same response in SD to elevated [CO2] is not recorded in every batch. This suggests that even in a controlled environment there is enough variation for non-treatment factors to influence SD. It is therefore not clear that these SD responses would be seen in a natural environment where many more factors are present.

Overall this data shows that the QTL analysis shows relatively good consistency between batches and positions, although there is some room for improvement. More consistency between batches may have resulted in more QTLs being found, however the consistency is enough that the QTL analysis went ahead.

### QTL mapping with Ler-0 haplotype on chromosome 2

A QTL analysis was run using all the MAGIC lines phenotyped. While 700 lines were sowed, only 451 lines germinated and many did not germinate in both conditions. The number of lines included in each analysis therefore varies depending on germination rate.

|  |
| --- |
|  |
|  |
|  |
| Figure 3.3.4 The QTL analysis results. These scans show the position on the chromosome and the logP value for the likelihood of a QTL. QTLs above the threshold are indicated with an orange dot. |

Table 3.3‑2 QTLs found by the QTL analysis including all MAGIC lines phenotyped. QTL name is how the QTL is referred to in this thesis. SNP refers to the SNP with the most significant LogP score. LogP is the log of the total LOD score from 500 permutations. Chr is the chromosome the QTL is found on. Genome wide P value is the probability the QTL is having an effect when other genes have been taken into account. Variance is how much variance the QTL explains of the phenotype. N is the number of MAGIC lines in the analysis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Phenotype | QTL  Name | SNP | Position | LogP | chr | Genome wide p value | Variance | N |
| Elevated CO2 Stomatal Density | QTL1 | PERL0174281 | 20682947 | 3.77663 | 1 | 0.055 | 9% | 386 |
|  | QTL2 | MN2\_11300378 | 11293296 | 39.4789 | 2 | 0 | 46% | 386 |
|  |  |  |  |  |  |  |  |  |
| Ambient CO2 Stomatal Density | QTL2 | MASC09221 | 11424767 | 38.6668 | 2 | 0 | 44% | 393 |
|  | QTL5 | NMSNP5\_6416385 | 6416385 | 3.82382 | 5 | 0.058 | 11% | 393 |

The QTL analysis, with all the MAGIC lines phenotyped included, resulted in 2 QTLs being identified for SD in 450ppm and 2 QTLs being identified for SD in 1000ppm. No QTLs were detected for the percentage difference in SD between treatments, the ratio of SD between treatments or the raw difference in SD per mm2 between treatments. This was true even after adjusting to a less stringent LOD threshold of 2.

Table 3.3‑2 shows the QTLs for SD in ambient and elevated [CO2], their positions on their respective chromosomes, LogP score, genome wide p values and variance explained.

Although no QTLs were detected in this QTL analysis on the effect of [CO2] on SD, it does not mean that there are no genes controlling the trait. The MAGIC lines and the accompanying software designed for it, predicts that with 460 lines all genes of 10% effect and 52% of 5% effect would be found (Kover et al, 2009). 333 lines have been used in the QTL analysis on difference in SD between [CO2] treatments which will reduce the analysis’s power to identify QTLs of small effect. This analysis would therefore suggest that there could be many genes of small effect that control how plants respond to elevated [CO2]. In Chapter 1.6 several genes (eg *HIC*, *CA1*, *CA4*, *CRSP*) were discussed as genes that control SD in response to [CO2] changes, however no QTLs for these genes has been found. It suggests that while these genes may be involved in the SD [CO2] response pathway as small contributors to variation, they are not major regulators of the variance. They may also not be required for these environmental conditions as many (*CRSP*, *CA1/CA4*) were genes found to control the change in SD between 150ppm and 500ppm, not 450ppm to 1000ppm [CO2]. It can only be concluded from this, however, that their contribution to control of the trait was below the detectable threshold, and that further genes are likely playing a part.

As the Ler-0 haplotype on the *ER* gene has such a dominant effect on the QTL analysis, explaining 44% of the phenotypic variance, it was decided to rerun the analysis with Ler-0 haplotype at *ER* excluded from the QTL. This resulted in the removal of 23 lines from the analysis.

### QTL Mapping without Ler-0 haplotype at ERECTA

The removal of the 23 lines found to have the Ler-0 haplotype at ER resulted in different QTLs being detected and different genome wide significance levels. The results of the QTL analysis are in

Table 3.3‑3

|  |
| --- |
|  |
|  |
|  |
| Figure 3.3.5 The QTL analysis results with Ler-0 haplotype at Erecta removed. These scans show the position on the chromosome and the logP value for the likelihood of a QTL. QTLs above the threshold are indicated with an orange dot. |

Table 3.3‑3 QTLs found by the QTL analysis of MAGIC lines with those with the Ler-0 haplotype at ER removed. QTL name is the name the QTL is referred to in this thesis. SNP refers to the SNP with the most significant LogP score. LogP is the log of the total LOD score from 500 permutations. Chr is the chromosome the QTL is found on. Genome wide P value is the probability the QTL is having an effect when other genes have been taken into account. Variance is how much variance the QTL explains of the phenotype. N is the number of MAGIC lines in the analysis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Phenotype | QTL  Name | SNP name | Position | LogP | chr | Genome wide p value | Variance explained | N |
| Elevated CO2 Stomatal Density | QTL1 | PERL0172350 | 20353910 | 4.91 | 1 | 0.004 | 9% | 359 |
|  |  |  |  |  |  |  |  |  |
| Ambient CO2 Stomatal Density | QTL1 | MN1\_21944762 | 21944762 | 5.836687 | 1 | 0.001 | 9% | 370 |
|  | QTL2 | ATC\_828 | 11773084 | 3.645014 | 2 | 0.043 | 10% | 370 |
|  | QTL5 | MASC02570 | 5422334 | 4.026973 | 5 | 0.013 | 13% | 370 |

The QTL mapping of SD with the Ler-0 haplotype at *ER* removed shows one QTL on chromosome 1 (QTL1) in 1000ppm [CO2] which is in approximately the same position as the QTL found on chromosome 1 with Ler-0 included. QTL1 did not have genome wide significance with Ler-0 included (P=0.055), however, without Ler-0 haplotype included it is statistically significant (P=0.004). The genome wide significance of each QTL can be seen in

Table 3.3‑2 for Ler-0 ER haplotype included and

Table 3.3‑3 with for Ler-0 ER haplotype excluded.

Ambient [CO2] SD shows 3 QTLs (QTL1, QTL2, QTL5) which can be seen in

Table 3.3‑3. The QTLs include a QTL approximately in the position of *ER (QTL2)* and a QTL in approximately the same position on chromosome 5 as one of the QTL found with Ler-0 *ER* haplotype included in the QTL analysis (QTL5). The other QTL (QTL1) found in 450ppm is in approximately the same position as QTL1 on chromosome 1 in 1000ppm.

The difference in the number of significant QTLs between [CO2] treatments suggest that there are different controls on QTLs in different [CO2] treatments. The difference in QTL number would appear to contradict the evidence of the phenotypes which suggested no GXE interaction and that no QTLs were found for the different in SD between treatments. The difference in QTL number between treatments suggests that the GXE effect is being masked by different genotypes responding in different directions. While the QTL on chromosome 1 is consistent across treatments and explains approximately 9% of the variation regardless of the [CO2] treatment, both the QTL on chromosome 2 and 5 are not significant in elevated [CO2] suggesting there are factors which reduce the significance of these genes, even though they may still have a small effect.

Overall, the maximum amount of variance explained by the QTL analysis is 32% leaving 68% unaccounted in 450ppm. In 1000ppm only 9% of variance is explained leaving 92% unaccounted for. This suggests that there are many genes in control of SD, but only very few have an effect large enough to detect.

### Haplotype Analysis of Ecotypes at each QTL with Ler-0

A haplotype analysis looks at the modelled phenotype of each parental lines genotype at that locus. It uses least squared differences from the QTL model to predict what the SD would be if a MAGIC line had each parental line’s haplotype. This gives information as to which parental line is most strongly affected by and in which direction by the QTL. This can provide information on which lines to look at for further fine mapping/gene identification as well identifying which QTLs are the same loci.

|  |
| --- |
| A) |
| B) |
| Figure 3.3.6 Predicted phenotypes of ecotype’s haplotype. Estimates based on least squares estimates from the fixed effects mulitple regression model of QTL. Y axis shows the predicted SD per mm2. The title shows the name given to the QTL, the name of the SNP identified by the QTL analysis, chromosome number and the LOD score. The predicted phenotypes can be seen for A) elevated [CO2] QTLs and B) ambient [CO2] QTLs. |

The haplotype analysis of ecotypes at each QTL with Ler-0 *ER* haplotype included reveals several features of the QTLs. The SD phenotype at the QTL on chromosome 2 is overwhelmingly highest in Ler-0. This QTL is highly likely to be ERECTA as it is 200kb downstream of *ER*, within the 300kb region genes are expected to be found within. ER is also known to heavily influence stomatal density (Torii et al, 1996). It is interesting to note that this QTL adds approximately 170 stomatal per mm2 consistently across treatments. This suggests that any MAGIC line with the Ler-0 haplotype is non-responsive to these [CO2] treatments in terms of SD change which is in line with what was found by Caine (2016) and also suggests that having functional ERECTA is important in SD responsiveness to [CO2].

The other QTLs show a range of SD phenotypes across ecotypes with no ecotype showing an extreme high or low SD. It can be seen on QTL5 that Wu-0 is the highest SD haplotype and Can-0 is the lowest SD haplotype but none are as overwhelming as Ler-0 in QTL2.

### Haplotype Analysis without Ler-0

Without Ler-0 at *ER*, the pheontypes of the lines haplotypes at the QTLs should become easier to distinguish. The haplotype model without Ler-0 at *ER* was therefore run again.

|  |
| --- |
| A) |
| B) |
| Figure 3.3.7 Predicted phenotypes of ecotype’s haplotype with Ler-0 haplotype at ER excluded. Estimates based on least squares estimates from the fixed effects mulitple regression model. Y axis shows the predicted SD per mm2. The title shows the name given to the QTL, the name of the SNP identified by the QTL analysis, chromosome number and the LOD score. The predicted phenotypes can be seen for A) elevated [CO2] QTLs and B) ambient [CO2] QTLs. |

The QTL found on chromosome 1, based on the analysis without the Ler-0 haplotype at *ER,* is found in both [CO2] treatments and is therefore likely to be the same QTL (QTL1). This is supported by the observation that the QTL1 haplotype analysis shows similar SD patterns across ecotypes in both treatments. For example, Bur-0 has the highest SD in both treatments and Can-0 shows the lowest SD in both treatments. This pattern is also observed when Ler-0 is included in the analysis suggesting they are all the same QTL.

Within these analyses for QTL1, the haplotype for Po-0 shows a very wide range of SDs as well as outliers. This is because very few MAGIC lines with Po-0 at the QTL’s position were phenotyped, with a maximum of 9 in each QTL run compared to 30 for Hi-0. This is not because of selection bias as lines were chosen at random, however it could be because of poor germination of Po-0 haplotype seed. This could be corrected with further phenotyping however it means that here the haplotype analysis for Po-0 at QTL1 should be disregarded.

The QTL on chromosome 2 is likely to be *ER*. This is supported by the observation that neither the presence nor absence of the Ler-0 haplotype at ER effects the haplotype phenotype patterns seen in the other ecotypes. For example, Hi-0 is the second highest haplotype phenotype in the first QTL and the highest once Ler-0 has been removed and Wil-2 is the lowest in both of these analysis. This is confirmed by ANOVA on the difference in computed SDs at the SNPs of the QTLs with Ler-0 haplotype included (ATC\_828) and excluded (MASC09221). ANOVA showed no significant effect of QTL on the computed SDs for each line (P=0.06) (Figure 3.3.8). This strongly suggests that this QTL is *ER* and that the haplotype of *ER* is important in controlling SD even once the non-functional Ler-0 haplotype has been removed.

|  |
| --- |
|  |
| Figure 3.3.8 The difference in computed SDs for the QTLs ATC\_828 and MASC09221 in ambient [CO2]. These are the QTLs on chromosome 2 with Ler-0 haplotype included (ATC\_828) and excluded (MASC09221). There is no significant difference between the QTLs (ANOVA, P=0.06) and the pattern of results is the same across the 2 QTLs suggesting that despite the distance between the SNPs, it is the same QTL. Dots represent outliers. |

Interestingly, *ER* does not appear to be as important in regulating SD at elevated [CO2] as there is no QTL at this position under 1000ppm [CO2]. It could be speculated that this is because the control of SD is achieved through a mechanism different to that which has been described by Engineer et al (2014), so not using the EPF2/ERECTA/YDA/MAPK cascade pathway. Alternatively, this pathway may play less of a role in these elevated [CO2] conditions.

The final QTL on chromosome 5 in 450ppm has the highest SD in Wu-0 and the lowest in Can-0. This is also true of the QTL on chromosome 5 in 450ppm with Ler-0 included and suggests this is the same QTL. Po-0 again shows a very wide range and this is likely to be for the same reason seen for QTL1. Po-0 haplotype occurred in a maximum of 13 lines for each analysis and for this reason Po-0 should again be disregarded.

### Gene Ontology analysis

The gene ontology of genes under a QTL can be used to identify candidate genes for the control of the trait studied. Some genes already have annotations while others can be placed into categories as to their predicted function. The genes underneath QTL1 and QTL5 were therefore analysed.

There are approximately 160 genes under the 600kB region around QTL1, all of which can be seen in the supplementary information in Table 1. None of the genes under QTL1 are currently known to be directly related to the control of stomatal development. Some are involved in auxin pathways such as *Auxin resistant 4* (*AXR4*) and this hormone has been shown to have an effect on SD (Tanaka et al, 2013). Other genes of note include *DREB* factors which relate to drought tolerance of which SD changes can play a factor (Lata and Prasad, 2011) and the master development gene *Grand Central* (*GCT*) which has control of developmental timing (Gillmor et al, 2010). The genes fall under a range of Gene Ontology (GO) terms which can be seen in Figure 3.3.9. There were no obvious candidate genes for chromosome 1 QTL

|  |  |  |
| --- | --- | --- |
| A) | B) | C) |
|  |  |  |
| Figure 3.3.9 Gene ontology analysis of 160 genes in 600kB region surrounding QTL1. The proportion of 160 genes in each category for biological process (A), Molecular Function (B) and Cellular component (C) are displayed. | | |

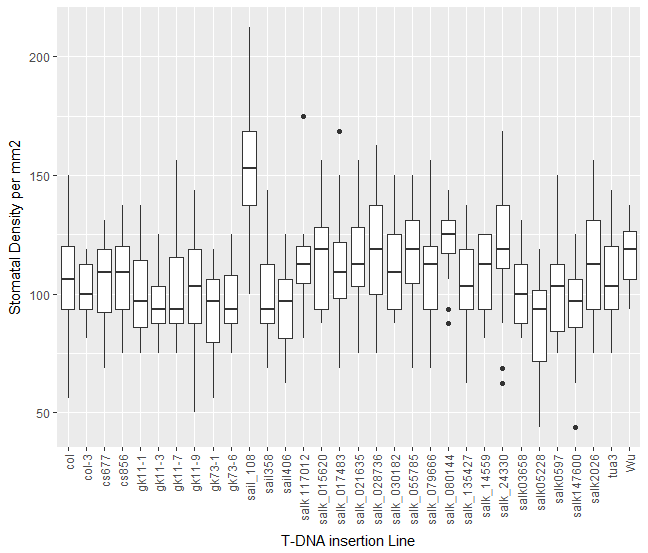
|  |  |  |
| --- | --- | --- |
| A) | B) | C) |
|  |  |  |
| Figure 3.3.10 Gene ontology analysis of 201 genes in 600kB region surrounding QTL5. The proportion of 201 genes in each category for biological process (A), Molecular Function (B) and Cellular component (C) are displayed. | | |

There are approximately 201 genes under the chromosome 5 QTL (See supplementary table 2). As with QTL1, no known genes involved in the control of stomatal development or SD are found in the surrounding 600kB of QTL5. There are many genes of interest however, including a MAPK (*MPK16*) which could be involved in the MAPK cascade which affects SD (see Section 1.5), tubulin genes involved in cell expansion (*TUA3*), a gene from the DREB subfamily related to drought (*AT5G18450*) and the auxin response gene *Bushy and Dwarf 2* (*BUD2*). Similar to QTL1, GO term analysis shows there to be a wide range of genes under QTL5.

As QTL5 is present in 450ppm and not 1000ppm [CO2] this suggests that the change could be related to why plants change their density in different [CO2] levels and it was chosen as the focus of further investigation.

### T-DNA Insertions

To find the gene responsible for the QTL on chromosome 5, a screen of T-DNA insertions in selected lines was carried out. The full list of candidate genes considered for phenotyping can be found in the supplementary information (Table 3) with the reason for each.



\*

Figure 3.3.11 Stomatal densities of T-DNA insertion lines of genes within the QTL region on chromosome 5. Box plots show T-DNA insertions and background lines grown in 450ppm [CO2] (N=16). \* indicated significant difference from its background line (See ) (P<0.05, Students T-test). Horizontal lines indicates median value and vertical lines the IQR. Dots indicate outliers of more than 1.5X the IQR outside the IQR

A screen of these selected T-DNA insertions revealed a line with a SD significantly different to its background wild type (Col-3) (T-test, P=0.001), with all others showing no significant difference (P>0.05). Not all of the candidate genes were phenotyped as phenotyping stopped when a significantly different lines was found. The T-DNA insertions that were phenotyped can be seen in Table 3.2‑3. The line, SAIL\_108\_D03C1, is a T-DNA insertion in the gene *ADG2*/*APL1*. *ADG2*/*APL1* is part of the rate limiting step in the starch biosynthesis pathway. It is the most highly expressed large subunit in the in the ADP-glucose pyrophoshorylase (AGP) complex (Crevillen et al, 2005).

To check that the T-DNA insertion line contained a T-DNA insertion in the expected locus, several SAIL\_108\_D03C1 plants were grown to maturity and 6 plants had leaf material taken for genotyping. These plants were genotyped for the presence of a T-DNA insertion in *ADG2* using the SAIL T-DNA left border and the ADG2F2 primer. To test for the WT locus, a PCR was performed using primers ADG2F2 and ADG2R2, which flank the expected T-DNA insertion site. By using both of these primer combinations it is possible to determine whether a T-DNA is present in the locus and whether it is a homozygous insertion line. The results (Figure 3.3.12) showed that all 6 plants had the sail T-DNA insertion, but some plants showed bands for the functional *ADG2* gene. It is possible that this was due to contamination during the PCR process, but it could also have been due to heterozygosity in the seed. Plants 2-4 did not show a band for the WT *ADG2* allele suggesting they were homozygous and so all future work studying *ADG2* was carried out on plants grown from the seed of these lines. This line was called *adg2-3* for the purpose of this thesis.

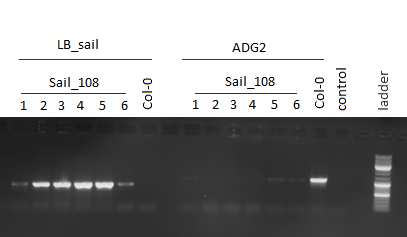


Figure 3.3.12 Gel checking the zygousity of SAIL\_108 plants. Gel shows the PCR products of 6 SAIL\_108 and Col-0 plants genotyped for presence of the sail T-DNA insertion (LB\_sail) and the WT allele (ADG2). Control lane contained no genomic template with ADG2 primers. Ladder shows size.

## Discussion and Conclusions

### Polygenetic Control of Stomatal Density

The spread of SD seen in the MAGIC lines compared to the parental lines is strongly suggestive of polygenetic control. Many quantitative traits are under polygenic control, including many metabolic compounds (Rowe et al, 2008) and root length (Lachowiec et al, 2015). It is not surprising that a complex trait that takes into account many environmental factors similar to the above traits would be found to also be polygenic. The polygenetic nature of the SD can be seen from the larger spread in phenotypes than the parental lines.

The range of SD phenotypes seen in the MAGIC lines is larger than that of the MAGIC parental lines, but it is similar to that found by Delgado et al (2011) in a natural population of 62 ecotypes. In control of SD, a plant must maintain the potential for maximum gas exchange capability while also minimising water loss. MAGIC lines contain phenotypic diversity outside of that seen in the parental lines because they do not contain adaption to a specific environment constraining the combination of alleles seen. With the 19 parental lines contributing to the genetic diversity, it provides a wider range of phenotypes that would be expected to be seen from 2 parental RILs.

The control of SD response to elevated CO2, also appeared to be polygenic. The range of responses in the MAGIC lines was larger than that seen in the parental lines. Similar to SD, this suggests that SD response to [CO2] is a polygenetically controlled trait. Most of the lines appeared to show an increase or no change in SD, reflecting the responses of the parental lines. There was limited differences in the overall mean SD between treatments however. GXE interactions can often be seen by a shift in phenotype between treatments (Ed-Soda et al, 2014) however as lines can change in opposite directions depending on genotype, the mean SD overall can hide GXE interactions.

The wide spread of data and the polygenic nature of the trait supported the use of the MAGIC lines in QTL analysis on both SD and the change in SD between [CO2] treatments and so a QTL analysis was carried out.

### Quantitative Trait Loci

In this analysis, QTL and GWAS analysis were run to find QTLs for SD, however GWAS showed nothing of significance and was therefore not included in this thesis.

The QTL analysis suggests that a total of 3 QTLs controlling for SD have been identified across QTL analysis, on chromosomes 1, 2 and 5. The evidence suggests that the QTL on chromosome 1 (QTL1) is the same QTL across [CO2] treatments and including Ler-0 haplotype. The QTL on chromosome 2 (QTL2) is very probably *ER* and has no significant effect on density in 1000ppm except in Ler-0. Finally, the QTL on chromosome 5 (QTL5) is present only at 450ppm [CO2]. In terms of areas of focus to find a gene of interest that could be responsible for the differences in SD between [CO2] treatments, it would suggest that *ER* (QTL2) and QTL5 are of the largest importance.

The QTLs found in the QTL analysis on SD in 450ppm were found to control approximately 32% of the trait in the MAGIC lines. This leaves approximately 68% of SD variance unaccounted for. The QTL model used here is only able to detect genes with an effect of 10% or more consistently with 450 lines (Kover et al, 2009), and this ability may be reduced given the smaller number of lines included (Max lines included = 393). It is therefore reasonable to assume that this analysis was unable to identify several genes that have a smaller effect size due to the experimental setup. The stomatal lineage pathway contains many genes (see section 1.5) and there are many more genes unmentioned that allow for the integration of signals from the environment (Reviewed in Casson and Gray, 2008). In addition to that, the genes that control cell size are relatively unknown and SD is affected by this as much as by stomatal initiation. There is therefore likely to be many more QTLs undetected.

The lack of any QTLs for SD response to [CO2] is disappointing but supports the findings of a previous study (Ferris et al, 2002). Ferris et al (2002) also found more QTLs for stomatal traits in ambient [CO2] than elevated [CO2] and did not show any QTLs for the control of the SD difference between [CO2] treatments. The Ferris et al (2002) data, and the result from this analysis suggests that QTLs for response are of small affect and so are hard to detect by means of QTL.

Both the Ferris et al (2002) study and this study looked at plants grown in ambient and elevated [CO2], although this study used 1000ppm while Ferris et al used 600ppm, which strongly supports that there is no major QTL for [CO2] increases above 450ppm. However neither study the response between sub-ambient and ambient [CO2]. Chapter 2.4.3 showed that there are differences in SD responses to increasing [CO2] in some lines and that there could be different sets of genes responsible for SD changes between sub-ambient and ambient [CO2] compared to ambient and elevated [CO2].

This study looked at the change between 450ppm and 1000ppm [CO2] as it aimed to look at genes and responses that can be expected from a future high [CO2] world however, it may be that the SD response to [CO2] is saturated at 450ppm. Plants have not been subject to [CO2] outside of 180-400ppm for the past 20,000 years before the industrial revolution (Augustin et al, 2004). While plants have need to maintain plasticity for the reasons explained in Chapter 1.3.4, they may not have capacity outside of a certain range which may already be exhausted at 450ppm. Studying the response between 180ppm and 450ppm [CO2] may therefore help to explain previous changes in SD, although it would not necessarily help us to understand future changes.

The 3 QTLs identified here do not have any stomatal development genes under them other than *ER*. This would suggest that in natural populations of plants, the genes that are varied to achieve differences in SD by adaption are not the core genes of stomatal development (See chapter 1.5), except *ER*. This does not mean that these genes may not exert their control through this core pathway. As only 19 ecotypes are used to make the MAGIC lines, there may be other ecotypes which vary stomatal development genes to achieve differences in SD but their genetics are not included. It could also be, however, that the stomatal development genes are buffered by feedback loops which prevent changes being effective. SPCH for example positively regulates EPF2 expression which in turn negatively regulates stomatal initiation and SPCH expression (Horst et al, 2015). Despite this, a GWAS of natural ecotypes would allow for a look at the genes of adaption in natural populations.

While here a G X E analysis was carried out by running a QTL on the difference between the two [CO2] treatments, it is also possible to run other kinds of G X E QTLs. The environmental interaction can be taken into account when running QTL models or separate models can be run after the initial QTL model has been run (Reviewed in El Soda et al, 2014). In the QTL analysis run in this thesis, there are differences in discovered QTLs in the different environments which would suggest there is a G X E effect to be studied. However, the overall difference in SD mean between the treatments is minimal suggesting that effects of [CO2] on genes are small, although it is also possible that the contrasting SD responses within the parental lines result in a small overall mean change as some reduce SD while some increase SD. The small overall effect [CO2] had on the mean SD of this population makes it difficult to study. A better method of studying differences, due to the small effect size, may be to choose very differently responding lines and create RILs of these for QTL analysis. However, this would also preclude the minor alleles found in the MAGIC lines, giving both advantages and disadvantages.

It is interesting to note the poor germination of seeds with Po-0 haplotype at QTL1 and QTL5 leading to few Po-0 haplotypes being included in the analysis at these loci. A single QTL with poor germination of a haplotype at that loci would be suggestive of a marker for germination in that line, however two QTLs seems less likely although still possible. Germination rate was not measured in this analysis however it would be possible to do a binary QTL analysis on germination as was done by Kover et al (2009) on the GLABROUS and ERECTA phenotypes. Further research into germination could therefore be examined to confirm if the poor germination was due to loci near these QTLs.

### Evaluating the QTL analysis

The QTL analysis here suffered from variation in SD introduced by the batch growing system. This was a consequence of limited growth space and ideally all plants would have been grown at the same time to ensure there was no batch effect. This could have been countered by growing lines in multiple batches to get an average across batches as well, however this would have required more cabinet space which was not available. A system such as that shown by Humplik et al (2015) or Granier et al (2006) which moves plants within their cabinet in order to ensure plants are all experiencing similar conditions would be ideal for QTL analysis but would still require growth in one batch. The method chosen here was therefore the best method with the resources available.

Further to this, this QTL analysis suffered from inclusion of lower than ideal numbers of MAGIC lines. Kover et al recommend the use of 450 lines. The seed in use for this analysis however showed poor germination in many lines, possibly due to its age. The poor germination meant that despite some 700 lines being sown, only around 400 were harvestable, and in addition, some were only harvestable in one treatment. Further phenotyping of the MAGIC lines would be required to more effectively detect genes of small effect. It would be particularly useful to phenotype more lines containing Po-0 as it is under-represented at 2 of the QTLs found here and may be underrepresented overall.

Despite this, 3 QTLs were discovered with strong genome wide significance and 2 differed between treatments suggesting that the QTL analysis was successful despite its limitations.

### Identification of QTL2 chromosome

The QTL on chromosome 2 is almost certainly *ER*. This can be seen from the similar chromosomal location and the SD pattern of the ecotypes in the haplotype analyses which are similar with and without Ler-0, except that Ler-0 no longer has a high SD. *ER* has long been known to affect SD, especially in Ler-0 (Torii et al, 1996) and it is not surprising that a haplotype that has such a strong effect on SD is found in this QTL analysis. It is however interesting to see what even with Ler-0 removed, *ER* is still identified as a regulator of SD, with Hi-0 haplotypes showing a higher SD and Wil-2 a lower SD, and that this QTL is not identified in elevated [CO2].

As the QTL for *ER* is not identified in elevated [CO2], it would suggest that the control of SD by *ER* is less or eliminated in elevated [CO2]. The [CO2] response pathway identified by Engineer et al is believed to work through the ER and MAPK cascade pathway and it is interesting that the variation in *ER* does not have an effect in elevated [CO2] given the work of Engineer et al. This could be because the response to [CO2] does not occur through ER, there are other ER like receptors which could be involved, or it could be that there are other pathways involved and the CRSP pathway is one of many small factors. The work identifying CRSP also took place between 150ppm and 500ppm [CO2] and it is possible that the 450ppm to 1000ppm [CO2] changes in SD are caused by a different pathway. How *ER* could be involved in regulating SD in the ecotypes is the subject of chapter 5.

### Identification of QTL5 chromosome

The QTL on chromosome 5 is present both with and without the Ler-0 *ER* haplotype, however it is only a significant QTL on a genome wide scale without Ler-0 haplotype at *ER* included. As QTL5 was the QTL which explained the highest percentage of variance (13%) in SD in ecotypes and it was significant in only one [CO2] treatment, it was chosen for further analysis.

There were 201 genes under the 600kb region identified by the QTL analysis. While fine mapping is the traditional approach, it is time consuming and costly. Fine mapping requires several months of producing lines between the most distinct ecotypes, as determined by the haplotype analysis. The small size of the genomic region identified by MAGIC QTL analysis allows for other approaches.

Phenotyping 450 lines in ambient and elevated [CO2] took approximately 14 months. There are approximately 200 genes in 600kB giving an approximate time to phenotype 200 lines of 6 months. It is possible to order from the Arabidopsis stock centre T-DNA insertions in the majority of genes in the Arabidopsis genome (Li et al, 2006). The decision was therefore taken to try to identify the gene causing the QTL on chromosome 5 by selecting likely candidates. Many of these T-DNA lines result in predicted loss-of-function (LOF) alleles of the target gene and so may have more distinct phenotypes than observed in the QTL analysis (unless similar LOF alleles exist in the MAGIC population).

There are many genes identified as being significant in the control of SD and response to [CO2] (See chapter 1.6), however none of these fall 300kb up or downstream of the QTL. There are, however, several datasets that can be consulted to find likely candidates for further research instead. The datasets include genomic data and transcriptional data, which inform of which genes are expressed between ecotypes (Gan et al, 2011), between stomatal development stages (Adrian et al, 2015), and between [CO2] treatments (Watson-Lazowski et al, 2016). These datasets (Gan et al 2011; Adrian et al, 2015; Watson-Lazowski et al, 2016), which are discussed in detail in Section 1.1.1, therefore formed the basis for the candidate approach that was taken.

This candidate gene approach using T-DNA knock-outs identified the gene ADG2 as a regulator of SD. Gel electrophoresis of genomic DNA confirmed that the SAIL line had a T-DNA insertion where expected and homozygous lines were identified.

While not all 200 genes have been phenotyped, which could mean that there remain other genes which could be the cause of the QTL, it had a large enough effect that it warranted further investigation as a controller of SD. It is therefore the subject of chapter 4.

### Conclusions

While this QTL analysis did not identify any genes as controllers of the difference in SD in different [CO2] levels, it did identify 3 QTLs which control SD, with only 1 being significant in elevated [CO2]. The QTLs did not explain a significant portion of variation (32%) and it is therefore likely that the control of SD is regulated by a number of genes rather than a few master regulators. Of the 3 QTLs found, candidate genes for 2 have been identified and are discussed further in chapters 4 and 5.

# ADG2, Starch Biosynthesis and Stomatal Density

### Starch, its synthesis and break down

Starch is the primary form of carbohydrate storage in plant leaves, synthesised in the daytime by chloroplasts and amyloplasts utilising sugars produced during photosynthesis and degraded at night when photosynthesis is subdued. Up to half of all photo-assimilated carbon is stored as starch (Zeeman et al, 2004), which accumulates in all plant cells with the exception of most cells of the root (Tsai et al, 2009)

Starch is synthesised by a combination of starch synthases and starch branching enzymes, which act together to make the storage molecule of a starch granule. The first step during synthesis is to fix sugars from glucose-1-phosphate into ADP-glucose by the AGP complex (See section 4.1.2). Following this, starch synthases (SS), branching enzymes (BE) and debranching enzymes such as isoamylase (ISA; Figure 3.4.1) assemble the starch molecule. The final outcome is a starch granule which has a specific shape and structure (Smith, 2001).

In its whole form, the granular structure of starch means only some enzymes are able to access it for break down (Zeeman et al, 2004). The inaccessibility of the unbroken down starch granule means that there must be a complex process through a series of enzymes to break it down and reduce it to sugars for utilisation. Starch is broken down primarily by hydrolases (such as amylases) phosphorylases (such as PHO1) and glucanotransferases (such as DPE1). These interact to break up a starch granule into sugars (Zeeman et al, 2004). The first step is carried out by amylases, usually α-amylase, with further break down following by the other enzymes, often outside of the chloroplast (Beck and Zeiger, 1989).

Mutants with altered starch biosynthesis and break down have been shown to have altered growth and fecundity (Zeeman et al, 2010; Andriotis et al, 2012), indicating that starch can impact development. In addition, plants with altered starch break down due to improper granule formation show altered growth and fecundity (Feike et al, 2016). Proper starch biosynthesis and breakdown is therefore essential to the growing plant, and any interruption can have serious implications for normal growth and development.

### The AGP complex

The ADP-glucose pyrophosphoryalse (AGP) complex fixes glucose-1-phosphate into ADP-glucose ready for starch synthesis by adding a glycosyl group. It is the first and rate limiting step in starch biosynthesis (Stark et al, 1992). The full starch biosynthesis pathway can be seen in Figure 3.4.1.

The AGP complex is made up of a large and small subunit. The small subunit of AGP is composed of two units of ADP-GLUCOSE PYROPHOSPHORYLASE SMALL SUBUNIT 1 (APS1). There is a secondary APS gene (*APS2*) (Crevillén et al, 2003), however due to mutations in its active site, it is believed to be inactive. The large subunit of AGP has 4 isoforms and analysis of their tissue expression patterns reveals that they are expressed differentially throughout the plant (Crevellin et al, 2004). The most abundant large subunit isoform of the AGP complex expressed in the leaves is ADG2/APL1 (Crevellin et al, 2004). ADG2 was previously identified as influencing stomatal density (SD) in chapter 3. By way of contrast, in fruits and roots the dominant isoform is APL3 and in stomata, APL4 is the most abundant isoform(Horrer, 2004).

|  |
| --- |
| https://static-content.springer.com/image/art%3A10.1007%2Fs11103-015-0293-2/MediaObjects/11103_2015_293_Fig1_HTML.gif |
| Figure 3.4.1The starch biosynthesis pathway and the genes associated with each step. ADG2/APL1 is highlighted with a black circle. Adapted from Schwarte et al, 2015. |

The large subunit of AGP is primarily involved in the regulating the activity of the complex however, it may also been demonstrate catalytic activity if the small subunit (APS1) has been knocked out (Ventriglia et al, 2008). *ADG2* is up regulated in the presence of sugars to facilitate the first step of starch biosynthesis (Sokolov et al, 1998), and knock-outs of *ADG2* have been shown to have a reduction in starch content of 40% compared with wildtype plants (Lin et al, 1988 *ADG2* is a plays a key role in starch biosynthesis, its position in the biosynthesis pathway can be seen in Figure 3.4.1. *ADG2* has not previously been implicated in the control of SD, only the regulation of ADP-glucose biosynthesis.

### Regulation of ADG2 and starch

Plants produce starch in a circadian manner, with production in daylight and utilisation of their stored starch at night (Figure 3.4.2; Lu et al, 2005; Smith et al, 2004). The expression of starch biosynthesis genes, including *ADG2*, also follow circadian rhythms (Smith et al, 2004). However, these transcriptional changes do not necessarily result in a change in enzyme levels suggesting that post-translational modification is important in regulating enzyme activity and levels (Lu et al, 2005; Smith et al, 2004).

*ADG2* expression can be seen in B as *APL1* alongside other *APL* genes. *ADG2* is the most highly expressed of the APL genes; expression is highest in the middle of the day and lowest at night (Smith et al, 2004). The change in expression between light and dark periods is not large, with an increase of around 50% and there is no difference in expression levels between dawn and dusk. Currently the relationship between ADG2 gene expression and protein abundance is not well understood and there is no indication of how consistent expression and enzyme levels are.

|  |
| --- |
| A)  B)) |
| Figure 3.4.2 Circadian regulation of starch and APL genes. The top bar indicates night (black) and day (white). A) Starch content of leaves across a 24h photoperiod expressed as mg g-1 fresh weight. Starch was extracted according to the methods of Critchley et al. (2001). 1 measurement was taken from5 plants. Error bars show standard error over 10%. B) Expression patterns of AGP complex genes. Expression data from single microarray for each time point with each array containing RNA from 3 leaves from 8 plants Adapted from Smith et al, 2004 |

*ADG2* expression is regulated by photosynthetic products and cytosol metabolism as well as circadian rhythms (Hendriks et al, 2003). Specifically, sucrose is an activator of *ADG2* expression and activity (Hendriks et al, 2003) while elevated inorganic phosphates (Pi) and elevated 3-phosphoglyceraldehyde (3-PGA), which are in turn regulated by triose phosphate levels, are a suppressor of ADG2 activity and expression (Mugford et al, 2014). Pi amd 3-PGA inhibit starch synthesis as free Pi and 3-PGA suggests low carbon export from the plastid because Pi and 3-PGA would ordinarily be used in photosynthetic activity. High Pi and 3-PGA, on the other hand, suggests a low carbon source to sink ratio, or that carbon is being used faster than it is being made (Ciereszko and Kleczkowski, 2003). High sucrose by contrast indicates more free sugar from photosynthetic activity than is being used leading to an increase in ADG2 level and AGPase activity and starch accumulation (Hendriks et al, 2003).

ADG2 is therefore regulated by a range of factors and expression and activity cannot be considered independent of time of day.

### Starch and regulation of stomatal function

BAM1 is the dominant β-amylase in guard cells and it breaks down starch into maltose (Valerio et al, 2011). AMY3 is a dominant α-AMYLASE of guard cells and breaks starch down into sugars (Thalmann et al, 2016). Their position in the starch biosynthesis and break down pathway can be seen in Figure 3.4.1. Previous experiments have shown that when starch degradation is impaired, through the knocking out of *β-amylase 1* (*BAM1*) and α-AMYLASE 3 (*AMY3*), stomatal opening is impaired.

