

Investigating the genetic basis of natural leaf shape
variation in *Arabidopsis thaliana*

Joseph Vaughan

PhD

University of York

Biology

September 2015

Abstract

Leaf shape varies considerably between natural accessions of *Arabidopsis thaliana*. In this project we aimed to identify the genetic basis of this naturally occurring leaf shape variation. To score natural variation in leaf shape, we used a geometric morphometric approach, recording the shape of leaves with co-ordinates. We identified and scored the major shape variations in datasets of leaf shape models using Principal Component Analysis (PCA). To identify loci associated with differences in leaf shape, we used genetic mapping. We scored leaf shape in the Bay-0 x Shahdara Recombinant Inbred Line (RIL) population and carried out Quantitative Trait Loci (QTL) mapping. We also scored leaf shape in multiple growth experiments at different harvest points for a collection of over three hundred natural *Arabidopsis* accessions, and carried out Genome Wide Association (GWA) mapping. We identified loci associated with variation in our leaf shape traits in both approaches. Indicative of leaf shape as a polygenic quantitative trait, associated loci were typically of small to medium effect. However, we did identify, associated with variation in a leaf size and margin morphology trait, one single large locus, suggesting that naturally occurring loci can also have a dramatic effect on leaf morphology. We also found correlations between leaf shape and hypocotyl length and leaf number, suggesting natural leaf shape variation may coincide with other changes in plant morphology. Several of the loci associated with leaf shape traits in our GWA mapping contained Nucleotide Binding Leucine Rich Repeat (NBLRR) genes, and so we tested for leaf shape differences in T-DNA insertion lines annotated for these associated NBLRR genes. We found several of these T-DNA lines had differences in leaf shape and hypocotyl length, and that some of these differences in morphology were specific to temperature and light conditions.

Table of contents

Abstract.....	2
Table of contents	3
List of figures	9
List of tables.....	13
Acknowledgements	14
Author's declaration	15
Chapter 1. Introduction	16
1.1.1 The diversity of leaf shape in flowering plants.....	16
1.1.2 Variation in leaf shape in differing environmental conditions.....	16
1.1.3 Leaf shape variation can be the result of genetic differences	18
1.1.4 Accurately quantifying differences in leaf shape	19
1.1.5 Mapping trait variation to natural genetic differences in plants	22
1.1.6 Naturally occurring leaf shape variation between <i>Arabidopsis</i> accessions	28
1.1.7 Investigating the genetic basis of leaf shape variation in <i>Arabidopsis</i>	30
Chapter 2. Materials and methods.....	32
2.1.1 Plant growth	32
2.1.2 Seed sterilisation and growth on agar plates	33
2.1.3 Crossing <i>Arabidopsis</i>	33
2.1.4 Epidermal cell size and shape analysis	34
2.1.5 PCR with genomic DNA and cDNA.....	35
2.1.6 Preparation of DNA.....	36
2.1.7 Preparation of RNA.....	36
2.1.8 Preparation of cDNA.....	37
2.1.9 Restriction digest for genotyping of PhyB-9 mutation.....	37
2.1.10 Gel electrophoresis.....	37

2.1.11	LeafAnalyser leaf shape analysis	38
2.1.12	Genetic mapping.....	39
2.1.13	Hypocotyl measurement	40
2.1.14	Error bars	40
2.1.15	Primer design.....	41
2.1.16	Plant material	42
Chapter 3.	Quantitative Trait Loci mapping in the Bay-0 x Shahdara population	44
3.1	Introduction.....	44
3.1.1	Quantitative Trait Loci mapping in the Bay-0 x Shahdara recombinant inbred line population to identify loci associated with natural leaf shape variation	44
3.1.2	Choosing growth conditions for the Bay-0 x Shahdara RILs.....	46
3.1.3	Software for QTL analysis	46
3.1.4	A comprehensive approach to leaf shape phenotyping.....	47
3.2	Results	48
3.2.1	Growth conditions for the Bay-0 x Shahdara population.....	48
3.2.2	Scoring leaf shape variation within the Bay-0 x Shahdara RIL population	48
3.2.3	Summary and Aims	49
3.2.4	High heritability of leaf shape suggests a genetic basis for this trait	53
3.2.5	Identifying loci significantly associated to variation in leaf shape	54
3.2.6	Several QTLs are associated with the average leaf shape score for each RIL ..	56
3.2.7	Identifying QTLs for heteroblasty variation.....	57
3.2.8	Margin morphology varies within the Bay-0 x Shahdara RILs	61
3.2.9	Comparing leaf shape at individual nodes of the leaf series.....	63
3.2.10	Differences in leaf shape and number between light and shade conditions...	64
3.2.11	Leaf shape analysis of the Bay-0 x Shahdara RIL population identifies several QTL	65
3.2.12	Epistatic interactions between leaf shape QTL	66
3.2.13	Confirming effect of QTLs and defining causative region with HIFs	68

3.2.14	HIF102 and HIF118 vary for leaf size and margin morphology	73
3.2.15	Variation in life history and silique number in HIF102 and HIF118.....	77
3.2.16	Variation in epidermal cell size and shape in HIF118 and HIF102.....	79
3.2.17	Network analysis to narrow list of candidate genes in the variant region between HIF102 and HIF118	80
3.2.18	A polymorphism in At5g43630 may be responsible for lIpPC2 variation in the Bay-0 x Shahdara RILs	83
3.3	Conclusions.....	86
3.3.1	Multiple QTLs detected for leaf shape traits within the Bay-0 x Shahdara RIL population	86
3.3.2	Some QTL effects are confirmed with a HIF approach.....	86
3.3.3	No QTLs were identified for leaf shape response to shade conditions.....	87
3.3.4	Light condition may affect the effect size and interaction between leaf shape QTL	88
3.3.5	Major trends in leaf size variation involve relative changes across the leaf series	88
3.3.6	Identification and characterisation of a margin morphology QTL	89
3.3.7	Loci controlling leaf shape in the Bay-0 x Shahdara RILs may have wider effects on plant morphology	91
3.3.8	Using a network analysis to identify candidate genes	92
Chapter 4.	Genome Wide Association mapping for leaf shape traits	94
4.1	Introduction.....	94
4.2	Results	97
4.2.1	Leaf shape varies considerably between natural accessions of <i>Arabidopsis</i> ...	97
4.2.2	Creating leaf shape datasets at new harvest points for GWA mapping	97
4.2.3	Mapping leaf shape differences between accessions at bolting to create the GWAS60 dataset.....	98
4.2.4	Scoring shape variation in late flowering accessions	102
4.2.5	Leaf number variation is associated with a commonly identified flowering time locus	103

4.2.6	A locus associated with leaf number variation is also associated with differences in leaf shape.....	106
4.2.7	Variation in shape between accessions at individual nodes of the leaf series	108
4.2.8	Mapping leaf shape variation in leaves harvested at bolting with the GWAS60 dataset	110
4.2.9	A leaf shape GWAS with a new harvest point	110
4.2.10	Capturing leaf shape variation early in the leaf series with the GWAS10 dataset	112
4.2.11	NBLRR genes underlie some loci associated with node specific leaf shape changes	112
4.2.12	Further investigation of shape variation at node six leaves.....	116
4.2.13	Investigating changes in shape across the leaf series with a heteroblasty PCA	116
4.2.14	NBLRR loci are associated with several phenotypes in GWAS10	119
4.2.15	A third growth experiment to further explore early leaf series shape variation	119
4.2.16	Variation in cotyledon shape and hypocotyl length in the natural accessions	122
4.2.17	NBLRR loci are associated with leaf shape traits in a separate growth experiment	123
4.2.18	Loci associated with variation in heteroblasty traits in GWAS09 contain NBLRR clusters	124
4.2.19	Multiple loci associated with variation in leaf shape across two separate growth experiments contain clusters of NBLRR genes	125
4.2.20	GWAS for leaf shape in <i>Arabidopsis</i> identifies loci for a variety of traits.....	126
4.3	Conclusions.....	127
4.3.1	Practical considerations for the scale of phenotyping experiments	127
4.3.2	Scoring leaf shape variation effectively in large datasets	128
4.3.3	A locus associated with variation in leaf shape and number	129