When *BAM1* and *AMY3* are both absent, starch is not degraded leading to a build-up of starch and a reduction in the amount of sugar osmolytes available for stomatal opening (Horrer et al, 2016). In addition, starch concentration was shown to be higher in the guard cells of plants with closed stomata compared with those which had their stomata induced to open by light or low CO2, suggesting the breakdown of starch is required for stomatal opening. This is believed to be because sugar osmolites are required for opening (Outlaw and Manchester, 1979).

Starch is essential for proper stomatal function (Outlaw and Manchester, 1979; Horrer et al, 2016; Valerio et al, 2011) however, no clear link between total guard cell starch content and stomatal function has been established. Additionally, the relationship between leaf starch content and SD, or between *ADG2* and other genes of the AGP complex and SD still remains unclear.

Adrian et al have shown that *ADG2* is a key stomatal entry priority gene (Adrian et al, 2015). Upon stomatal initiation by expression of the *SPCH* gene, *ADG2* is down regulated in guard cells and this is reflected in the changes in expression upon stomatal initiation. This therefore suggests that changes to ADG2 are of importance during entry into the stomatal lineage.

### Aims

The aims of this chapter were therefore:

1. To determine the SD and stomatal index of the *adg2-3* mutant.
2. To relate the *adg2* SD data to the QTL model results in order to determine whether it is the QTL on chromosome 5.
3. To identify the mechanism by which the knockout of *ADG2* affects SD.

## Methods

### Plant Material

Seed for T-DNA insertions and *Col-3* and *Col-2* background seed was ordered from the Nottingham Arabidopsis Stock Centre (NASC). Seed for *fba*2-2 was donated by Peter Venn (The University of Sheffield). Full list of seed can be seen in Table 4.2‑1. ADG2 seed was from confirmed homozygous plants identified in Chapter 3. Parental ecotype seed is stock from that donated by Dr Paula Kover (Bath University).

|  |  |  |  |
| --- | --- | --- | --- |
| Table 4.2‑1 germplasm used for experiments and their respective interrupted genes and background lines | | | |
| Name | Gene interrupted | Gene name | Background |
| *starch excess* | At1g10760 | *SEX1* | Col-0 |
| *pgm-1* | At5g51820 | *PGM* | Col-2 |
| *fba2-2* | AT4G38970 | *FBA2* | Col-0 |
| *atsweet11;12* | AT3G48740 , AT5G23660 | *ATSWEET 11, ATSWEET12* | Col-0 |
| SALK\_208532C | AT5G19220.1 | *ADG2* | Col-0 |
| SALK\_139228C | AT1G27680 | *APL2* | Col-0 |
| SALK\_203188 | AT4G39210 | *APL3* | Col-0 |
| SALK\_108632C | AT2G21590 | *APL4* | Col-0 |
| SALK\_059083C | AT5G48300 | *APS1* | Col-0 |
| SAIL\_108\_D03C1 (*adg2-3*) | AT5G19220 | *ADG2* | Col-3 |
| Col-0 | - | - | - |
| Col-3 | - | - | - |
| Col-2 | - | - | - |
| Wu-0 | - | - | - |
| Hi-0 | - | - | - |
| Can-0 | - | - | - |
| Wil-2 | - | - | - |

### Plant Growth

Plants were grown in the same manner as chapter 2.2.2, except experiment 4.3.9 in which the following lines were grown in a Sanyo Fititron SGC970H/C/R in the same conditions as chapter 2.2.2 with 10hrs of 200 μmol m−2 s−1 light, 70% humidity, and 450ppm [CO2] :

* *adg2-3*
* Col3
* Col0
* Col2
* *pgm-1*
* *atsweet11;12*
* *fba2-2*
* *starch excess*

In addition, plants grown for starch staining of leaves (1.3.11) were grown for 21 days instead of 18.

### DNA extractions

For genotyping, genomic DNA was extracted from leaf or seedlings using the Edwards prep method (Edwards et al. 1991). A leaf disc, of approximately 1cm in diameter, or several 7-14-day old seedlings were ground in 400 μl Edward’s Solution (200mM Tris-HCl (pH7.5), 250mM NaCl, 25mM EDTA, 0.5% SDS) using a vortexer and ball bearing and centrifuged at 13,500rpm for 5-10mins. The supernatant was transferred to a new 1.5ml Eppendorf and 400μl isopropanol (Fisher Scientific Laboratory Grade Propan-2-ol, 1067432) was added. The sample was mixed and centrifuged at 13,500rpm for 5-10mins to pellet the DNA. The supernatant was aspirated using a pipette without disturbing the pellet and the pellet was air dried for 5mins. The pellet was then reconstituted in 100μl sterile H2O and vortexed to mix. DNA was then stored at -20°C.

### Polymerase Chain Reaction (PCR)

The method for PCR was adapted from Sigma Aldrich JumpStart™ RedTaq® ReadyMix™ PCR Reaction protocol. PCR components, including RNA free, sterile H2O, forward and reverse primers (100μM stocks), DNA template and RedTaq (Sigma-Aldrich JumpStart™ RedTaq® ReadyMix™ PCR Reaction Mix with MgCl2) were first melted at room temperature. The forward and reverse primers for the genes and exons of interest are shown in .

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | Table 4.2‑2 Primers and their direction, target and use | | | | |
| Gene | Exon if specific target | | Primer Name | Direction | Type | Sequence |
| *ADG2* | - | | ADG2F1 | F | qPCR | AAGGGATCCAAGGACAGTTGC |
| *ADG2* | - | | ADG2R1 | R | qPCR | GATAGGAACGGCAGGCTTGG |
| *ADG2* | 7 | | ADG2 R2 | R | PCR | ACTCCCATTGAAGCTATGTATGGT |
| *ADG2* | 8 | | ADG2 F2 | F | PCR | GTGCCTCAGATTTTGGGCTT |
| *ADG2* | - | | ADG2 F3 | F | Sequencing | tgcaagacaggtcaagacaa |
| *ADG2* | - | | ADG2F4 | F | Sequencing/PCR | attgcagGTTCCTATCGGGG |
| *ADG2* | - | | ADG2F5 | F | Sequencing | AGGAGACGACCTGAAAGCAA |
| *ADG2* | - | | ADG2F6 | F | Sequencing | AGgtaagtaaaaccagccatga |
| *ADG2* | - | | ADG2R3 | R | Sequencing | tccagacaccacagcctaatg |
| *ADG2* | - | | ADG2R4 | R | Sequencing | CAGTTTGGCTCGTAATGGATG |
| *ADG2* | - | | ADG2R5 | R | Sequencing/PCR | TCTGAACCAAAGTCGTTTGC |
| *UBC21* |  | | UBC21F | F | qPCR | GAATGCTTGGAGTCCTGCTTG |
| *UBC21* |  | | UBC21R | R | qPCR | CTCAGGATGAGCCATCAATGC |

2X master mix was prepared according to the number of samples. 22.5μl of master mix was pipetted in to each PCR tube (0.2ml). 2.5μl of template DNA was in to each tube, mixed and briefly centrifuged to spin down the contents and eliminate air bubbles.

For example, a representative mix for one sample would contain:

12.5 µl RedTaq 2x mix

2.5 µl gDNA

0.5 µl For primer (10pmole/ul)

0.5 µl Rev primer (10pmol/ul)

9 µl H2O.

Samples were placed in a thermal cycler following the program shown in .3. Reaction volume was set at 25μl and run for 35 cycles.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4.2.3 The thermal profile, incubation temperature and times periods for PCR. | | | | | | |  |
|  | Step 1 | Step 2 | Step 3 | Step 4 | Step 5 | Step 6 | Step 7 |
| Temperature(°C) | 95 | 95 | 55 | 72 | 35 repeats of steps 2-4 | 72 | 12 |
| Time (minutes) | 3 | 0.5 | 0.5 | 1 | 5 minutes | end |

### Agarose Gel Electrophoresis

DNA separation via agarose gel electrophoresis was used to visualise the products of PCR and to verify the quality of RNA. Gels were either 2% agarose if PCR product was below 200bp or 1% agarose if above 200bp. Actin was used a positive control to verify the presence of gDNA. Agarose powder (Sigma Agarose Gel powder) was mixed with 1 X Tris-acetate-EDTA (TAE )buffer and dissolved by boiling. The fluorescent indicator Ethidium Bromide (stock 10mg/ml) was added to the liquid gel to a final concentration of 1μg/ml. The gel was then poured into a gel tray fitted with a comb.

The gel was placed in a tank filled with 1X TAE buffer and 10µl of samples loaded in to each well. A DNA ladder (2.5μl GeneRuler DNA LadderMix; Fisher Scientific 10181070) was loaded alongside to allow for fragment size comparisons. Gels were run at 100-120V for 20-40 minutes using a BioRad mini sub-cell and power supply. Gels were visualised using GelDoc-It™ system (UVP LLC) and images were taken using VisionWorks® LS analysis software (UVP LLC).

### RNA Extraction

RNA extractions for RT-qPCR were taken in the following manner. Samples were taken from 14 day old seedlings, with 3 samples of approximately 15 seedlings for each line used. Extractions were performed according to the manufacturer’s instruction of the Quick-RNA™ MiniPrep (Zymo Research, Cambridge Biosciences R1055a) with the additional inclusion of an on column DNase treatment. RNA concentrations were measured at 595nm using the ‘Nucleic Acid’, ‘RNA-40’ setting on the NANODROP-8000 Spectrophotometer V1.1 (ThermoScientific). RNase-free H2O was used as a blank and 1μl extracts was loaded on to the reading pin.

### cDNA Synthesis

cDNA synthesis was performed according to the protocol from Applied Biosystems High Capacity cDNA Reverse Transcription Kit. 500ng-2μg of total RNA per 20μl was used per reaction and an RNase inhibitor (RiboLock, Fisher Scientific 10859710) was included. The volumes of each component for the master mix were calculated according to the number of reactions and controls. Additional reactions were added to the master mix to account for any potential discrepancies that could occur during reagent transfers between tubes. Master mix contents per reaction can be seen in Table 4.2‑3.The 2X RT master mix was placed on ice and mixed gently. When master mux was ready, 10μL of 2X RT master mix was pipetted into each tube.

Table 4.2‑3 components of cDNA synthesis master mix

|  |  |  |
| --- | --- | --- |
| Component | Volume per reaction | Example for 6 |
| 10XRT buffer | 2ul | 13ul |
| 25xdNTP mix | 0.8ul | 5.2ul |
| 10x random primers | 2ul | 13ul |
| Reverse transcriptase | 1ul | 6.5ul |
| RNase inhibitor | 0.5ul | 3.25ul |
| RNAse free water | 3.7ul | 24.05ul |

Concentrations of RNA were equalised based on nano-drop concentrations by adding calculated amounts of RNase-free water and RNA into the tube so they equalled 500-2000ng RNA, depending on the lowest yielding sample. 10μL (usually 2μg of RNA) of RNA sample was pipetted in each tube. Tubes were briefly centrifuged to eliminate any air bubbles or material on the sides. To perform Reverse Transcription reactions, the thermal cycler was programmed with the thermal profile shown in .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 4.2‑4 cDNA synthesis thermocycler cycle | | | | |
|  | Step 1 | Step 2 | Step 3 | Step 4 |
| Temperature(°C) | 25 | 37 | 85 | 4 |
| Time (minutes) | 10 | 120 | 5 | end |

Following the end of the program, 10µl of the cDNA was diluted 20-50X in preparation for qPCR. The remaining cDNA stock was stored at -20°C in case of further need.

### qPCR Analysis

qPCR was performed according to the manufacturer’s instructions (SYBR® Green JumpStart™ Taq ReadyMix™ Sigma Aldrich S5193). Components were combined to make a master mix and forward and reverse primers to target the gene of interest were added (Table 4.2‑5). The primer mix concentration was 7.5pmol/µl. Housekeeping gene *UBC21* was used as a reference gene in all analysis as *UBC21* is stable across treatments (Czechowski et al. 2005). All qPCR primers and their gene or part of gene can be seen in .

Table 4.2‑5 Components of qPCR master mix

|  |  |  |
| --- | --- | --- |
| Component | Volume per reaction | Example for 12 |
| MgCl2 | 2.8ul | 36.4ul |
| dye | 0.2ul | 2.6ul |
| sterile water | 1ul | 13ul |
| SYBR green 2x mix | 10ul | 130ul |
| primer stock | 1ul | 13ul |

The 2-ΔΔCT method, as described by Livak and Schmittgen (2001) was used to determine relative expression levels. This method requires that one sample be set as a calibrator sample to which all other samples are compared. It also requires that a reference gene is used, in this case UBC21, the expression of which does not change across samples. To find relative expressions levels, raw Ct levels were first established by removing the reference gene ROX values. Disassociation curves were then viewed to check for anomalies. Finally relative fold change in gene expression was found using excel software using the layout in Table 4.2‑6

Table 4.2‑6 Calculating fold change in expression. The table shows the layout of a comparison of a control and test sample with 2^-ΔΔCt showing the final result

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sample | Sample Name | Ct (Gene of interest) | CT (Reference gene) | ΔCt | average Δ | ΔΔCt | 2^-ΔΔCt |
| Test | 1 | 2 | 1 | 1 | 1.33 | -3.67 | 12.72858 |
|  | 2 | 3 | 1 | 2 |  |  |  |
|  | 3 | 2 | 1 | 1 |  |  |  |
| Control | 1 | 6 | 1 | 5 | 5 |  |  |
|  | 2 | 6 | 1 | 5 |  |  |  |
|  | 3 | 6 | 1 | 5 |  |  |  |

The Ct value refers to the cycle in which florescence rose above the background threshold florescence level. The amount of target nucleic acid within the sample is proportional to the Ct values. Delta (Δ) is the sample Ct value minus the control Ct value (UBC21 in this case). Average delta (average Δ) is the mean value between biological and technical replicates. Delta delta (ΔΔ) normalises expression between a test sample and the chosen calibrator sample. Test values for each biological rep are then inverted to produce the opposite sign, either positive or negative. 2Ln (natural log) is then used to establish gene expression of the test sample relative to the control; no change in relative expression to the control is shown as 1.

### Sequencing

Sequencing was carried out by The University of Sheffield Core Genomics Facility. The primers used for sequencing can be seen in table Table 4.2‑2.

### Stomatal Impressions

Stomatal impressions were taken in the same manner as chapter 2.2.3

### Starch Assay – Staining

To quantify the starch content of leaves, two methods were used. In the first method, starch content was quantified by staining with Lugol’s solution, which turns from brown to blue upon contact with starch. Plants were cleared in 80% ethanol in a 90°C water bath for 20-40 minutes. Ethanol was changed every 10 minutes. Once cleared, plants were stained using a pasteur pipette to drop 1M Lugols Solution (8g potassium iodide, 2.54g iodine and 100ml water). The stain was left for two minutes then plants were rinsed in water for two minutes to remove excess stain. Plants were placed on a glass slide and then on a light box and photos taken using Panasonic Lumix DMC-FZ200.

Quantification of staining was carried out in ImageJ software (1.51K). Pictures were loaded and image colour was quantified using mean grey value with freehand selection used to draw around the leaf area. The darker the leaf, the more starch is present and the lower the mean grey value for the leaf is. For quantification, 5-10 leaves were stained.

### Starch Assay – Enzyme Quantification

The second method for quantification of starch involves the enzymatic hydrolysis of starch to glucose, which is then quantified by spectrophotometry. An adapted version of the method described by Bellasio et al (2014) was used for this experiment and is described below.

To quantify starch, four samples of each treatment were used. 100-400mg plant material was ground in liquid nitrogen in a Qiagen TissueLyser ball mill. Plant material was then weighed into screw cap lid Eppendorfs and frozen until ready for assay.

The plant material was first mixed with 1.5ml of 80% ethanol and heated at 100°C for 3 minutes and then spun down in a micro-centrifuge for 5 minutes at 12000RPM. The supernatant was removed. The pellet was then resuspended in 1.5ml 80% ethanol and the heating step was repeated to ensure complete removal of free sugars. Samples were then spun down in a micro-centrifuge again and the supernatant discarded.

The pellet was then dried and then resuspended in 1.5ml H20 using a vortex. Following this, 0.5ml of the sample/H2O mix was taken using a 1ml syringe and placed into a seperate screw cap Eppendorf and labelled ‘test sample’. Another 0.5ml mix was added to a second Eppendorf labelled ‘control sample’. To the test sample, 0.5ml 200mM NA acetate, 160 µl of Pure amyloglucosidase from *Aspergillus niger* in 50% glycerol and 0.02% sodium azide (E-AMGDF, Megazyme international, IR), and 5µl of amylase (E-BLAAM, Megazyme international, IR) was added to carry out the starch digestion. To the control tube, 665µl of 200mM NA acetate was added. Tubes were then incubated at 37°C for 2 hours. Tubes were mixed via inversion every half an hour. Samples were then centrifuged at 12000RPG for 5 minutes to pellet material.

A glucose assay was then carried out by using the same method as described by Bellasio et al. (2014) however 10ml tubes were used in the place of 5ml tubes. First, three glucose standards and a standard blank were made by adding 20 µl of glucose standard solution (Glucose solution 1 µg µl−1 in benzoic acid 0.1%) and 580 µl of water to 3 tubes and 600 µl water to one tube. For each sample and control sample incubated above, 540 µl of water was added and 60 µl of sample to each tube.

To each of these tubes (samples, controls, glucose standards and standard blank), 2 mls of the following solution was added: 500 U of glucose oxidase (Aspergillus niger) and 100 purpogallin units of peroxidase (horseradish) in 0.1 M phosphate buffer pH 7.0 plus o-dianisidine (See Bellasio et al, 2014 for details of how to make this solution). The tubes were incubated at 37°C for 45 minutes in a water bath. Reaction was stopped by adding 400 ml of H2SO4 75% to each tube in the same order the reaction was started in the tubes. Absorbance was then read at 530 nm using a spectrophotometer (CECIL CE1020). Glucose concentration was found using the equation found in Bellasio et al, 2014 and the spreadsheet layout seen in .

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4.2‑7 Example of spreadsheet for calculating starch content with example | | | | | | | | | |
| Sample | Final Weight of sample | Spec Colour | Blank Colour | Final colour | Solution concentration of sample | g in 0.5ml assay | µg glucose in 60µl | µg glucose in 1.165 | µg glucose per g weight |
| Can 1 9am | 0.1894 | 0.1260 | 0 | 0.1210 | 0.1263 | 0.042 | 4.9086 | 95.3088 | 2264.4643 |

### Exon Usage analysis

Analysis of exon usage was carried out using the Bioconductor packages SubRead and DEXSeq. First, in SubRead, an index was built using the Tair10 Arabidopsis genome as reference. Then the fastq files from the study by Gan et al, (2011) for Can-0 and Wu-0 were aligned to the index to the exon level using SubRead. FeatureCounts was then used to assign mapped reads to exons and produce counts of reads for each exon.

DEXSeq was then used to look at differential exon expression (Anders et al, 2012). A DEXSeqDataSet object from the SubRead counts analysis was created. Then data was normalised by estimating size factors which estimate sequencing depth of each sample. The variability of the data was then estimated to distinguish for technical and biological noise. This is done by assessing how genes vary in their expression level and in their repetitions and compares how other genes of similar expression level vary. A curve is fitted to the data and gene expression is shrunk to the curve to account for the dispersion. Finally, differential exon usage is estimated using a generalised linear model. Differential expression was accepted at P<0.05.

Code can be seen at <https://github.com/CrookedY/ExonUsage>

### Promotor Analysis

Sequences for promotor analysis were taken from the 1001 genomes project (http://1001genomes.org/). The sequence 8,000bp upstream from the start site for each of the 19 parental lines was examined. A list of cis acting promotor sites was found on The Arabidopsis Gene Regulatory Information Server (AGRIS) (Yilmaz et al, 2011) and Python was used to search for the number of times the elements appeared in the coding strand sequences.

### Statistical Analysis

All statistics were carried out using R statistical programming language. Student’s T-test was used to assess the differences between T-DNA insertions and their background lines stomatal density. Two-way ANOVA was performed to determine the effect of line and PPM on SD and SI and cotyledon size. Pearson’s correlation coefficient was used to investigate whether there was a correlation between cotyledon size and SD and SI. Line of regression on graphs was calculated using a linear model. ANOVA was used to look at time series data with a Tukey’s test performed as a post hoc test for all analyses. Significance was accepted at P<0.05.

Graphics were produced using R using ggplot2 with the exception of the graphs for exon expression which were produced using the Bioconductor R package DexSeq.

## Results

### *ADG2* Phenotyping

The experiment shown in Chapter 3, Figure 3.3.11, that identified *adg2-3* as having a phenotype of interest, was carried out in 450ppm CO2. In order to confirm the phenotype and investigate the effect of elevated [CO2], plants were regrown in both 450ppm and 1000ppm to determine whether the phenotype was consistent across CO2 concentrations. The QTL was only significant at 450ppm (see chapter 3.3.4), however its value is close to the significance threshold in 1000ppm and therefore the phenotype at this higher CO2 concentration was examined.

|  |  |
| --- | --- |
| A) | B) |
| Figure 4.3.2 adg2-3 has altered stomatal density and index. Violin plots of SD (A) and SI (B) of adg2-3 mutant grown in 450ppm and 1000ppm (N=16). Letter indicate significantly similar groups (P<0.05, ANOVA with post hoc Tukeys test). Dots indicate outliers of 1.5x IQR outside IQR. | |

The *adg2* mutant demonstrated a significant increase in SD in both [CO2] treatments (P=0.001), as seen in a, and showed an increase in SI compared with Col-3 in ambient [CO2] (P=0.001) as seen in b. The change in SI in elevated [CO2] was not significant (P=0.2). The significant difference in SI in 450ppm suggests that a change in stomatal initiation may account for the increase in SD under these conditions however, the same is not true for plants grown in 1000ppm. The higher SD, but similar SI, implies that plants in elevated [CO2] have smaller epidermal cells than in 450ppm but this has no major impact on stomatal initiation.

As the significant QTL was only found under 450ppm conditions, it was decided to focus on plants grown in 450ppm for the majority of further experiments to allow for more thorough phenotyping.

### Cotyledon Size and Stomatal Density

The cotyledons of the *adg2-3* mutant were observably smaller than that of Col-3, suggesting the size difference may have been responsible for the change in SD. The size difference was confirmed via cotyledon area measurements (, P=0.002). Younger cotyledons have higher stomatal densities than fully expanded ones, as shown in Chapter 2, Figure 2.3.3. To ensure that the change in SD in adg2-3 wasn’t a result of the cotyledon size, size of the cotyledon was also controlled for by individually determining the SD, SI and cotyledon size for each cotyledon.

|  |
| --- |
| c  b  c  b  a  a  b |
| Figure 4.3.3 Size of Arabidopsis adg2-3 and Col-0 cotyledons in mm2. Cotyledons were grown in both 450ppm and 1000ppm [CO2]. adg2-3 cotyledons are smaller than Col-3. Letters indicate statistically similar groups (P<0.05, ANOVA, Tukeys Test), N=10. |

|  |  |
| --- | --- |
| A) | B) |
| Figure 4.3.4 Cotyledon Size and SD correlate. Correlation between stomatal density (A) and stomatal index (B) and cotyledon size are shown. Line of best fit represents linear regression mode. Shaded area shows 95% confidence intervals. There is a significant negative relationship between size and density (P<0.01, Person's correlation) but not for stomatal index (P=0.076, Person’s correlation). However, clustering of the data can be seen between lines with large and small cotyledons. | |

There is a strong negative relationship between cotyledon size and SD (, P< 0.001) showing that SD is reduced in larger cotyledons. The same trend was also observed for SI, although this is not significant (Figure 4.3.2, P= 0.076). A potential alternative explanation for these trends could be because the large cotyledons are all *Col-3*, the line with the lower SD. In order to control for this potential bias the data was split into two groups based on line and the analysis run again

|  |  |
| --- | --- |
| A) | B) |
| Figure 4.3.5 Effect of size on SD and SI within lines. Relationship between stomatal density (A) and stomatal index (B) shown. There is no significant difference (P<0.05, pearson's correlation) for either line. Line of best fit is linear regression and shading shows 95% confidence intervals. | |

In Figure 4.3.5 A, the two genotypes are plotted separately. Splitting the data shows that cotyledons of similar size have a higher SD in *adg2-3* than in Col-3. Splitting the data also results in non-statistically significant trends ( A, Col-3: P=0.198, *adg2-3*: P=0.67). Similarly, SI does not show significant correlation when data is split ( B, Col-3: p=0.97, *adg2-3*: P=0.81). This suggests that the phenotype is not due to a still expanding cotyledon, warranting further investigation into the control of the *adg2* mutant phenotype.

### Phenotype in Mature Leaves

All previous phenotyping experiments were carried out in cotyledons however, some phenotypes are only seen in specific developmental stages of Arabidopsis and some genes are linked to developmental stage. For example, growth rate at different developmental stages of Arabidopsis can be different and are controlled by a different QTL (Bac-Molenaar et al, 2015) and trichome coverage has been shown to be linked to leaf developmental stage ([Lièvre](https://nph.onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Li%C3%A8vre%2C+Maryline) et al, 2016). This SD phenotype could therefore be cotyledon specific. To see if the increased SD phenotype is observed in mature leaves as well as cotyledons, SD and SI measurements were taken for 6 week old plants.

|  |  |
| --- | --- |
| \*  A) | \* |
| Figure 4.3.6 The effect of adg2-3 on mature leaves. Stomatal Density (A) and Stomatal Index (B) or mature leaves of adg2 mutant grown in 450ppm is shown. \* indicates significant difference (P<0.05, T Test), N=36 | |

There was a significant difference in SD between Col-3 and *adg2-3* (P<0.001) and a significant difference in SI (P=0.03). This confirms that the phenotypes of *adg2-3* seen in cotyledons are also present in mature leaves. There was no observable difference in leaf size. This research is therefore applicable to mature leaves as well as cotyledons.

### *ADG2* Sequence in 19 parental lines

Examining the sequence of *ADG2* and the sequence 5’ and 3’ to the translational start and stop codons can indicate whether a SNP or deletion is responsible for the phenotype. To be of interest, any SNPs must be unique, or mostly unique, between Can-0 and Wu-0 as these are the haplotypes with the largest predicted differences in the haplotype analysis (Chapter 3 Figure 3.3.7). shows the 1001 genome browser display of *ADG2* in the 19 parental ecotypes as sequenced in Gan et al, (2011). It can be observed from this display that there are SNPs throughout the gene as well as deletions in the 5’ promotor region of the gene in some ecotypes. Between Can-0 and Wu-0, the two with the largest predicted haplotype difference, there is one SNP difference in the exons of *ADG2*, a T in exon one, and a SNP in intron 7. Upstream there are many SNPs in the promotor region indicating the possibility of differential gene expression due to SNPs in binding elements.

It can also be noted that there is a predicted anti-sense RNA overlapping the gene. Anti-sense RNAs have been shown to affect plant phenotypes, including gene expression in flowering, specifically the methylation of *Flowering Locus C* (*FLC*) (Letswaart, 2012). The T-DNA insertion in *adg2-3* occurs within the anti-sense RNA as well as within the gene, suggesting the possibility that it is the interruption of the anti-sense RNA that is causing the phenotype rather than the interruption of the gene itself.

Viewing the amino acid track for the genome browser reveals that the SNP observed in exon one is a non-synonymous mutation. This suggests this amino acid change is not responsible for the observed difference in haplotype phenotype. No other non-synonymous changes between these ecotypes and Col-0 are unique to either ecotype suggesting they are not the cause of the QTL (Figure 4.3.8).

|  |
| --- |
|  |
| Figure 4.3.7 Graphical representation of SNPs in 19 parental ecotypes of the MAGIC (Multiparent Advanced Generation Inter-Cross) Lines in ADG2 (AT5G19220). Graphic shows the structure of ADG2, the structure of the anti-sense RNA (AT5G19221) and the SNPs present in the gene (coloured lines indicate SNPs) and the untranslated regions and the deletions (grey bars). 5’ to 3’ runs from right to left. Differences in SNPs between Wu-0 and Can-0 highlighted with black circles on Can-0. From 1001 genomes project browser (<http://signal.salk.edu/atg1001/3.0/gebrowser.php>) |
|  |
| Figure 4.3.8 Graphical representation of amino acid changes caused by non-synonymous SNPS of the MAGIC (Multiparent Advanced Generation Inter-Cross) Lines in ADG2 (AT5G19220). Graphic shows the structure of ADG2, the structure of the anti-sense RNA (AT5G19221) and the amino acid changes present in the gene (red lines indicate amino acid changes, green the location of synonymous SNPs) and the untranslated regions and the deletions (grey bars). 5’ to 3’ runs from right to left. Changes are represented as lines with new amino acid letter next to it. Box around letter indicates it is in an exon. Green means the SNP causes no amino acid change, red mean there is a non-synonymous change. Grey boxes indicate deletions. From 1001 genomes project browser (<http://signal.salk.edu/atg1001/3.0/gebrowser.php>) |

### 

### *ADG2* Sequence in Can-0 and Wu-0

Whole genome sequencing by Gan et al (2011) of the 19 parental lines was carried out to 27-60 fold sequencing coverage. Following advice by from Dr P. Kover (University of Bath) ADG2 in Can-0, Wil-2, Hi-0 and Wu-0 was resequenced to confirm interesting SNPs, especially in introns or promotors. Hi-0 and Wil-2 were sequenced to compare SNPs across lines and confirm uniqueness in Can-0 and Wu-0. SNPs were determined to be of interest if they were only found in one ecotype and were not synonymous mutations. These criteria found two SNPs of interest which can be seen in Figure 4.3.9. The first SNP was a single bp deletion in the promotor region of *ADG2* of Can-0 (Figure 4.3.9A) 78 bp upstream from the transcriptional start site. The second was a SNP in intron 7 in Can-0.

Changes to promotors can result in changes to expression, either in terms of level or timing or response to transcription factors and signals. The deletion in Can-0 promotor suggested studying transcription was a promising avenue of research. SNPs in introns can also cause changes to expression if introns are regulatory and can also cause changes in splicing. Splicing was therefore also investigated.

|  |  |
| --- | --- |
| Wu-0 |  |
| Can-0  A) |
| Wu-0  B))  Can-0 | A  B |
| A  C) | |
| Figure 4.3.9 Changes in ADG sequence between Can-0 and Wu-0. A single base pair deletion in the promotor site of Can-0 -78 bp from start site (A) and a single SNP in intron 7 in Can-0 (B) were identified. C shows the approximate position of the SNP ADG2 with 5’ end on the right and 3’ on the left. Boxes indicate exons and lines introns. | |

### Promotor analysis

To study the changes in the *ADG2* promotor site, an analysis of the number of promotor cis elements in each ecotype of interest was conducted. The list of expected promotor elements was downloaded from AGRIS and the programming language Python was used to search for the elements. The results

Table 4.3‑1) showed that while there were differences in the number of each element in each ecotype, there was no unique pattern that marked Wu-0 as being different from the other ecotypes, leading to its differing haplotype phenotypes. Can-0 does appear to have a different number of tgataa and actttg binding sites (see Table 4.3‑2 to match sequence to bind site description) giving it a unique combination of upstream sites. This sequencing found is a one bp deletion in promotor at site -78 upstream of the promotor start site however this is not in either of the missing transcription factor bind sites mentioned above. This suggests that the deletion found with sequencing is not of a known transcription factor bind-site. It is possible that this change could still have an effect on expression. Based on differences in promotor, expression analyses were performed and are detailed below in section 4.3.10.

Table 4.3‑1 Number of times each promotor element appears in the 8000bp upstream of ADG2 start site. Promotor element names can be found in Table **4.3.2**. Can-0 stands out as having a unique combination of promotor elements among ecotypes (genome sequences collected from 1001 genomes project and analysed using AGRIS (Yilmaz et al, 2011))

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Promotor element | | | | | | | | | | | | | | | | |
| Line | actcat | actttg | tgataa | agatag | agatag | tgccgactt | Tgtctc | ccattg | aacaaac | caaca | ccagtg | accaaac | ttgacc | ttgact | tgtctc | aaccaaac | aactaac |
| bur | 2 | 5 | 11 | 6 | 6 | 1 | 5 | 7 | 4 | 20 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| can | 2 | 4 | 10 | 7 | 7 | 0 | 5 | 7 | 4 | 20 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| col | 2 | 5 | 11 | 7 | 7 | 1 | 5 | 8 | 4 | 20 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| ct | 2 | 4 | 11 | 6 | 6 | 0 | 5 | 7 | 3 | 19 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| edi | 2 | 5 | 10 | 7 | 7 | 1 | 5 | 7 | 4 | 20 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| hi | 2 | 5 | 9 | 7 | 7 | 1 | 5 | 7 | 3 | 19 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| kn | 2 | 5 | 10 | 7 | 7 | 1 | 5 | 8 | 4 | 20 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| ler | 2 | 4 | 11 | 7 | 7 | 0 | 5 | 7 | 4 | 19 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| mt | 2 | 5 | 10 | 7 | 7 | 1 | 5 | 7 | 3 | 19 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| no | 2 | 4 | 11 | 6 | 6 | 0 | 5 | 7 | 3 | 19 | 6 | 3 | 10 | 5 | 5 | 1 | 2 |
| oy | 2 | 5 | 11 | 6 | 6 | 1 | 5 | 7 | 3 | 19 | 6 | 3 | 10 | 6 | 5 | 1 | 3 |
| po | 2 | 5 | 11 | 7 | 7 | 1 | 5 | 7 | 3 | 20 | 6 | 4 | 9 | 6 | 5 | 2 | 3 |
| rsch | 2 | 5 | 10 | 7 | 7 | 1 | 5 | 7 | 3 | 19 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| sf | 2 | 5 | 11 | 7 | 7 | 1 | 5 | 7 | 4 | 20 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| tsu | 2 | 5 | 10 | 7 | 7 | 1 | 5 | 7 | 4 | 20 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| wil | 2 | 5 | 11 | 7 | 7 | 1 | 5 | 7 | 4 | 20 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| ws | 2 | 5 | 11 | 7 | 7 | 1 | 5 | 7 | 4 | 20 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| wu | 2 | 5 | 11 | 7 | 7 | 1 | 5 | 7 | 4 | 20 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |
| zu | 2 | 5 | 11 | 7 | 7 | 1 | 5 | 7 | 4 | 20 | 6 | 3 | 9 | 6 | 5 | 1 | 3 |

Table 4.3‑2 List of promotor bind sites found upstream of ADG2. The respective sequences, start and end sites and bind site family are shown

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bind Site Name | BS Genome Start | BS Genome End | Binding Site Sequence | Binding Site Family/TF |
| **ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH** | 6467001 | 6467006 | actcat | [bZIP](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=bZIP) |
| **W-box promoter motif** | 6467493 | 6467498 | ttgacc | [WRKY](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=WRKY) |
| **W-box promoter motif** | 6466985 | 6466990 | ttgact | [WRKY](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=WRKY) |
| **ARF1 binding site motif** | 6467124 | 6467129 | tgtctc | [ARF](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=ARF) |
| **MYB binding site promoter** | 6467255 | 6467262 | aaccaaac | [MYB](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=MYB) |
| **MYB4 binding site motif** | 6466996 | 6467002 | aactaac | [MYB](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=MYB) |
| **MYB4 binding site motif** | 6467255 | 6467261 | accaaac | [MYB](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=MYB) |
| **MYB4 binding site motif** | 6466859 | 6466865 | aacaaac | [MYB](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=MYB) |
| **RAV1-A binding site motif** | 6466989 | 6466993 | caaca | [ABI3VP1](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=ABI3VP1) |
| **RAV1-A binding site motif** | 6467161 | 6467165 | caaca | [ABI3VP1](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=ABI3VP1) |
| **RAV1-A binding site motif** | 6466862 | 6466866 | caaca | [ABI3VP1](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=ABI3VP1) |
| **LFY consensus binding site motif** | 6467278 | 6467283 | ccagtg | [LFY](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=LFY) |
| **LFY consensus binding site motif** | 6467366 | 6467371 | ccattg | [LFY](http://agris-knowledgebase.org/AtTFDB/tfbrowse.html?fam=LFY) |
| **ARF binding site motif** | 6467124 | 6467129 | tgtctc | ... |
| **DRE-like promoter motif** | 6467169 | 6467177 | tgccgactt | ... |
| **GATA promoter motif [LRE]** | 6467387 | 6467392 | agatag | ... |
| **GATA promoter motif [LRE]** | 6467568 | 6467573 | agatag | ... |
| **GATA promoter motif [LRE]** | 6467235 | 6467240 | tgataa | ... |
| **T-box promoter motif** | 6467174 | 6467179 | actttg | ... |

### Anti-Sense RNA

To rule out the possibility than the anti-sense RNA (asRNA) present in the same position was responsible for the phenotype, a T-DNA insertion line, SALK-208532, with an insertion in *ADG2* but not in the asRNA was grown (Figure 4.3.10). This showed the same phenotype as the *adg2-3* insertion line with a statistically significant increase in SD (T-test, P=0.001) and SI (T-test, P=0.001). It is still possible that the expression of the anti-sense RNA could still be disrupted by the SALK insertion line however, as the insertion is in the 3’-UTR, a more likely explanation is that it is the interruption of the gene having an effect. In line with the sequence analysis discussed above, this would also support a scenario whereby a SNP or deletion in the Can-0 allele leads to reduced expression or activity of ADG2 in that ecotype.

|  |  |
| --- | --- |
| A)  \* | B)  \* |
| Figure 4.3.10 Stomatal Density (A) and Index (B) of Col-0 and T-DNA insertion line SALK-208532 in box plots. Box indicates Inter-quartile range, line indicates median. Dots indicate outliers more than 1.5X IQR outside of IQR. \* indicates significant difference (P<0.05, T-test) N=16. | |

### The AGP Complex

As ADG2 is part of a complex, it is possible that it is the overall disruption of the complex which has a regulatory impact on SD. To determine if the effect on SD was specifically caused by ADG2, or whether other genes involved in the complex also have an effect, T-DNA insertion knock-outs were grown for the small subunit, and the other isoforms of the large subunit (). Analysis on the data showed that only *ADG2* knockouts influenced SD (P=0.001) compared with the wild-type Col-3. No other lines showed a significantly different SD compared with their background line (Col-0) (T-test, P>0.05). This suggests that SD regulation is not controlled by the complex as a whole and also that ADG1, the small subunit, does not have a regulatory effect. It is possible that the other large subunit isoforms could an effect on SD regulation however, as these isoforms are not highly expressed in the leaf, it may not be possible to determine their impact through knock-out.

|  |
| --- |
| \* |
| Figure 4.3.11 Effect of AGPase gene knockouts on SD. SD shown for T-DNA knock outs of the genes that make up the ADP-glucose pyrophosphorylase (AGP) complex and their background lines. \* indicates significantly different from background line (P<0.05, T-test) N=16. Dots indicate outliers more than 1.5 times IQR outside the IQR |

Knocking out the small subunit of the AGP complex is expected to result in a similar reduced starch phenotype to that of ADG2 (Ragel et al, 2013). That the small subunit should result in deficient starch but did not have an SD change would therefore suggest that it is the gene itself and not the starch deficiency having an effect. To check this, starch staining was used to determine the starch content of each line at dusk.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Col-0 | APS1 | APL2 | APL3 | Col-3 | ADG2 |
|  |  |  |  |  |  |
| Figure 4.3.12 Starch staining of a leaf of 3 week old T-DNA insertion lines of the genes that make up the ADP-glucose pyrophosphorylase (AGP) complex at dusk. Only ADG2 (adg2-3) shows a different in colour. | | | | | |

Starch staining suggests that, with the exception ofADG2, other isoforms in the complex and the small subunit do not have an observable reduction in starch content (Figure 4.3.12). Though staining does not give an accurate picture of starch content, the technique can show clear differences in relative starch content. As APS1 does not appear to show a reduced starch content, it was possible that starch is the cause of the phenotype, and this required further investigation.