4.3.4	Incorporating our GWAS results into the wider project.....	130
Chapter 5.	Analysis of T-DNA insertion lines for NBLRR genes associated with natural leaf shape variation	134
5.1	Introduction.....	134
5.2	Results	137
5.2.1	Leaf shape varies amongst T-DNA insertion lines for candidate genes within GWAS loci	137
5.2.2	The T-DNA insertion line At1g72850-1 may have a condition specific phenotype.....	138
5.2.3	Characterisation of further T-DNA insertion lines for genes At5g45240 and At1g72840	141
5.2.4	PCR genotyping to confirm presence of T-DNA insertion in lines	146
5.2.5	Association of T-DNA line leaf shape differences to T-DNA insertions in F2 populations.....	154
5.2.6	T-DNA insertion lines for homologues of At1g72850 and At5g45240 vary in leaf shape and size.....	158
5.2.7	T-DNA insertions in some NBLRR genes are associated with differences in leaf shape	159
5.2.8	The At5g45240-1 insertion does not decrease llpPC2 score in a Phytochrome B mutant background	160
5.2.9	Hypocotyl length of T-DNA lines for some NBLRR genes may vary in response to different wavelengths of light	163
5.2.10	Leaf shape of some T-DNA lines for NBLRR genes vary between temperatures	168
5.2.11	Expression of At5g45240 in Col-0 and four T-DNA insertion lines	175
5.3	Conclusions.....	181
5.3.1	Analysis of T-DNA insertion lines for NBLRR genes associated with leaf shape traits in a GWAS.....	181
5.3.2	Leaf shape and size vary between different T-DNA insertion lines for At5g45240	182

5.3.3	Effect of the At5g45240-1 T-DNA insertion in a At1g72850-1 and PhyB-9 background	184
5.3.4	Further characterisation of NBLRR genes with possible leaf shape function	186
Chapter 6.	Discussion	189
6.1.1	Natural leaf shape variation	189
6.1.2	Strategies for mapping leaf shape variation in <i>Arabidopsis</i>	191
6.1.3	Leaf shape and variations in wider plant morphology	193
6.1.4	NBLRR genes associated with differences in leaf shape	193
6.1.5	Effects of environmental variation on leaf shape in T-DNA lines for NBLRR genes	194
6.1.6	Natural genetic variation in NBLRR genes across <i>Arabidopsis</i> accessions	196
Appendices	197
Abbreviations	211
References	213

List of figures

Figure 1-1 Principal Components capture major trends in variation	22
Figure 1-2 Linking genetic variation to trait variation	27
Figure 1-3 Natural Arabidopsis accessions	27
Figure 3-1 Rosettes of Bay-0 and Shahdara accessions	45
Figure 3-2 Difference in llpPC2 between Bay-0 and Shahdara.....	50
Figure 3-3 Difference in llpPC3 between Bay-0 and Shahdara.....	50
Figure 3-4 Difference in size between Bay-0 and Shahdara.....	51
Figure 3-5 Heteroblasty PCA on Bay-0 x Shahdara population	52
Figure 3-6 Leaf library PCs	54
Figure 3-7 QTLs for leaf shape and number traits in the Bay-0 x Shahdara RILs	55
Figure 3-8 Average llpPC2 score in light and shade treatment RILs.....	56
Figure 3-9 Difference in llpPC3 score between RILs.....	57
Figure 3-10 QTL plots for the first four heteroblasty PCs.....	59
Figure 3-11 Margin morphology variation	60
Figure 3-12 QTL for margin morphology	60
Figure 3-13 QTLs for llpPC2 variation at individual nodes.....	62
Figure 3-14 llpPC2 score between RILs grouped by genotype at a QTL.....	64
Figure 3-15 Relationship between leaf number and days to bolt.....	65
Figure 3-16 Days to bolt in Bay-0 Shahdara RILs between growth experiments	65
Figure 3-17 Epistatic relationship between QTL 1a and QTL 5a.....	67
Figure 3-18 Differences in llpPC2 within HIFs.....	69
Figure 3-19 Genetically varying regions within HIFs	70
Figure 3-20 llpPC2 effect of QTLs across RILs	72
Figure 3-21 HIF118 plants.....	75
Figure 3-22 HIF102 plants.....	75
Figure 3-23 Difference in leaf size between HIF118 plants.....	76
Figure 3-24 Difference between HIF102 plants.....	76
Figure 3-25 Genotype of HIF102 and HIF118 at margin morphology QTL.....	77
Figure 3-26 Difference in silique number within HIF102 and HIF118	78
Figure 3-27 Difference in days to bolt within HIF102 and HIF118	78
Figure 3-28 Epidermal cell area between HIFs.....	79
Figure 3-29 Shape differences between HIF abaxial cells	80
Figure 3-30 Shape differences between HIF abaxial cells	80