### Starch Deficiency

To establish whether the phenotype seen is independent of the *ADG2* gene and related to the gene’s starch deficient phenotype, mutants that were deficient in starch accumulation (*fba2-2*, *pgm-1* and *adg2-3*; Lu et al, 2012; Pal et al, 2013), break-down (*sex1*) and sucrose transport (*atsweet11:12*) were studied These plants included the starch excess mutant (*sex1*) (Yu et al, 2001) to study if an excess of starch caused the opposite of the phenotype seen in *adg*2 as well as the starch over accumulator in leaves *atsweet11:12* (Chen et al, 2012).

|  |  |
| --- | --- |
| A) | B)  \*  \*  \* |
| Figure 4.3.13 Stomatal density (A) and stomatal index (B) of starch deficient (fba2, pgm1, adg2-3) or starch excess (sex1, sweet) plants and their background lines. \* indicates significantly different to its background line (T-test). \*\* indicates significant different at P<0.01 and \*\*\* indicates P<0.001. Dots indicate outliers more than 1.5 times IQR outside the IQR. N=16 | |

All three starch deficient mutants (*fba2-1*, *pgm-1*, *adg2-3*) showed a significant increase in both SD and SI (P<0.05). This suggests that the phenotype seen in response to reduced starch is independent of the gene *ADG2*, though the impact of cotyledon size may need to be accounted for. The lines with an excess of starch (*sex1, atsweet11:12*) however did not show a significant change in SD (T-test, P=0.52) indicating that having more starch does not affect SD, only a reduction in starch content.

### *ADG2* Circadian Expression Analysis

Regulatory elements can be found a long way from transcriptional start sites, meaning the analysis of promotor bind sites in section may not account for all regulatory regions. An analysis of expression throughout the day of *ADG2* was therefore carried out in Can-0, Wu-0, ADG2-3 and Col-0 to investigate expression level. *ADG2* has previously been shown to be circadian regulated with an approximately 50% increase in expression (see ) in the middle of the day as compared to dawn expression levels (Smith et al, 2004). Three time points (dawn, midday and end of day) were therefore chosen to study how expression varies between ecotypes.

|  |
| --- |
|  |
| Figure 4.3.14 The expression of ADG2 relative to Col-3 (blue) throughout the light period of the day (white box). Can-0 (red) and Wu-0 (green) have no significant difference at any point in the day (P<0.05, ANOVA, Tukeys test) suggesting that this expression of ADG2 is not related to the QTL. adg2-3 was significantly different at all 3 time points to all other lines. Black boxes represents dark. Time points were taken at 09:00, 14:00 and 19:00, N=3. |

Variation in gene expression occurred throughout the day, with expression rising by around 20% in the middle of the day as compared to dawn (9:00), and falling in the evening to 30-40% lower than at dawn in the 3 ecotypes (). There was no significant difference in expression of *ADG2* between ecotypes at any time point (P=0.466) but there was a significant difference between *adg2-3* and all 3 ecotypes (P<0.05). While not significant, it is possible that Wu-0 has slightly reduced expression in the middle of the day in compared with other ecotypes, however the large variation between samples makes this difficult to ascertain.

The knock out *adg2-3* does still show some expression. This is likely to be because the fragment measured occurs before the T-DNA insertion. It too follows a circadian rhythm and the expression level is significantly lower than the ecotypes at all times of day (P<0.001).

The changes to *ADG2* are comparable to those seen by Smith et al, 2004. Their experiments revealed a modest increase in expression 4 hours post dawn followed by a fall to dawn levels. The present experiment however, showed expression level falling below dawn levels at dusk, while the 2004 study showed a more modest drop. This difference could be due to different experimental set ups, for example the difference in photoperiods used. However, Mugford et al, 2014, showed that light period does not have a significant effect on *ADG2* expression between 6:18 and 12:12 light:dark conditions, suggesting that photoperiod is not responsible for the changes. Other changes in factors such as light intensity, CO2, humidity etc were also present though and could have affected expression level.

As there are no significant differences in *ADG2* expression between ecotypes, if this gene is the basis for the QTL on chromosome 5, it is not likely to be due to any significant change in either expression level or circadian pattern. The haplotype model in chapter 3 suggests that Can-0 and Wu-0 haplotypes result in the largest difference in phenotype meaning they should show the greatest difference in expression if they phenotype was due to differences in expression in *ADG2*. As there is not difference in gene expression level between lines, gene expression of *ADG2* can be ruled out as the cause of the QTL on chromosome 5.

### Circadian Starch Content of Ecotypes and *adg2-3*

Reduced starch has been shown in 0 to result in increased SD in multiple lines. It is therefore possible the starch content of leaves has an effect on the ecotypes resulting in differences in SD. Starch is produced in a circadian manner, therefore starch accumulation over the course of the day was studied.

|  |
| --- |
|  |
| Figure 4.3.15 The change in starch content as measured using lugol’s solution. Mean grey value of cotyledons stained with Lugol's solution throughout the light period and 2 hours predawn shown. The darker the leaf from staining, the lower the grey value and the more starch the leaf contains. Point represents mean and error bars standard error. |

The changes in expression are consistent with diurnal changes to starch production. Starch production occurs during the day when photosynthesis is occurring and falls at night as starch break down begins (Lu et al, 2005). These changes are evident in the starch content seen in the 4 lines studied, the 3 ecotypes and *adg2-3*. Two methods of starch quantification were used to look at variation throughout the day.

Staining with Lugols solution can give a semi-quantitative assay of starch content. By measuring grey value, it was possible to turn a qualitative assay into a semi-quantitative assay. While earlier in the day, when starch content is lower, staining did not give a clear picture of starch accumulating, towards the end of the day, it was possible to observe differences(). The starch content of the mutant was significantly reduced compared with Can-0, Wu-0 and Col-3 leaves at dusk (P<0.01). In addition, the starch content of Can-0 was significantly decreased compared with Wu-0 and Col-3 (P<0.05).

It cannot be said how these changes in colour directly relate to total starch content as no plant material was collected for enzymatic quantification at the same time. Stain colour may not have a linear relationship with starch content and therefore relative starch content cannot be inferred from this. These results however, do not follow what would be expected in the ecotypes based on how the mutant behaves. Can-0 has the lowest SD but has the lowest starch content at dusk. If starch was the driver of SD in the ecotypes, Can-0 would be expected to have the highest content as low starch content in mutants causes higher stomatal densities.

These data do not appear to support the hypothesis that starch content is a driver of SD, this may be because in Can-0 there are other controls of SD that have a larger effect than starch content. It is also possible that SD is not related to total starch content, but is instead related to starch break down, or starch granule shape or size. It is also possible that the assay used is not accurate enough to look at differences in starch content. The QTL was previously predicted to account for approximately 13% of variance leaving a further 87% that can control for SD outside of this QTL. These 87% of controls could be shuffled out in the MAGIC QTL lines to allow the QTL on chromosome 5 to be seen.

This analysis was done as a pilot study to look at the best timings for sampling times for enzymatic quantification. It was decided to take a measurement at dawn and dusk and a midway measurement to see if Can-0, Wu-0 and Col-3 accumulated starch in a similar manner throughout the day as suggested here.

|  |
| --- |
|  |
| Figure 4.3.16 The starch content of lines throughout the light period of the day. Can-0 (red) and Wu-0 (purple) have no significant difference at the beginning or end of the day (P>0.05, ANOVA, Tukey’s test) but do show a difference in rate of accumulation throughout the day (p<0.05).Col-3 (Green) and adg2-3 (Blue) are significantly different to all other lines at the end of the day (P<0.05). |

It can be seen from that while there are differences in starch accumulation throughout the day, there is no significant difference between Can-0 and Wu-0 in starch accumulation at the end of the day (P>0.05). If starch content at the end of the day was the driver of the QTL on chromosome 5, a significant difference between these two ecotypes could be expected. Col-3, which is closely related to Col-0, had an intermediate SD in the model but a lower starch content than Can-0 and Wu-0 in this study, although this was not significant (P=0.271). The lower starch content of Col-3 observed does not fit with expectations of the haplotype model, as Col-0 has an intermediate SD phenotype and would therefore be expected to be found in the middle of Can-0 and Hi-0. The mutant line *adg2-3* had a reduced starch accumulation compared to its wild type (P=0.0082) which is what was expected. Low starch accumulation in other *adg2* mutants was previously observed (Lin et al, 1988; Smith et al, 2004).

Overall, there does not seem to be an observable correlation between SD and total starch accumulation in these ecotypes. It is possible that the effect of total starch accumulation on SD in these lines is masked by other genes and environmental signals but that cannot be elucidated from this experiment. With few lines it makes it difficult to draw any conclusions except that this data does not support that total starch accumulation is responsible for the chromosome 5 QTL.

Dawn measurements also show very similar starch content for all lines, in particular Can-0 and Wu-0, implying that dawn starch content is also not responsible for this QTL. All lines are statistically similar at dawn (9:00) except *adg2-3* which is only similar to Col-3. This suggests that plants aim for the same amount of depletion at the end of the night even if they have accumulated less. It is also possible that the assay is not sensitive enough to detect differences at very similar starch contents, further samples would be needed. There do appear to be differences in starch accumulation rates throughout the day with Can-0 accumulating starch more rapidly than Wu-0 or Col-0 and this could perhaps be a reason for the differences in SD although this was not significant (P=0.076).

Can-0 accumulated starch more rapidly and has a lower SD, however Wu-0 is similar to Col-0 but their stomatal densities are significantly different (See section 2.3.4). This study therefore does not provide enough evidence to determine whether the rate of accumulation is a factor affecting SD. Further time points would create a more complete picture of starch accumulation throughout the day and provide more evidence of differences in accumulation rate throughout the day.

There is no significant correlation between starch content at dusk and *ADG2* expression at 14:00 (P=0.16) suggesting that the level of gene expression for *ADG2* does not necessarily directly control starch accumulation. As it is the complex as a whole that is the rate limiting step, it could be that the *APS1* gene could also be taken into account when considering *ADG2* and starch. The previous experiment looking at APS1 starch content with staining did not show a difference in starch content, but it is possible differences would be observed with an assay. There are also only 3 data points that can be considered for this correlation (because *adg2-3* is a mutant is does not represent an ecotype) and more data from further ecotypes may show correlation.

This data does not support the hypothesis that total starch is a factor in SD regulation, however, rate of starch accumulation could have a role. Additionally, starch break down and utilisation at night could have an effect but were not studied in this experiment. Overall the findings of this study do not give support to the theory that differences in starch content are the cause of the QTL on chromosome 5.

### Alternative Splicing

As neither *ADG2* expression nor starch content appear to explain the QTL, it was decided to investigate alternative splicing. Different splice forms can have a major effect on phenotypes; for example the flowering gene *FCA*, for example, has 4 possible different transcripts and increased expression of the γ transcript, results in an earlier flowering plant (Eckardt, 2002).

As previously reported in 1.1.1, there is little difference in genomic sequence between ecotypes, however there is a single SNP in an intron in between exons 7 and 8 (Figure 4.3.9). It does not affect the splice motif GT|AG however. Graphical viewing of *ADG2* in the GBrowse instance for *Arabidopsis thaliana* (19 MAGIC founder lines) (), appears to show expression of some exons, such as exon7, in some ecotypes to be present, such as Hi-0, while not in others, such as Zu-0. This is especially true around intron 7 making alternative splicing a good candidate for further analysis for control of this phenotype.

|  |
| --- |
|  |
| Figure 4.3.17 GBrowse instance for Arabidopsis thaliana (19 MAGIC founder lines) at ADG2. RNA seq data is derived from study by Gan et al, 2011. Found at https://gbrowse.inf.ethz.ch/gb/gbrowse/thaliana-19magic/ |

A B C

A C B

Wil-2

Can-0

|  |
| --- |
|  |
|  |
| Figure 4.3.18 Alternative splicing in ADG2. A) The location of the SNP in intron 7 indicated by a black line surrounded by exons 7 and 8 (labelled 7&8). Gene 5’ to 3’ runs right to left. B) Gel revealing two different length PCR products for the ADG2 gene at exons 7 and 8 shown with arrows in Can-0. Gel shows products of Can-0, Wu-0, Hi-0 and Wil-2 between exons 7 and 8 (ADG2), the actin control (ACT) and the no template genomic DNA control (gDNA). Bands of the same size as the second band in the ADG2 lanes can be seen in the genomic controls of Wil-2, Wu-0 and Hi-0 indicated with \* suggesting there is some genomic contamination in the cDNA. Can-0 however has no contamination in its gDNA but has bands in the cDNA. |

To look for evidence of alternative splicing, PCR of cDNA of 4 ecotypes was examined using gel electrophoresis to look for potential splice variants. A region spanning exons 7 and 8 was amplified using PCR to look for intron inclusion. Two bands were observed in all 4 ecotypes suggesting evidence of alternative splicing in all (). Negative controls of RNA with no reverse transcriptase used during cDNA synthesis also showed bands in 3 of 4 ecotypes, suggesting that this band could be due to contaminating genomic DNA. Can-0 had no band in the negative control, indicating that the second band in the cDNA sample exists without the genomic contamination and is a candidate for further investigation.

|  |
| --- |
|  |
| Figure 4.3.19 The expression of each exon in ADG2 in Can-0 and Wu-0. The gene model can be seen beneath the normalised expression data for each exon. There was no significant difference in expression of any exon (P>0.05, GLM). Data from Gan, et al, 2011. |

To look for further evidence of alternative splicing, a bioinformatics approach was used. RNA-seq data produced by Gan et al (2011) was reanalysed using the dexseq package in R (Anders et al, 2012). While there is evidence in Figure 4.3.17 to suggest there could be alternative splicing, no evidence of alternative isoforms or differential expression of alternative isoforms could be found between Can-0 and Wu-0 (Figure 4.3.20, P=0.99). It is possible that alternative splicing does differ between ecotypes but cannot be seen via this approach due to the short read length of the chosen library prep for the RNA-seq set, and the low statistical power available as there were only two replicates for each line. However, the differential exon usage analysis shows incredibly similar patterns and the adjusted p value is close to 1.

In addition, viewing of the raw reads in genome browser was not suggestive of alternative splicing. There is no exon skipping visible, however there is minor evidence of intron inclusion which could be due to unfinished splicing. This suggests that if alternative splicing does exist, it is minor and does not appear to be different among the two ecotypes.

A))

B)

|  |  |
| --- | --- |
|  |  |
| Figure 4.3.20 Interactive Genome Viewer image of ADG2 in A) Can-0 and B) Wu-0. No evidence of alternative splicing can be seen. Some intron inclusion can be seen highlighted by black boxes | |

## Discussion

### QTL5 and ADG2

While *ADG2* does appear to play a role in regulating SD, these data do not clearly demonstrate that variation in *ADG2* explains the QTL on chromosome 5. There is no data that consistently differentiates between Can-0 and Wu-0 in sequence, expression, splicing or starch content. As these are the two most different ecotypes in the haplotype models, there is suggestion that the phenotype found here is unrelated to the QTL and was found by coincidence. However, as the QTL on chromosome 5 accounts for only 13% of variation in SD, it is also possible that differences in expression, splicing or starch content between Can-0 and Wu-0 cannot be seen due to confounding effects of their genetic background. In the MAGIC lines, this confounding genetic background may be ‘shuffled out’ through recombination allowing for the role of ADG2 to be more clearly seen. Further study to find MAGIC lines with differential expression of ADG2 may be a potential future avenue for study to further study the role of ADG2 in SD.

To study specifially whether the different ADG2 haplotypes of Can-0 and Wu-0 influence stomatal density, transformation of the ADG2-3 mutant with the genomic sequence of ADG2 from Can-0 and Wu-0 would most effectively prove whether ADG2 has an effect on stomatal density. If the stomatal density of the two transformed lines differed it would show that the allele of ADG2 present had an effect on SD and that ADG2 was the gene responsible for the QTL on chromosome 5. This could be an exciting future avenue of research.

In addition, non-synonymous polymorphisms in the *ADG2* gene could be transformed into *E. coli* and heterologous over-expression of the different variants could be assessed by the activity of the ADP-glucose pyrophosphorylase produced *in vitro* in the manner of Salamone et al (2000). This would allow for analysis of the effect of specific SNPs in ADG2 and the effect these have on plants.

Starch content has been clearly shown here to play a role in SD, but it is not clear to what extent or why. The previous association with stomatal aperture shows that starch is necessary for the proper functioning of stomata (Outlaw and Manchester, 1979; Horrer et al, 2016; Valerio et al, 2011), and the research of this thesis suggests that it may play a part in regulating SD. These data suggest starch deficiency may lead to increases in SD, which we hypothesise could be a starvation response or a response to excess free sugars.

### Starch and Stomatal density

The growth and development of a plant are tightly controlled by carbon availability (Gibson, 2005) and as such it is reasonable to consider that stomatal development could be controlled by starch availability. A plant lacking in starch may consider itself to be starving and so increase the number of stomata to try and acquire more carbon for starch biosynthesis. A plant with an excess of starch however does not try to reduce the number of stomata, perhaps because having enough starch does not mean the plant could not utilise more.

In a cotyledon, the initial rate of growth is limited by starch content of the seeds (Andriotis et al, 2012) and it is possible that rather than starch production of the cotyledon affecting SD, it is the starch content of the seed. However, as mature leaves of the *adg2-3* mutant showed a similar SD phenotype to *adg2-3* cotyledons, the reduced density phenotype of the cotyledons is still likely an effect of the differences in starch production of the cotyledon leaves. As cotyledons have another starch source from the seed than photosynthesis, it is also possible that they are partially shielded from the effects of reduced starch production, or alternatively strongly affected by the absence of the seed starch if the maternal mutant did not lay down a high amount of starch during seed maturation. A starvation response in cotyledons from lack of starch could therefore be exacerbated or reduced by maternal factors.

The *sex1* mutant produces up to seven times the amount of starch compared with its wild type. However, this is not achieved through increased photosynthesis, but rather through a reduced rate of starch degradation (Yu et al, 2001). This limits carbon availability for the plant with Glucose 3-phosphate levels at around one-fourth of that of the wild type (Yu et al, 2001). This means the *sex1* mutant is not a true over accumulator of starch and the plant could still be perceiving itself as lacking in available carbon.

It has also been shown that *sex1* mutants have elevated expression of starvation response genes at night, suggesting that night time starvation is not responsible for the increased SD phenotype (Adriontis et al, 2012). However, this does not rule out the possibility that specific starvation genes, which were not measured by Andriotis et al, (2012), are responsible. The starvation response could mask the effects of an increase in starch content and may also suggest the possibility that the adg2-3 phenotype is not due to a starvation response.

Carbon efflux effects are also present in *atSweet11:12*, where starch accumulates in leaves but sucrose transport around the plant restricted, leading to other parts of the plant lacking carbon (Chen et al, 2012). The effects of increased carbon could also be masked by other parts of the phenotype in his mutant too.

PGM and ADG2, in contrast to SEX1, fix sugars for starch synthesis and there is an increase in sugar availability in their mutant lines (Ragel et al, 2013). This suggests that sugar availability may also have effect on SD, rather than a lack of starch specifically. It could also mean that the *sex1* mutant is not the best method to investigate a plants response to an increase in starch as there are many factors to consider alongside its increased starch accumulation, such as lack of sugar availability. Further research is needed to establish which part of the metabolic change in these mutants causes the change in SD phenotype.

Further investigation using plants with different starch contents in different growth conditions is needed to further elucidate the full effect starch content can have on SD. It has previously been shown that there is an increase in starch content in elevated CO2 (Teng et al, 2006) and it is possible that this work and the role of starch content in SD does relate back to the [CO2]SD response phenotype. Further studies looking at carbon availability should also be undertaken to ascertain exactly how impaired starch biosynthesis and altered carbon flow affects SD and stomatal development, and to determine whether sugar content has an effect.

Sugars have previously been shown to affect SD. Akita et al, (2013) demonstrated that adding sucrose, glucose or fructose results in stomatal clustering and the expression of stomatal lineage cell-specific genes in cells that were non-stomatal lineage. This suggests that sugar signalling is involved in stomatal patterning and that it’s possible that the increase in sugars available, due to the decrease in fixation, could be the cause of the SD change. Further research investigating sugar content and flux is needed to find whether starch content, carbon availability or sugar content is responsible for the effect on SD in adg2-3.

Finally, starch degradation at night has not been considered during these experiments. Rate of starch use at night has been shown to be tightly controlled as premature depletion results in starvation (Scialdone et al, 2013). As a depleted starch content has been shown to affect SD, it is also possible that night time starvation could cause a change in SD.

Early depletion of starch at night can be a result of starch granule size. It has been shown that Early Starvation1 (ESV1), a gene which controls matrix organisation in starch granules (Feike et al, 2016), increases the availability of starch polymers to degradation enzymes, causing premature depletion at night. Changes to granule structure could therefore cause changes to the availability of sugars from starch and affect SD without affecting starch content.

Starch turnover of the guard cells has been shown to affect stomatal opening (Horrer et al, 2016) and the specific starch content of guard cells and the starch turnover in them has not been examined here. Using the methods of Horrer et al (2016) it would be possible to examine the direct correlation of SD and guard cell starch content and turnover to see if location of starch plays an effect in SD . Localisation of starch and starch turnover is therefore a possible future avenue of examination.

Further investigation is needed to confirm that none of the ecotypes experience night starvation and that starch granule formation is regular. To investigate why *adg2-3* affects SD, investigations should focus on whether starvation response genes are upregulated in *adg2-3*, if free sugars affect SD or if increased starch in *adg2-3* counters the phenotype seen.

### Conclusions

In conclusion, ADG2 is responsible for a starch deficiency phenotype which causes an increase in SD. The gene *ADG2* falls under the 600Kb region around QTL5 however the evidence presented here does not support that it is responsible for QTL5. The haplotype model for the parental lines at QTL5 suggests that Can-0 and Wu-0 haplotypes at this locus result in the most different phenotypes and as such differences between the two lines should be evident. This cannot be seen however, and instead it appears that starch content deficiency leads to an increase in SD but an increase in starch does not. The reason for this has not been established but it can be theorised that this is either due to a starvation response or because of an increase in free sugars.

# The Effect of ERECTA Transcript and Splicing on the Regulation of Stomatal Density

While it cannot be confirmed by the haplotype analysis that the QTL on chromosome 2 is caused by *ER*, the patterns of the haplotypes for the QTL with and without Ler-0 haplotype at *ER* suggest that it is the same gene. This is supported by the ANOVA showing no significant difference of QTL analysis effect on estimated SDs of the ecotypes. It was therefore decided to investigate the role of ER in SD and SD response to [CO2], especially as there is evidence of ER playing a role in the [CO2] response pathway elsewhere (Engineer et al, 2014; Caine, 2016)

### ERECTA is a pleiotropic gene

ERECTA (ER) is a member of a small family of receptor kinases found in many plant species including in Arabidopsis. It is encoded by the most highly expressed member of the *ERECTA* gene family, which also includes *ERECTA-like 1* (*ERL1*) and *ERECTA-like 2* (*ERL2*) (Shpak, 2005). Loss of function *er* plants, for example the Landsberg erecta ecotype (L*er*-0), have many developmental deformations including changes to cotyledon development (Chen et al, 2014), compact inflorescence, shortened siliques (Redei, 1992) and an over-abundance of stomata that are improperly spaced with many smaller cells in-between (Shpak et al, 2005).

In the L*er*-0 and in Col-0 loss of function mutant *er* knock-out mutants, Water Use Efficiency (WUE) is greatly reduced. This is associated with an increase in stomatal density (Masle et al, 2005). In *ER* over expression lines of Arabidopsis, tomato and rice, stomatal density was decreased, epidermal cell sizes are larger, WUE is increased and heat tolerance is increased (Shen et al, 2015). Similarly, in poplar, over expression of *ER* led to increased WUE through reduced stomatal density (Xing et al, 2011). This indicates that over expression of *ER*, which leads to decreased stomatal density could be a good method of creating more WUE plants.

*ER* was the only known stomatal development gene to be identified through QTL analysis in Chapter 3 further indicating its importance in regulating SD. Understanding control of ER gene expression and receptor activity is therefore important in understanding WUE and plant development.

### QTLs and ERECTA

*ER* has been identified in numerous previous QTL studies and eQTLs in plants as being associated with particular traits. For examples, it has been found to be associated with plant height (Kearsey et al, 2003), WUE (Masle et al, 2005), seed size (Moore et al, 2013) and insect resistance ([Kliebenstein](javascript:searchAuthor('Kliebenstein,%20D.')) et al, 2002). ER is proposed to be a key controller of development (reviewed in Schpak, 2013) and it is therefore not surprising that the associated gene should be identified in multiple QTL analyses, particularly those which incorporate L*er*-0 into the germplasm because it carries a non-functional allele that results in drastically different morphology (Torii et al, 1996). *ER* was also identified as being a potential regulator of stomatal density by the QTL analysis carried out in this study (Chapter 3.3.3) both with and without the Ler-0 haplotype at *ER* removed. This confirmed that sequence or expression diversity in *ER* is important in controlling stomatal traits even background in lines other than L*er*-0.

### Regulation of ERECTA through introns

The ER protein is encoded by a complex gene with 26 introns and 27 exons. The regulation of ER transcription and translation is relatively unknown. It has however been shown that ER requires introns in the genetic sequence in order to be translated into protein. Constructs that lack introns are poorly translated, leading to a loss of function *er* phenotype with ER protein accumulation 500–900 times less in comparison to constructs with introns (Karve et al, 2011). It is believed that the introns act to stabilise the mRNA for translation. Not all introns are required, but multiple introns are required in specific locations for ER protein expression. Specifically, it appears that introns in the leucine rich repeat area of the gene are needed, as without them there is no protein expression (See Figure 4.4.1). In addition the loss of function *er* phenotype is not rescued by expression of constructs of the *ER* gene with introns only in the kinase domains (Karve et al, 2011). This leads to the possibility that changes to the intronic sequences, for example SNPs and deletions, could have a large effect on ER protein level and therefore stomatal phenotypes.

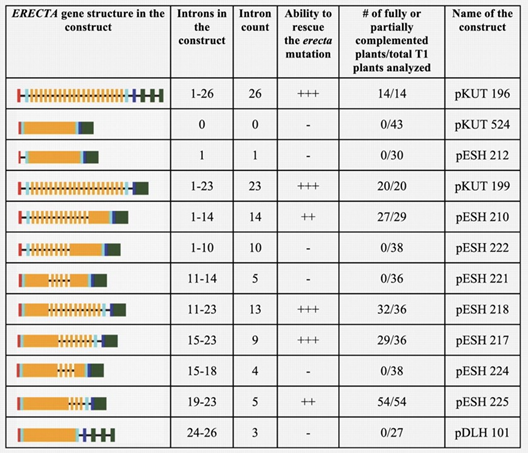


Figure 4.4.1 Introns play a role in translation of ERECTA. Intronic sequences are required across the gene and between specific exons for proper gene and protein expression. In this figure from Karve et al (2011), how the inclusion of different introns affects ER protein level is shown. In the gene structure column, exons are represented as boxes and introns as lines. The red boxes indicate signal sequences, blue indicates segments with paired cysteines, orange represents Leucine Rich Repeat regions, and dark blue represents transmembrane domains and dark green represents a kinase domain. Whether introns restore Col-0 er loss of function mutant ER protein levels are shown by the ability to rescue column. +++ means complete complementation, ++ partial complementation, and – no complementation. (from Karve et al, 2011)

### Regulation of ERECTA through epigenetics

Control of *ER* expression has been suggested to be under epigenetic control. Cytosine methylation and histone modification are epigenetic modifications which can affect gene expression (Pecinka and Lin, 2014). Wang et al (2015) showed that in the Col-0 *ibm1* methylation mutant, hyper methylation of the *ER* histones and cytosines results in an increase in stomatal density similar to that of Ler-0. They show that this is accompanied by a reduction in *ER* transcript levels (Wang et al, 2016). While transcript levels do not necessarily result in increased protein level, they are often correlated with this (Liu et al, 2016). The finding of reduced ER expression with increased methylation in line with what would be expected of an *ER* knockout. Hyper-methylated *ER* shows low *ER* expression, leading to a reduced stomatal density. The experiments of Wang et al (2016) suggest that methylation and epigenetics have a role in regulation of *ER*.

### ERECTA and [CO2]

ER is a central part of the genetic pathway that controls stomatal initiation (See Chapter 1 section 1.5). As genes products such as EPF2 and Stomagen exert their effect through interactions with ER, and these are further affected by environmental signals, loss of ER might be expected to result in a loss of responsiveness to some signals. This loss of response can be seen in response to [CO2]. Engineer et al (2014), found that *er* mutants grown in elevated [CO2] show an increase in SD in comparison to the wild type background and Caine (2016) found a non-response phenotype in *er* mutants. ER can therefore be expected to play an important role in the SD response to [CO2].

### Aims

The aims of this chapter were to

1. Relate the *ER* phenotype of a subset of the MAGIC Arabidopsis parental ecotypes to the QTL model results. To determine further how ER is responsible for the QTL on chromosome 2 which is responsible for the control of SD. Particularly, differences between Can-0 and Wil-2, the ecotypes with the highest and lowest SD phenotype in the haplotype model for QTL2 were hypothesised to exist.
2. To determine how *ER* expression and/or splicing changes in different environmental conditions.

## Methods

### Plant Material

Seed of Can-0, Wu-0, Hi-0 and Wil-2 were donated by Dr Paula Kover (Bath University) and grown to seed to increase stock volume. Col-0 was from grown to seed and collected from lab stock originally obtained from NASC.

### Plant Growth

The ecotypes of Can-0, Wu-0, Hi-0 and Wil-2 were grown as described in Chapter 2.2.2 in 20°C, 70% humidity, 200 μmol m−2 s−1 light, and either 450ppm (ambient) or 1000ppm (elevated) [CO2] in a Sanyo-Gallenkamp SGC970/C/HQI cabinet.

Plants for short term response to [CO2] experiments (Figure 5.3.10) were grown in F2+S Levington, Everris Professional compost mixed with the insecticide (NilNat) after stratification at 4°C for 2-3 days. They had an 11hr photoperiod of 250 μmol m−2 s−1, 20°C, 65% humidity and a night period of 16°C at 65% humidity. Plants were initally germinated and grown in ambient [CO2] in a Sanyo-Gallenkamp SGC970/P/PLL with an additive CO2 injection to ensure 450ppm. Plants were grown for 14 days in 450ppm and treatment plants were placed in either 200ppm [CO2] for 6 hours (subambient) or 1000ppm [CO2] for 6 hours (Elevated [CO2]). Treatment conditions of 200ppm were achieved in a Conviron BDR 16 cabinet with a soda-lime [CO2] scrub and an additive CO2 injection to achieve 200ppm. Elevated [CO2] conditions were achieved by additive CO2 injection to the Conviron BDR 16 cabinet. All CO2 was supplied from CO2 cylinders (BOC, UK).

Plants used in experiments to investigate short term response to light were grown in 50 μmol m−2 s−1 (low light) for 10 days, and treatment plants were transferred to 250 μmol m−2 s−1 (high light)for 6 hrs. They were grown in 12 hour day, 65% relative humidity and 22°C constant temperature for in a Sanyo-Gallenkamp SGC2352/FM in F2+S Levington, Everris Professional soil mixed with the insecticide (NilNat) after stratification at 4°C for 2-3 days.

Plants for RNA extractions were collected in Eppendorf tubes with a single ball bearing and frozen in liquid nitrogen. They were ground to powder in Qiagen TissueLyser II and kept frozen at -80°C until RNA extraction.

### RNA Extraction

Seedling RNA extractions were carried out as described in Chapter 4.2.6. RNA extractions for short term response experiments (Section 5.3.8) were carried out by Jordan Brown (University of Sheffield) and Dr Nicholas Zoulias (University of Sheffield)

### cDNA synthesis

cDNA synthesis was carried out as described in Chapter 4.2.7

### qPCR

qPCR was carried out in the same manner as chapter 4.2.8. Oligonucleotide primers for qPCR of *ERECTA* were designed using PrimerBlast and synthesised by Sigma Aldrich and can be seen in Table 5.2‑1 along with ubiquitin control primers.

Table 5.2‑1 List of primers used in qPCR and their respective exons targeted

|  |  |  |
| --- | --- | --- |
| Primer Name | *ER* Exon Target | Sequence |
| Er\_ex26+27F | 26+27 | GGTAGCCTCTGGGATCTTCTTC |
| Er\_ex26+27R | 26+27 | TGTGCTGCACCATATGCTATC |
| ERECTAex1-2For | 1+2 | TGGGTTTCTCTTCTGCTTGAG |
| ERECTAex1-2Rev | 1+2 | AGAAGGTGAAGTTGTCCAGTCA |
| ERECTAex17-18For | 17+18 | GGTTGAGCTATCTCGTATCGGT |
| ERECTAex17-18Rev | 17+18 | TCACCAAGGGAAGAAGGAATG |
| UBC21F | * UBC21 control | GAATGCTTGGAGTCCTGCTTG |
| UBC21R | * UBC21 control | CTCAGGATGAGCCATCAATGC |

### Exon Usage analysis

Exon usage analysis was carried out using DEX-Seq in R in the same manner as chapter 4.2.13, except RNA-seq data for Hi-0 and Wil-2 from Gan et al, 2011 were also incorporated.

### Statistics

All statistical analyses were carried out using the statistical programming language R. Two-way ANOVA was used to investigate the effect of ecotype and [CO2] on ER gene expression. One way ANOVA was used to investigate the relationship between *ER* expression and ecotype. Pearson’s correlation coefficient was used to investigate the correlation between stomatal density and *ER* expression. Tukey’s test was used as post hoc test for all ANOVAs. Student’s T-test was used to investigate the difference between *ER* expression in short term transfer experiments. Statistical Significance of relationship or effect was accepted at P<0.05.

Graphics were also produced in R using ggplot2 or base plot except for the graphs demonstrating exon expression, which were produced using the Bioconductor R package DexSeq.

## Results

### *ER* sequence in 19 parental lines

Differences in the sequences between Arabidopsis ecotypes can result in differential gene function and gene expression. The genome sequences of the 19 parental ecotypes of the MAGIC lines were therefore viewed and compared on the 1001 genomes browser (http://signal.salk.edu/atg1001/3.0/gebrowser.php) at the *ER* locus (Figure 5.3.1 and Figure 5.3.2). The 19 parental genomes show considerable diversity at the *ER* locus both within the exons and in the non-coding introns and untranslated regions. L*er*-0 has been included in Figure 5.3.1 and Figure 5.3.2, however, the L*er*-0 ER haplotype contains the amino acid change of leucine to isoleucine at amino acid 750 in exon 26 which makes ER non-functional. Exon 26 is one of the domains required for protein kinase activity (Torii et al, 1995). The leucine to isoleucine change is a result of a single SNP of A to T, the position of which is highlighted in Figure 5.3.1 and Figure 5.3.2. This change highlighted ER’s importance in the control of stomatal density in Chapter 3, however this change was disregarded in the analysis of the other ecotype’s ER phenotype as the SNP is not the cause of natural variation in stomatal density.

There are other non-synonymous amino acid changes that can be identified in the ER gene sequence in Figure 5.3.1. In Oy-0, a non-synonymous change in exon 26 was found, which is also within a kinase domain. However, in the QTL model, the Oy-0 haplotype has an SD in approximately the middle of the 19 parental lines stomatal density in the QTL2 haplotype model (See Chapter 3.3.6) and appears unaffected by this SNP. Can-0 and Bur-0 contain non-synonymous mutations in exon 1 and exon 2, however neither of their haplotypes are at the extremes in the SD phenotype haplotype model.

Some introns are noticeable for containing more SNPs than others, for example intron 1 has 10 possible SNPs across ecotypes, while others appear to be more conserved, for example intron 5 which has no SNPs. Despite there being considerable variation in *ER* sequence, there is no SNP that appears significant in the Hi-0 *ER* gene (Hi-0 has the highest stomatal density in the haplotype model) or the Wil-2 *ER* gene (The lowest stomatal density in the haplotype model). Therefore there is no obvious SNP in or around the *ER* locus in Hi-0 or Wil-2 to suggest the cause of the different stomata density phenotypes of the ecotypes.

|  |
| --- |
|  |
| Figure 5.3.1 Comparison of SNPs present in the ERECTA gene and untranslated regions in the 19 parental ecotypes. Coloured lines with letters indicate SNP in the sequence of the genome (R means purine and M pyramidine representing uncertainty in the sequencing), and grey boxes indicate deletions. Gene model is shown above with pink boxes representing exons and black lines representing introns. Of particular note is the SNP in exon 26 which leads to the non-synonymous mutant that renders the L*er*-0 haplotype non-functional as a kinase (circled in black). Image from <http://signal.salk.edu/atg1001/3.0/gebrowser.php>. Derived from sequence data from by Gan et al, 2011 |

|  |
| --- |
| Figure 5.3.2 Amino Acid changes in ER caused by non-synonymous SNPs. Red lines indicate non-synonymous changes and the letter the encoded amino acid is indicated alongside. Green lines indicate location of synonymous SNP changes and the amino acid encoded. Boxes indicate SNP change is in an exon. Grey boxes indicate deletions. The single amino acid change from leucine to isoleucine in Ler-0 in exon 26 is highlighted. Image from <http://signal.salk.edu/atg1001/3.0/gebrowser.php> |

### Epigenetic regulation of *ERECTA* gene expression

Wang et al (2016) showed that increased epigenetic regulation, from increased methylation, resulted in increased stomatal density in Col-0 mutants. The gene *IBM1*, a histone demethylase, results in increased gene body and histone methylation when it is knocked out (Miura et al, 2009). Knock-outs of *imb1* show increased SD because of decreased expression of *ER* transcript*.* This suggests that increased SD may be accompanied by increased methylation in ecotypes. The gene body methylation from data by Kawakatsu et al (2016) of the 19 parental lines was therefore examined (Figure 5.3.3).