Figure 3-31 Example of network analysis results	82
Figure 3-32 Correlation of llpPC2 and hypocotyl data	84
Figure 3-33 Difference in llpPC2 score between HIF350 and rHIF47	85
Figure 3-34 Coincidence of QTLs for llpPC2 and hypocotyl traits	85
Figure 4-1 Leaves collected for GWAS60 dataset.....	99
Figure 4-2 Leaf shape changes in GWAS60 leaf shape dataset.....	100
Figure 4-3 Leaf library PCs	101
Figure 4-4 Correlations between leaf shape traits in the GWAS60 dataset.....	102
Figure 4-5 Locus associated with leaf number variation on chromosome 4	105
Figure 4-6 Locus associated with leaf number variation on chromosome 5	105
Figure 4-7 Locus associated with G60_pPC2 on chromosome 4.....	106
Figure 4-8 Difference in G60_pPC2 across the leaf series.....	107
Figure 4-9 Locus associated with llpPC2 variation on chromosome 1	107
Figure 4-10 Locus associated with llpPC3 variation on chromosome 5	109
Figure 4-11 Number of leaves collected for GWAS10 dataset.....	111
Figure 4-12 Locus associated with llpPC2 variation on chromosome 1	113
Figure 4-13 Node 6 specific leaf shape variation	113
Figure 4-14 Locus associated with variation in node 6 shape variation on chromosome 4 .	115
Figure 4-15 Locus associated with node 6 shape variation on chromosome 1	115
Figure 4-16 Heteroblasty PCs	117
Figure 4-17 A chromosome 5 locus associated with HetPC1 variation.....	118
Figure 4-18 Leaves harvested for GWAS09 dataset	121
Figure 4-19 Correlation of shape traits between leaves and cotyledons.....	121
Figure 4-20 Locus associated with llpPC3 variation in node 6 leaves	123
Figure 4-21 Major shape variations in leaves at nodes 1 and 2	124
Figure 5-1 Difference in llpPC2 between Col-0 and T-DNA lines At5g45240-1 and At1g72840-1	138
Figure 5-2 Difference in llpPC2 score between Col-0 and At1g72850-1.....	139
Figure 5-3 llpPC2 score of Col-0 and At1g72850-1 in two growth experiments	139
Figure 5-4 Rosettes of Col-0 and T-DNA line At1g72850-1	140
Figure 5-5 llpPC2 score for three T-DNA lines and Col-0 in two light conditions.....	140
Figure 5-6 llpPC2 score and size differences between Col-0 and T-DNA lines annotated for At1g72840	142
Figure 5-7 Rosettes of Col-0 and T-DNA lines At1g72840-1 and At1g72840-2	143
Figure 5-8 At5g45240 gene model	143

Figure 5-9 Difference in llpPC2 score between Col-0 and T-DNA lines for At5g45240.....	144
Figure 5-10 Difference in size between Col-0 and T-DNA lines for At5g45240.....	145
Figure 5-11 Image of Col-0 and T-DNA lines At5g45240-1 and At5g45240-2	146
Figure 5-12 Testing At1g72850-1 and At5g45240-1 insertion position	148
Figure 5-13 Testing At1g72840-1 insertion position	149
Figure 5-14 Testing At5g45240-2 insertion position	149
Figure 5-15 Testing insertion position in T-DNA lines At5g45240-2, At5g45240-3 and At5g45240-4	150
Figure 5-16 Testing insertion position in T-DNA lines annotated for At5g45240	151
Figure 5-17 Testing insertion position in T-DNA lines annotated for At5g45240	152
Figure 5-18 Difference in llpPC2 score between Col-0 and At5g45240-1 control lines	156
Figure 5-19 llpPC2 score in F2 plants with or without At5g45240-1 insertion	156
Figure 5-20 Difference in size between Col-0 and At5g45240-2 control lines.....	157
Figure 5-21 Difference in llpPC3 score between Col-0 and At5g45240-2 control lines	157
Figure 5-22 Size in F2 plants with or without At5g45240-2 insertion.....	158
Figure 5-23 llpPC3 score in F2 plants with or without At5g45240-2 insertion	158
Figure 5-24 Difference in llpPC2 score between Col-0 and At4g09420-1	159
Figure 5-25 llpPC2 score in At5g45240-1 PhyB-9 double mutant	162
Figure 5-26 llpPC3 score in At5g45240-1 PhyB-9 double mutant	162
Figure 5-27 Leaf size in At5g45240-1 PhyB-9 double mutant	162
Figure 5-28 Hypocotyl growth in blue light	165
Figure 5-29 Hypocotyl growth in blue light	165
Figure 5-30 Hypocotyl growth in red light.....	166
Figure 5-31 Hypocotyl growth in red light.....	166
Figure 5-32 Hypocotyl length in dark grown seedlings	167
Figure 5-33 Hypocotyl length in dark grown seedlings	168
Figure 5-34 Hypocotyl length in dark grown seedlings	168
Figure 5-35 Difference in llpPC2 between Col-0 and At5g45240-1 at two temperatures	171
Figure 5-36 Difference in llpPC3 between Col-0 and At5g45240-1 at two temperatures	171
Figure 5-37 Centroid size difference between Col-0 and At5g45240-1 at two temperatures	172
Figure 5-38 Centroid size difference between Col-0 and At5g45240-2 at two temperatures	172
Figure 5-39 Difference in llpPC2 between Col-0 and At4g09420-1 at two temperatures	173
Figure 5-40 Difference in llpPC3 between Col-0 and At4g09420-1 at two temperatures	173