Data for gene body methylation in Can-0, which has one of the lowest stomatal densities of the ecotypes, shows one of the lowest methylations of the 19 parental lines (Figure 5.3.3), in line with the findings of Wang et al. However the low methylation does not fit with the QTL2 haplotype model of ecotype phenotypes (Chapter 3.3.6), which would suggest Can-0 should have a mid-level of methylation leading to a mid-level of *ER* expression and a medium stomatal density in comparison to the other 19 parental lines.

The highest ecotype in the haplotype model, Hi-0, does not stand out as being highly methylated, however it probable that the methylation of specific residues is responsible for reduction in expression. Wil-2, the lowest in the haplotype model (Chapter 3.3.6), similarly does not show any obvious difference in methylation in comparison to the other 19 ecotypes. Thus, from these analyses there does not appear to be a correlation between methylations and in the *ER* gene and SD. In addition, the parental lines plants used for the QTL analysis and SD phenotyping will have been grown under different conditions than the plants used for the experiments by Kawakatsu et al (2016). The SD phenotype data from the haplotype model may therefore not match the methylation status by Kawakatsu et al, (2016), as epigenetic changes vary with environmental changes. Regardless of these differences, gene body methylation of *ER* does not show strong evidence of influencing stomatal density in this analysis and is not considered further

|  |
| --- |
|  |
| Figure 5.3.3 Methylated sites of *ERECTA* in 19 Ecotypes of Arabidopsis. Green lines indicate methylated site found by MethylC-seq bisulfite sequencing (by Kawakatsu et al, 2016) in the ER gene in 19 ecotypes. Gene model of ER can be seen at the top with green indicating exon, black indicating intron and red indicating UTR. Variation can be seen along the gene with Can-0 showing the lowest methylation. Image from <http://neomorph.salk.edu/1001.aj.php>. |

### *ERECTA* gene expression levels in 19 parental ecotypes

Changes to gene expression can have a strong effect on phenotype. Gene expression changes can be temporal, spatial or overall expression level changes. RNA-seq can be used to examine overall expression level changes between ecotypes. Published RNA seq data of the 19 parental lines from Gan et al (2011) was therefore examined for evidence of expression level changes.

|  |
| --- |
|  |
| Figure 5.3.4 ER expression varies between ecotypes. Reads Per Kilobase of transcript per Million mapped reads (RPKM) of ER for 19 ecotypes of Arabidopsis. Data from RNA-seq by Gan et al, (2011). Dots indicate mean, error bars indicate standard error, letters and colours indicate statistically similar groups (ANOVA P=0.002; Tukeys Post Hoc, P<0.05; N=2) |

The level of expression of *ER* varies between ecotypes in the RNA-seq experiment carried out by Gan et al, (2011). The reads were analysed by Gan et al (2014) to find Reads Per Kilobase of transcript per Million mapped reads (RPKM), which takes into account all of the reads mapped to all exons across the whole gene and standardises for overall number of reads for the dataset. This mode of analysis is carried out to remove bias, however several studies have shown that it can introduce bias and is therefore not the best method to identify differential expression between datasets (Bullard et al, 2010; Wagner et al, 2012).Despite this, it is the measurement offered by Gan et al and is still in wide use in the bioinformatics community and RPKM of *ER* was examined as a preliminary examination of *ER* gene expression.

There is a significant difference in *ER* expression level between ecotypes (ANOVA, P=0.0006) with Wu-0 showing the highest expression and Bur-0 the lowest expression of *ER*. The majority of the lines, however, are statistically similar to each other which may be a result of the low number of replicates for each line (N=2). In Figure 5.3.4 it can be seen that Wu-0, Zu-0 and Ws-0 *ER* expression levels are significantly different to Can-0 and Bur-0, but that overall there is more similarity than difference.

Regardless of haplotype model, the relationship between *ER* expression level and stomatal density of the parental ecotypes were examined. There was no correlation between ecotype stomatal density and ecotype RPKM (R=0.262, P=0.27). However it should also be noted that as the QTL on chromosome 2 has an estimated effect of 10% on the observed stomatal densities in MAGIC lines (or 44% when Ler-0 is included) it is highly likely that combinations of genetic variation at other loci will be having a stronger overall effect on SD, except in Ler-0 where the phenotype is dominant. This data therefore suggests that overall expression level of the whole *ER* gene does not affect stomatal density but that other factors may also influence this.

|  |
| --- |
|  |
| Figure 5.3.5 The relationship between ER transcript expression and stomatal density. Scatter plot shows Reads Per Kilobase of transcript per Million mapped reads (RPKM) of ER and stomatal density per mm2 for each of the 19 parental lines of MAGIC Arabidopsis. There is no significant correlation between RPKM of ER and stomatal density (Pearson’s Correlation, P=0.27, N=2). |

### Expression of *ER* Exons

Analysis of the exon expression data of Gan et al, (2011) in the online gbrowse for the 19 parental ecotypes of MAGIC Arabidopsis (), indicated that some regions of the gene are more highly expressed than others. It can be seen that the kinase encoding domains (exons 24-27) have high expression in comparison to the other exons (Figure 5.3.6). The exons towards the 3’ end of the *ER* gene (where the kinase domains are encoded) and middle of the transcript appear to be more highly expressed. However this bias towards the 3’ end could have arisen as an artefact of the cDNA synthesis for RNA-seq, as annealing of random primers on long genes can result in 3’ bias (see discussion for further details).

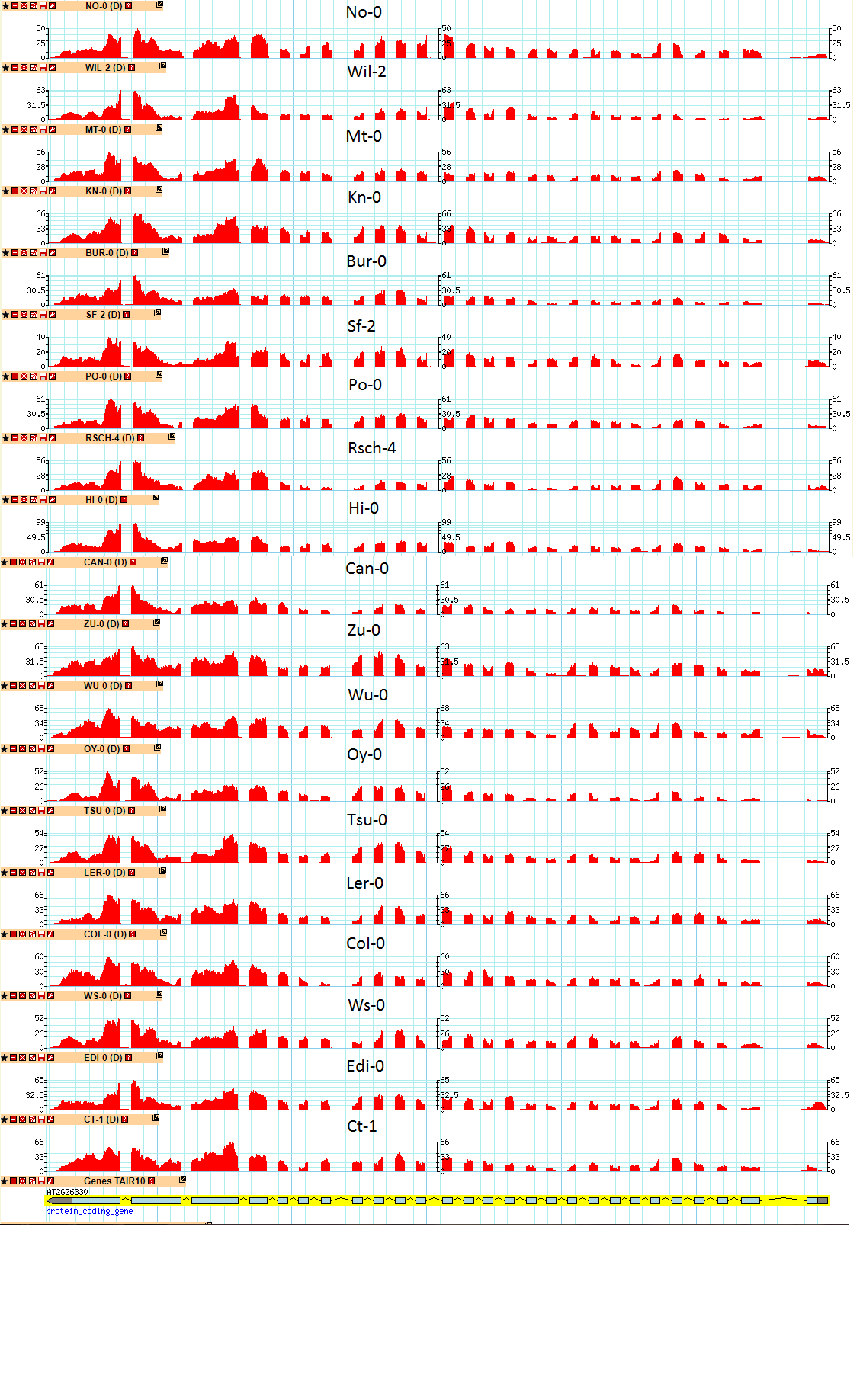


Figure 5.3.6 RNA-seq suggests higher levels of transcription of ER at the 3’ end of the ERECTA gene. Much higher alignment to the 3’ end of the RNA can be seen as well as variation in the highest read number for the kinase domains 25+26 (indicated by black box on gene model). RNA-seq read data from Gan et al (2011) and data displayed and analysed in gbrowse (https://gbrowse.inf.ethz.ch/)

Despite the possibility of bias, it can be seen that some parental ecotypes have higher expression levels of the *ER* kinase domains than others. Hi-0 has a peak of 99 sequence reads compared to the next highest of 68 in Wu-0, and Sf-2 has the lowest at 40 reads. Hi-0 is predicted to give the highest stomatal density haplotype in the haplotype model (Chapter 3.3.6) and the expression data of the kinase domains suggests that it may be because of increased expression of the ER kinase domains. Higher *ER* expression in Hi-0 does not fit well with data showing that *er* knock out mutants have a relatively high stomatal density (Torii et al, 1996) or data showing that ER over expressing lines have low stomatal density (Shen et al, 2015). Developing stomata are known to express *ER*, and it could be that higher stomatal numbers result in increased *ER* expression (Wang, 2007) which could explain why this data is difficult to interpret. Hi-0 however is not the parental line with the highest SD, although it is still relatively high. It is therefore not clear whether elevated *ER* expression is due to increased SD however it is a strong hypothesis.

The data shown in was not normalised to account for total number of reads, which could account for the differences in which ecotype has the highest expression in comparison to the data presented in . Further research was required to look at these differences. For this reason, two experimental approaches were chosen to investigate. RT-qPCR was used to look at individual *ER* exon expression across ecotypes and [CO2] treatments, and a bioinformatics splicing analysis was used to look for evidence that differences in exon expression being due to alternative splicing.

### Expression of ER exons between [CO2] treatments

The above use of published RNA-seq data provides valuable insights into the effect of *ER expression* on SD, however it does not look at the effect of [CO2] treatments, and is based on only 2 samples. RT-qPCR on 4 ecotypes was therefore carried out to study *ER* expression across ecotypes and treatments.

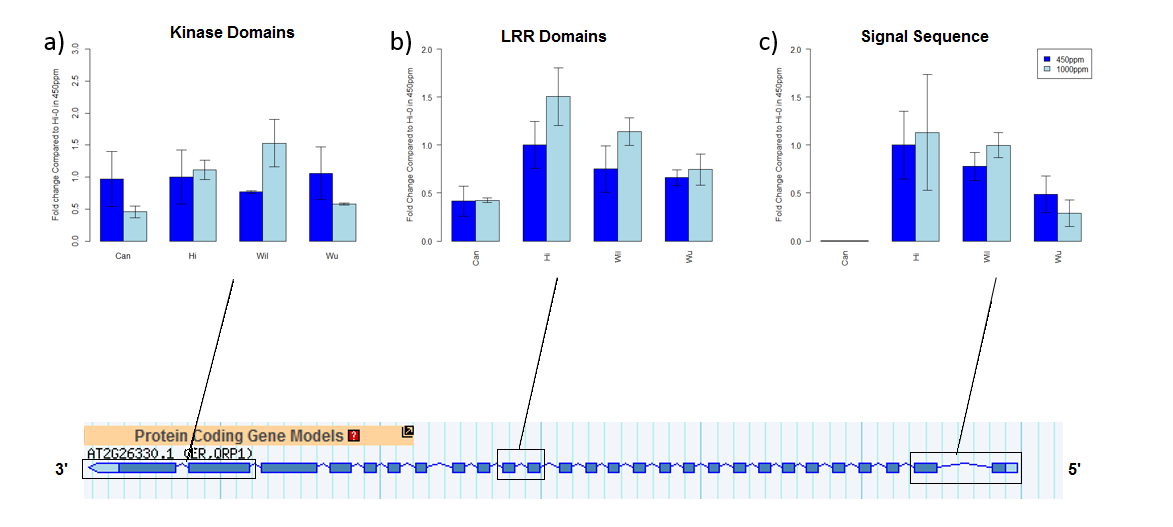


Figure 5.3.7 Relative expression of ER exons in 4 ecotypes of arabidopsis. Can-0, Hi-0, Wil-2 and Wu-0 ER expression in 450ppm (dark blue) and 1000ppm (light blue) shown. A) Shows exons 25 and 26, the kinase domains B) shows exons 17 and 18, Leucine Rich Repeats C) shows exons 1+2, a signal sequence. All expression was normalised to Hi-0 in 450ppm in each condition. Location of exons on gene model can be seen below. \* indicates significantly different to other ecotypes (P<0.05, N=3)

Analysis of RT-qPCR data showed that different ecotypes may express different ER exons at different levels relative to each other, and that growth in different [CO2] levels resulted in different levels of expression of *ER* exons (Figure 5.3.7).

Analysis of expression levels of ER exons 17+18 showed significant differences between both treatments and ecotypes. Within ecotypes there was no significant difference between ambient and elevated [CO2] *ER* relative expression levels, there was only a significant difference over all ecotypes, with elevated [CO2] consistently inducing higher ER exon 17+18 expression levels than ambient [CO2] (P=0.04). Between ecotypes, Can-0 had significantly lower expression than the other ecotypes (ANOVA P<0.05), the other ecotypes had similar relative levels of ER exons 17+18.

For ER exons 1+2 relative expression levels, there was no effect of [CO2] treatment on expression level. Ecotype, however, showed a significant effect on ER exons 1+2 expression level, with Can-0 exon 1+2 expression level showing close to zero relative expression and showing significant difference to the other ecotypes. While this could be evidence of alternative splicing of ER through the loss of the expression of the exon 1+2 secretory and signalling domains, it is also possible that it is a fault with the oglionucleotide primers for ER exon 1+2. Oglionucleotide sequences of primers were designed based on the sequences of the 19 parental genomes by Gan et al (2011). It is possible that a SNP is present in Can-0 that was not found by the Gan et al (2011) sequences and other primers were not tried.

There were no significant differences in expression of *ER* exons 26+27 between [CO2] treatments or ecotypes (ANOVA, P>0.05). Exons 26+27 encode the kinase domains of ER and the similar level of expression across ecotypes and treatments could show the importance of the kinase domains in all lines and conditions. This area of the gene sequence encodes part of the kinase domains in which Ler-0 has a SNP which affects SD (Torii et al, 1996) and this data suggests proper *ER* kinase domain expression may be important in ecotypes.

This data therefore suggests there is some variation in expression level of ER exons, with Can-0 having a lower expression level of exon 1+2 and exons 17+18 compared to Hi-0, Wil-2 and Wu-0 (P<0.05). In addition, ER exons 17 and 18 appear to be more highly expressed in elevated [CO2] suggesting exons 17+18 have increased levels of expression in elevated [CO2].

### Expression of ER correlates with SD

To study the effect changes in *ER* expression level had on stomatal density, stomatal density and *ER* expression were examined across 4 ecotypes.

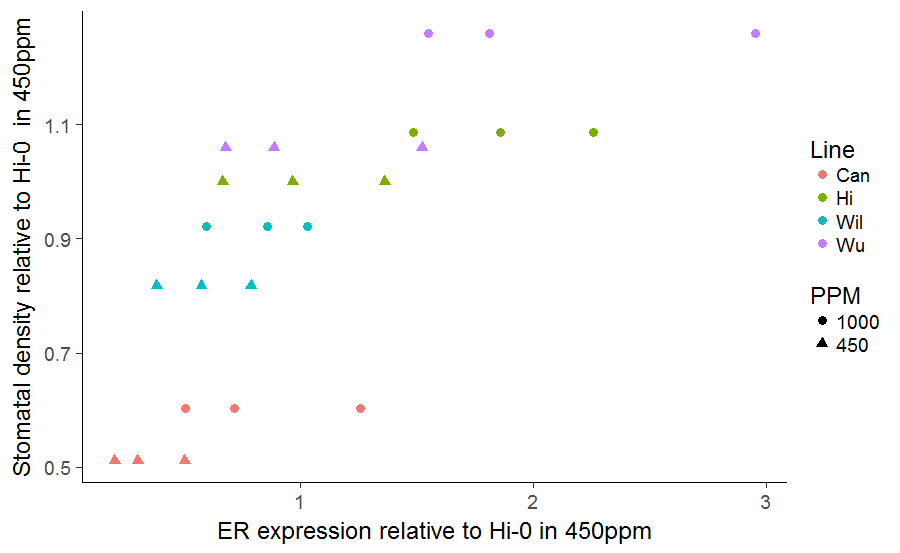


Figure 5.3.8 A correlation between ERECTA expression and stomatal density. Relative expression of ER exons 17+18 from RT-qPCR data in 4 ecotypes (Can-0, Hi-0, Wu-0 and Wil-2) in 450ppm and 1000ppm [CO2] and the stomatal density of the ecotypes (relatives to Hi-0 in 450ppm) shown. Colours represent ecotypes and shape shows [CO2] concentration (circle represents 1000ppm and triangle represents 450ppm). There is statistically significant correlation between ER expression and stomatal density (Pearson’s Correlation, R=0.72, P<0.001), N=3)

The data in Figure 5.3.8 suggests evidence of correlation between *ER* expression levels from RT-qPCR data of exon 17 and 18 and the stomatal density of Can-0, Hi-0, Wu-0 and Wil-2. Ecotypes with higher *ER* exon 17+18 expression levels show higher stomatal density (Pearson’s correlation. R2=0.72 P<0.0001). This is not what would be expected behaviour of Arabidopsis plants responding to an increase in *ER* transcript expression level based on information from *ER* knock-out lines. Lines with less *ER* transcript expression usually show increased stomatal density (Torii et al, 1996) whereas here lower expression level appears to cause a decrease in SD.

It is could be that the correlation between RT-qPCR expression levels of ER and SD in ecotypes is actually due to the increase in developing stomata in different conditions and in different ecotypes, rather than exposing the role of *ER* in affecting stomatal density. In developing stomata, there is an upregulation of *ER* expression level (Wang et al, 2007) and if there are more stomata from an increased stomatal density, ER expression level appears upregulated across the whole leaf. This would mean it is stomatal density affecting *ER* expression level rather than ER expression level affecting stomatal density.

There was no significant correlation between exons 26+27 and SD or between exon 1+2 and SD however (Pearson’s P>0.05). This could suggest that it is certain exons which are expressed differently in developing stomata and that changes to the expression levels of other exons is independent of stomata numbers. Studying alternative splicing of ER across different conditions and ecotype lines could therefore provide evidence or explanations as to why some exons are correlated with SD while others appear unaffected.

### Alternative Splicing in ERECTA

While ER has never been shown to have alternative splicing, alternative gene splicing has been shown to have an effect phenotypes such as flowering (Eckardt et al, 2002) and as such is worth exploring when looking for other causes of phenotypes. Alternative splicing of *ER* was therefore examined in Hi-0, Can-0 and Wil-2 using the Gan et al (2011), RNA-seq data set using a bioinformatics approach. The Gan et al, (2011) dataset was analysed using RSubRead and DexSeq packages.

|  |
| --- |
| a) |
| b) |
| Figure 5.3.9 Exon expression of ER in 3 arabdopsis ecotypes. Expression (raw read counts) of Can and Hi (a) and Hi and Wil (b) for each exon and the corresponding gene model of ER below (exons in grey). Differences in exon expression can be seen however these are not significant (GLM, P>0.05, N=2) |

There was no significant difference in the levels of expression of ER exons between Hi-0 and Can-0 when adjusted for biological and technical variation (P=0.99). To account for biological and technical variation, dispersion estimates were made. First, different sequencing depths were accounted for by estimating size factors. Next, dispersion estimation is carried out to look at variation between replicates and conditions. In short this methodology examines how genes vary in their expression level in their repetitions and looks at how other genes of similar expression level vary. A curve is fitted to the data and gene expression is shrunk to the curve to account for the dispersion. The dispersion estimate thereby reduces the effect of technical and biological noise.

The difference in expression between Hi-0 and Wil-2 exons, the two most different ecotypes in the haplotype analysis, also do not show any significant difference once biological and technical variation has been adjusted for (P>0.05). However, before adjustment, *ER* exons 3, 4, 5, 15 and 16 show significant differences in expression, and is possible that the low number of samples (N=2) or the manner in which the cDNA was synthesised (see section ) is concealing significant results. Further research into alternative splicing of *ER* transcripts in different ecotypes, through RT-PCR of specific regions of *ER* may identify alternative products. Further RNA-seq work could be used to identify alternative splicing in this transcript. However, the issues associated with creating cDNA of long genes, like *ER*, may remain.

### Changes to *ER* expression level in response to short term environmental change

To study whether *ER* expression level is affected by changing the steady state environmental conditions of the growth environment, Arabidopsis plants of the ecotype Col-0 were moved between [CO2] or light treatments and then examined. Response of *ER* exons 17+18 expression level to elevated (1000ppm) [CO2] and sub-ambient (200ppm) [CO2]and high light were examined.

Plants that were moved between ambient (450ppm CO2)and elevated (1000ppm CO2) for 6 hours showed a significant reduction in *ER* expression level (T-test, P=0.001) (Figure 5.3.10A). This is in line with the hypothesis that *ER* expression levels correlate with stomatal density and development. Development of stomata is down regulated in Col-0 in elevated [CO2] and so *ER* expression level is reduced. It is worth noting that at 6 hours of exposure to elevated [CO2] there would still be developing stomata, but further progression through the stomatal lineage may be down-regulated leading to changes in expression level. The decrease in ER expression after 6 hours of elevated [CO2] exposure suggests there is quick feedback between the [CO2] response and *ER* expression level.

a)

b)

|  |  |
| --- | --- |
| \* | \* |
| \*  c) |  |
| Figure 5.3.10 Changes in ER expression levels in different short term [CO2] and light treatments. Change in relative expression of ER exon 17+18 after 6 hours of treatment of A) elevated [CO2] (1000ppm), B) subambient [CO2] (200ppm)), and C) high light (250 μmol m−2 s−1). LL is 50 μmol m−2 s−1, and Control indicates plants grown continuously at 450ppm [CO2] and high light. \* indicates significantly different expression levels (T-test, P<0.05, N=3). | |

There was no significant difference between plants moved to sub-ambient [CO2] for 6 hours and those kept at ambient [CO2] (Figure 5.3.10B). This does not fit with the hypothesis that stomata affect ER expression, as there is an increase in SD in elevated [CO2] in Col-0. That there is no significant difference in expression in subambient [CO2] could suggest that the down-regulation of stomatal development takes longer to take effect in sub-ambient [CO2] than elevated [CO2]. However, no change in ER expression in subambient [CO2] could provide evidence that the change in *ER* is due to the environment modulating *ER* expression level rather than the environment changing stomatal initiation and therefore ER expression level. Future research to investigate expression levels of further genes in the stomatal development pathway would be required to confirm how long subambient [CO2] takes to cause other transcript levels to change and therefore change stomatal initiation.

Finally, the response of plants moved from low light (50 μmol m−2 s−1) to high light (250 μmol m−2 s−1) was examined. Plants moved from low light to high light exhibit higher SD than those grown in low light (Gay and Hurd, 1975). In this RT-qPCR, plants showed increased *ER* expression level in high light compared to those in low light (Figure 5.3.10C). In this case, the significant increase in *ER* expression level in plants moved to high light would suggest that changes to stomatal development affect *ER* expression levels as previously hypothesised.

Overall, the changes in ER expression in different treatments broadly support the hypothesis that changes in ER level are due to changes in developing stomata. The results of *ER* expression following transfer to subambient [CO2] however do not.

## Discussion and Conclusions

### ER Expression and Stomatal Density

There is a strong correlation between stomatal density and *ER* exons 17+18 expression level, with Arabidopsis parental ecotype lines that show relatively high *ER* expression level showing high stomatal density. This would ordinarily seem counterintuitive to the *er* knock-out mutant phenotype, however when the evidence of Wang et al, (2007) is take into account, that when there are increased numbers of developing stomata there is a corresponding increase in *ER* expression as developing stomata express *ER*, it fits with previous data. As it is only exons 17+18 which show this correlation however, it could suggest that it is a particular splice variant that increases with increasing SD.

The increase in *ER* expression level in ecotypes in elevated [CO2] is also in agreement with that found by Watson-Lazowski et al (2016) in their analysis of *Plantago lanceolata* grown in ambient and elevated [CO2]. Plantago shows an increase in stomatal density in elevated [CO2] and there is an accompanying increase in *ER* expression. Similarly in this analysis, plants grown in elevated [CO2] generally had increased stomatal density and increased *ER* expression.

The evidence concerning ER expression level gained from the short-term treatment transfer experiments (Section 5.3.7) largely supported the evidence from the experiments with parental ecotypes grown in different [CO2] levels correlating SD. The change in ER in short term transfer experiments also fits with data from Watson-Lazowski et al (2016), except that of plants exposed to sub-ambient [CO2].

It was observed that increases in light led to an increase in SD and a concomitant increase in *ER* expression level, while increases in [CO2] lead to a decrease in SD (in Col-0) and a decrease in *ER* expression level was observed. These data would suggest that the changes to stomatal development result from changes in these two environmental conditions are initiated quickly, within 6 hours, although developmental effects may not be observed till several days later due to the time taken to form a new stoma.

In contrast, plants transferred to sub-ambient [CO2] did not show a change in *ER* expression level. Chapter 2 showed that Col-0 has an increased SD in sub-ambient [CO2] compared to ambient [CO2], and as such it might be expected that an increase in *ER* expression level would result from the SD increase. The fact this increase does not occur could suggest that either changes to SD and initiation take longer to achieve in sub-ambient [CO2] than in elevated [CO2] or high light treatments, or that a change in *ER* expression levels is not necessary to alter SD, and that *ER* expression level does not always correlate with SD.

mRNA levels do is not always directly correlate with protein levels (Liu et al, 2016) and no change in mRNA expression does not mean there has not been changes in protein expression. To investigate further whether the changes in *ER* expression level between ecotypes and treatments result in changes to ER protein expression, the ER protein levels could be quantified in several ways. There is no available anti-body for ER, instead a reporter such as luciferase can be fused to ER and florescence level detected in the manner used by Karve et al (2011), or a protein tag could be attached to the protein for detection by anti-body as was used by Chen et al, 2015, or peptide levels could be detected by mass spectrometry. For further examination of protein levels in Arabidopsis ecotypes, a similar approach to these would need to be taken.

The Arabidopsis parental ecotypes studied here did show some differences in the levels of expression of specific exons in RT-qPCR experiments, however, there was no consistent relative increase or decrease in Hi-0 or Wil-2, the two most different ecotypes from the haplotype model for the QTL on chromosome 2 (See chapter 3.3.6). RNA-seq data suggested that the higher stomatal density in Hi-0 haplotypes of MAGIC lines could be a result of increased expression of the kinase domains (Section 5.3.7), however RT-qPCR experiments did not support this hypothesis.

Further research would be required to identify if and when expression level of *ER* has an effect on stomatal density. The data presented in this chapter suggested it is more probable that the stomatal density affects *ER* expression level, rather than *ER* expression level affecting stomatal density. Nonetheless, plants lacking *ER* expression or over-expressing *ER* exhibit changes to stomatal density, which suggests that changing *ER* expression level can have an effect on SD. Here it is hypothesised however that it is stomatal density which affects *ER* expression, rather than the other way round.

### Alternative Splicing

The analysis of *ER* transcript splicing in the RNA-seq data set from Gan et al (2011) and between RT-qPCR data of specific exons revealed evidence of alternative splicing of *ER*, but the level of this was largely non-significant between ecotype treatments. There was a significant difference in expression when looking at expression of exons 17+18 of *ER* in elevated [CO2] and between ecotypes (Section 5.3.7) and this could suggest that in different environmental conditions and in different ecotypes, different isoforms of ER transcript are expressed. There were no significant changes in expression of *ER* exons in different [CO2] levels within ecotypes however. This is more likely that different ecotypes of Arabidopsis splice *ER* transcripts in alternative ways than the environmental factors having an effect on *ER* splicing.

While it may appear that some *ER* exons are more highly expressed than others in the RNA-seq data set (Gan et al, 2011), it is possible that this is an artefact of the cDNA synthesis process. It has been shown that cDNA synthesis can create alternative splicing like events through ‘template switching’ in which the synthesising DNA disassociates from the DNA template and re-associates with another DNA template elsewhere (Roy and Imirimia, 2008). This particularly happens in repetitive sequence regions and in *ER*, the leucine rich repeat encoding exons are repetitive. Template switching can create what appears to be exon skipping which is potentially what may have happened during cDNA synthesis here.

It has also been shown that there can particularly be a 3’ end bias in cDNA synthesis that should be accounted for in studies. This is because, generally, more primers anneal to the 3’ end of a transcript and in long genes DNA polymerase can ‘fall off’ before the end of the transcript resulting in shortened transcripts which align to the 3’ end (Roberts et al, 2011). Furthermore, specific structural features of RNA transcripts, such as hair-pin loops, can cause DNA polymerase to falter resulting in poor synthesis and shortened transcripts (Roberts et al, 2011). While it may therefore appear that there could be differences in *ER* exon expression levels, these could be due to artefacts of the cDNA synthesis technique in both the RNA-seq and the RT-qPCR. A potential solution to this is direct RNA sequencing (Ozsolak et al, 2009) which is available but not commonly carried out.

Despite this, while there is currently no published research on alternative splicing of *ER*, there is ample evidence of plants using alternative splicing to create different isoforms of genes in different environmental conditions. Calixto et al (2018) showed that in cold treatments, alternative splicing is induced in 2,442 gene transcripts compared to the control. The non-coding RNA COOLAIR has been shown to have an effect on flowering time due to differing isoforms in different temperatures being produced causing it to modulate expression of flowering locus C differently (Li et al, 2015). When under salt stress, it has been shown that 49% of all intron containing genes undergo alternative splicing (Ding et al, 2014). These examples illustrate that alternative splicing is a common response to environmental conditions and it is possible that alternative splicing of ER also occurs. It was suggested by Karve et al (2011) that the requirement for the presence of introns in the transcript of ER for translation might be related to splicing. Although these authors did not link translational efficiency to alternative splicing and they failed to find any alternatively spliced *ER* mRNA.

In the analysis presented here there was no clear evidence of alternative splicing of *ER* in Hi-0 or Wil-2. The analysis of Gan et al (2011) RNA-seq data did not show significant differences between these Arabidopsis ecotype lines. The differences in relative expression levels of different *ER* exons between ecotypes in the RT-qPCR data suggested that there could be some alternative splicing occurring. However, the cDNA synthesis carried out for RT-qPCR is likely to have the same synthesis biases, which are discussed above, as in the RNA seq data which makes drawing conclusions difficult.

The evidence examined here was not conclusive in linking *ER* expression levels of splicing to altering the stomatal density, and further research using PCR and gel electrophoresis to find different length transcripts should be carried out to confirm that alternative splicing is present in this gene.

### Conclusion

In conclusion, the work presented here could not associate expression or alternative splicing of *ER* with an effect on stomatal density in Arabidopsis. This study was unable to explain why Hi-0 has an increased density in the haplotype model and Wil-2 a low SD, as neither *ER* expression levels nor splicing are consistently different between ecotypes. There is evidence in the RT-qPCR data that changes in expression of particular exons may occur between ambient and elevated [CO2]which could explain why the QTL on chromosome 2 is not significant in elevated [CO2], but is significant in ambient CO2. There is also evidence that different *ER* exons are expressed at different levels in different [CO2] treatments and ecotypes, suggesting that at least under some conditions alternative splicing of *ER* occurs.

Further research to investigate the level of gene and exon expression of different ecotypes in different environmental conditions would be needed to better understand alternative splicing and expression and to further examine how ER affects stomatal density. In addition to this, *ER* over-expression and knock-out mutant plants could be generated to study the effect of increases or absence of ER expression on different ecotypes. This would allow for better understanding as to why the Hi-0 haplotype in the haplotype model had an increased stomatal density and Wil-2 a decreased SD which is not answered here.

# General discussion

## Introduction

Anthropogenically caused rising atmospheric [CO2] levels are anticipated to impact on plant growth, development and behaviour, due to changes in atmospheric composition and the resulting climatic changes (Zhao et al, 2017; Zhao and Li, 2015). Whilst these changes are occurring over the next 50 years, there is a need to double global agricultural output to meet growing populations. As such, it is important for crops to be developed that are adapted to the changing environment (Tilman et al, 2011). To understand how to future-proof crops, it is necessary to know how plants will change phenotypically in response to changes in [CO2].

Changes to [CO2] have been shown to affect many plant traits, but in particular, they have been shown to affect stomatal densities (SD) (Woodward and Kelly, 1995). SD is responsible for establishing the gas exchange capacity of the plant (Franks et al, 2009; Dow and Bergman, 2014) and can have an effect on transpiration, thermo-regulation and carbon dioxide uptake (Franks et al, 2015). Throughout this thesis, the response of SD to changing [CO2] levels and the underlying genetics relating to SD have been extensively discussed, and some evidence as to the genetic control of SD have been found.

## The Stomatal Density in 19 Arabidopsis ecotypes

Arabidopsis plants have previously been collected from many geographic locations, and these ecotypes have been demonstrated to possess a variety of adaptations to their environments, including differences in stomatal density. The objective of this study was to associate these phenotypic differences in SD with their underlying genotypic variation(s) by analysing quantitative trait loci (QTLs). However, the generation of robust data from QTL analysis requires the sampling of many genotypes, with adequate experimental replication. Before high-throughput analysis of SD could be carried out, a method of phenotyping had to be developed that would give quick and accurate data. Arabidopsis cotyledons were shown to adjust their SD phenotypes in a similar manner to mature leaves and so were the organ chosen for large scale phenotypic screening. In addition, it was shown that between 15 to 20 measurements of the cotyledons from a particular genotype would give an accurate measure of SD. As such, measurements of SD from 16 cotyledon areas was chosen as the phenotypic screening method to maximise accuracy and efficiency.

Analysis of the 19 parental ecotypes used to create the Arabidopsis MAGIC lines revealed a wide range in their cotyledon SD. There were also differences in the extent and direction of the change in SD in response to [CO2] between ecotypes. Most of the parental ecotypes showed an increase, or no response in SD, following growth at elevated [CO2] in comparison to SD growth at ambient [CO2]. This increase in SD in most ecotypes is contrary to the reduction in SD most regularly described for Arabidopsis (which are most usually studies of the Col-0 ecotype), showing that there is variance in SD response to [CO2]  both within and across species. There was clear evidence that the traits of SD, and the change in SD between [CO2] conditions, were sufficiently different across the 19 parental lines for the derived MAGIC recombinant inbred lines (RILs) to be used in a QTL analysis.

The SD responses to [CO2] suggested that there were potentially two GXE interactions occurring in the lines, as lines generally reduced their SD between sub-ambient and ambient [CO2] and increased their SD between ambient and elevated [CO2]. As the interest of this thesis was to investigate the changes in SD in response to future [CO2] levels, the differences between ambient and elevated [CO2] were chosen for further examination.

Chapter 2 thus confirmed the proposed hypotheses, that there would be a range of SD responses to [CO2] across Arabidopsis ecotypes; that cotyledons can be used as an organ of study for SD; and that the 19 MAGIC parental lines would show enough diversity for QTL analysis in MAGIC lines in Chapter 1 6.1.1.

## QTLs in MAGIC lines

Following the assessment of cotyledon stomatal densities across 19 genotypes from the Arabidopsis stomata lines a QTL analysis was carried out using 451 MAGIC lines and with the kind help of Dr Kover (Bath University). Three QTLs were found to be associated with the regulation of SD, although together they only accounted for a small proportion of the SD variance (31% in 450ppm) and so other genetic loci are likely also involved. Surprisingly, no QTLs were identified that were associated with the difference in SD between [CO2] environments. Whilst no QTLs for the [CO2]-induced difference in SD were found, this does not preclude this trait from genetic regulation, rather it implies that there may be many genes of small effect involved. Only one QTL was discovered that was associated with SD in elevated [CO2] in comparison to three QTLs associated with SD in ambient [CO2]. This strongly implies that the genes responsible for SD variation between lines are affected by [CO2] concentration. QTL1 was found to be significantly associated with stomatal density in both ambient and elevated [CO2]. It explained similar amounts of variance in both [CO2] treatments, suggesting that this locus is associated with SD regulation independently of GXE interactions. QTL2 and QTL5 however were not significantly associated with SD in elevated [CO2], showing that these QTL are affected by [CO2] concentration. The association with QTL2 is extremely likely to be explained by the presence of the gene *ER* at this locus, which has been previously shown to encode a receptor protein that acts to regulate stomatal development. QTL5 was mapped to an 600kb region.  For genes of interest encoded within the  QTL5 locus, corresponding T-DNA insertion mutants were obtained and studied. Further work suggested, through the screening of Arabidopsis T-DNA insertion mutations, that the observed variation in SD could be caused by the starch biosynthesis gene *ADG2*.

Of the hypotheses proposed in Chapter 1 6.1.2, only some were shown to be robust. There was a normal distribution of phenotypes between the MAGIC lines making them suitable for QTL analysis. However, the QTLs identified as being associated with the regulation of SD suggest that the hypothesis that genes such as *CRSP*, *EPF2* and *CA1/CA4* would be found to be controllers of SD should be rejected. No QTLs in the vicinity of these genes were found in this study. In addition to this, no QTL for the difference in SD between [CO2] treatments was found. The QTL analysis results presented here have similarities to that of a prior QTL analysis carried out in Poplar in which controls for SD in ambient [CO2] were found, but none for SD in elevated and no QTLs for the difference in SD were found (Ferris et al, 2001). The role of *ER* in SD has been previously extensively researched and shown to be of importance in regulations of SD (Torii et al, 1996; Schpak, 2004; Shpak et al, 2005; ) and in addition has been shown to be important in SD changes in [CO2] conditions (Caine, 2016; Engineer et al, 2014) and was therefore investigated further to attempt to establish how the role of ER varied between [CO2] levels. The identification of ADG2 as a factor affecting SD was a novel finding that was also decided to be of importance to investigate further.

## The role of starch in SD regulation

The starch biosynthesis gene *ADG2* was found to be associated with SD in both elevated and ambient [CO2], showing an effect regardless of [CO2] treatment, thus the role of starch was investigated further. Additional Arabidopsis T-DNA insertion mutants with starch biosynthesis deficiencies were sourced and their stomatal phenotypes observed. These studies showed that other plants impacted in starch biosynthesis, in addition to *ADG2* mutants, also show an increase in SD. The change in SD in other starch biosynthesis mutants suggests that the starch content or free sugar content of a plant has an effect on SD. Several defects leading to a decrease in starch biosynthesis resulted in enhanced stomatal development but Arabidopsis T-DNA insertion mutant lines that supported an increase in starch concentration, due to improper starch break down or transport, did not show a corresponding reduction in SD. This indicates that a decrease in starch biosynthesis (or increase in sugar levels) results in an increase in SD but that an increase in starch levels does not result in a decrease in SD.

It was originally suggested that the SD increase observed in ADG2 defective mutants could be a response to starvation - that the plant perceives low carbohydrate reserves and produces more stomata to enable it to fix more CO2. However, the *sex1* mutant also perceives itself to be starving at night and it was found that there was no accompanying increase in SD in *sex1*. An excess of sugar has previously been shown to affect SD and cause stomatal clustering, and the starch biosynthesis mutants *adg2-3* and *pgm1* have also been shown to contain an increase in free sugars (Ragel et al, 2013). Starch content has been shown to increase in elevated [CO2] and free sugars have also been shown to be affected in some plant lines (Van Der Kooij et al, 2000). If free sugars can cause an increase in SD, it could be that an increase in free sugars in elevated [CO2] causes the increase in SD observed in some Arabidopsis ecotypes, meaning that the rate of starch synthesis could affect how a plant responds to [CO2] .

Starch has previously been found to be required for the proper aperture responses of stomata, with stomatal opening impaired when there is a reduction in starch content. Stomatal aperture and gas exchange were not been examined in this thesis, however, Azoulay-Shemer et al (2018) showed that *adg2* mutants show reduced aperture response to [CO2]. Thus it appears that a reduced level of starch biosynthesis is associated with an increase stomatal density but a reduction in stomatal opening capacity.

Examination of the ecotypes suggested to have haplotypes that result in increased or decreased SDs did not result in any statistically significant results to confirm that *ADG2* is the gene underlying the association of QTL5 with SD. There were no significant differences in *ADG2* gene sequence, gene expression nor splicing found between the parental ecotypes. Further research could still identify *ADG2* as being the cause of QTL5 though. QTL5 is only responsible for 13% of the control of SD and it is possible that its effect on SD if therefore unable to be picked up through use of ecotypes which have other SD controls. Further experiments to examine this have been suggested in Chapter 4.4.1.

Overall, while it was possible to find some evidence as to how *ADG2* affects stomatal density, it was not possible to fully understand the mechanism of how *ADG2* affected SD. It was hypothesised in Chapter 1 6.1.3 that it would be possible to relate the specific alleles in the parental lines to the QTL analysis and explain why it was a QTL, however this was not the case.

## The role of ER in the control of SD

ERECTA is encoded by a particularly long gene sequence with a predicted 26 introns. In this study the *ER* gene was shown to have complex gene expression pattern that varies between [CO2] treatments and Arabidopsis ecotypes. The transcripts of several *ER* exons were shown to accumulate to relatively different levels, with some exon transcripts levels correlating with SD while other exons showed no correlation. Neither the overall level of ER expression, nor particular exon transcript levels could explain why Hi-0 ER haplotype MAGIC lines had higher SD and the Wil-2 haplotype had a lower SD as indicated by the haplotype analysis model described in Chapter 3. No consistent association patterns could be discerned between *ER* genomic sequences, expression levels or splicing data, and ecotype SD.

The strongest observed correlation was between exon 17+18 transcripts levels and ecotype SD. It was concluded that this correlation potentially arises because developing stomata express relatively high ER levels (Wang et al, 2007) but other ER exon transcript levels did not correlate. Expression levels of of *ER* also appeared to correlate with stomatal development in plants exposed to short term environmental changes. In high light, where SD is increased, there was an up-regulation of *ER* expression level. In elevated [CO2] there was a down-regulation of *ER* expression level in the Col-0 ecotype which correlated with a decrease in SD in elevated [CO2]. However, plant grown at sub-ambient [CO2] did not show a significant change in *ER* expression. Further research into the time taken for stomatal developmental changes to take place in different environmental conditions would help to elucidate whether these changes in *ER* expression are caused by developmental changes, or are a result of an altered number of developing stomata each expressing *ER*.

Overall, the results provide little support for the hypotheses outlined at the beginning of Chapter 1 6.1.4. There was no evidence of how changes to *ER* sequence, expression level or transcript splicing could affect SD, nor how these changes could be related to the Arabidopsis ecotype haplotypes. It is clear that ER is involved in the control of stomatal development and it remains possible that further research could confirm the hypothesis that variation in the *ER* transcript level or nucleotide sequence explains the observed differences in ecotype SD, but the research presented here is not currently able to do so.

## Future directions

While there was not conclusive proof that ADG2 is the gene causing the QTL on chromosome 5, there is still evidence of its effect on SD in chapter 4, and this could be further examined in the future. This further examination could be done through looking at how different haplotypes of ecotypes affect SD of a chosen line, for example transforming *adg2-3* with Wil-2 *ADG2* haplotype. This could conclusively prove whether ADG2 is the cause of this QTL. Suggestions for further possible experiments relating to ADG2 have been discussed in Chapter 4.4.1.

Further genes which regulate the direction and extent of SD changes following growth of the MAGIC population plants at ambient or elevated [CO2] are still unknown. The QTL analysis identified gene loci that regulate SD, but not genes which cause differences in SD response to elevated [CO2]. Further QTL analysis could be carried out to identify genes of interest for SD response to [CO2] but it seems likely from this QTL, and the previous QTL analysis by Ferris et al (2001), that this method of genetic analysis is not an effective way of finding these genes. Instead RNA-seq could perhaps be used to look for differences between lines in [CO2] treatments. It would be particularly advantageous if RNA-seq of MAGIC lines in different [CO2] treatments could be carried out for eQTL analysis (see Chapter 1.7.2). This would provide further information as to the gene expression changes which might exert an effect on the SD phenotype. Micro-arrays of Arabidopsis in different [CO2] treatments have previously been run without success in finding genes associated with SD response to [CO2] (Coupe et al., 2005), however RNA-seq gives more comprehensive information, including information such as alternative splicing (Chen et al 2017).

Of the 3 Arabidopsis QTLs identified here as being associated with stomatal density, QTL2 has been identified as *ER*, QTL5 is suggested to be ADG2, and QTL1 remains unknown. The use of T-DNA insertions was successful in the identification of a gene that causes changes to SD for QTL5, and screening of T-DNA insertions is likely to be successful at finding other genes of interest. Screening of further T-DNA insertion lines in the 600kb of genomic sequence surrounding QTL1 is therefore an important future avenue of research in the search for genes responsible for variation in SD. If T-DNA insertion lines cannot be utilised then a fine-mapping population could be created to more finely map the area under the QTL. Using the haplotype phenotype model derived in Chapter 3, it would be possible to identify the most different parental lines and produce crosses between them. Further to this, it may be possible to examine the reconstructed sequences of the MAGIC lines and find MAGIC lines with desired haplotypes at different QTL, making it possible to study loci without other high effect QTLs affecting the SD phenotype. It may also be possible to produce fine mapping populations from the identified MAGIC lines of interest.

The study of *ER*, and especially its splicing variants, could be further examined to dissect the role of genomic, expression and splicing variation in *ER* in the control of SD, and to establish why it is only significant as a regulator of SD in ambient [CO2]. It would be of interest to establish whether the changes in *ER* expression between ecotypes and treatments are a result of stomatal development changes or a cause of changes to stomatal development. This could be examined by studying the expression of other stomatal development genes such as the transcription regulators *SPCH*, *MUTE* and *FAMA* to examine the timings of their expression and see if changes in *ER* expression are independent.

The reason for *ER* variation derived changes in SD between parental ecotypes was not established and it would be of interest to carry out further analysis into *ER* expression patterns in different ecotypes. Temporal and spatial patterns of *ER* expression, particularly difference between guard cell expression and epidermal pavement cell expression, could be examined. It would also be of interest to further examine the transcript levels of each exon in different environmental conditions. A further examination of how the haplotypes of different parental lines affect responsiveness of SD to different environmental conditions should be carried out. This could be done in a transgenic manner by looking at different ecotype haplotypes in the Col-0 *er* mutant or in L*er*-0.

*adg2* and other starch mutants have been previously been shown to have altered SD and the mechanisms that cause this change are still largely unknown. It would be advantageous for the role of free sugars to be more thoroughly investigated through the measurement of sugars in starch mutants and through the study of other mutants deficient in the proper metabolism or detection of sugars. The expression of starvation genes should also be examined to see if the SD response is related to the starvation response. However, the lack of SD change in the *sex1* mutant suggests starvation is unlikely to be the cause of the SD change in *adg2* mutants. A study of the expression levels of stomatal development genes in response to changes in starch and free sugar changes could provide information as to how SD is changed by changes to starch biosynthesis.

In addition, starch content was only examined in the light period in this thesis. Disturbances to starch break down can have implications in night time starch utilisation and it is possible that changes to this could affect stomata. Starch granule shape has proven to be important in a previous study by Feike et al (2016) in the proper night time break down of starch, leading to starvation if incorrect. Starch granule formation could therefore also be an avenue for further research. Starch biosynthesis and breakdown to sugars are under circadian control and it would be of interest to study if disturbances to circadian rhythms result in changed SD. A more thorough examination of the effects of starch would therefore be of interest to establish the full effects starch metabolism can has on SD.

## Potential applications

Plants created with a reduction in SD have previously been shown to have increased WUE and are therefore more drought tolerant (Doheny-Adams et al, 2012). The reduction in stomatal density can also have other effects however, such as changes to leaf temperature (Doheny-Adams et al, 2012) which are discussed in 1.2.1. A study using the genes identified in the QTL analysis to manipulate SD could therefore be of interest, however, it would be inadvisable to use *ADG2* for SD manipulation. This is due to the likely adverse impact on yield by destabilising starch biosynthesis. However, if identified the downstream components of this sugar signalling may be of use in manipulating SD. If SD is controlled by starch content, it could be beneficial to uncouple the two processes in order to allow better manipulation of SD, for traits such as drought tolerance and improved WUE, without starch content affecting SD. This will make manipulated phenotypes more robust to changes in environment. Understanding how starch and sugar content affect plant environmental responses may also lead to a better understanding of how plants will respond to [CO2] in the future and thus enable better modelling of global plant responses to [CO2] and the accompanying changes in gas exchange capacity.

*ER* has been shown to be a regulator of SD variation between ecotypes, potentially labelling it a good target for manipulation. Other genes in the stomatal development pathway may be less suitable due to their regulation via feedback loops. Over-expression or knock-out of some stomatal development genes such as *SPCH*, *EPF2*, and *STOMAGEN*, results in strong phenotypic changes which can be deleterious (or in the case of *SPCH* or *STOMAGEN* even lethality) and also impact a plants proper response to environmental changes. A plant with reduced stomatal density through changes to *SPCH* or *EPF2* may also experience overheating in warm temperatures due to an inability to lose enough heat through transpiration. A *SPCH* or *EPF2* mutant would not be able to increase its SD as the genes required to increase stomatal initiation have been manipulated. Manipulating *ER* expression may therefore create a plant that is still able to respond to environmental variables, but has a decreased SD and enhanced water use efficiency.

The results of this thesis may therefore provide information on potential applications in plant and crop development.

## Concluding Remarks

From the analyses carried out in this thesis it is clear that the control of stomatal density by [CO2] is under the control of many genes and varies between ecotypes. That control of a stomatal phenotype is complicated is not surprising, given the large number of environmental inputs and developmental changes that interact to control stomatal development and spacing. Whilst no clear mechanism that causes reductions / increases in SD was found, it could be concluded that there are different controls on SD in different [CO2] treatments. Therefore, further analysis and experiments would help to elucidate the manner in which the controls take effect.

# References

Abrash, E.B., Davies, K.A., and Bergmann, D.C. (2011). Generation of Signaling Specificity in Arabidopsis by Spatially Restricted Buffering of Ligand-Receptor Interactions. Plant Cell *23*, 2864-2879.

Adrian, J., Chang, J., Ballenger, C.E., Bargmann, B.O.R., Alassimone, J., Davies, K.A., Lau, O.S., Matos, J.L., Hachez, C., Lanctot, A.*, et al.* (2015). Transcriptome Dynamics of the Stomatal Lineage: Birth, Amplification, and Termination of a Self-Renewing Population. Developmental Cell *33*, 107-118.

Ainsworth, E.A., and Long, S.P. (2005). What have we learned from 15 years of free-air CO2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy. New Phytologist *165*, 351-371.

Ainsworth, E.A., and Rogers, A. (2007). The response of photosynthesis and stomatal conductance to rising CO2 : mechanisms and environmental interactions. Plant Cell and Environment *30*, 258-270.

Akita, K., Hasezawa, S., and Higaki, T. (2013). Breaking of Plant Stomatal One-Cell-Spacing Rule by Sugar Solution Immersion. Plos One *8*, 5.

Allen, L.H. (1992). Free-Air CO-2 Enrichment Field Experiments - An Historical Overview. Critical Reviews in Plant Sciences *11*, 121-134.

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K.M., Cao, J., Chae, E., Dezwaan, T.M., Ding, W.*, et al.* (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell *166*, 481-491.

Amiguet-Vercher, A., Santuari, L., Gonzalez-Guzman, M., Depuydt, S., Rodriguez, P.L., and Hardtke, C.S. (2015). The IBO germination quantitative trait locus encodes a phosphatase 2C-related variant with a nonsynonymous amino acid change that interferes with abscisic acid signaling. New Phytologist *205*, 1076-1082.

Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome Research *22*, 2008-2017.

Andriotis, V.M.E., Pike, M.J., Schwarz, S.L., Rawsthorne, S., Wang, T.L., and Smith, A.M. (2012). Altered Starch Turnover in the Maternal Plant Has Major Effects on Arabidopsis Fruit Growth and Seed Composition. Plant Physiology *160*, 1175-1186.

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

Apple, M.E., Olszyk, D.M., Ormrod, D.P., Lewis, A., Southworth, D., and Tingey, D.T. (2000). Morphology and stomatal function of Douglas fir needles exposed to climate chance: Elevated CO2 and temperature. International Journal of Plant Sciences *161*, 127-132.

Araujo, W.L., Nunes-Nesi, A., Osorio, S., Usadel, B., Fuentes, D., Nagy, R., Balbo, I., Lehmann, M., Studart-Witkowski, C., Tohge, T.*, et al.* (2011). Antisense Inhibition of the Iron-Sulphur Subunit of Succinate Dehydrogenase Enhances Photosynthesis and Growth in Tomato via an Organic Acid-Mediated Effect on Stomatal Aperture. Plant Cell *23*, 600-627.

Augustin, L., Barbante, C., Barnes, P.R.F., Barnola, J.M., Bigler, M., Castellano, E., Cattani, O., Chappellaz, J., DahlJensen, D., Delmonte, B.*, et al.* (2004). Eight glacial cycles from an Antarctic ice core. Nature *429*, 623-628.

Azoulay‐Shemer, T., Schwankl, N., Rog, I., Moshelion, M., and Schroeder, J.I.J.F.l. (2018). Starch biosynthesis by AGP ase, but not starch degradation by BAM 1/3 and SEX 1, is rate‐limiting for CO 2‐regulated stomatal movements under short‐day conditions.

Bac-Molenaar, J.A., Vreugdenhil, D., Granier, C., and Keurentjes, J.J.B. (2015). Genome-wide association mapping of growth dynamics detects time-specific and general quantitative trait loci. Journal of Experimental Botany *66*, 5567-5580.

Bandaranayake, C.K., Koumproglou, R., Wang, X.Y., Wilkes, T., and Kearsey, M.J. (2004). QTL analysis of morphological and developmental traits in the Ler x Cvi population of Arabidopsis thaliana - QTL analysis in Arabidopsis. Euphytica *137*, 361-371.

Bandillo, N., Raghavan, C., Muyco, P.A., Sevilla, M.A.L., Lobina, I.T., Dilla-Ermita, C.J., Tung, C.W., McCouch, S., Thomson, M., Mauleon, R.*, et al.* (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice *6*, 15.

Beck, E., and Ziegler, P. (1989). BIOSYNTHESIS AND DEGRADATION OF STARCH IN HIGHER-PLANTS. Annual Review of Plant Physiology and Plant Molecular Biology *40*, 95-117.

Beerling, D.J. (1997). Carbon isotope discrimination and stomatal responses of mature Pinus sylvestris L trees exposed in situ for three years to elevated CO2 and temperature. Acta Oecologica-International Journal of Ecology *18*, 697-712.

Beerling, D.J., and Chaloner, W.G. (1993). The impact of atmospheric CO2 and temperature-change on stomatal density - observations from quercus-robur lammas leaves. Annals of Botany *71*, 231-235.

Bellasio, C., Fini, A., and Ferrini, F. (2014). Evaluation of a High Throughput Starch Analysis Optimised for Wood. Plos One *9*, 11.

Bereiter, B., Eggleston, S., Schmitt, J., Nehrbass-Ahles, C., Stocker, T.F., Fischer, H., Kipfstuhl, S., and Chappellaz, J. (2015). Revision of the EPICA Dome C CO2 record from 800 to 600kyr before present. Geophysical Research Letters *42*, 542-549.

Bergmann, D.C., Lukowitz, W., and Somerville, C.R. (2004). Stomatal development and pattern controlled by a MAPKK kinase. Science *304*, 1494-1497.

Boccalandro, H.E., Rugnone, M.L., Moreno, J.E., Ploschuk, E.L., Serna, L., Yanovsky, M.J., and Casal, J.J. (2009). Phytochrome B Enhances Photosynthesis at the Expense of Water-Use Efficiency in Arabidopsis. Plant Physiology *150*, 1083-1092.

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J., and Roux, F. (2010). Linkage and Association Mapping of Arabidopsis thaliana Flowering Time in Nature. Plos Genetics *6*, 17.

Brem, R.B., and Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proceedings of the National Academy of Sciences of the United States of America *102*, 1572-1577.

Brooks, J.R., Flanagan, L.B., Varney, G.T., and Ehleringer, J.R. (1997). Vertical gradients in photosynthetic gas exchange characteristics and refixation of respired CO2 within boreal forest canopies. Tree Physiology *17*, 1-12.

Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. Bmc Bioinformatics *11*, 13.

Caine, B., (2016), Towards the identification and characterisation of toolkit genes responsible for stomatal development and CO2 response in the non-vascular land plant Physcomitrella patens.PhD, University of Sheffield, Sheffield.

Caine, R.S., Yin, X., Sloan, J., Harrison, E.L., Mohammed, U., Fulton, T., Biswal, A.K., Dionora, J., Chater, C.C., Coe, R.A. and Bandyopadhyay, A., 2018. Rice with reduced stomatal density conserves water and has improved drought tolerance under future climate conditions. *New Phytologist*.

Caldera, H.I.U., De Costa, W., Woodward, F.I., Lake, J.A., and Ranwala, S.M.W. (2017). Effects of elevated carbon dioxide on stomatal characteristics and carbon isotope ratio of Arabidopsis thaliana ecotypes originating from an altitudinal gradient. Physiologia Plantarum *159*, 74-92.

Calixto, C.P.G., Guo, W., James, A.B., Tzioutziou, N.A., Entizne, J.C., Panter, P.E., Knight, H., Nimmo, H., Zhang, R., and Brown, J.W.S. (2018). Rapid and dynamic alternative splicing impacts the Arabidopsis cold response transcriptome. The Plant Cell.

Carins Murphy, M.R., Jordan, G.J., and Brodribb, T.J. (2016). Cell expansion not cell differentiation predominantly co-ordinates veins and stomata within and among herbs and woody angiosperms grown under sun and shade. Annals of Botany *118*, 1127-1138.

Ceulemans, R., Vanpraet, L., and Jiang, X.N. (1995). Effects Of CO2 Enrichment, Leaf Position And Clone On Stomatal Index And Epidermal-Cell Density In Poplar (Populus). New Phytologist *131*, 99-107.

Chater, C., Peng, K., Movahedi, M., Dunn, J.A., Walker, H.J., Liang, Y.K., McLachlan, D.H., Casson, S., Isner, J.C., Wilson, I.*, et al.* (2015). Elevated CO2-Induced Responses in Stomata Require ABA and ABA Signaling. Current Biology *25*, 2709-2716.

Chen, J.X., Tresenrider, A., Chia, M.H., McSwiggen, D.T., Spedale, G., Jorgensen, V., Liao, H., van Werven, F.J., and Unal, E. (2017). Kinetochore inactivation by expression of a repressive mRNA. Elife *6*, 31.

Chen, L.Q., Qu, X.Q., Hou, B.H., Sosso, D., Osorio, S., Fernie, A.R., and Frommer, W.B. (2012). Sucrose Efflux Mediated by SWEET Proteins as a Key Step for Phloem Transport. Science *335*, 207-211.

Ciais, P., Denning, A.S., Tans, P.P., Berry, J.A., Randall, D.A., Collatz, G.J., Sellers, P.J., White, J.W.C., Trolier, M., Meijer, H.A.J.*, et al.* (1997). A three-dimensional synthesis study of delta O-18 in atmospheric CO2 .1. Surface fluxes. Journal of Geophysical Research-Atmospheres *102*, 5857-5872.

Ciereszko, I., and Kleczkowski, L.A. (2005). Expression of several genes involved in sucrose/starch metabolism as affected by different strategies to induce phosphate deficiency in Arabidopsis. Acta Physiologiae Plantarum *27*, 147-155.

Clark, R.M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T.T., Fu, G., Hinds, D.A.*, et al.* (2007). Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. Science *317*, 338-342.

Coupe, S.A., Palmer, B.G., Lake, J.A., Overy, S.A., Oxborough, K., Woodward, F.I., Gray, J.E., and Quick, W.P. (2006). Systemic signalling of environmental cues in Arabidopsis leaves. Journal of Experimental Botany *57*, 329-341.

Crawford, A.J., McLachlan, D.H., Hetherington, A.M., and Franklin, K.A. (2012). High temperature exposure increases plant cooling capacity. Current Biology *22*, R396-R397.

Crevillen, P., Ventriglia, T., Pinto, F., Orea, A., Merida, A., and Romero, J.M. (2005). Differential pattern of expression and sugar regulation of Arabidopsis thaliana ADP-glucose pyrophosphorylase-encoding genes. Journal of Biological Chemistry *280*, 8143-8149.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiology *139*, 5-17.

Dahlman, L. (2018). *Climate Change: Global Temperature | NOAA Climate.gov*. [online] Climate.gov. Available at: https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature [Accessed 24 Jun. 2018].

Delgado, D., Alonso-Blanco, C., Fenoll, C., and Mena, M. (2011). Natural variation in stomatal abundance of Arabidopsis thaliana includes cryptic diversity for different developmental processes. Annals of Botany *107*, 1247-1258.

Dell'Acqua, M., Gatti, D.M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hlaing, A.L., Aung, H.H., Nelissen, H., Baute, J.*, et al.* (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biology *16*, 23.

Dewitte, W., Riou-Khamlichi, C., Scofield, S., Healy, J.M.S., Jacqmard, A., Kilby, N.J., and Murray, J.A.H. (2003). Altered cell cycle distribution, hyperplasia, and inhibited differentiation in arabidopsis caused by the D-type cyclin CYCD3. Plant Cell *15*, 79-92.

Ding, F., Cui, P., Wang, Z.Y., Zhang, S.D., Ali, S., and Xiong, L.M. (2014). Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in Arabidopsis. Bmc Genomics *15*, 14.

Dittberner, H., Korte, A., Mettler-Altmann, T., Weber, A., Monroe, G., and de Meaux, J. (2018). Natural variation in stomata size contributes to the local adaptation of water-use efficiency in Arabidopsis thaliana. bioRxiv.

Doerge, R.W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. Nature Reviews Genetics *3*, 43-52.

Doheny-Adams, T., Hunt, L., Franks, P.J., Beerling, D.J., and Gray, J.E. (2012). Genetic manipulation of stomatal density influences stomatal size, plant growth and tolerance to restricted water supply across a growth carbon dioxide gradient. Philosophical Transactions of the Royal Society B-Biological Sciences *367*, 547-555.

Dow, G.J., and Bergmann, D.C. (2014). Patterning and processes: how stomatal development defines physiological potential. Current Opinion in Plant Biology *21*, 67-74.

Dow, G.J., Berry, J.A., and Bergmann, D.C. (2014). The physiological importance of developmental mechanisms that enforce proper stomatal spacing in Arabidopsis thaliana. New Phytologist *201*, 1205-1217.

Drake, B.G., GonzalezMeler, M.A., and Long, S.P. (1997). More efficient plants: A consequence of rising atmospheric CO2? Annual Review of Plant Physiology and Plant Molecular Biology *48*, 609-639.

Driscoll, S.P., Prins, A., Olmos, E., Kunert, K.J., and Foyer, C.H. (2006). Specification of adaxial and abaxial stomata, epidermal structure and photosynthesis to CO2 enrichment in maize leaves. Journal of Experimental Botany *57*, 381-390.

Eckardt, N.A. (2002). Alternative splicing and the control of flowering time. Plant Cell *14*, 743-747.

Edwards, K., Johnstone, C., and Thompson, C. (1991). A Simple And Rapid Method For The Preparation Of Plant Genomic Dna For PCR Analysis. Nucleic Acids Research *19*, 1349-1349.

El-Soda, M., Malosetti, M., Zwaan, B.J., Koornneef, M., and Aarts, M.G.M. (2014). Genotype x environment interaction QTL mapping in plants: lessons from Arabidopsis. Trends in Plant Science *19*, 390-398.

Engineer, C.B., Ghassemian, M., Anderson, J.C., Peck, S.C., Hu, H.H., and Schroeder, J.I. (2014). Carbonic anhydrases, EPF2 and a novel protease mediate CO2 control of stomatal development. Nature *513*, 246-+.

Engineer, C.B., Hashimoto-Sugimoto, M., Negi, J., Israelsson-Nordstrom, M., Azoulay-Shemer, T., Rappel, W.J., Iba, K., and Schroeder, J.I. (2016). CO2 Sensing and CO2 peculation of Stomatal Conductance: Advances and Open Questions. Trends in Plant Science *21*, 16-30.

FAO (2015a) FAO statistical pocketbook 2015. <http://www.fao.org/3/a-i4691e.pdf>

Ferris, R., Long, L., Bunn, S.M., Robinson, K.M., Bradshaw, H.D., Rae, A.M., and Taylor, G. (2002). Leaf stomatal and epidermal cell development: identification of putative quantitative trait loci in relation to elevated carbon dioxide concentration in poplar. Tree Physiology *22*, 633-640.

Fick, S.E., and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology *37*, 4302-4315.

Feike, D., Seung, D., Graf, A., Bischof, S., Ellick, T., Coiro, M., Soyk, S., Eicke, S., Mettler-Altmann, T., Lu, K.J.*, et al.* (2016). The Starch Granule-Associated Protein EARLY STARVATION1 Is Required for the Control of Starch Degradation in Arabidopsis thaliana Leaves. Plant Cell *28*, 1472-1489.

Fournier-Level, A., Perry, E.O., Wang, J.A., Braun, P.T., Migneault, A., Cooper, M.D., Metcalf, C.J.E., and Schmitt, J. (2016). Predicting the evolutionary dynamics of seasonal adaptation to novel climates in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America *113*, E2812-E2821.

Frank, D.C., Poulter, B., Saurer, M., Esper, J., Huntingford, C., Helle, G., Treydte, K., Zimmermann, N.E., Schleser, G.H., Ahlstrom, A.*, et al.* (2015). Water-use efficiency and transpiration across European forests during the Anthropocene. Nature Climate Change *5*, 579-+.

Franks, P.J. and Beerling, D.J., 2009. CO2‐forced evolution of plant gas exchange capacity and water‐use efficiency over the Phanerozoic. Geobiology, *7*(2), pp.227-236.

Franks, P.J., and Beerling, D.J. (2009). Maximum leaf conductance driven by CO2 effects on stomatal size and density over geologic time. Proceedings of the National Academy of Sciences of the United States of America *106*, 10343-10347.

Franks, P.J., Doheny-Adams, T.W., Britton-Harper, Z.J., and Gray, J.E. (2015). Increasing water-use efficiency directly through genetic manipulation of stomatal density. New Phytologist *207*, 188-195.

Franks, P.J., Drake, P.L., and Beerling, D.J. (2009). Plasticity in maximum stomatal conductance constrained by negative correlation between stomatal size and density: an analysis using Eucalyptus globulus. Plant Cell and Environment *32*, 1737-1748.

Franks, P.J., and Farquhar, G.D. (2007). The mechanical diversity of stomata and its significance in gas-exchange control. Plant Physiology *143*, 78-87.

Furukawa, A. (1997). Stomatal frequency of Quercus myrsinaefolia grown under different irradiances. Photosynthetica *34*, 195-199.

Gan, X.C., Stegle, O., Behr, J., Steffen, J.G., Drewe, P., Hildebrand, K.L., Lyngsoe, R., Schultheiss, S.J., Osborne, E.J., Sreedharan, V.T.*, et al.* (2011). Multiple reference genomes and transcriptomes for Arabidopsis thaliana. Nature *477*, 419-423.

Gardner, K.A., Wittern, L.M., and Mackay, I.J. (2016). A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. Plant Biotechnology Journal *14*, 1406-1417.

Gay, A.P., and Hurd, R.G. (1975). Influence of light on stomatal density in tomato. New Phytologist *75*, 37-46.

Gerhart, L.M., and Ward, J.K. (2010). Plant responses to low CO2 of the past. New Phytologist *188*, 674-695.

Gibson, S.I. (2005). Control of plant development and gene expression by sugar signaling. Current Opinion in Plant Biology *8*, 93-102.

Gilbert, W., 1978. Why genes in pieces?. Nature, *271*(5645), p.501.

Gillmor, C.S., Park, M.Y., Smith, M.R., Pepitone, R., Kerstetter, R.A., and Poethig, R.S. (2010). The MED12-MED13 module of Mediator regulates the timing of embryo patterning in Arabidopsis. Development *137*, 113-122.

Gnan, S., Priest, A., and Kover, P.X. (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-+.

Granier, C., Aguirrezabal, L., Chenu, K., Cookson, S.J., Dauzat, M., Hamard, P., Thioux, J.J., Rolland, G., Bouchier-Combaud, S., Lebaudy, A.*, et al.* (2006). PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. New Phytologist *169*, 623-635.

Granier, C., Turc, O., and Tardieu, F. (2000). Co-ordination of cell division and tissue expansion in sunflower, tobacco, and pea leaves: Dependence or independence of both processes? Journal of Plant Growth Regulation *19*, 45-54.

Gray, J.E., Holroyd, G.H., van der Lee, F.M., Bahrami, A.R., Sijmons, P.C., Woodward, F.I., Schuch, W., and Heterington, A.M. (2000). The HIC signalling pathway links CO2 perception to stomatal development. Nature *408*, 713-716.

Gray, S.B., Dermody, O., Klein, S.P., Locke, A.M., McGrath, J.M., Paul, R.E., Rosenthal, D.M., Ruiz-Vera, U.M., Siebers, M.H., Strellner, R.*, et al.* (2016). Intensifying drought eliminates the expected benefits of elevated carbon dioxide for soybean. Nature Plants *2*, 8.

Guo, C., Spinelli, M., Ye, C.T., Li, Q.S.Q., and Liang, C. (2017). Genome-Wide Comparative Analysis of Miniature Inverted Repeat Transposable Elements in 19 Arabidopsis thaliana Ecotype Accessions. Scientific Reports *7*, 12.

Hanschen, F.S., Pfitzmann, M., Witzel, K., Stutzel, H., Schreiner, M., and Zrenner, R. (2018). Differences in the enzymatic hydrolysis of glucosinolates increase the defense metabolite diversity in 19 Arabidopsis thaliana accessions. Plant Physiology and Biochemistry *124*, 126-135.

Hara, K., Kajita, R., Torii, K.U., Bergmann, D.C., and Kakimoto, T. (2007). The secretory peptide gene EPF1 enforces the stomatal one-cell-spacing rule. Genes & Development *21*, 1720-1725.

Hara, K., Yokoo, T., Kajita, R., Onishi, T., Yahata, S., Peterson, K.M., Torii, K.U., and Kakimoto, T. (2009). Epidermal Cell Density is Autoregulated via a Secretory Peptide, EPIDERMAL PATTERNING FACTOR 2 in Arabidopsis Leaves. Plant and Cell Physiology *50*, 1019-1031.

Hashimoto, M., Negi, J., Young, J., Israelsson, M., Schroeder, J.I., and Iba, K. (2006). Arabidopsis HT1 kinase controls stomatal movements in response to CO2. Nature Cell Biology *8*, 391-U352.

Heckman, D.S., Geiser, D.M., Eidell, B.R., Stauffer, R.L., Kardos, N.L., and Hedges, S.B. (2001). Molecular evidence for the early colonization of land by fungi and plants. Science *293*, 1129-1133.

Hendriks, J.H.M., Kolbe, A., Gibon, Y., Stitt, M., and Geigenberger, P. (2003). ADP-glucose pyrophosphorylase is activated by posttranslational redox-modification in response to light and to sugars in leaves of Arabidopsis and other plant species. Plant Physiology *133*, 838-849.

Hetherington, A.M., and Woodward, F.I. (2003). The role of stomata in sensing and driving environmental change. Nature *424*, 901-908.

Hofstra, G., and Hesketh, J.D. (1969). EFFECTS OF TEMPERATURE ON GAS EXCHANGE OF LEAVES IN LIGHT AND DARK. Planta *85*, 228-&.

Holland, J.B. (2007). Genetic architecture of complex traits in plants. Current Opinion in Plant Biology *10*, 156-161.

Holloway, S., Pearce, J.M., Ohsumi, T. and Hards, V.L.(2005). A Review of Natural CO2 Occurrences and their Relevance to CO2 Storage. British Geological Survey External Report, CR/05/104, 117pp.

Hollerer, I., Barker, J.C., Jorgensen, V., Tresenrider, A., Dugast-Darzacq, C., Darzacq, X., Chan, L.Y., Tjian, R., Ünal, E., and Brar, G.A. (2018). mRNA expression-mediated gene repression in human cells. bioRxiv.

Horrer, D., Flutsch, S., Pazmino, D., Matthews, J.S.A., Thalmann, M., Nigro, A., Leonhardt, N., Lawson, T., and Santelia, D. (2016). Blue Light Induces a Distinct Starch Degradation Pathway in Guard Cells for Stomatal Opening. Current Biology *26*, 362-370.

Horst, R.J., Fujita, H., Lee, J.S., Rychel, A.L., Garrick, J.M., Kawaguchi, M., Peterson, K.M., and Torii, K.U. (2015). Molecular Framework of a Regulatory Circuit Initiating Two-Dimensional Spatial Patterning of Stomatal Lineage. Plos Genetics *11*, 23.

Hu, H.H., Boisson-Dernier, A., Israelsson-Nordstrom, M., Bohmer, M., Xue, S.W., Ries, A., Godoski, J., Kuhn, J.M., and Schroeder, J.I. (2010). Carbonic anhydrases are upstream regulators of CO2-controlled stomatal movements in guard cells. Nature Cell Biology *12*, 87-U234.

Huang, B.E., Verbyla, K.L., Verbyla, A.P., Raghavan, C., Singh, V.K., Gaur, P., Leung, H., Varshney, R.K., and Cavanagh, C.R. (2015). MAGIC populations in crops: current status and future prospects. Theoretical and Applied Genetics *128*, 999-1017.

Huang, X.Q., Paulo, M.J., Boer, M., Effgen, S., Keizer, P., Koornneef, M., and 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.

Humplik, J.F., Lazar, D., Husickova, A., and Spichal, L. (2015). Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses - a review. Plant Methods *11*, 10.

Hunt, L., and Gray, J.E. (2009). The Signaling Peptide EPF2 Controls Asymmetric Cell Divisions during Stomatal Development. Current Biology *19*, 864-869.

Huynh, B.L., Ehlers, J.D., Huang, B.E., Munoz-Amatriain, M., Lonardi, S., Santos, J.R.P., Ndeve, A., Batieno, B.J., Boukar, O., Cisse, N.*, et al.* (2018). A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (Vigna unguiculata L. Walp.). Plant Journal *93*, 1129-1142.

IA, A.-S., and SL, O.K. (2002). Taxonomy and Phylogeny of Arabidopsis (Brassicaceae). (*The Arabidopsis Book / American Society of Plant Biologists*., *American Society of Plant Biologists*).

Idso, C.D., Idso, S.B., and Balling, R.C. (2001). An intensive two-week study of an urban CO2 dome in Phoenix, Arizona, USA. Atmospheric Environment *35*, 995-1000.