Figure 5-41 Size difference between Col-0 and At4g09420-1 at two temperatures	174
Figure 5-42 Difference in lIpPC2 between Col-0 and At1g72850-1.....	174
Figure 5-43 At5g45240 gene model	176
Figure 5-44 Transcript levels between exons 1-2.....	177
Figure 5-45 Transcript levels between exons 2-3.....	178
Figure 5-46 Transcript levels between exons 3-4.....	179
Figure 5-47 Transcript levels between exons 5-6.....	179
Figure 5-48 Transcript levels between exons 8-9.....	180
Figure 5-49 Transcript levels between exons 9-10.....	180

List of tables

Table 2-1 Arabidopsis lines.....	42
Table 2-2 Primer sequences	43
Table 3-1 Estimated heriability.....	51
Table 3-2 QTLs identified.....	55
Table 3-3 Candidate genes identified with network analysis.....	82
Table 4-1 Loci identified in leaf shape GWAS.....	109
Table 4-2 Loci associated with leaf shape traits in GWAS10 and GWAS09.....	114
Table 5-1 T-DNA insertion lines annotated for NBLRR genes.....	137
Table 5-2 Confirming presence of insertions in At5g45240 T-DNA lines	153
Table 5-3 Summary of RT-PCR results using exon spanning primers over at5g45240.....	177

Acknowledgements

Thanks to my supervisor Richard Waites for general support, but also generally being a steady and calming influence throughout the PhD. Thanks to Petra for knowing the answer to every question and for her assistance in the project. Thanks to all the people who've been happy to entertain my fumbling scientific questions, Seth Davis, Mike Haydon and Calvin Dytham all offered great insights into their experiences and approaches to biology. Michael Schultze, Frans Maaithus and Lihong all also provided a place for amusing discussions during coffee time. Thanks to my TAP panel, Betsy Pownall and Gonzalo Blanco for helping me to improve in my report writing and presentation structuring. Thanks to the *Arabidopsis* community generally, as it would not have been possible to complete a project of this scope without the resources that are publicly available for this organism. I'd also like to thank all the fellow PhD students I've known over my time in York, it's been a really nice place to study in large part down to the friendly, welcoming and inquisitive nature of the students. In particular, for introducing to plant science research and helping me at the beginning of my PhD, Jo Hepworth, Vera Matser and Tom Brabbs. For being there whenever you need someone to fall into a bush or carry you home, Simon Fellgett. For keeping me company either at 45 Dale street or the Trafalgar, Tom Smith. For being a perfectly reasonable, if somewhat lacklustre replacement for the previous two Toms, Tom Hartley. For cathartic miserliness, and also upbeat enthusiasm, Hannah Brunson. For some of the best football and worst drinking, Adrian Speakman and Dan Bawdon. Also generally everyone who played for YSBL or Biology and managed to incorporate my below average ability into a greater and often surprisingly successful whole. I'd also like to thank Emma Lindsay, Eleanor Walton, Richard Maguire, Waheed Arshad, Juan Patishtan, Tom Irving and Rachael Oakenfull for general chatter and making L2 a fun place to work. Thanks also to my girlfriend Alicia for stopping me getting scurvy during the more intensive periods of my PhD, and my parents, for always seeming pleased with whatever I'm doing generally.

Author's declaration

The work presented in this thesis was carried out by Joseph Vaughan, except where otherwise stated.

Petra Stirnberg assisted with the harvesting of the Bay-0 x Shahdara population. The 10 accession leaf library reference set, used to create a set of Procrustes fitted Principal Components in this thesis was a product of Vera Matsler's PhD thesis and has been published in Kieffer et al., 2011. The Actin 2 primers used were designed by Michael Schultze.