Ietswaart, R., Wu, Z., and Dean, C. (2012). Flowering time control: another window to the connection between antisense RNA and chromatin. Trends in Genetics *28*, 445-453.

IPCC (2007). Climate Change 2007: Working Group I: The Physical Science Basis. (Cambridge, Cambridge University Press).

Jauregui, I., Aparicio-Tejo, P.M., Avila, C., Rueda-Lopez, M., and Aranjuelo, I. (2015). Root and shoot performance of Arabidopsis thaliana exposed to elevated CO2: A physiologic, metabolic and transcriptomic response. Journal of Plant Physiology *189*, 65-76.

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., and Dean, C. (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science *290*, 344-347.

Kammer, P.M., Steiner, J.S., and Schob, C. (2015). Arabis alpina and Arabidopsis thaliana have different stomatal development strategies in response to high altitude pressure conditions. Alpine Botany *125*, 101-112.

Kanaoka, M.M., Pillitteri, L.J., Fujii, H., Yoshida, Y., Bogenschutz, N.L., Takabayashi, J., Zhu, J.K., and Torii, K.U. (2008). SCREAM/ICE1 and SCREAM2 specify three cell-state transitional steps leading to Arabidopsis stomatal differentiation. Plant Cell *20*, 1775-1785.

Kang, C.Y., Lian, H.L., Wang, F.F., Huang, J.R., and Yang, H.Q. (2009). Cryptochromes, Phytochromes, and COP1 Regulate Light-Controlled Stomatal Development in Arabidopsis. Plant Cell *21*, 2624-2641.

Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A., Heckerman, D., Daly, M.J., and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. Genetics *178*, 1709-1723.

Karve, R., Liu, W.S., Willet, S.G., Torii, K.U., and Shpak, E.D. (2011). The presence of multiple introns is essential for ERECTA expression in Arabidopsis. Rna-a Publication of the Rna Society *17*, 1907-1921.

Kawakatsu, T., Huang, S.S.C., Jupe, F., Sasaki, E., Schmitz, R.J., Urich, M.A., Castanon, R., Nery, J.R., Barragan, C., He, Y.P.*, et al.* (2016). Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell *166*, 492-505.

Kearsey, M.J., Pooni, H.S., and Syed, N.H. (2003). Genetics of quantitative traits in Arabidopsis thaliana. Heredity *91*, 456-464.

Kimball, B.A. (2016). Crop responses to elevated CO2 and interactions with H2O, N, and temperature. Current Opinion in Plant Biology *31*, 36-43.

Kirschbaum, M.U.F., and McMillan, A.M.S. (2018). Warming and Elevated CO2 Have Opposing Influences on Transpiration. Which is more Important? Current Forestry Reports *4*, 51-71.

Kliebenstein, D., Pedersen, D., Barker, B., and Mitchell-Olds, T. (2002). Comparative analysis of quantitative trait loci controlling glucosinolates, myrosinase and insect resistance in Arabidapsis thaliana. Genetics *161*, 325-332.

Kolla, V.A., Vavasseur, A., and Raghavendra, A.S. (2007). Hydrogen peroxide production is an early event during bicarbonate induced stomatal closure in abaxial epidermis of Arabidopsis. Planta *225*, 1421-1429.

Koornneef, M., Alonso-Blanco, C., and Vreugdenhil, D. (2004). Naturally occurring genetic variation in Arabidopsis thaliana. Annual Review of Plant Biology *55*, 141-172.

Kover, P.X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I.M., Purugganan, M.D., Durrant, C., and Mott, R. (2009). A Multiparent Advanced Generation Inter-Cross to Fine-Map Quantitative Traits in Arabidopsis thaliana. Plos Genetics *5*, 15.

Kover, P.X., and Mott, R. (2012). Mapping the genetic basis of ecologically and evolutionarily relevant traits in Arabidopsis thaliana. Current Opinion in Plant Biology *15*, 212-217.

Lachowiec, J., Shen, X., Queitsch, C., and Carlborg, O. (2015). A Genome-Wide Association Analysis Reveals Epistatic Cancellation of Additive Genetic Variance for Root Length in Arabidopsis thaliana. Plos Genetics *11*, 21.

Lake, J.A., Woodward, F.I., and Quick, W.P. (2002). Long-distance CO2 signalling in plants. Journal of Experimental Botany *53*, 183-193.

Lander, E.S., and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using rflp linkage maps. Genetics *121*, 185-199.

Lata, C., and Prasad, M. (2011). Role of DREBs in regulation of abiotic stress responses in plants. Journal of Experimental Botany *62*, 4731-4748.

Lau, O.S., Davies, K.A., Chang, J., Adrian, J., Rowe, M.H., Ballenger, C.E., and Bergmann, D.C. (2014). Direct roles of SPEECHLESS in the specification of stomatal self-renewing cells. Science *345*, 1605-1609.

Laurie, C.C., Chasalow, S.D., LeDeaux, J.R., McCarroll, R., Bush, D., Hauge, B., Lai, C.Q., Clark, D., Rocheford, T.R., and Dudley, J.W. (2004). The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics *168*, 2141-2155.

Le, J., Liu, X.G., Yang, K.Z., Chen, X.L., Zou, J.J., Wang, H.Z., Wang, M., Vanneste, S., Morita, M., Tasaka, M.*, et al.* (2014). Auxin transport and activity regulate stomatal patterning and development. Nature Communications *5*, 8.

Leakey, A (@Leakey77). "High CO2 "hospital" is keeping mutant with no stomata alive. We don't think you'll be able to pick it apart from WT!" 27 Jan 2016, 5:53 p.m. Tweet. https://twitter.com/leakey77/status/692390035950272512

Lee, E., Liu, X.G., Eglit, Y., and Sack, F. (2013). FOUR LIPS and MYB88 conditionally restrict the G1/S transition during stomatal formation. Journal of Experimental Botany *64*, 5207-5219.

Lesk, C., Rowhani, P., and Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. Nature *529*, 84-+.

Levine, L.H., Richards, J.T., and Wheeler, R.M. (2009). Super-elevated CO2 interferes with stomatal response to ABA and night closure in soybean (Glycine max). Journal of Plant Physiology *166*, 903-913.

Li, P.J., Tao, Z., and Dean, C. (2015). Phenotypic evolution through variation in splicing of the noncoding RNA COOLAIR. Genes & Development *29*, 696-701.

Li, Y., Rosso, M.G., Ulker, B., and Weisshaar, B. (2006). Analysis of T-DNA insertion site distribution patterns in Arabidopsis thaliana reveals special features of genes without insertions. Genomics *87*, 645-652.

Lievre, M., Granier, C., and Guedon, Y. (2016). Identifying developmental phases in the Arabidopsis thaliana rosette using integrative segmentation models. New Phytologist *210*, 1466-1478.

Lin, T.P., Caspar, T., Somerville, C.R., and Preiss, J. (1988). A Starch Deficient Mutant Of Arabidopsis-Thaliana With Low ADPglucose Pyrophosphorylase Activity Lacks One Of The 2 Subunits Of The Enzyme. Plant Physiology *88*, 1175-1181.

Lin, G.Z., Zhang, L., Han, Z.F., Yang, X.R., Liu, W.J., Li, E.T., Chang, J.B., Qi, Y.J., Shpak, E.D., and Chai, J.J. (2017). A receptor-like protein acts as a specificity switch for the regulation of stomatal development. Genes & Development *31*, 927-938.

Liu, L., Shen, F.Z., Xin, C.P., and Wang, Z. (2016). Multi-scale modeling of Arabidopsis thaliana response to different CO2 conditions: From gene expression to metabolic flux. Journal of Integrative Plant Biology *58*, 2-11.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods *25*, 402-408.

Lu, W., Tang, X.L., Huo, Y.Q., Xu, R., Qi, S.D., Huang, J.G., Zheng, C.C., and Wu, C.A. (2012). Identification and characterization of fructose 1,6-bisphosphate aldolase genes in Arabidopsis reveal a gene family with diverse responses to abiotic stresses. Gene *503*, 65-74.

Lu, Y., Gehan, J.P., and Sharkey, T.D. (2005). Daylength and circadian effects on starch degradation and maltose metabolism. Plant Physiology *138*, 2280-2291.

Lukowitz, W., Roeder, A., Parmenter, D., and Somerville, C. (2004). A MAPKK kinase gene regulates extra-embryonic cell fate in Arabidopsis. Cell *116*, 109-119.

Malmberg, R.L., Held, S., Waits, A., and Mauricio, R. (2005). Epistasis for fitness-related quantitative traits in Arabidopsis thaliana grown in the field and in the greenhouse. Genetics *171*, 2013-2027.

MacAlister, C.A., Ohashi-Ito, K., and Bergmann, D.C. (2007). Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. Nature *445*, 537-540.

Masle, J., Gilmore, S.R., and Farquhar, G.D. (2005). The ERECTA gene regulates plant transpiration efficiency in Arabidopsis. Nature *436*, 866-870.

McDonald, E.P., Erickson, J.E., and Kruger, E.L. (2002). Can decreased transpiration limit plant nitrogen acquisition in elevated CO2? Functional Plant Biology *29*, 1115-1120.

Menard, G.N., Moreno, J.M., Bryant, F.M., Munoz-Azcarate, O., Kelly, A.A., Hassani-Pak, K., Kurup, S., and Eastmond, P.J. (2017). Genome Wide Analysis of Fatty Acid Desaturation and Its Response to Temperature. Plant Physiology *173*, 1594-1605.

Meng, L.J., Zhao, X.Q., Ponce, K., Ye, G.Y., and Leung, H. (2016). QTL mapping for agronomic traits using multi-parent advanced generation inter-cross (MAGIC) populations derived from diverse elite indica rice lines. Field Crops Research *189*, 19-42.

Miglietta, F., Raschi, A., Bettarini, I., Resti, R., and Selvi, F. (1993). Natural Co2 Springs In Italy - A Resource For Examining Long-Term Response Of Vegetation To Rising Atmospheric CO2 Concentrations. Plant Cell and Environment *16*, 873-878.

Miller, G., Schlauch, K., Tam, R., Cortes, D., Torres, M.A., Shulaev, V., Dangl, J.L., and Mittler, R. (2009). The Plant NADPH Oxidase RBOHD Mediates Rapid Systemic Signaling in Response to Diverse Stimuli. Science Signaling *2*, 10.

Miura, A., Nakamura, M., Inagaki, S., Kobayashi, A., Saze, H., and Kakutani, T. (2009). An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. Embo Journal *28*, 1078-1086.

Monchgesang, S., Strehmel, N., Schmidt, S., Westphal, L., Taruttis, F., Muller, E., Herklotz, S., Neumann, S., and Scheel, D. (2016). Natural variation of root exudates in Arabidopsis thaliana-linking metabolomic and genomic data. Scientific Reports *6*, 11.

Moore, C.R., Gronwall, D.S., Miller, N.D., and Spalding, E.P. (2013). Mapping Quantitative Trait Loci Affecting Arabidopsis thaliana Seed Morphology Features Extracted Computationally From Images. G3-Genes Genomes Genetics *3*, 109-118.

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z.H., Schneider, H., and Donoghue, P.C.J. (2018). The timescale of early land plant evolution. Proceedings of the National Academy of Sciences of the United States of America *115*, E2274-E2283.

Mouchel, C.F., Briggs, G.C., and Hardtke, C.S. (2004). Natural genetic variation in Arabidopsis identifies BREVIS RADIX, a novel regulator of cell proliferation and elongation in the root. Genes & Development *18*, 700-714.

Mugford, S.T., Fernandez, O., Brinton, J., Flis, A., Krohn, N., Encke, B., Feil, R., Sulpice, R., Lunn, J.E., Stitt, M.*, et al.* (2014). Regulatory Properties of ADP Glucose Pyrophosphorylase Are Required for Adjustment of Leaf Starch Synthesis in Different Photoperiods. Plant Physiology *166*, 1733-U1877.

Nadeau J.A., Sack F.D.(2002) Stomatal development in Arabidopsis. In: Somerville C.R., Meyerowitz E.M., Meyerowitz E.M., editors. The Arabidopsis book. American Society of Plant Biologists; Rockville, MD. <http://www.aspb.org/publications/arabidopsis>.

Nakano, H., Makino, A. and Mae, T., 1997. The effect of elevated partial pressures of CO2 on the relationship between photosynthetic capacity and N content in rice leaves. *Plant physiology*, *115*(1), pp.191-198.

Nations, F.a.A.O.o.t.U. (2012). Global agriculture towards 2050. A.D.E. Division, ed. (http://www.fao.org/fileadmin/templates/wsfs/docs/Issues\_papers/HLEF2050\_Global\_Agriculture.pdf, Food and Agriculture Organization).

Nepstad, D.C., Stickler, C.M., Soares, B., and Merry, F. (2008). Interactions among Amazon land use, forests and climate: prospects for a near-term forest tipping point. Philosophical Transactions of the Royal Society B-Biological Sciences *363*, 1737-1746.

NOAA. (2018). *ESRL Global Monitoring Division - Global Greenhouse Gas Reference Network*. [online] Esrl.noaa.gov. Available at: https://www.esrl.noaa.gov/gmd/ccgg/trends/ [Accessed 24 Jun. 2018].

Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H.G., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R.*, et al.* (2005). The pattern of polymorphism in Arabidopsis thaliana. Plos Biology *3*, 1289-1299.

Ohashi-Ito, K., and Bergmann, D.C. (2006). Arabidopsis FAMA controls the final proliferation/differentiation switch during stomatal development. Plant Cell *18*, 2493-2505.

Ongom, P.O., and Ejeta, G. (2018). Mating Design and Genetic Structure of a Multi-Parent Advanced Generation Intercross (MAGIC) Population of Sorghum (Sorghum bicolor (L.) Moench). G3-Genes Genomes Genetics *8*, 331-341.

Outlaw, W.H., and Manchester, J. (1979). Guard-Cell Starch Concentration Quantitatively Related To Stomatal Aperture. Plant Physiology *64*, 79-82.

Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., McInerney, P., Thompson, J.F., Bowers, J., Jarosz, M., and Milos, P.M. (2009). Direct RNA sequencing. Nature *461*, 814-U873.

Pal, S.K., Liput, M., Piques, M., Ishihara, H., Obata, T., Martins, M.C.M., Sulpice, R., van Dongen, J.T., Fernie, A.R., Yadav, U.P.*, et al.* (2013). Diurnal Changes of Polysome Loading Track Sucrose Content in the Rosette of Wild-Type Arabidopsis and the Starchless pgm Mutant. Plant Physiology *162*, 1246-1265.

Paudel, I., Halpern, M., Wagner, Y., Raveh, E., Yermiyahu, U., Hoch, G., and Klein, T. (2018). Elevated CO2 compensates for drought effects in lemon saplings via stomatal downregulation, increased soil moisture, and increased wood carbon storage. Environmental and Experimental Botany *148*, 117-127.

Pecinka, A., and Liu, C.H. (2014). Drugs for Plant Chromosome and Chromatin Research. Cytogenetic and Genome Research *143*, 51-59.

Pillitteri, L.J., Sloan, D.B., Bogenschutz, N.L., and Torii, K.U. (2007). Termination of asymmetric cell division and differentiation of stomata. Nature *445*, 501-505.

Pillitteri, L.J., and Torii, K.U. (2012). Mechanisms of Stomatal Development. In Annual Review of Plant Biology, Vol 63, S.S. Merchant, ed. (Palo Alto: Annual Reviews), pp. 591-614.

Poole, I., Weyers, J.D.B., Lawson, T., and Raven, J.A. (1996). Variations in stomatal density and index: Implications for palaeoclimatic reconstructions. Plant Cell and Environment *19*, 705-712.

Prenger, J.J. and Ling, P.P., (2001). Greenhouse condensation control understanding and using vapor pressure deficit (VPD).

Price, A.H. (2006). Believe it or not, QTLs are accurate! Trends in Plant Science *11*, 213-216.

Purcell, C., Batke, S.P., Yiotis, C., Caballero, R., Soh, W.K., Murray, M., and McElwain, J.C. (2018). Increasing stomatal conductance in response to rising atmospheric CO2. Annals of Botany *121*, 1137-1149.

Rae, A.M., Tricker, P.J., Bunn, S.M., and Taylor, G. (2007). Adaptation of tree growth to elevated CO2: quantitative trait loci for biomass in Populus. New Phytologist *175*, 59-69.

Ragel, P., Streb, S., Feil, R., Sahrawy, M., Annunziata, M.G., Lunn, J.E., Zeeman, S., and Merida, A. (2013). Loss of Starch Granule Initiation Has a Deleterious Effect on the Growth of Arabidopsis Plants Due to an Accumulation of ADP-Glucose. Plant Physiology *163*, 75-85.

**Rédei GP** (1992) A heuristic glance at the past of Arabidopsis genetics *In* Koncz, C., N.H. Chua and J. Schell (ed.) Methods in Arabidopsis Research. World Scientific, New Jersey, NJ pp 1-15

Reid, C.D., Maherali, H., Johnson, H.B., Smith, S.D., Wullschleger, S.D., and Jackson, R.B. (2003). On the relationship between stomatal characters and atmospheric CO2. Geophysical Research Letters *30*, 4.

Retallack, G.J. (2002). Carbon dioxide and climate over the past 300 Myr. Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences *360*, 659-673.

Rigden, A.J., and Salvucci, G.D. (2017). Stomatal response to humidity and CO2 implicated in recent decline in US evaporation. Global Change Biology *23*, 1140-1151.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., and Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biology *12*, 14.

Rowe, H.C., Hansen, B.G., Halkier, B.A., and Kliebenstein, D.J. (2008). Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome. Plant Cell *20*, 1199-1216.

Roy, S.W., and Irimia, M. (2008). When good transcripts go bad: artifactual RT-PCR 'splicing' and genome analysis. Bioessays *30*, 601-605.

Rui, Y., Xiao, C.W., Yi, H.J., Kandemir, B., Wang, J.Z., Puri, V.M., and Anderson, C.T. (2017). POLYGALACTURONASE INVOLVED IN EXPANSION3 Functions in Seedling Development, Rosette Growth, and Stomatal Dynamics in Arabidopsis thaliana. Plant Cell *29*, 2413-2432.

Sablowski, R. (2016). Coordination of plant cell growth and division: collective control or mutual agreement? Current Opinion in Plant Biology *34*, 54-60.

Sager, J.C., and McFarlane, J.C. (1997). Radiation. Plant growth chamber handbook, 1-29.

Sachs T (1991) Pattern formation in plant tissues. Cambridge University Press, Cambridge, UK.

Salamone, P. R., Greene, T. W., Kavakli, I. H., & Okita, T. W. (2000). Isolation and characterization of a higher plant ADP‐glucose pyrophosphorylase small subunit homotetramer. *Febs Letters*, *482*(1-2), 113-118.

Sax, K., and McPhee, H.C. (1923). Color factors in bean hybrids. Journal of Heredity *14*, 205-208.

Scholze, M., Knorr, W., Arnell, N.W., and Prentice, I.C. (2006). A climate-change risk analysis for world ecosystems. Proceedings of the National Academy of Sciences of the United States of America *103*, 13116-13120.

Schneider, C. A.; Rasband, W. S. & Eliceiri, K. W. (2012), NIH Image to ImageJ: 25 years of image analysis" Nature methods **9(7)**: 671-675, PMID 22930834

Schwarte, S., Wegner, F., Havenstein, K., Groth, D., Steup, M., and Tiedemann, R. (2015). Sequence variation, differential expression, and divergent evolution in starch-related genes among accessions of Arabidopsis thaliana. Plant Molecular Biology *87*, 489-519.

Scialdone, A., and Howard, M. (2015). How plants manage food reserves at night: quantitative models and open questions. Frontiers in Plant Science *6*, 7.

Shen, H., Zhong, X.B., Zhao, F.F., Wang, Y.M., Yan, B.X., Li, Q., Chen, G.Y., Mao, B.Z., Wang, J.J., Li, Y.S.*, et al.* (2015). Overexpression of receptor-like kinase ERECTA improves thermotolerance in rice and tomato. Nature Biotechnology *33*, 996-U153.

Shpak, E.D. (2013). Diverse Roles of ERECTA Family Genes in Plant Development. Journal of Integrative Plant Biology *55*, 1238-1250.

Shpak, E.D., McAbee, J.M., Pillitteri, L.J., and Torii, K.U. (2005). Stomatal patterning and differentiation by synergistic interactions of receptor kinases. Science *309*, 290-293.

Singh, B., and El Maayar, M. (1998). Potential impacts of greenhouse gas climate change scenarios on sugar cane yields in Trinidad. Tropical Agriculture *75*, 348-354.

"single-nucleotide polymorphism / SNP | Learn Science at Scitable". www.nature.com. Archived from the original on 2015-11-10*. Retrieved 2019-01-17*

Smeekens, S. (2000). Sugar-induced signal transduction in plants. Annual Review of Plant Physiology and Plant Molecular Biology *51*, 49-81.

Smith, S.M., Fulton, D.C., Chia, T., Thorneycroft, D., Chapple, A., Dunstan, H., Hylton, C., Zeeman, S.C., and Smith, A.M. (2004). Diurnal changes in the transcriptome encoding enzymes of starch metabolism provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in arabidopsis leaves. Plant Physiology *136*, 2687-2699.

Sokolov, L.N., Dejardin, A., and Kleczkowski, L.A. (1998). Sugars and light/dark exposure trigger differential regulation of ADP-glucose pyrophosphorylase genes in Arabidopsis thaliana (thale cress). Biochemical Journal *336*, 681-687.

Springer, C.J., and Ward, J.K. (2007). Flowering time and elevated atmospheric CO2. New Phytologist *176*, 243-255.

Stark, D.M., Timmerman, K.P., Barry, G.F., Preiss, J., and Kishore, G.M. (1992). Regulation Of The Amount Of Starch In Plant-Tissues By ADP Glucose Pyrophosphorylase. Science *258*, 287-292.

Stoynova-Bakalova, E., Karanov, E., Petrov, P., and Hall, M.A. (2004). Cell division and cell expansion in cotyledons of Arabidopsis seedlings. New Phytologist *162*, 471-479.

Sugano, S.S., Shimada, T., Imai, Y., Okawa, K., Tamai, A., Mori, M., and Hara-Nishimura, I. (2010). Stomagen positively regulates stomatal density in Arabidopsis. Nature *463*, 241-U130.

Sun, B.N., Xiao, L., Xie, S.P., Deng, S.H., Wang, Y.D., Jia, H., and Turner, S. (2007). Quantitative analysis of paleoatmospheric CO2 level based on stomatal characters of fossil Ginkgo from Jurassic to Cretaceous in China. Acta Geologica Sinica-English Edition *81*, 931-939.

Sun, B.N., Wang, Q.J., Konrad, W., Ma, F.J., Dong, J.L., and Wang, Z.X. (2017). Reconstruction of atmospheric CO2 during the Oligocene based on leaf fossils from the Ningming Formation in Guangxi, China. Palaeogeography Palaeoclimatology Palaeoecology *467*, 5-15.

Swamy, B.P.M., Vikram, P., Dixit, S., Ahmed, H.U., and Kumar, A. (2011). Meta-analysis of grain yield QTL identified during agricultural drought in grasses showed consensus. Bmc Genomics *12*, 18.

Tateda, C., Obara, K., Abe, Y., Sekine, R., Nekoduka, S., Hikage, T., Nishihara, M., Sekine, K.T. and Fujisaki, K., 2018. The host stomatal density determines resistance to Septoria gentianae in Japanese gentian. Molecular Plant-Microbe Interactions, (ja).

Tanaka, Y., Nose, T., Jikumaru, Y., and Kamiya, Y. (2013). ABA inhibits entry into stomatal-lineage development in Arabidopsis leaves. Plant Journal *74*, 448-457.

Teng, N.J., Wang, J., Chen, T., Wu, X.Q., Wang, Y.H., and Lin, J.X. (2006). Elevated CO2 induces physiological, biochemical and structural changes in leaves of Arabidopsis thaliana. New Phytologist *172*, 92-103.

Thalmann, M., Pazmino, D., Seung, D., Horrer, D., Nigro, A., Meier, T., Kolling, K., Pfeifhofer, H.W., Zeeman, S.C., and Santelia, D. (2016). Regulation of Leaf Starch Degradation by Abscisic Acid Is Important for Osmotic Stress Tolerance in Plants. Plant Cell *28*, 1860-1878.

Thomas, D. (2010). Methods for Investigating Gene-Environment Interactions in Candidate Pathway and Genome-Wide Association Studies. In Annual Review of Public Health, Vol 31, J.E. Fielding, R.C. Brownson, and L.W. Green, eds. (Palo Alto: Annual Reviews), pp. 21-36.

Tilman, D., Balzer, C., Hill, J., and Befort, B.L. (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences of the United States of America *108*, 20260-20264.

Tisne, S., Reymond, M., Vile, D., Fabre, J., Dauzat, M., Koornneef, M., and Granier, C. (2008). Combined genetic and modeling approaches reveal that epidermal cell area and number in leaves are controlled by leaf and plant developmental processes in Arabidopsis. Plant Physiology *148*, 1117-1127.

Tognetti, R., Sebastiani, L., Vitagliano, C., Raschi, A., and Minnocci, A. (2001). Responses of two olive tree (Olea europaea L.) cultivars to elevated CO2 concentration in the field. Photosynthetica *39*, 403-410.

Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F., and Komeda, Y. (1996). The arabidopsis ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. Plant Cell *8*, 735-746.

Tsai, H.L., Lue, W.L., Lu, K.J., Hsieh, M.H., Wang, S.M., and Chen, J. (2009). Starch Synthesis in Arabidopsis Is Achieved by Spatial Cotranscription of Core Starch Metabolism Genes. Plant Physiology *151*, 1582-1595.

Tsukaya, H., Tsuge, T., and Uchimiya, H. (1994). THE COTYLEDON - A SUPERIOR SYSTEM FOR STUDIES OF LEAF DEVELOPMENT. Planta *195*, 309-312.

Underwood, W., Melotto, M., & He, S. Y. (2007). Role of plant stomata in bacterial invasion. Cellular microbiology, *9*(7), 1621-1629.

Vahisalu, T., Kollist, H., Wang, Y.F., Nishimura, N., Chan, W.Y., Valerio, G., Lamminmaki, A., Brosche, M., Moldau, H., Desikan, R.*, et al.* (2008). SLAC1 is required for plant guard cell S-type anion channel function in stomatal signalling. Nature *452*, 487-U415.

Valvekens, D., Vanmontagu, M., and Vanlijsebettens, M. (1988). Agrobacterium-Tumefaciens-Mediated Transformation Of Arabidopsis-thaliana Root Explants By Using Kanamycin Selection. Proceedings of the National Academy of Sciences of the United States of America *85*, 5536-5540.

van der Kooi, C.J., Reich, M., Low, M., De Kok, L.J., and Tausz, M. (2016). Growth and yield stimulation under elevated CO2 and drought: A meta-analysis on crops. Environmental and Experimental Botany *122*, 150-157.

van Eeuwijk, F.A., Bink, M., Chenu, K., and Chapman, S.C. (2010). Detection and use of QTL for complex traits in multiple environments. Current Opinion in Plant Biology *13*, 193-205.

Vandewater, P.K., Leavitt, S.W., and Betancourt, J.L. (1994). Trends In Stomatal Density And C-13/C-12 Ratios Of Pinus-Flexilis Needles During Last Glacial-Interglacial Cycle. Science *264*, 239-243.

Venable, D.L. (1992). Size-Number Trade-Offs And The Variation Of Seed Size With Plant Resource Status. American Naturalist *140*, 287-304.

Ventriglia, T., Kuhn, M.L., Ruiz, M.T., Ribeiro-Pedro, M., Valverde, F., Ballicora, M.A., Preiss, J., and Romero, J.M. (2008). Two Arabidopsis ADP-glucose pyrophosphorylase large subunits (APL1 and APL2) are catalytic. Plant Physiology *148*, 65-76.

Verbyla, A.P., Cavanagh, C.R., and Verbyla, K.L. (2014). Whole-Genome Analysis of Multienvironment or Multitrait QTL in MAGIC. G3-Genes Genomes Genetics *4*, 1569-1584.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in Biosciences *131*, 281-285.

Wang, Z., Gerstein, M. and Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews genetics, *10*(1), p.57.

Wang, Y.S., Wang, C.K., Guo, X.Q., Liu, G.R., and Huang, Y. (2002). Trend, seasonal and diurnal variations of atmospheric CO2 in Beijing. Chinese Science Bulletin *47*, 2050-2055.

Wang, H.C., Ngwenyama, N., Liu, Y.D., Walker, J.C., and Zhang, S.Q. (2007). Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in Arabidopsis. Plant Cell *19*, 63-73.

Wang, Y.H., X.Y., Zhu, J.K., and Dong, J. (2016). Demethylation of ERECTA receptor genes by IBM1 histone demethylase affects stomatal development. Development *143*, 4452-4461.

Ward, J.K., and Strain, B.R. (1997). Effects of low and elevated CO2 partial pressure on growth and reproduction of Arabidopsis thaliana from different elevations. Plant Cell and Environment *20*, 254-260.

Watson-Lazowski, A., Lin, Y.A., Miglietta, F., Edwards, R.J., Chapman, M.A., and Taylor, G. (2016). Plant adaptation or acclimation to rising CO2? Insight from first multigenerational RNA-Seq transcriptome. Global Change Biology *22*, 3760-3773.

Webb, A.A.R., and Hetherington, A.M. (1997). Convergence of the abscisic acid, CO2, and extracellular calcium signal transduction pathways in stomatal guard cells. Plant Physiology *114*, 1557-1560.

Wheeler, T.R., Craufurd, P.Q., Ellis, R.H., Porter, J.R., and Prasad, P.V.V. (2000). Temperature variability and the yield of annual crops. Agriculture Ecosystems & Environment *82*, 159-167.

Wild, A., and Wolf, G. (1980). The Effect Of Different Light Intensities On The Frequency And Size Of Stomata, The Size Of Cells, The Number, Size And Chlorophyll Content Of Chloroplasts In The Mesophyll And The Guard-Cells During The Ontogeny Of Primary Leaves Of Sinapis-Alba. Zeitschrift Fur Pflanzenphysiologie *97*, 325-342.

Woodward, F.I., and Bazzaz, F.A. (1988). THE RESPONSES OF STOMATAL DENSITY TO CO2 PARTIAL-PRESSURE. Journal of Experimental Botany *39*, 1771-1781.

Woodward, A.W., and Bartel, B. (2005). Auxin: Regulation, action, and interaction. Annals of Botany *95*, 707-735.

Woodward, F.I. (1987). Stomatal Numbers Are Sensitive To Increases In Co2 From Preindustrial Levels. Nature *327*, 617-618.

Woodward, F.I., and Kelly, C.K. (1995). THE INFLUENCE OF CO2 CONCENTRATION ON STOMATAL DENSITY. New Phytologist *131*, 311-327.

Wu, G., Liu, H., Hua, L., Luo, Q., Lin, Y., He, P., Feng, S., Liu, J. and Ye, Q., (2018). Differential responses of stomata and photosynthesis to elevated temperature in two co-occurring subtropical forest tree species. *Frontiers in plant science*, *9*, p.467.

Xing, H.T., Guo, P., Xia, X.L., and Yin, W.L. (2011). PdERECTA, a leucine-rich repeat receptor-like kinase of poplar, confers enhanced water use efficiency in Arabidopsis. Planta *234*, 229-241.

Xue, S.W., Hu, H.H., Ries, A., Merilo, E., Kollist, H., and Schroeder, J.I. (2011). Central functions of bicarbonate in S-type anion channel activation and OST1 protein kinase in CO2 signal transduction in guard cell. Embo Journal *30*, 1645-1658.

Yan, W.M., Zhong, Y.Q.W., and Shangguan, Z.P. (2017). Contrasting responses of leaf stomatal characteristics to climate change: a considerable challenge to predict carbon and water cycles. Global Change Biology *23*, 3781-3793.

Yang, J., Zhu, J., and Williams, R.W. (2007). Mapping the genetic architecture of complex traits in experimental populations. Bioinformatics *23*, 1527-1536.

Yang, M., and Sack, F.D. (1995). The too many mouths and four lips mutations affect stomatal production in arabidopsis. Plant Cell *7*, 2227-2239.

Yilmaz, A., Mejia-Guerra, M.K., Kurz, K., Liang, X.Y., Welch, L., and Grotewold, E. (2011). AGRIS: the Arabidopsis Gene Regulatory Information Server, an update. Nucleic Acids Research *39*, D1118-D1122.

Yu, T.S., Kofler, H., Hausler, R.E., Hille, D., Flugge, U.I., Zeeman, S.C., Smith, A.M., Kossmann, J., Lloyd, J., Ritte, G.*, et al.* (2001). The Arabidopsis sex1 mutant is defective in the R1 protein, a general regulator of starch degradation in plants, and not in the chloroplast hexose transporter. Plant Cell *13*, 1907-1918.

Zhang, R.P., Liang, T.G., Guo, J., Xie, H.J., Feng, Q.S., and Aimaiti, Y. (2018). Grassland dynamics in response to climate change and human activities in Xinjiang from 2000 to 2014. Scientific Reports *8*, 11.

Zhao, C., Liu, B., Piao, S.L., Wang, X.H., Lobell, D.B., Huang, Y., Huang, M.T., Yao, Y.T., Bassu, S., Ciais, P.*, et al.* (2017). Temperature increase reduces global yields of major crops in four independent estimates. Proceedings of the National Academy of Sciences of the United States of America *114*, 9326-9331.

Zhao, D.L., and Li, Y.R. (2015). Climate Change and Sugarcane Production: Potential Impact and Mitigation Strategies. International Journal of Agronomy, 10.

Zeeman, S.C., Kossmann, J., and Smith, A.M. (2010). Starch: Its Metabolism, Evolution, and Biotechnological Modification in Plants. In Annual Review of Plant Biology, Vol 61, S. Merchant, W.R. Briggs, and D. Ort, eds. (Palo Alto: Annual Reviews), pp. 209-234.

Zeeman, S.C., Smith, S.M., and Smith, A.M. (2004). The breakdown of starch in leaves. New Phytologist *163*, 247-261.

Zoulias, N., Harrison, E.L., Casson, S.A., and Gray, J.E. (2018). Molecular control of stomatal development. The Biochemical journal *475*, 441-454.

# Supplementary Information

Table 1 Genes under QTL1. All genes found within 300Kb of QTL1 according to arabidopsis.org bulk downloader

|  |  |  |  |
| --- | --- | --- | --- |
| **Locus Identifier** | **Gene Description** | **Primary Gene Symbol** | **All Gene Symbols** |
| AT1G55340 | hypothetical protein (DUF1639);(source:Araport11) |  |  |
| AT1G55430 | Cysteine/Histidine-rich C1 domain family protein;(source:Araport11) |  |  |
| AT1G55160 | WAS/WASL-interacting family protein;(source:Araport11) |  |  |
| AT1G55928 | nuclear speckle splicing regulatory-like protein (DUF2040);(source:Araport11) |  |  |
| AT1G55475 | hypothetical protein;(source:Araport11) |  |  |
| AT1G55000 | peptidoglycan-binding LysM domain-containing protein;(source:Araport11) |  |  |
| AT1G55690 | Sec14p-like phosphatidylinositol transfer family protein;(source:Araport11) |  |  |
| AT1G55980 | FAD/NAD(P)-binding oxidoreductase family protein;(source:Araport11) |  |  |
| AT1G55175 | hypothetical protein;(source:Araport11) |  |  |
| AT1G55820 | lysine-specific demethylase, putative (DUF1296);(source:Araport11) |  |  |
| AT1G55915 | zinc ion binding protein;(source:Araport11) |  |  |
| AT1G54920 | hypothetical protein;(source:Araport11) |  |  |
| AT1G55280 | Lipase/lipooxygenase, PLAT/LH2 family protein;(source:Araport11) |  |  |
| AT1G56030 | RING/U-box protein;(source:Araport11) |  |  |
| AT1G55950 | DNA-binding storekeeper protein-related transcriptional regulator;(source:Araport11) |  |  |
| AT1G56000 | FAD/NAD(P)-binding oxidoreductase family protein;(source:Araport11) |  |  |
| AT1G55152 | hypothetical protein;(source:Araport11) |  |  |
| AT1G56090 | Tetratricopeptide repeat (TPR)-like superfamily protein;(source:Araport11) |  |  |
| AT1G55530 | RING/U-box superfamily protein;(source:Araport11) |  |  |
| AT1G55535 | transmembrane protein;(source:Araport11) |  |  |
| AT1G54790 | GDSL-motif esterase/acyltransferase/lipase. Enzyme group with broad substrate specificity that may catalyze acyltransfer or hydrolase reactions with lipid and non-lipid substrates. |  |  |
| AT1G54680 | translation initiation factor 3 subunit I;(source:Araport11) |  |  |
| AT1G56040 | HEAT/U-box protein;(source:Araport11) |  |  |
| AT1G55210 | Disease resistance-responsive (dirigent-like protein) family protein;(source:Araport11) |  |  |
| AT1G56020 | serine/arginine repetitive matrix-like protein;(source:Araport11) |  |  |
| AT1G55710 | hypothetical protein;(source:Araport11) |  |  |
| AT1G54720 | early-responsive to dehydration protein-related / ERD protein-like protein;(source:Araport11) |  |  |
| AT1G54930 | GRF zinc finger / Zinc knuckle protein;(source:Araport11) |  |  |
| AT1G55546 | 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase;(source:Araport11) |  |  |
| AT1G54610 | Protein kinase superfamily protein;(source:Araport11) |  |  |
| AT1G54820 | Protein kinase superfamily protein;(source:Araport11) |  |  |
| AT1G55240 | proteinase inhibitor I4, serpin (DUF716);(source:Araport11) |  |  |
| AT1G55680 | Transducin/WD40 repeat-like superfamily protein;(source:Araport11) |  |  |
| AT1G55880 | Pyridoxal-5-phosphate-dependent enzyme family protein;(source:Araport11) |  |  |
| AT1G56045 | Ribosomal protein L41 family;(source:Araport11) |  |  |
| AT1G55550 | P-loop containing nucleoside triphosphate hydrolases superfamily protein;(source:Araport11) |  |  |
| AT1G54740 | FANTASTIC four-like protein (DUF3049);(source:Araport11) |  |  |
| AT1G55440 | Cysteine/Histidine-rich C1 domain family protein;(source:Araport11) |  |  |
| AT1G54950 | transmembrane protein;(source:Araport11) |  |  |
| AT1G55200 | kinase with adenine nucleotide alpha hydrolases-like domain-containing protein;(source:Araport11) |  |  |
| AT1G54650 | Methyltransferase family protein;(source:Araport11) |  |  |
| AT1G55170 | DNA double-strand break repair protein;(source:Araport11) |  |  |
| AT1G55930 | CBS domain-containing protein / transporter associated domain-containing protein;(source:Araport11) |  |  |
| AT1G55960 | Polyketide cyclase/dehydrase and lipid transport superfamily protein;(source:Araport11) |  |  |
| AT1G54590 | pre-mRNA-splicing factor-like protein;(source:Araport11) |  |  |
| AT1G54640 | F-box family protein-like protein;(source:Araport11) |  |  |
| AT1G55365 | hypothetical protein;(source:Araport11) |  |  |
| AT1G55700 | Cysteine/Histidine-rich C1 domain family protein;(source:Araport11) |  |  |
| AT1G55750 | BSD domain (BTF2-like transcription factors, Synapse-associated proteins and DOS2-like proteins);(source:Araport11) |  |  |
| AT1G56080 | interactor of constitutive active ROPs protein;(source:Araport11) |  |  |
| AT1G54980 | Plant invertase/pectin methylesterase inhibitor superfamily protein;(source:Araport11) |  |  |
| AT1G55760 | BTB/POZ domain-containing protein;(source:Araport11) |  |  |
| AT1G56060 | cysteine-rich/transmembrane domain protein B;(source:Araport11) |  |  |
| AT1G55270 | Galactose oxidase/kelch repeat superfamily protein;(source:Araport11) |  |  |
| AT1G55890 | Tetratricopeptide repeat (TPR)-like superfamily protein;(source:Araport11) |  |  |
| AT1G55800 | hypothetical protein;(source:Araport11) |  |  |
| AT1G55390 | Cysteine/Histidine-rich C1 domain family protein;(source:Araport11) |  |  |
| AT1G55380 | Cysteine/Histidine-rich C1 domain family protein;(source:Araport11) |  |  |
| AT1G55840 | Sec14p-like phosphatidylinositol transfer family protein;(source:Araport11) |  |  |
| AT1G54870 | ChlADR is an aldehyde reductase that catalyzes the reduction of the aldehyde carbonyl groups on saturated and alpha,beta-unsaturated aldehydes with more than 5 carbons in vitro. The N-terminal region of this protein directs GFP to the chloroplast where where ChlADR likely helps to maintain the photosynthetic process by detoxifying reactive carbonyls formed during lipid peroxidation. |  |  |
| AT1G55205 | hypothetical protein;(source:Araport11) |  |  |
| AT1G54620 | Plant invertase/pectin methylesterase inhibitor superfamily protein;(source:Araport11) |  |  |
| AT1G55770 | Plant invertase/pectin methylesterase inhibitor superfamily protein;(source:Araport11) |  |  |
| AT1G55220 | hypothetical protein;(source:Araport11) |  |  |
| AT1G55590 | RNI-like superfamily protein;(source:Araport11) |  |  |
| AT1G55450 | S-adenosyl-L-methionine-dependent methyltransferases superfamily protein;(source:Araport11) |  |  |
| AT1G55830 | coiled-coil protein;(source:Araport11) |  |  |
| AT1G55990 | glycine-rich protein;(source:Araport11) |  |  |
| AT1G54730 | Major facilitator superfamily protein;(source:Araport11) |  |  |
| AT1G55790 | ferredoxin-fold anticodon-binding domain protein;(source:Araport11) |  |  |
| AT1G54860 | Glycoprotein membrane precursor GPI-anchored;(source:Araport11) |  |  |
| AT1G56085 | Cyclophilin;(source:Araport11) |  |  |
| AT1G54770 | Fcf2 pre-rRNA processing protein;(source:Araport11) |  |  |
| AT1G55230 | proteinase inhibitor I4, serpin (DUF716);(source:Araport11) |  |  |
| AT1G55630 | Pentatricopeptide repeat (PPR) superfamily protein;(source:Araport11) |  |  |
| AT1G54700 | hypothetical protein;(source:Araport11) |  |  |
| AT1G54890 | Late embryogenesis abundant (LEA) protein-like protein;(source:Araport11) |  |  |
| AT1G55090 | carbon-nitrogen hydrolase family protein;(source:Araport11) |  |  |
| AT1G55207 | hypothetical protein;(source:Araport11) |  |  |
| AT1G55030 | RNI-like superfamily protein;(source:Araport11) |  |  |
| AT1G55070 | F-box and associated interaction domains-containing protein;(source:Araport11) |  |  |
| AT1G55360 | tRNA-splicing ligase (DUF239);(source:Araport11) |  |  |
| AT1G54775 | Encodes a Plant thionin family protein |  |  |
| AT1G55675 | transmembrane protein;(source:Araport11) |  |  |
| AT1G55050 | hypothetical protein;(source:Araport11) |  |  |
| AT1G55265 | DUF538 family protein, putative (Protein of unknown function, DUF538);(source:Araport11) |  |  |
| AT1G54850 | HSP20-like chaperones superfamily protein;(source:Araport11) |  |  |
| AT1G54575 | hypothetical protein;(source:Araport11) |  |  |
| AT1G56050 | GTP-binding protein-like protein;(source:Araport11) | (ENGD-2) | (ENGD-2) |
| AT1G55290 | encodes a protein whose sequence is similar to oxidoreductase, 2OG-Fe(II) oxygenase | (F6'H2) | (F6'H2) |
| AT1G55650 | HMG (high mobility group) box protein with ARID/BRIGHT DNA-binding domain-containing protein;(source:Araport11) | (HMGB11) | (HMGB11) |
| AT1G55460 | DNA/RNA-binding protein Kin17, conserved region;(source:Araport11) | (KIN17) | (KIN17) |
|  |
| (ATKIN17) |
| AT1G55080 | mediator of RNA polymerase II transcription subunit-like protein;(source:Araport11) | (MED9) | (MED9) |
| AT1G55190 | PRA1 (Prenylated rab acceptor) family protein;(source:Araport11) | (PRA7) | PRENYLATED RAB ACCEPTOR 1.F2 (PRA1.F2) |
| AT1G55140 | Encodes one of two chloroplast Mini-RNase III-like enzymes in Arabidopsis. Double mutants display imprecise maturation of 23S rRNA and other rRNAs. | (RNC3) | (RNC3) |
| AT1G55900 | component of a translocase in the mitochondrial inner membrane | (TIM50) | EMBRYO DEFECTIVE 1860 (emb1860) |
|  |
|  |
| AT1G55940 | member of CYP708A | "CYTOCHROME P450, FAMILY 708, SUBFAMILY A, POLYPEPTIDE 1" (CYP708A1) | "CYTOCHROME P450, FAMILY 708, SUBFAMILY A, POLYPEPTIDE 1" (CYP708A1) |
| AT1G54830 | Encodes a NUCLEAR FACTOR-Y C (NF-YC) homologue NF-YC3. NF-YC3., NF-YC4 and NF-YC9 redundantly modulate GA- and ABA-mediated seed germination. | "NUCLEAR FACTOR Y, SUBUNIT C3" (NF-YC3) | "NUCLEAR FACTOR Y, SUBUNIT C3" (NF-YC3) |
| AT1G55870 | Encodes a poly(A)-specific ribonuclease, AtPARN. Expression of AtPARN is upregulated by ABA or stress treatment. Mutant is hypersensitivity to salicylic acid as well as ABA. Functions with AGS1 to regulate the poly(A) status of mitochondrial mRNA. | ABA-HYPERSENSITIVE GERMINATION 2 (AHG2) | ABA-HYPERSENSITIVE GERMINATION 2 (AHG2) |
|  |
| ARABIDOPSIS THALIANA POLY(A) RIBONUCLEASE (ATPARN) |
| AT1G54580 | encodes an acyl carrier protein expressed in leaves, roots, and dry seeds. Gene expression is not regulated by light but downregulated by starvation and upregulated by increased level of exogenous sucrose. The mRNA is cell-to-cell mobile. | ACYL CARRIER PROTEIN 2 (ACP2) | (ATACP2) |
|  |
| ACYL CARRIER PROTEIN 2 (ACP2) |
| AT1G54630 | encodes an acyl carrier protein expressed in leaves, roots, and dry seeds. Gene expression is not regulated by light. | ACYL CARRIER PROTEIN 3 (ACP3) | ACYL CARRIER PROTEIN 3 (ACP3) |
|  |
| (ATACP3) |
| AT1G55320 | Encodes a protein with similarity to acyl activating enzymes. AAE18 is localized to the peroxisome where it may be involved in metabolism of auxin precursors to active auxins. | ACYL-ACTIVATING ENZYME 18 (AAE18) | ACYL-ACTIVATING ENZYME 18 (AAE18) |
| AT1G54760 | AGAMOUS-like 85;(source:Araport11) | AGAMOUS-LIKE 85 (AGL85) | AGAMOUS-LIKE 85 (AGL85) |
| AT1G55330 | Encodes a putative arabinogalactan-protein (AGP21). | ARABINOGALACTAN PROTEIN 21 (AGP21) | ARABINOGALACTAN PROTEIN 21 (AtAGP21) |
|  |
| ARABINOGALACTAN PROTEIN 21 (AGP21) |
| AT1G54990 | auxin response mutant (AXR4) The mRNA is cell-to-cell mobile. | AUXIN RESISTANT 4 (AXR4) | REDUCED ROOT GRAVITROPISM (RGR) |
|  |
| AUXIN RESISTANT 4 (AXR4) |
|  |
| REDUCED ROOT GRAVITROPISM 1 (RGR1) |
| AT1G55120 | Encodes a protein with fructan exohydrolase (FEH) activity acting on levan-type fructans (6-FEH, levanase). The enzyme does not have invertase activity. | BETA-FRUCTOFURANOSIDASE 5 (FRUCT5) | BETA-FRUCTOFURANOSIDASE 5 (ATFRUCT5) |
|  |
| 6-FRUCTAN EXOHYDROLASE (AtcwINV3) |
|  |
| BETA-FRUCTOFURANOSIDASE 5 (FRUCT5) |
| AT1G55510 | branched-chain alpha-keto acid decarboxylase E1 beta | BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE E1 BETA SUBUNIT (BCDH BETA1) | BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE E1 BETA SUBUNIT (BCDH BETA1) |
| AT1G55610 | mutant has Altered vascular cell differentiation; LRR Receptor Kinase | BRI1 LIKE (BRL1) | BRI1 LIKE (BRL1) |
| AT1G55730 | member of Low affinity calcium antiporter CAX2 family | CATION EXCHANGER 5 (CAX5) | CATION EXCHANGER 5 (CAX5) |
|  |
| (ATCAX5) |
| AT1G55720 | member of Low affinity calcium antiporter CAX2 family | CATION EXCHANGER 6 (CAX6) | CATION EXCHANGER 6 (CAX6) |
|  |
| (ATCAX6) |
| AT1G55850 | encodes a protein similar to cellulose synthase The mRNA is cell-to-cell mobile. | CELLULOSE SYNTHASE LIKE E1 (CSLE1) | CELLULOSE SYNTHASE LIKE E1 (CSLE1) |
|  |
| (ATCSLE1) |
| AT1G55490 | encodes the beta subunit of the chloroplast chaperonin 60, a homologue of bacterial GroEL. Mutants in this gene develops lesions on its leaves, expresses systemic acquired resistance (SAR) and develops accelerated cell death to heat shock stress. The protein has molecular chaperone activity for suppressing protein aggregation in vitro. | CHAPERONIN 60 BETA (CPN60B) | CHAPERONIN 60 BETA (CPN60B) |
|  |
| LESION INITIATION 1 (LEN1) |
|  |
| CHAPERONIN-60BETA1 (CPN60BETA1) |
|  |
| (CPNB1) |
| AT1G55620 | Encodes a chloride channel protein that has been localized to the chloroplast and golgi apparatus. Complements yeast gef1 mutant and therefor may function to acidify the golgi lumen. | CHLORIDE CHANNEL F (CLC-F) | (ATCLC-F) |
|  |
| CHLORIDE CHANNEL F (CLC-F) |
| AT1G55350 | Similar to maize DEK1, a gene encoding a membrane protein of the calpain gene superfamily required for aleurone cell development in the endosperm of maize grains. A key component of the embryonic L1 cell-layer specification pathway. It localizes to membranes and undergoes intramolecular autolytic cleavage events that release the calpain domain into the cytoplasm. | DEFECTIVE KERNEL 1 (DEK1) | DEFECTIVE KERNEL 1 (DEK1) |
| AT1G55420 | Cysteine/Histidine-rich C1 domain family protein;(source:Araport11) | EMBRYO SAC DEVELOPMENT ARREST 11 (EDA11) | EMBRYO SAC DEVELOPMENT ARREST 11 (EDA11) |
| AT1G55500 | evolutionarily conserved C-terminal region 4;(source:Araport11) | EVOLUTIONARILY CONSERVED C-TERMINAL REGION 4 (ECT4) | EVOLUTIONARILY CONSERVED C-TERMINAL REGION 4 (ECT4) |
| AT1G55660 | FOF2, is the F-box protein family. Overexpression of FOF2 results in delayed transitions to flowering under both LD and SD conditions. | F-BOX OF FLOWERING 2 (FOF2) | F-BOX OF FLOWERING 2 (FOF2) |
| AT1G54690 | Encodes HTA3, a histone H2A protein. H2AX is a meiosis-specific isoform of histone H2A. Upon DSB formation, rapid accumulation of phosphorylated H2AX (γ-H2AX) occurs around the break site. H2AX foci accumulate in early G2. Immunolocalization studies in spread preparations of wild-type meiocytes at G2/early leptotene revealed the accumulation of numerous rather diffuse γ-H2AX foci throughout the chromatin. However, their accumulation is not contemporaneous with that of AtSPO11-1. At 3 h post-S, no γ-H2AX foci are detected. During the 3- to 5-h window when AtSPO11-1 foci rapidly disappear, there is an equally swift accumulation of γ-H2AX to a maximum of >50 diffuse foci. The level of γH2AX then remains constant for a further 13 h before undergoing a gradual decrease to 10?20 foci in the 18- to 24-h post-S period. By 30 h the foci have disappeared from the chromatin. | GAMMA HISTONE VARIANT H2AX (GAMMA-H2AX) | HISTONE H2A 3 (HTA3) |
|  |
| (H2AXb) |
|  |
| GAMMA HISTONE VARIANT H2AX (GAMMA-H2AX) |
|  |
| GAMMA H2AX (G-H2AX) |
| AT1G55260 | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein;(source:Araport11) | GLYCOSYLPHOSPHATIDYLINOSITOL-ANCHORED LIPID PROTEIN TRANSFER 6 (LTPG6) | GLYCOSYLPHOSPHATIDYLINOSITOL-ANCHORED LIPID PROTEIN TRANSFER 6 (LTPG6) |
| AT1G55325 | Encodes the Arabidopsis homolog of the transcriptional regulator MED13, is dynamically expressed during embryogenesis and regulates both developmental timing and the radial pattern formation. | GRAND CENTRAL (GCT) | GRAND CENTRAL (GCT) |
|  |
| MACCHI-BOU 2 (MAB2) |
|  |
| MEDIATOR OF RNA POLYMERASE II TRANSCRIPTION SUBUNIT 13 (MED13) |
| AT1G55970 | HAC4 is most likely to be an expressed pseudogene that lacks HAT function. there is a single nucleotide deletion in both the HAC4 genomic and cDNA sequences relative to its homologs. The resulting frameshift within the open reading frame causes a stop codon to occur within the predicted acetyltransferase catalytic domain. | HISTONE ACETYLTRANSFERASE OF THE CBP FAMILY 4 (HAC4) | (HAC04) |
|  |
| HISTONE ACETYLTRANSFERASE OF THE GNAT/MYST SUPERFAMILY 04 (HAG04) |
|  |
| (HAC02) |
|  |
| HISTONE ACETYLTRANSFERASE OF THE GNAT/MYST SUPERFAMILY 4 (HAG4) |
|  |
| ARABIDOPSIS THALIANA P300/CBP ACETYLTRANSFERASE-RELATED PROTEIN 2 (ATHPCAT2) |
|  |
| HISTONE ACETYLTRANSFERASE OF THE CBP FAMILY 4 (HAC4) |
|  |
| HISTONE ACETYLTRANSFERASE OF THE CBP FAMILY 6 (HAC6) |
| AT1G55250 | Encodes one of two orthologous E3 ubiquitin ligases in Arabidopsis that are involved in monoubiquitination of histone H2B. | HISTONE MONO-UBIQUITINATION 2 (HUB2) | HISTONE MONO-UBIQUITINATION 2 (HUB2) |
| AT1G55805 | BolA-like family protein;(source:Araport11) | HOMOLOG OF E.COLI BOLA 1 (BOLA1) | HOMOLOG OF E.COLI BOLA 2 (BOLA2) |
|  |
| HOMOLOG OF E.COLI BOLA 1 (BOLA1) |
| AT1G54710 | autophagy 18h-like protein;(source:Araport11) | HOMOLOG OF YEAST AUTOPHAGY 18 (ATG18) H (ATG18H) | HOMOLOG OF YEAST AUTOPHAGY 18 (ATG18) H (ATATG18H) |
|  |
| HOMOLOG OF YEAST AUTOPHAGY 18 (ATG18) H (ATG18H) |
| AT1G54840 | Encodes an atypical member of the sHSP20 family that is involved in histone demethylation. Loss of function mutations show increased methylation. IMD2 co-localizes to the nucleus with, and physically interacts with, IMD1, a protein involved in RNA directed DNA methylation. IMD2 contains an alpha crystallin domain , that is required for its function. | INCREASED DNA METHYLATION 2 (IDM2) | (ROS5) |
|  |
| INCREASED DNA METHYLATION 2 (IDM2) |
| AT1G55110 | indeterminate(ID)-domain 7;(source:Araport11) | INDETERMINATE(ID)-DOMAIN 7 (IDD7) | (REDSTART) |
|  |
| INDETERMINATE(ID)-DOMAIN 7 (AtIDD7) |
|  |
| INDETERMINATE(ID)-DOMAIN 7 (IDD7) |
| AT1G55580 | Encodes a member of the GRAS family of putative transcriptional regulators. It is involved in the initiation of axillary meristems during both the vegetative and reproductive growth phases and functions upstream of REV and AXR1 in the regulation of shoot branching. | LATERAL SUPPRESSOR (LAS) | SCARECROW-LIKE 18 (SCL18) |
|  |
| LATERAL SUPPRESSOR (LAS) |
| AT1G55020 | lipoxygenase, a defense gene conferring resistance Xanthomonas campestris The mRNA is cell-to-cell mobile. | LIPOXYGENASE 1 (LOX1) | LIPOXYGENASE 1 (LOX1) |
|  |
| ARABIDOPSIS LIPOXYGENASE 1 (ATLOX1) |
| AT1G55540 | Nuclear pore complex protein;(source:Araport11) | LONO 1 (LNO1) | LONO 1 (LNO1) |
|  |
| EMBRYO DEFECTIVE 1011 (emb1011) |
| AT1G56070 | encodes a translation elongation factor 2-like protein that is involved in cold-induced translation. Mutations in this gene specifically blocks low temperature-induced transcription of cold-responsive genes but induces the expression of CBF genes and mutants carrying the recessive mutations fail to acclimate to cold and is freezing sensitive. | LOW EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 1 (LOS1) | LOW EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 1 (LOS1) |
|  |
| LOW EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 1 (AT1G56075.1) |
| AT1G56010 | Encodes a transcription factor involved in shoot apical meristem formation and auxin-mediated lateral root formation. The gene is thought not to be involved in stress responses (NaCl, auxins, ethylene). NAC1 (NAC1) | NAC DOMAIN CONTAINING PROTEIN 1 (NAC1) | ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN 22 (ANAC022) |
|  |
| NAC DOMAIN CONTAINING PROTEIN 1 (NAC1) |
| AT1G55370 | NDH-dependent cyclic electron flow 5;(source:Araport11) | NDH-DEPENDENT CYCLIC ELECTRON FLOW 5 (NDF5) | NDH-DEPENDENT CYCLIC ELECTRON FLOW 5 (NDF5) |
| AT1G54960 | member of MEKK subfamily | NPK1-RELATED PROTEIN KINASE 2 (NP2) | NPK1-RELATED PROTEIN KINASE 2 (ANP2) |
|  |
| NPK1-RELATED PROTEIN KINASE 2 (NP2) |
|  |
| MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASES 2 (MAPKKK2) |
| AT1G55180 | member of C2-PLD. subfamily Represents a phospholipase D (PLD) gene with four exons, hence it is a member of the alpha class. Its amino acid sequence is quite different from other PLDs, therefore it might possess unique structural and/or catalytic properties. | PHOSPHOLIPASE D ALPHA 4 (PLDEPSILON) | (PLDALPHA4) |
|  |
| PHOSPHOLIPASE D ALPHA 4 (PLDEPSILON) |
| AT1G55670 | Encodes subunit G of photosystem I, an 11-kDa membrane protein that plays an important role in electron transport between plastocyanin and PSI and is involved in the stability of the PSI complex. PSI-G subunit is bound to PSI-B and is in contact with Lhca1. The protein inserts into thylakoids by a direct or "spontaneous" pathway that does not involve the activities of any known chloroplast protein-targeting machinery. PSI-G appears to be directly or indirectly involved in the interaction between Photosystem I and plastocyanin. | PHOTOSYSTEM I SUBUNIT G (PSAG) | PHOTOSYSTEM I SUBUNIT G (PSAG) |
| AT1G55010 | Predicted to encode a PR (pathogenesis-related) protein. Belongs to the plant defensin (PDF) family with the following members: At1g75830/PDF1.1, At5g44420/PDF1.2a, At2g26020/PDF1.2b, At5g44430/PDF1.2c, At2g26010/PDF1.3, At1g19610/PDF1.4, At1g55010/PDF1.5, At2g02120/PDF2.1, At2g02100/PDF2.2, At2g02130/PDF2.3, At1g61070/PDF2.4, At5g63660/PDF2.5, At2g02140/PDF2.6, At5g38330/PDF3.1 and At4g30070/PDF3.2. | PLANT DEFENSIN 1.5 (PDF1.5) | PLANT DEFENSIN 1.5 (PDF1.5) |
| AT1G54940 | Encodes a xylan glucuronosyltransferase. | PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN 4 (PGSIP4) | GLUCURONIC ACID SUBSTITUTION OF XYLAN 4 (GUX4) |
|  |
| PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN 4 (PGSIP4) |
| AT1G55640 | prenylated RAB acceptor 1.G1;(source:Araport11) | PRENYLATED RAB ACCEPTOR 1.G1 (PRA1.G1) | PRENYLATED RAB ACCEPTOR 1.G1 (PRA1.G1) |
| AT1G54970 | encodes a proline-rich protein that is specifically expressed in the root. The mRNA is cell-to-cell mobile. | PROLINE-RICH PROTEIN 1 (PRP1) | PROLINE-RICH PROTEIN 1 (ATPRP1) |
|
| AT1G55480 | Encodes a member of a novel plant protein family containing a PDZ, a K-box, and a TPR motif. mRNA but not protein levels decrease after wounding. ZKT is phosphorylated at Thr and Ser residues after wounding. The mRNA is cell-to-cell mobile. | PROTEIN CONTAINING PDZ DOMAIN, A K-BOX DOMAIN, AND A TPR REGION (ZKT) | PROTEIN CONTAINING PDZ DOMAIN, A K-BOX DOMAIN, AND A TPR REGION (ZKT) |
| AT1G55150 | DEA(D/H)-box RNA helicase family protein;(source:Araport11) | RNA HELICASE 20 (RH20) | RNA HELICASE 20 (RH20) |
|  |
| (ATRH20) |
| AT1G55310 | Encodes a SR spliceosome protein that is localized to nuclear specks, interacts with SR45 and the U1-70K protein of the U1 snRNP, has sequence similar to human SC35 protein. Barta et al (2010) have proposed a nomenclature for Serine/Arginine-Rich Protein Splicing Factors (SR proteins): Plant Cell. 2010, 22:2926. | SC35-LIKE SPLICING FACTOR 33 (SCL33) | SC35-LIKE SPLICING FACTOR 33 (SCL33) |
|  |
| SC35-LIKE SPLICING FACTOR 33 (At-SCL33) |
|  |
| SC35-LIKE SPLICING FACTOR 33 (SR33) |
| AT1G55740 | seed imbibition 1;(source:Araport11) | SEED IMBIBITION 1 (SIP1) | SEED IMBIBITION 1 (AtSIP1) |
|  |
| RAFFINOSE SYNTHASE 1 (RS1) |
|  |
| SEED IMBIBITION 1 (SIP1) |
| AT1G55920 | Encodes a chloroplast/cytosol localized serine O-acetyltransferase involved in sulfur assimilation and cysteine biosynthesis. Expressed in the vascular system. The mRNA is cell-to-cell mobile. | SERINE ACETYLTRANSFERASE 2;1 (SERAT2;1) | SERINE ACETYLTRANSFERASE 1 (SAT1) |
|  |
| SERINE ACETYLTRANSFERASE 2 |
|  |
| 1 (SERAT2 |
|  |
| 1) |
|  |
| SERINE ACETYLTRANSFERASE 2 |
|  |
| SERINE ACETYLTRANSFERASE 5 (SAT5) |
| AT1G55570 | SKU5 similar 12;(source:Araport11) | SKU5 SIMILAR 12 (sks12) | SKU5 SIMILAR 12 (sks12) |
| AT1G55560 | SKU5 similar 14;(source:Araport11) | SKU5 SIMILAR 14 (sks14) | SKU5 SIMILAR 14 (sks14) |
| AT1G55040 | SED1 is a protein of unknown function that is located in the mitochondrion. sed1 mutants are embryo lethal. | SLOW EMBRYO DEVELOPMENT1 (SED1) | SLOW EMBRYO DEVELOPMENT1 (SED1) |
| AT1G55520 | TATA-box binding protein. Required for basal transcription. Acts facilitating the recruitment of TFIID to the promoter, which together with the RNA polymerase form the preinitiation complex. | TATA BINDING PROTEIN 2 (TBP2) | A. THALIANA TATA BINDING PROTEIN 2 (ATTBP2) |
|  |
| TATA BINDING PROTEIN 2 (TBP2) |
| AT1G55300 | TBP-associated factor 7;(source:Araport11) | TBP-ASSOCIATED FACTOR 7 (TAF7) | TBP-ASSOCIATED FACTOR 7 (TAF7) |
| AT1G54780 | Encodes a thylakoid lumen protein regulating photosystem II repair cycle. Has acid phosphatase activity. The mRNA is cell-to-cell mobile. | THYLAKOID LUMEN PROTEIN 18.3 (TLP18.3) | (ATTLP18.3) |
|  |
| THYLAKOID LUMEN PROTEIN 18.3 (TLP18.3) |
| AT1G55130 | Encodes an Arabidopsis Transmembrane nine (TMN) protein. Transmembrane nine (TM9) proteins are localized in the secretory pathway of eukaryotic cells and are involved in cell adhesion and phagocytosis. | TRANSMEMBRANE NINE 6 (TMN6) | TRANSMEMBRANE NINE 6 (TMN6) |
|  |
| (AtTMN6) |
| AT1G55060 | Ubiquitin-like gene, believed to be a pseudogene because of amino acid substitutions in 3 of the 5 ubiquitin repeats found in the UBQ12 gene product | UBIQUITIN 12 (UBQ12) | UBIQUITIN 12 (UBQ12) |
| AT1G55860 | encodes a ubiquitin-protein ligase containing a HECT domain. There are six other HECT-domain UPLs in Arabidopsis. | UBIQUITIN-PROTEIN LIGASE 1 (UPL1) | UBIQUITIN-PROTEIN LIGASE 1 (UPL1) |
| AT1G55810 | One of the homologous genes predicted to encode proteins with UPRT domains (Uracil phosphoribosyltransferase). Five of these genes (At5g40870, At3g27190, At1g55810, At4g26510 and At3g27440) show a high level of identity, and are annotated as also containing a N-terminal uracil kinase (UK) domain. These genes are referred to as UKL1 (UK-like 1), UKL2, UKL3, UKL4 and UKL5, resAAectively. | URIDINE KINASE-LIKE 3 (UKL3) | URIDINE KINASE-LIKE 3 (UKL3) |
| AT1G55600 | member of WRKY Transcription Factor; Group I. It has WRKY domain at its N terminal end and zinc-finger like motif. | WRKY DNA-BINDING PROTEIN 10 (WRKY10) | WRKY DNA-BINDING PROTEIN 10 (WRKY10) |
|  |
| MINISEED 3 (MINI3) |
|  |
|  |
| AT1G55910 | member of Putative zinc transporter ZIP2 - like family | ZINC TRANSPORTER 11 PRECURSOR (ZIP11) | ZINC TRANSPORTER 11 PRECURSOR (ZIP11) |

Table 2 Genes surrounding QTL5. All genes found surrounding QTL5 300Kb up or downstream of significant SNP bulk downloaded from arabdiopsis.org