This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1. Introduction

1.1.1 The diversity of leaf shape in flowering plants

Massive variation in leaf shape exists within flowering plants. Leaves can be relatively simple, formed as single lamina attached to a petiole such as for birch leaves. Leaves can be lobed, like those of oak trees, formed of a single lamina with an undulating margin. They can be compound, made up of multiple laminas formed along a central petiole or rachis as in horse chestnut trees and tomato plants. They can be complex, such as the intricate patterns found in carrot leaves. These are broad categorisations, and leaf shape varies dramatically within each of these groups. Not only do great differences in shape exist between species of plants, but leaf shape within a species or even a single individual also varies considerably.

Natural leaf shape variation is particularly intriguing due to the sessile nature of plants and the inherent link between the form of a leaf and its function. Leaves intercept light, harvesting this energy for growth. As such, interception of optimal light is critical. As most plants are rooted to a single position, they are limited in their range of movement, and exposed to great environmental variation. It may be that some variation in leaf shape is adaptive for different environmental and ecological conditions, and so it is possible variation in leaf shape among plants reflects a combination of evolutionary history and environmental circumstance. Leaf shape could be related to environmental properties such as light capture, heat retention or loss, wind resistance and air circulation, or water droplet dispersion. Trends between leaf size and such environmental properties appear more common in the literature than for leaf shape, for which relatively few possible adaptive trends have been identified (Nicotra et al., 2011). The striking array of leaf shapes amongst plants invites questions on the significance of leaf shape variation from developmental, evolutionary, environmental and genetic perspectives.

1.1.2 Variation in leaf shape in differing environmental conditions

Studies in leaf morphology over environmental gradients largely report trends for size (Nicotra et al., 2011). This may be a result of measurement bias due to the complexity of accurately measuring leaf shape differences, though some studies do include measurements of leaf length and width as proxies for shape. Across 386 perennial plant species in south eastern Australia, narrow leaves are associated with greater mass per leaf area, although no environmental correlations were found with soil phosphorous levels or rainfall and leaf

shape (Fonseca et al., 2000). In another study, leaf size, width and length were all found to decrease with rainfall amongst 690 south east Australian plant species (McDonald et al., 2003). The trend is strongest for leaf size, and there was no consistent pattern of size reduction between the length and width traits, for example size reduction by leaf narrowing was as common as by leaf shortening (McDonald et al., 2003). A more dissected leaf shape has been shown to be associated with faster carbon gain and water loss in eight *Pelargonium* species, though this trait did not show any climatic trends with species site of origin (Nicotra et al., 2008). Although the shape of four *Helichrysum* species leaves collected from wild grown plants did not vary significantly with climate, a trend between smaller narrower leaves in cooler, drier conditions was described by the authors (Glennon and Cron, 2015). A relatively striking association of leaf shape variation and latitude has been identified in Ivy-leaf morning glory, *Ipomoea hederacea*. The leaf shape of *I. hederacea* is polymorphic between lobed and heart shaped leaves and shows a latitudinal pattern of variation, with heart shaped leaves restricted to southern populations (Campitelli and Stinchcombe, 2013). Work on *Acer rubrum* has identified changes in leaf margin serration in response to environmental variation. The leaves of *A. rubrum* plants grown in growth cabinets at cooler temperatures have more margin serrations and more dissected leaves than those at higher temperatures (Royer, 2012). Similar trends have been found for wild *A. rubrum* trees across collection sites in the eastern U.S.A. and in common gardens experiments where in both cases serration was increased for trees grown in colder locations (Royer et al., 2008, 2009).

Correlations of leaf morphology to climate have also been investigated for *Arabidopsis*. Differences in leaf shape, measured as leaf and petiole length and width, were measured across 21 accessions grown in a common garden experiment, and leaf length was found to decrease with latitude of origin for the accessions (Hopkins et al., 2008). In plants grown from seed collected from natural populations of *Arabidopsis* in the Swiss Alps and grown in a common garden experiment, ratio of leaf dry weight to leaf area was found to decrease with increasing altitude of seed collection site (Luo et al., 2015). This trend was only observed in plants that underwent a vernalisation treatment (Luo et al., 2015), highlighting the potential need to grow plants in conditions similar to what plants would experience in nature to observe possible associations of traits and conditions at origin.