|  |  |  |  |
| --- | --- | --- | --- |
| **Locus Identifier** | **Gene Description** | **Primary Gene Symbol** | **All Gene Symbols** |
| AT5G19510 | Translation elongation factor EF1B/ribosomal protein S6 family protein;(source:Araport11) |  |  |
| AT5G18490 | vacuolar sorting-associated protein (DUF946);(source:Araport11) |  |  |
| AT5G18970 | AWPM-19-like family protein;(source:Araport11) |  |  |
| AT5G18633 | transposable\_element\_gene;(source:Araport11);non-LTR retrotransposon family (LINE), has a 1.8e-39 P-value blast match to GB:NP\_038602 L1 repeat, Tf subfamily, member 18 (LINE-element) (Mus musculus);(source:TAIR10) |  |  |
| AT5G18760 | RING/U-box superfamily protein;(source:Araport11) |  |  |
| AT5G19540 | thermosome subunit gamma;(source:Araport11) |  |  |
| AT5G18600 | Thioredoxin superfamily protein;(source:Araport11) |  |  |
| AT5G19830 | Peptidyl-tRNA hydrolase family protein;(source:Araport11) |  |  |
| AT5G18630 | alpha/beta-Hydrolases superfamily protein;(source:Araport11) |  |  |
| AT5G19420 | Regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain-containing protein;(source:Araport11) |  |  |
| AT5G19015 | transposable\_element\_gene;(source:Araport11);CACTA-like transposase family (Tnp2/En/Spm), has a 8.7e-16 P-value blast match to GB:CAA40555 TNP2 (CACTA-element) (Antirrhinum majus);(source:TAIR10) |  |  |
| AT5G18661 | This gene encodes a small protein and has either evidence of transcription or purifying selection. |  |  |
| AT5G19025 | Ribosomal protein L34e superfamily protein;(source:Araport11) |  |  |
| AT5G18750 | DNAJ heat shock N-terminal domain-containing protein;(source:Araport11) |  |  |
| AT5G18850 | Low-density receptor-like protein;(source:Araport11) |  |  |
| AT5G19590 | DUF538 family protein (Protein of unknown function, DUF538);(source:Araport11) |  |  |
| AT5G19710 | histidine containing phosphotransfer protein;(source:Araport11) |  |  |
| AT5G18740 | hypothetical protein (DUF3444);(source:Araport11) |  |  |
| AT5G19750 | Peroxisomal membrane 22 kDa (Mpv17/PMP22) family protein;(source:Araport11) |  |  |
| AT5G18636 | Ta11-like non-LTR retrotransposon;(source:Araport11) |  |  |
| AT5G18720 | DNAJ heat shock amino-terminal domain protein, putative (DUF3444);(source:Araport11) |  |  |
| AT5G19640 | Major facilitator superfamily protein;(source:Araport11) |  |  |
| AT5G18880 | RNA-directed DNA polymerase (reverse transcriptase)-related family protein;(source:Araport11) |  |  |
| AT5G18780 | F-box/RNI-like superfamily protein;(source:Araport11) |  |  |
| AT5G18748 | This gene encodes a small protein and has either evidence of transcription or purifying selection. |  |  |
| AT5G19570 | transmembrane protein;(source:Araport11) |  |  |
| AT5G18640 | alpha/beta-Hydrolases superfamily protein;(source:Araport11) |  |  |
| AT5G18460 | carboxyl-terminal peptidase (DUF239);(source:Araport11) |  |  |
| AT5G19810 | Proline-rich extensin-like family protein;(source:Araport11) |  |  |
| AT5G18770 | F-box/FBD-like domains containing protein;(source:Araport11) |  |  |
| AT5G19095 | pre-tRNA tRNA-Gly (anticodon: GCC);(source:Araport11, TAIR10) |  |  |
| AT5G19670 | Exostosin family protein;(source:Araport11) |  |  |
| AT5G18980 | ARM repeat superfamily protein;(source:Araport11) |  |  |
| AT5G18590 | Galactose oxidase/kelch repeat superfamily protein;(source:Araport11) |  |  |
| AT5G19730 | Pectin lyase-like superfamily protein;(source:Araport11) |  |  |
| AT5G19473 | RPM1-interacting protein 4 (RIN4) family protein;(source:Araport11) |  |  |
| AT5G19240 | Glycoprotein membrane precursor GPI-anchored;(source:Araport11) |  |  |
| AT5G19370 | rhodanese-like domain-containing protein / PPIC-type PPIASE domain-containing protein;(source:Araport11) |  |  |
| AT5G19221 | Potential natural antisense gene, locus overlaps with AT5G19220 |  |  |
| AT5G18840 | Major facilitator superfamily protein;(source:Araport11) |  |  |
| AT5G18550 | Zinc finger C-x8-C-x5-C-x3-H type family protein;(source:Araport11) |  |  |
| AT5G18920 | Cox19-like CHCH family protein;(source:Araport11) |  |  |
| AT5G19120 | Eukaryotic aspartyl protease family protein;(source:Araport11) |  |  |
| AT5G19090 | Heavy metal transport/detoxification superfamily protein;(source:Araport11) |  |  |
| AT5G18755 | pre-tRNA tRNA-Pro (anticodon: TGG);(source:Araport11, TAIR10) |  |  |
| AT5G19030 | RNA-binding (RRM/RBD/RNP motifs) family protein;(source:Araport11) |  |  |
| AT5G18710 | DNAJ heat shock amino-terminal domain protein, putative (DUF3444);(source:Araport11) |  |  |
| AT5G19100 | Eukaryotic aspartyl protease family protein;(source:Araport11) |  |  |
| AT5G18800 | Cox19-like CHCH family protein;(source:Araport11) |  |  |
| AT5G19110 | Eukaryotic aspartyl protease family protein;(source:Araport11) |  |  |
| AT5G19340 | hypothetical protein;(source:Araport11) |  |  |
| AT5G18790 | Ribosomal protein L33 family protein;(source:Araport11) |  |  |
| AT5G18510 | Aminotransferase-like, plant mobile domain family protein;(source:Araport11) |  |  |
| AT5G19350 | RNA-binding (RRM/RBD/RNP motifs) family protein;(source:Araport11) |  |  |
| AT5G18990 | Pectin lyase-like superfamily protein;(source:Araport11) |  |  |
| AT5G19760 | Encodes a novel mitochondrial carrier capable of transporting both dicarboxylates (such as malate, oxaloacetate, oxoglutarate, and maleate) and tricarboxylates (such as citrate, isocitrate, cis-aconitate, and trans-aconitate). |  |  |
| AT5G19480 | mediator of RNA polymerase II transcription subunit;(source:Araport11) |  |  |
| AT5G18940 | Mo25 family protein;(source:Araport11) |  |  |
| AT5G19500 | Encodes a putative amino acid transporter that localizes to the chloroplast inner envelope membrane. |  |  |
| AT5G18900 | 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein;(source:Araport11) |  |  |
| AT5G19210 | P-loop containing nucleoside triphosphate hydrolases superfamily protein;(source:Araport11) |  |  |
| AT5G19172 | Encodes a defensin-like (DEFL) family protein. |  |  |
| AT5G19580 | glyoxal oxidase-related protein;(source:Araport11) |  |  |
| AT5G19175 | Encodes a defensin-like (DEFL) family protein. |  |  |
| AT5G19060 | cytochrome P450 family protein;(source:Araport11) |  |  |
| AT5G18470 | Curculin-like (mannose-binding) lectin family protein;(source:Araport11) |  |  |
| AT5G19170 | NEP-interacting protein, putative (DUF239);(source:Araport11) |  |  |
| AT5G19315 | Encodes a defensin-like (DEFL) family protein. |  |  |
| AT5G18475 | Pentatricopeptide repeat (PPR) superfamily protein;(source:Araport11) |  |  |
| AT5G19050 | alpha/beta-Hydrolases superfamily protein;(source:Araport11) |  |  |
| AT5G19860 | transmembrane protein, putative (Protein of unknown function, DUF538);(source:Araport11) |  |  |
| AT5G19230 | Glycoprotein membrane precursor GPI-anchored;(source:Araport11) |  |  |
| AT5G19490 | Histone superfamily protein;(source:Araport11) |  |  |
| AT5G19720 | Ribosomal protein L25/Gln-tRNA synthetase, anti-codon-binding domain-containing protein;(source:Araport11) |  |  |
| AT5G19070 | SNARE associated Golgi protein family;(source:Araport11) |  |  |
| AT5G18730 | DNAJ heat shock amino-terminal domain protein;(source:Araport11) |  |  |
| AT5G19250 | Glycoprotein membrane precursor GPI-anchored;(source:Araport11) |  |  |
| AT5G19850 | alpha/beta-Hydrolases superfamily protein;(source:Araport11) |  |  |
| AT5G18500 | Protein kinase superfamily protein;(source:Araport11) |  |  |
| AT5G19165 | transposable\_element\_gene;(source:Araport11);copia-like retrotransposon family, has a 1.1e-237 P-value blast match to GB:CAA72989 open reading frame 1 (Ty1\_Copia-element) (Brassica oleracea);(source:TAIR10) |  |  |
| AT5G18910 | Protein kinase superfamily protein;(source:Araport11) |  |  |
| AT5G19680 | Leucine-rich repeat (LRR) family protein;(source:Araport11) |  |  |
| AT5G19270 | reverse transcriptase-like protein;(source:Araport11) |  |  |
| AT5G19840 | 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein;(source:Araport11) |  |  |
| AT5G19300 | methyltransferase C9orf114 protein;(source:Araport11) |  |  |
| AT5G19630 | alpha/beta-Hydrolases superfamily protein;(source:Araport11) |  |  |
| AT5G18950 | Tetratricopeptide repeat (TPR)-like superfamily protein;(source:Araport11) |  |  |
| AT5G19151 | hypothetical protein;(source:Araport11) |  |  |
| AT5G19485 | transferases/nucleotidyltransferase;(source:Araport11) |  |  |
| AT5G19130 | GPI transamidase component family protein / Gaa1-like family protein;(source:Araport11) |  |  |
| AT5G18540 | E3 ubiquitin-protein ligase;(source:Araport11) |  |  |
| AT5G19190 | hypothetical protein;(source:Araport11) |  |  |
| AT5G19097 | transposable\_element\_gene;(source:Araport11);copia-like retrotransposon family, has a 1.4e-284 P-value blast match to GB:CAA72989 open reading frame 1 (Ty1\_Copia-element) (Brassica oleracea);(source:TAIR10) |  |  |
| AT5G19440 | similar to Eucalyptus gunnii alcohol dehydrogenase of unknown physiological function (GI:1143445), apple tree, PIR:T16995; NOT a cinnamyl-alcohol dehydrogenase |  |  |
| AT5G19150 | AT5G19150 is a dehydratase that converts (S)-NAD(P)HX to NAD(P)H. |  |  |
| AT5G18450 | encodes a member of the DREB subfamily A-2 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are eight members in this subfamily including DREB2A AND DREB2B that are involved in response to drought. |  |  |
| AT5G19140 | aluminum induced protein with YGL and LRDR motifs;(source:Araport11) | (AILP1) | (AILP1)  (AtAILP1) |
| AT5G19430 | Encodes a C3HC4-type RING finger E3 ubiquitin ligase of the RING/U-box superfamily whose expression is responsive to both phosphate (Pi) and phosphite (Phi) in both roots and shoots. | (C3HC4) | (C3HC4) |
| AT5G19290 | alpha/beta-Hydrolases superfamily protein;(source:Araport11) | (MAGL16) | (MAGL16) |
| AT5G18610 | Encodes a receptor-like cytoplasmic kinase that is an immediate downstream component of the chitin receptor CERK1 and contributes to the regulation of chitin-induced immunity. | (PBL27) | (PBL27) |
| AT5G19390 | Encodes a protein with similarity to REN1, a Rho GTPase activating protein.It is cytoplasmic and plasma membrane associated in interphase, but during mitosis localizes to the CDZ/CDS in a POK-dependent manner. | (PHGAP2) | (PHGAP2) |
| AT5G18560 | Encodes PUCHI, a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole. PUCHI is required for morphogenesis in the early lateral root primordium of Arabidopsis. Expressed in early floral meristem (stage 1 to 2). Required for early floral meristem growth and for bract suppression. Triple mutant with bop1 and bop2 displays a strong defect in the determination of floral meristem identity with reduced LFY expression and the lack of AP1 expression. | (PUCHI) | (PUCHI) |
| AT5G18700 | Encodes a microtubule-associated kinase-like protein RUNKEL (RUK). Contains a putative serine/threonine kinase domain and a microtubule-binding domain. RUK directly binds to microtubules in vitro and colocalizes with mitotic preprophase band, spindle, and phragmoplast in vivo. Required for cell plate expansion in cytokinesis. | (RUK) | RUNKEL (RUK)  (RUK)  EMBRYO DEFECTIVE 3013 (EMB3013) |
| AT5G19400 | Encodes SMG7, a protein that possesses an evolutionarily conserved EST1 domain and exhibits strong homology to human SMG6 (EST1A) and SMG7 (EST1C) proteins. SMG7 plays an evolutionarily conserved role in nonsense-mediated RNA decay (NMD). Required for exit from meiosis. Hypomorphic smg7 alleles render mutant plants sterile by causing an unusual cell-cycle arrest in anaphase II that is characterized by delayed chromosome decondensation and aberrant rearrangement of the meiotic spindle. Disruption of SMG7 causes embryonic lethality. | (SMG7) | (SMG7) |
| AT5G19820 | Encodes an importin that transports HYL1, a component of the microprocessor, from the cytoplasm to the nucleus to constitute functional microprocessor, thereby affecting miRNA processing. Knockdown amiR mutants significantly reduced nuclear portion of HYL1 protein and correspondingly compromised the pri-miRNA processing in the nucleus. | (KARYOPHERIN ENABLING THE TRANSPORT OF THE CYTOPLASMIC HYL1 (KETCH1) | EMBRYO DEFECTIVE 2734 (emb2734)  (KARYOPHERIN ENABLING THE TRANSPORT OF THE CYTOPLASMIC HYL1 (KETCH1) |
| AT5G19530 | Encodes a spermine synthase. Required for internode elongation and vascular development, specifically in the mechanism that defines the boundaries between veins and nonvein regions. This mechanism may be mediated by polar auxin transport. Though ACL5 has been shown to function as a spermine synthase in E. coli, an ACL5 knockout has no effect on the endogenous levels of free and conjugated polyamines in Arabidopsis, suggesting that ACL5 may have a very specific or altogether different in vivo function. | ACAULIS 5 (ACL5) | ACAULIS 5 (ACL5) |
| AT5G19220 | Encodes the large subunit of ADP-glucose pyrophosphorylase which catalyzes the first, rate limiting step in starch biosynthesis. The large subunit plays a regulatory role whereas the small subunit (ApS) is the catalytic isoform. Four isoforms (ApL1-4) have been identified. ApL1 is the major large subunit isoform present in leaves. Mutational analysis of APS1 suggests that APL1 and APL2 can compensate for loss of APS1 catalytic activity,suggesting both have catalytic as well as regulatory functions. | ADP GLUCOSE PYROPHOSPHORYLASE LARGE SUBUNIT 1 (APL1) | ADP GLUCOSE PYROPHOSPHORYLASE LARGE SUBUNIT 1 (APL1)  ADP GLUCOSE PYROPHOSPHORYLASE 2 (ADG2) |
| AT5G18690 | arabinogalactan protein 25;(source:Araport11) | ARABINOGALACTAN PROTEIN 25 (AGP25) | ARABIDOPSIS THALIANA ARABINOGALACTAN PROTEINS 25 (ATAGP25)  ARABINOGALACTAN PROTEIN 25 (AGP25) |
| AT5G19330 | Encodes an armadillo repeat protein involved in the abscisic acid response. The protein interacts with a transcription factor, ABF2, which controls ABA-dependent gene expression via the G-box-type ABA-responsive elements. | ARM REPEAT PROTEIN INTERACTING WITH ABF2 (ARIA) | ARM REPEAT PROTEIN INTERACTING WITH ABF2 (ARIA) |
| AT5G19550 | Nitrogen metabolism. Major cytosolic isoenzyme controlling aspartate biosynthesis in the light. The mRNA is cell-to-cell mobile. | ASPARTATE AMINOTRANSFERASE 2 (ASP2) | ASPARTATE AMINOTRANSFERASE 2 (ASP2)  ASPARTATE AMINOTRANSFERASE 2 (AAT2) |
| AT5G19410 | ABC-2 type transporter family protein;(source:Araport11) | ATP-BINDING CASSETTE G23 (ABCG23) | ATP-BINDING CASSETTE G23 (ABCG23) |
| AT5G18670 | putative beta-amylase BMY3 (BMY3) | BETA-AMYLASE 3 (BMY3) | BETA-AMYLASE 9 (BAM9)  BETA-AMYLASE 3 (BMY3) |
| AT5G19000 | Encodes a member of the MATH-BTB domain proteins (BPMs) that directly interact with and target for proteasomal degradation the class I homeobox-leucine zipper (HD-ZIP) transcription factor ATHB6. Known members include AT5G19000 (BPM1), AT3G06190 (BPM2), AT2G39760 (BPM3), AT3G03740 (BPM4), AT5G21010 (BPM5) and AT3G43700 (BPM6). | BTB-POZ AND MATH DOMAIN 1 (BPM1) | BTB-POZ AND MATH DOMAIN 1 (BPM1) |
| AT5G18930 | Adenosylmethionine decarboxylase family protein;(source:Araport11) | BUSHY AND DWARF 2 (BUD2) | (SAMDC4)  BUSHY AND DWARF 2 (BUD2) |
| AT5G19450 | calcium-dependent protein kinase (CDPK19) mRNA, complete | CALCIUM-DEPENDENT PROTEIN KINASE 19 (CDPK19) | (CPK8)  CALCIUM-DEPENDENT PROTEIN KINASE 19 (CDPK19) |
| AT5G19360 | member of Calcium Dependent Protein Kinase | CALCIUM-DEPENDENT PROTEIN KINASE 34 (CPK34) | (ATCPK34)  CALCIUM-DEPENDENT PROTEIN KINASE 34 (CPK34) |
| AT5G18520 | Encodes a candidate G-protein Coupled Receptor that is involved in the regulation of root growth by bacterial N-acyl-homoserine lactones (AHLs) and plays a role in mediating interactions between plants and microbes. The mRNA is cell-to-cell mobile. | CANDIDATE G-PROTEIN COUPLED RECEPTOR 7 (CAND7) | CANDIDATE G-PROTEIN COUPLED RECEPTOR 7 (CAND7) |
| AT5G18820 | Encodes a subunit of chloroplasts chaperonins that are involved in mediating the folding of newly synthesized, translocated, or stress-denatured proteins. Cpn60 subunits are: Cpn60alpha1 (At2g28000), AtCpn60alpha2 (At5g18820), AtCpn60beta1 (At1g55490), AtCpn60beta2 (At3g13470), AtCpn60beta3 (At5g56500), AtCpn60beta4 (At1g26230). | CHAPERONIN-60ALPHA2 (CPN60ALPHA2) | (CPNA2)  CHAPERONIN-60ALPHA2 (CPN60ALPHA2)  EMBRYO DEFECTIVE 3007 (EMB3007) |
| AT5G19310 | Encodes CHR23. Overexpression results in increased variability of growth and gene expression. | CHROMATIN REMODELING 23 (CHR23) | (ATCHR23)  CHROMATIN REMODELING 23 (CHR23) |
| AT5G18620 | Encodes a member of the A. thaliana imitation switch (AtISWI) subfamily of chromatin remodeling factors. Double mutation in CHR17 and CHR11 results in the loss of the evenly spaced nucleosome pattern in gene bodies, but does not affect nucleosome density. | CHROMATIN REMODELING FACTOR17 (CHR17) | CHROMATIN REMODELING FACTOR17 (CHR17) |
| AT5G19380 | Encodes one of the CRT-Like transporters (CLT1/AT5G19380, CLT2/AT4G24460, CLT3/AT5G12170). Required for glutathione homeostasis and stress responses. Mutants lacking these transporters are heavy metal-sensitive, glutathione(GSH)-deficient, and hypersensitive to Phytophthora infection. | CRT (CHLOROQUINE-RESISTANCE TRANSPORTER)-LIKE TRANSPORTER 1 (CLT1) | CRT (CHLOROQUINE-RESISTANCE TRANSPORTER)-LIKE TRANSPORTER 1 (CLT1) |
| AT5G19180 | Encodes a subunit of a RUB-activating enzyme analogous to the E1 ubiquitin-activating enzyme. ECR1 functions as a heterodimer with AXR1 to activate RUB, a ubiquitin-related protein. | E1 C-TERMINAL RELATED 1 (ECR1) | E1 C-TERMINAL RELATED 1 (ECR1) |
| AT5G19700 | Encodes a MATE transporter involved in leaf senescence and iron homeostasis. | EARLY LEAF SENESCENCE 1 (ELS1) | EARLY LEAF SENESCENCE 1 (ELS1) |
| AT5G18570 | Encodes AtObgC, a plant ortholog of bacterial Obg. AtObgC is a chloroplast-targeting GTPase essential for early embryogenesis. Mutations in this locus result in embryo lethality. The protein is dually localized in the stroma and the inner envelope membrane and is involved in thylakoid membrane biogenesis and functions primarily in plastid ribosome biogenesis during chloroplast development. | EMBRYO DEFECTIVE 269 (EMB269) | EMBRYO DEFECTIVE 269 (EMB269) (ATOBGC) (OBG A-2) EMBRYO DEFECTIVE 3138 (EMB3138) OBG-LIKE PROTEIN (ATOBGL) CHLOROPLASTIC SAR1 (CPSAR1) |
| AT5G19260 | A member of the FAF family proteins encoded by the FANTASTIC FOUR (FAF) genes: AT4G02810 (FAF1), AT1G03170 (FAF2), AT5G19260 (FAF3) and AT3G06020 (FAF4). FAFs have the potential to regulate shoot meristem size in Arabidopsis thaliana. FAFs can repress WUS, which ultimately leads to an arrest of meristem activity in FAF overexpressing lines. | FANTASTIC FOUR 3 (FAF3) | FANTASTIC FOUR 3 (FAF3) |
| AT5G18960 | FAR1-related sequence 12;(source:Araport11) | FAR1-RELATED SEQUENCE 12 (FRS12) | FAR1-RELATED SEQUENCE 12 (FRS12) |
| AT5G18580 | fass mutants have aberrant cell shapes due to defects in arrangement of cortical microtubules. Encodes a protein highly conserved in higher plants and similar in its C-terminal part to B' regulatory subunits of type 2A protein phosphatases. Interacts with an Arabidopsis type A subunit of PP2A in the yeast two-hybrid system. | FASS 1 (FASS) | FASS 1 (FASS) TONNEAU 2 (TON2) (FASS 2) (FS1) GORDO (GDO) EMBRYO DEFECTIVE 40 (EMB40) |
| AT5G19610 | GNOM-like 2;(source:Araport11) | GNOM-LIKE 2 (GNL2) | GNOM-LIKE 2 (GNL2) |
| AT5G19855 | Encodes a chloroplast stromal localized RbcX protein that acts as a chaperone in the folding of Rubisco. | HOMOLOGUE OF CYANOBACTERIAL RBCX 2 (RbcX2) | (AtRbcX2)  HOMOLOGUE OF CYANOBACTERIAL RBCX 2 (RbcX2) |
| AT5G19800 | hydroxyproline-rich glycoprotein family protein;(source:Araport11) | HYDROXYPROLINE-RICH GLYCOPROTEIN 2 (HRGP2) | HYDROXYPROLINE-RICH GLYCOPROTEIN 2 (HRGP2)  (ATHRGP2) |
| AT5G19040 | Encodes cytokinin synthase. | ISOPENTENYLTRANSFERASE 5 (IPT5) | ARABIDOPSIS THALIANA ISOPENTENYLTRANSFERASE 5 (ATIPT5)  ISOPENTENYLTRANSFERASE 5 (IPT5) |
| AT5G19280 | kinase associated protein phosphatase composed of three domains: an amino-terminal signal anchor, a kinase interaction (KI) domain, and a type 2C protein phosphatase catalytic region | KINASE ASSOCIATED PROTEIN PHOSPHATASE (KAPP) | ROOT ATTENUATED GROWTH 1 (RAG1)  KINASE ASSOCIATED PROTEIN PHOSPHATASE (KAPP) |
| AT5G19740 | LAMP is an AMP paralog that overlaps in expression within the vascular system. Along with LAMP it suppresses meristem activity within the peripheral zone of the shoot apical meristem. LAMP is localized to the endoplasmic reticulum. | LIKE AMP 1 (LAMP1) | LIKE AMP 1 (LAMP1) |
| AT5G19080 | Paralog of LOG2 (At3g09770), a ubiquitin ligase that regulates amino acid export. | LOG2-LIKE UBIQUITIN LIGASE3 (LUL3) | VRUBC1 BINDING PARTNER 1 (VBP1)  LOG2-LIKE UBIQUITIN LIGASE3 (LUL3)  (ATVBP1) |
| AT5G19520 | mechanosensitive channel of small conductance-like 9;(source:Araport11) | MECHANOSENSITIVE CHANNEL OF SMALL CONDUCTANCE-LIKE 9 (MSL9) | MECHANOSENSITIVE CHANNEL OF SMALL CONDUCTANCE-LIKE 9 (MSL9)  (ATMSL9) |
| AT5G19020 | Encodes a pentatricopeptide repeat protein (PPR) protein involved in mitochondrial mRNA editing. | MITOCHONDRIAL EDITING FACTOR 18 (MEF18) | MITOCHONDRIAL EDITING FACTOR 18 (MEF18) |
| AT5G19010 | member of MAP Kinase | MITOGEN-ACTIVATED PROTEIN KINASE 16 (MPK16) | MITOGEN-ACTIVATED PROTEIN KINASE 16 (MPK16) |
| AT5G18650 | Encodes a RING-type E3 ubiquitin ligase that interacts with and ubiquitinates MYB30, leads to MYB30 proteasomal degradation and downregulation of its transcriptional activity. Since MYB30 is a positive regulator of Arabidopsis HR and defence responses, MIEL1 is involved in the negative regulation of these processes. The mRNA is cell-to-cell mobile. | MYB30-INTERACTING E3 LIGASE 1 (MIEL1) | MYB30-INTERACTING E3 LIGASE 1 (MIEL1) |
| AT5G18860 | Encodes a purine nucleoside hydrolase active in the apoplast. It might play a role in salvaging extracellular ATP. NSH3 transcript levels rise in response to jasmonic acid and wounding. | NUCLEOSIDE HYDROLASE 3 (NSH3) | NUCLEOSIDE HYDROLASE 3 (NSH3) |
| AT5G18890 | Inosine-uridine preferring nucleoside hydrolase family protein;(source:Araport11) | NUCLEOSIDE HYDROLASE 4 (NSH4) | NUCLEOSIDE HYDROLASE 4 (NSH4) |
| AT5G18870 | Inosine-uridine preferring nucleoside hydrolase family protein;(source:Araport11) | NUCLEOSIDE HYDROLASE 5 (NSH5) | NUCLEOSIDE HYDROLASE 5 (NSH5) |
| AT5G19460 | nudix hydrolase homolog 20;(source:Araport11) | NUDIX HYDROLASE HOMOLOG 20 (NUDT20) | NUDIX HYDROLASE HOMOLOG 20 (NUDT20) NUDIX HYDROLASE HOMOLOG 20 (atnudt20) |
| AT5G19470 | nudix hydrolase homolog 24;(source:Araport11) | NUDIX HYDROLASE HOMOLOG 24 (NUDT24) | NUDIX HYDROLASE HOMOLOG 24 (atnudt24) NUDIX HYDROLASE HOMOLOG 24 (NUDX24) NUDIX HYDROLASE HOMOLOG 24 (NUDT24) |
| AT5G19650 | ovate family protein 8;(source:Araport11) | OVATE FAMILY PROTEIN 8 (OFP8) | (ATOFP8) OVATE FAMILY PROTEIN 8 (OFP8) |
| AT5G18660 | Encodes a protein with 3,8-divinyl protochlorophyllide a 8-vinyl reductase activity. Mutants accumulate divinyl chlorophyll rather than monovinyl chlorophyll. | PALE-GREEN AND CHLOROPHYLL B REDUCED 2 (PCB2) | PALE-GREEN AND CHLOROPHYLL B REDUCED 2 (PCB2) |
| AT5G18480 | Encodes an IPC (inositol phosphorylceramide) glucuronosyltransferase. Defects in transmission via the pollen are evident but the defect in transmission through the male gametophyte is not due to improper pollen development or inability of pollen tubes to germinate and grow. Using a pollen specific complementation strategy to obtain homozygotes, loss of function results in constitutive hypersensitive response and severe growth defects. | PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN 6 (PGSIP6) | INOSITOL PHOSPHORYLCERAMIDE GLUCURONOSYLTRANSFERASE 1 (IPUT1)  PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN 6 (PGSIP6) |
| AT5G19320 | Encodes RAN GTPase activating protein 2. The protein is localized to the nuclear envelope during interphase. | RAN GTPASE ACTIVATING PROTEIN 2 (RANGAP2) | RAN GTPASE ACTIVATING PROTEIN 2 (RANGAP2) |
| AT5G19790 | encodes a member of the ERF (ethylene response factor) subfamily B-6 of ERF/AP2 transcription factor family (RAP2.11). The protein contains one AP2 domain. There are 12 members in this subfamily including RAP2.11. | RELATED TO AP2 11 (RAP2.11) | RELATED TO AP2 11 (RAP2.11) |
| AT5G19560 | Encodes a member of KPP-like gene family, homolog of KPP (kinase partner protein) gene in tomato. Also a member of the RopGEF (guanine nucleotide exchange factor) family, containing the novel PRONE domain (plant-specific Rop nucleotide exchanger), which is exclusively active towards members of the Rop subfamily. | ROP (RHO OF PLANTS) GUANINE NUCLEOTIDE EXCHANGE FACTOR 10 (ROPGEF10) | ARABIDOPSIS THALIANA ROP UANINE NUCLEOTIDE EXCHANGE FACTOR 10 (ATROPGEF10) GUANINE EXCHANGE FACTOR 10 (GEF10) ROP (RHO OF PLANTS) GUANINE NUCLEOTIDE EXCHANGE FACTOR 10 (ROPGEF10) |
| AT5G18810 | encodes an SC35-like splicing factor of 28 kD localized to the nuclear specks. Barta et al (2010) have proposed a nomenclature for Serine/Arginine-Rich Protein Splicing Factors (SR proteins): Plant Cell. 2010, 22:2926. | SC35-LIKE SPLICING FACTOR 28 (SCL28) | SC35-LIKE SPLICING FACTOR 28 (At-SCL28) SC35-LIKE SPLICING FACTOR 28 (SCL28) |
| AT5G19660 | S1P appears to function as a Golgi-localized subtilase and to help protect seedlings against salt and osmotic stress. The roots of s1p-3 mutants are hypersensitive to NaCl, KCl, LiCl, and mannitol. Several salt-stress responsive genes show weaker induction in an s1P-3 mutant background. The proteolytic cleavage of the bZIP17 transcription factor depends on S1P in vitro. And there is evidence that S1P can cleave bZIP17 in vitro. | SITE-1 PROTEASE (S1P) | (ATSBT6.1) SITE-1 PROTEASE (S1P) SITE-1 PROTEASE (ATS1P) |
| AT5G18830 | Encodes a member of the Squamosa Binding Protein family of transcriptional regulators. SPL7 is expressed highly in roots and appears to play a role in copper homeostasis. Mutants are hypersensitive to copper deficient conditions and display a retarded growth phenotype. SPL7 binds to the promoter of the copper responsive miRNAs miR398b and miR389c. | SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 7 (SPL7) | SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 7 (SPL7) |
| AT5G19690 | encodes an oligosaccharyl transferase involved response to high salt. Mutants are hypersensitive to high salt conditions The mRNA is cell-to-cell mobile. | STAUROSPORIN AND TEMPERATURE SENSITIVE 3-LIKE A (STT3A) | STAUROSPORIN AND TEMPERATURE SENSITIVE 3-LIKE A (STT3A) |
| AT5G19600 | Encodes sulfate transporter Sultr3;5. | SULFATE TRANSPORTER 3;5 (SULTR3;5) | SULFATE TRANSPORTER 3 5 (SULTR35) |
| AT5G19200 | Encodes one of the Arabidopsis proteins (At3g06060/TSC10A and At5g19200/TSC10B) with significant similarity to the yeast 3-ketodihydrosphinganine (3-KDS) reductase, Tsc10p. Both TSC10A and TSC10B are bona fide 3-KDS reductase as shown by complementation experiment in yeast. | TSC10B (TSC10B) | TSC10B (TSC10B) |
| AT5G18680 | Member of TLP family | TUBBY LIKE PROTEIN 11 (TLP11) | TUBBY LIKE PROTEIN 11 (TLP11) TUBBY LIKE PROTEIN 11 (AtTLP11) |
| AT5G19770 | tubulin 3 | TUBULIN ALPHA-3 (TUA3) | TUBULIN ALPHA-3 (TUA3) |
| AT5G19780 | Encodes an isoform of alpha tubulin. Closely related to adjacent gene TUA3 suggesting recent duplication. The mRNA is cell-to-cell mobile. | TUBULIN ALPHA-5 (TUA5) | TUBULIN ALPHA-5 (TUA5) |

Table 3 Candidate genes around QTL5 and why of interest. Descriptions from arabidopsis.org

|  |  |  |  |
| --- | --- | --- | --- |
| **locus** | **Decription** | **Other names** | **why of note** |
| AT5G18450 | encodes a member of the DREB subfamily A-2 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are eight members in this subfamily including DREB2A AND DREB2B that are involved in response to drought. |  | DREB subfamily |
| AT5G18470 | Curculin-like (mannose-binding) lectin family protein; FUNCTIONS IN: sugar binding; INVOLVED IN: response to karrikin; LOCATED IN: plant-type cell wall; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Curculin-like (mannose-binding) lectin (InterPro:IPR001480); BEST Arabidopsis thaliana protein match is: lectin protein kinase family protein (TAIR:AT1G67520.1) |  | carbohydrate binding |
| AT5G18480 | Encodes an IPC (inositol phosphorylceramide) glucuronosyltransferase. Defects in transmission via the pollen are evident but the defect in transmission through the male gametophyte is not due to improper pollen development or inability of pollen tubes to germinate and grow. | INOSITOL PHOSPHORYLCERAMIDE GLUCURONOSYLTRANSFERASE 1; IPUT1; PGSIP6; PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN 6 | carbohydrate biosynthesis |
| AT5G18510 | Aminotransferase-like, plant mobile domain family protein; CONTAINS InterPro DOMAIN/s: Aminotransferase-like, plant mobile domain (InterPro:IPR019557); BEST Arabidopsis thaliana protein match is: Aminotransferase-like, plant mobile domain family protein (TAIR:AT1G50830.1) |  | amino acid transferase |
| AT5G18580 | fass mutants have aberrant cell shapes due to defects in arrangement of cortical microtubules. Encodes a protein highly conserved in higher plants and similar in its C-terminal part to B' regulatory subunits of type 2A protein phosphatases. Interacts with an Arabidopsis type A subunit of PP2A in the yeast two-hybrid system. | EMB40; EMBRYO DEFECTIVE 40; FASS; FASS 1; FASS 2; FS1; GDO; GORDO; TON2; TONNEAU 2 | Cell shape |
| AT5G18620 | Encodes a member of the A. thaliana imitation switch (AtISWI) subfamily of chromatin remodeling factors. Double mutation in CHR17 and CHR11 results in the loss of the evenly spaced nucleosome pattern in gene bodies, but does not affect nucleosome density. | CHR17; CHROMATIN REMODELING FACTOR17 | RNA seq co2 spring |
| AT5G18660 | Encodes a protein with 3,8-divinyl protochlorophyllide a 8-vinyl reductase activity. Mutants accumulate divinyl chlorophyll rather than monovinyl chlorophyll. | PALE-GREEN AND CHLOROPHYLL B REDUCED 2; PCB2 | chlorophyl |
| AT5G18670 | putative beta-amylase BMY3 (BMY3) | BAM9; BETA-AMYLASE 3; BETA-AMYLASE 9; BMY3 | Starch |
| AT5G18700 | Encodes a microtubule-associated kinase-like protein RUNKEL (RUK). Contains a putative serine/threonine kinase domain and a microtubule-binding domain. RUK directly binds to microtubules in vitro and colocalizes with mitotic preprophase band, spindle, and phragmoplast in vivo. Required for cell plate expansion in cytokinesis. | EMB3013; EMBRYO DEFECTIVE 3013; RUK; RUNKEL | cell cycle, cell expansion |
| AT5G18840 | Major facilitator superfamily protein; FUNCTIONS IN: carbohydrate transmembrane transporter activity, sugar:hydrogen symporter activity; INVOLVED IN: transport, transmembrane transport; LOCATED IN: integral to membrane, membrane; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 7 growth stages; CONTAINS InterPro DOMAIN/s: Major facilitator superfamily (InterPro:IPR020846), General substrate transporter (InterPro:IPR005828), Sugar/inositol transporter (InterPro:IPR003663), Major facilitator superfamily, general substrate transporter (InterPro:IPR016196); BEST Arabidopsis thaliana protein match is: Major facilitator superfamily protein (TAIR:AT2G48020.1) |  | carbohydrate transporter |
| AT5G18930 | BUSHY AND DWARF 2 (BUD2); FUNCTIONS IN: adenosylmethionine decarboxylase activity; INVOLVED IN: spermidine biosynthetic process, spermine biosynthetic process, polyamine biosynthetic process; LOCATED IN: cellular\_component unknown; EXPRESSED IN: 11 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: S-adenosylmethionine decarboxylase, core (InterPro:IPR016067), S-adenosylmethionine decarboxylase (InterPro:IPR001985), S-adenosylmethionine decarboxylase, conserved site (InterPro:IPR018166), S-adenosylmethionine decarboxylase subgroup (InterPro:IPR018167); BEST Arabidopsis thaliana protein match is: S-adenosylmethionine decarboxylase (TAIR:AT3G02470.4) | BUD2; BUSHY AND DWARF 2; SAMDC4 | Auxin response |
| AT5G18960 | FAR1-related sequence 12 (FRS12); FUNCTIONS IN: zinc ion binding; INVOLVED IN: response to red or far red light; LOCATED IN: cellular\_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: MULE transposase, conserved domain (InterPro:IPR018289), Transcription factor, FAR1-related (InterPro:IPR004330), Zinc finger, PMZ-type (InterPro:IPR006564), Zinc finger, SWIM-type (InterPro:IPR007527); BEST Arabidopsis thaliana protein match is: FAR1-related sequence 7 (TAIR:AT3G06250.1) | FAR1-RELATED SEQUENCE 12; FRS12 | response to light |
| AT5G18990 | Pectin lyase-like superfamily protein; FUNCTIONS IN: pectinesterase activity; INVOLVED IN: cell wall modification; LOCATED IN: endomembrane system, cell wall, plant-type cell wall; EXPRESSED IN: central cell, flower; CONTAINS InterPro DOMAIN/s: Pectin lyase fold/virulence factor (InterPro:IPR011050), Pectinesterase, catalytic (InterPro:IPR000070), Pectin lyase fold (InterPro:IPR012334); BEST Arabidopsis thaliana protein match is: Pectin lyase-like superfamily protein (TAIR:AT3G24130.1) |  | cell wall modification |
| AT5G19010 | member of MAP Kinase | MITOGEN-ACTIVATED PROTEIN KINASE 16; MPK16 | MPK16 |
| AT5G19040 | Encodes cytokinin synthase. | ARABIDOPSIS THALIANA ISOPENTENYLTRANSFERASE 5; ATIPT5; IPT5; ISOPENTENYLTRANSFERASE 5 | cytokinin sythesis |
| AT5G19080 | Paralog of LOG2 (At3g09770), a ubiquitin ligase that regulates amino acid export. | ATVBP1; LOG2-LIKE UBIQUITIN LIGASE3; LUL3; VBP1; VRUBC1 BINDING PARTNER 1 | amino acid |
| AT5G19095 | tRNA-Gly (anticodon: GCC) |  | amino acid |
| AT5G19140 | AILP1; FUNCTIONS IN: molecular\_function unknown; INVOLVED IN: response to auxin stimulus, response to aluminum ion; LOCATED IN: plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; BEST Arabidopsis thaliana protein match is: Aluminium induced protein with YGL and LRDR motifs (TAIR:AT5G43830.1) | AILP1; ATAILP1 | Auxin response |
| AT5G19150 | pfkB-like carbohydrate kinase family protein; INVOLVED IN: biological\_process unknown; LOCATED IN: chloroplast; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Uncharacterised protein family, carbohydrate kinase-related (InterPro:IPR000631) |  | carbohydrate kinase |
| AT5G19160 | Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene family containing a plant-specific DUF231 (domain of unknown function) domain. TBL gene family has 46 members, two of which (TBR/AT5G06700 and TBL3/AT5G01360) have been shown to be involved in the synthesis and deposition of secondary wall cellulose, presumably by influencing the esterification state of pectic polymers. A nomenclature for this gene family has been proposed (Volker Bischoff & Wolf Scheible, 2010, personal communication). | TBL11; TRICHOME BIREFRINGENCE-LIKE 11 | Cell wall synthesis |
| AT5G19220 | Encodes the large subunit of ADP-glucose pyrophosphorylase which catalyzes the first, rate limiting step in starch biosynthesis. The large subunit plays a regulatory role whereas the small subunit (ApS) is the catalytic isoform. Four isoforms (ApL1-4) have been identified. ApL1 is the major large subunit isoform present in leaves. Mutational analysis of APS1 suggests that APL1 and APL2 can compensate for loss of APS1 catalytic activity,suggesting both have catalytic as well as regulatory functions. | ADG2; ADP GLUCOSE PYROPHOSPHORYLASE 2; ADP GLUCOSE PYROPHOSPHORYLASE LARGE SUBUNIT 1; APL1 | starch biosynthesis |
| AT5G19260 | A member of the FAF family proteins encoded by the FANTASTIC FOUR (FAF) genes: AT4G02810 (FAF1), AT1G03170 (FAF2), AT5G19260 (FAF3) and AT3G06020 (FAF4). FAFs have the potential to regulate shoot meristem size in Arabidopsis thaliana. FAFs can repress WUS, which ultimately leads to an arrest of meristem activity in FAF overexpressing lines. | FAF3; FANTASTIC FOUR 3 | meristem size |
| AT5G19320 | Encodes RAN GTPase activating protein 2. The protein is localized to the nuclear envelope during interphase. | RAN GTPASE ACTIVATING PROTEIN 2; RANGAP2 | cell cycle |
| AT5G19330 | Encodes an armadillo repeat protein involved in the abscisic acid response. The protein interacts with a transcription factor, ABF2, which controls ABA-dependent gene expression via the G-box-type ABA-responsive elements. | ARIA; ARM REPEAT PROTEIN INTERACTING WITH ABF2 | ABA response |
| AT5G19390 | Encodes a protein with similarity to REN1, a Rho GTPase activating protein.It is cytoplasmic and plasma membrane associated in interphase, but during mitosis localizes to the CDZ/CDS in a POK-dependent manner. | PHGAP2 | cell cycle |
| AT5G19400 | Encodes SMG7, a protein that possesses an evolutionarily conserved EST1 domain and exhibits strong homology to human SMG6 (EST1A) and SMG7 (EST1C) proteins. SMG7 plays an evolutionarily conserved role in nonsense-mediated RNA decay (NMD). Required for exit from meiosis. Hypomorphic smg7 alleles render mutant plants sterile by causing an unusual cell-cycle arrest in anaphase II that is characterized by delayed chromosome decondensation and aberrant rearrangement of the meiotic spindle. Disruption of SMG7 causes embryonic lethality. | SMG7 | cell cycle |
| AT5G19450 | calcium-dependent protein kinase (CDPK19) mRNA, complete | CALCIUM-DEPENDENT PROTEIN KINASE 19; CDPK19; CPK8 | CDPK |
| AT5G19500 | Encodes a putative amino acid transporter that localizes to the chloroplast inner envelope membrane. |  | amino acid |
| AT5G19530 | Encodes a spermine synthase. Required for internode elongation and vascular development, specifically in the mechanism that defines the boundaries between veins and nonvein regions. This mechanism may be mediated by polar auxin transport. Though ACL5 has been shown to function as a spermine synthase in E. coli, an ACL5 knockout has no effect on the endogenous levels of free and conjugated polyamines in Arabidopsis, suggesting that ACL5 may have a very specific or altogether different in vivo function. | ACAULIS 5; ACL5 | auxin |
| AT5G19540 | unknown protein; FUNCTIONS IN: molecular\_function unknown; INVOLVED IN: biological\_process unknown; LOCATED IN: chloroplast; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages |  | CO2 spring rna seq |
| AT5G19550 | Nitrogen metabolism. Major cytosolic isoenzyme controlling aspartate biosynthesis in the light. The mRNA is cell-to-cell mobile. | AAT2; ASP2; ASPARTATE AMINOTRANSFERASE 2 | aspartate synthesis |
| AT5G19660 | S1P appears to function as a Golgi-localized subtilase and to help protect seedlings against salt and osmotic stress. The roots of s1p-3 mutants are hypersensitive to NaCl, KCl, LiCl, and mannitol. Several salt-stress responsive genes show weaker induction in an s1P-3 mutant background. The proteolytic cleavage of the bZIP17 transcription factor depends on S1P in vitro. And there is evidence that S1P can cleave bZIP17 in vitro. | ATS1P; ATSBT6.1; S1P; SITE-1 PROTEASE | osmotic stress |
| AT5G19680 | Leucine-rich repeat (LRR) family protein; CONTAINS InterPro DOMAIN/s: Leucine-rich repeat (InterPro:IPR001611); BEST Arabidopsis thaliana protein match is: Leucine-rich repeat (LRR) family protein (TAIR:AT5G22320.2) |  | leucine rich repeats |
| AT5G19730 | Pectin lyase-like superfamily protein; FUNCTIONS IN: pectinesterase activity; INVOLVED IN: cell wall modification; LOCATED IN: cell wall, plant-type cell wall, cytoplasm; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Pectinesterase, active site (InterPro:IPR018040), Pectin lyase fold/virulence factor (InterPro:IPR011050), Pectinesterase, catalytic (InterPro:IPR000070), Pectin lyase fold (InterPro:IPR012334); BEST Arabidopsis thaliana protein match is: Pectin lyase-like superfamily protein (TAIR:AT2G36710.1) |  | cell wall |
| AT5G19760 | Encodes a novel mitochondrial carrier capable of transporting both dicarboxylates (such as malate, oxaloacetate, oxoglutarate, and maleate) and tricarboxylates (such as citrate, isocitrate, cis-aconitate, and trans-aconitate). |  | dicarboxylate transporter |
| AT5G19770 | tubulin 3 | TUA3; TUBULIN ALPHA-3 | cell expansion |
| AT5G19780 | Encodes an isoform of alpha tubulin. Closely related to adjacent gene TUA3 suggesting recent duplication. The mRNA is cell-to-cell mobile. | TUA5; TUBULIN ALPHA-5 | tubulin |
| AT5G19810 | Proline-rich extensin-like family protein; LOCATED IN: endomembrane system; BEST Arabidopsis thaliana protein match is: hydroxyproline-rich glycoprotein family protein (TAIR:AT5G19800.1) |  | extensin, cell size |
| AT5G19830 | Peptidyl-tRNA hydrolase family protein; FUNCTIONS IN: aminoacyl-tRNA hydrolase activity; INVOLVED IN: translation; LOCATED IN: mitochondrion; EXPRESSED IN: 16 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Peptidyl-tRNA hydrolase, conserved site (InterPro:IPR018171), Peptidyl-tRNA hydrolase (InterPro:IPR001328); BEST Arabidopsis thaliana protein match is: Peptidyl-tRNA hydrolase family protein (TAIR:AT5G16140.2) |  | t-rna function |