As well as such leaf shape variations between accessions and related species, which will have both a genetic and environmental component when not grown in a common garden, differences in leaf shape have been identified within individual plants. Trees make a useful

system for identifying such differences, as they produce a large number of leaves that occupy different areas of the canopy, and so are exposed to often quite different conditions. Leaves from the upper canopy of *Quercus rubra* trees have been found to be smaller and more lobed than leaves lower in the canopy (Zwieniecki et al., 2004). These differences in size and shape were not present in the first week of leaf emergence, suggesting the differences may be a response to the local conditions at respective canopy positions (Zwieniecki et al., 2004). A trend for increased leaf area in lower parts of the canopy has also been reported for other tree species, such as *Tilia japonica* and *Ulmus davidiana* (Koike et al., 2001), and in six arboretum grown deciduous species, including *Acer*, *Ginkgo* and Oak species (Sack et al., 2006). This suggests that leaf shape within individuals may vary in part through response to specific growth conditions.

1.1.3 Leaf shape variation can be the result of genetic differences

There can be considerable variation in leaf shape even between closely related species. The genera *Pelargonium* and *Mimulus* contain species with a variety of different leaf shapes, species within these genera show simple, lobed or compound leaves (Ferris et al., 2015; Nicotra et al., 2008). In three mapping populations created using three different *Mimulus* species, Quantitative Trait Loci (QTL) for leaf shape were found to coincide in position on the genome, suggesting common loci may control shape differences between the species (Ferris et al., 2015). A polymorphism for lobed or heart shaped leaves of Ivy-leaf morning glory, *Ipomoea hederacea*, segregates in a three to one ratio, suggesting a single locus is responsible for this difference in leaf shape (Elmore, 1986). Leaf shape variation has also been mapped to genetic loci in crop species, such as *Brassica rapa* and maize (Baker et al., 2015; Ku et al., 2012), and in trees, in *Populus* and Oak species (Drost et al., 2015; Gailing, 2008; Wu et al., 1997). A close relative of *A. thaliana*, *Cardamina hirsuta*, has compound leaves of leaflets along a central rachis. Differences in regulation of KNOX genes and the gene SIL3 are thought to have a function in leaf dissection and leaflet formation in this species, potentially indicating some of the genetic differences that result in some of the differences between leaves of this species and the simple ovate leaves of *A. thaliana* (Hay and Tsiantis, 2006; Kougioumoutzi et al., 2012).

When comparing leaf shape differences between species or different genetic lines, it is important to consider that leaf shape within individual plants is likely to vary with developmental or growth phases. Changes in plant morphology associated with growth stages are often described as heteroblasty, and though this term was originally used to

describe abrupt differences in form it is often now used more generally to describe more gradual differences plant morphology through developmental phases (Zotz et al., 2011). Differences in the heteroblastic morphology changes within individual plants between genetic lines can be quantified and has been investigated, and in some cases mapped to genetic loci, in tomato, *Antirrhinum*, *Cardamina hirsuta* and *Arabidopsis thaliana* (Cartolano et al., 2015; Chitwood et al., 2014; Costa et al., 2012; Telfer et al., 1997).

Heteroblastic changes in shape can be quantified by comparing shape across multiple leaves of the leaf series of a plant at once. Through this approach it is possible to compare changes in shape and size common across the leaf series, but also for relative differences in shape and size between leaves in the leaf series to be compared between plants. Such heteroblastic changes in shape and size may be genetically controlled separately from changes in shape and size common across the leaf series. In tomato, morphology changes common across the leaf series map to different genetic loci than do heteroblastic changes shape and size across the leaf series (Chitwood et al., 2014). Heteroblastic differences in leaf shape and size contribute to the differences in leaf morphology between *Antirrhinum* species. *Antirrhinum majus* and *Antirrhinum charidemi* were crossed to create an F2 population, in which heteroblastic differences in leaf shape and size could be mapped to genetic loci (Costa et al., 2012). In *Cardimina Hirsuta*, natural variation in the *FLC* gene is associated with variation in flowering time but also differences in changes in leaf shape size and number (Cartolano et al., 2015). This highlights the importance of not considering leaves in isolation from the rest of the plant, as leaf morphology and number are likely to differ with developmental timing. Such heteroblastic differences in shape and size also suggest that multiple leaves per plant should be sampled to accurately describe the leaf shape variation of a plant, and that studies using a single leaf to represent the shape of an individual will be unable to describe the full shape variation within an individual plant.

1.1.4 Accurately quantifying differences in leaf shape

Leaf shape has often been measured using dimensions across the leaf as proxies, such as length, width or length to width ratio. Although these methods can be accurate in describing one aspect of a leaf, they are limited in the shape information they can record. For example variation in the acuteness of a leaf tip or curvature of the leaf would not be recorded by taking length and width measurements, and would require an additional measurement. Applying a number of individually recorded measurements across a large dataset can also become time consuming. Many studies of leaf shape are limited to measuring one or two of

the total leaves grown per plant (Hopkins et al., 2008; Jiménez-Gómez et al., 2010; Juenger et al., 2005), and so may be unable to identify differences in within plant leaf shape variation between lines.

Geometric morphometrics is an alternative approach to capturing and analysing shape differences that uses co-ordinates to record the shapes of samples (Adams et al., 2004; Rohlf, 1986). Co-ordinates may be placed over a shape to record the position of landmarks, common features across sampled shapes, or using semi landmarks, equally spaced co-ordinates around the entire margin or a set region of a shape. Geometric morphometrics has been used to study shape variation in a wide variety of organisms for different experimental applications. For example, developmental instability in fly wing shape has been studied using co-ordinate landmark analysis (Klingenberg and McIntyre, 1998), differences in mouse jawbones have been identified and mapped to genetic loci using co-ordinate landmarks (Klingenberg et al., 2001), and shape analysis has been used to corroborate molecular evidence on the evolution of great apes (Lockwood et al., 2004).

The typical form of a leaf lends itself toward analysis as a two dimensional shape, as most species leaves are relatively flat and can be pressed. Pressed leaves can then be conveniently imaged on a flatbed scanner. When analysing leaf shape in two dimensions, semi-landmarks spaced over the leaf margin can accurately capture the shape of a leaf. As the shape of a leaf is accurately recorded through this approach, further measurements can often be applied to the leaf co-ordinate shape model and so avoiding the need to rescore samples themselves for any measurements that may be more latterly required that were not part of an initial design. Once a dataset of these shape models has been recorded, the major shape variations between these co-ordinate shape models can be analysed and scored with Principal Component Analysis (PCA). PCA of shape co-ordinates identifies correlated variations in the co-ordinates as Principal Components (PCs), which correspond to the major shape changes within the shape models, and can be used to score the shape changes within the shape model dataset.

We chose to use this approach to measure leaf shape, harvesting each leaf from the plant and creating a two dimensional image of the leaf with a scanner, after pressing leaves flat. To record the shape of each leaf, semi-landmarks would be placed around the leaf margin automatically, using LeafAnalyser, software designed previously for this purpose (Weight et al., 2008). Recording leaf shape from a digital image in this way takes seconds, and makes experiments requiring shape measurements of thousands of leaves feasible.

Once a dataset of co-ordinate shape models is created, the major shape variations can be identified with PCA. To understand PCA on a dataset of shape models, it can be helpful to consider each shape model as a point in a multidimensional space. As each shape model has the same number of semi-landmarks, in this example 48, and is two dimensional, each shape model can be considered a point in a 96 dimensional space, where each axis of this space describes variation in either the x or y co-ordinate of one of the 48 semi landmarks. A dataset of many shape models would form a cloud of points in this space.

PCA captures the major shape variations in such datasets by defining new axes through this point cloud. These axes are the PCs produced by a PCA. Whilst each original axis of this multidimensional space describes a single aspect of variation in the x or y co-ordinate of each of the 48 semi-landmarks, PCs run through the major axes of the point cloud itself, and so describe the major variations in the positions of the shape model points within the cloud. See Figure 1-1 for a two dimensional example. The first PC crosses the widest dimension of the cloud, and so describes the largest portion of variation in the position of the shape models in the multi dimensional space, the following PCs are orthogonal to all previous PCs and describe the next greatest axis of variation across the point cloud. As each of the points in the cloud is a shape model, the PCs therefore describe the major correlated variations amongst the semi landmark co-ordinates making up these shape models. In this way PCA can be used to identify the major leaf shape variations, capturing them as PCs.

Shape variation in these co-ordinate shape models can be scored using PCs created from a PCA on that particular dataset, or using PCs created from a separate dataset, allowing the same PCs to be applied to measure shape in different experiments. Previous work in the group created a reference set of *Arabidopsis* leaves harvested from ten early flowering accessions, referred to as the leaf library, and used a PCA on these leaves to produce a set of PCs describing the major leaf shape variations (Matser, 2014). These leaf library PCs have been used to describe shape variation in the characterisation of *TCP* and *MAX1* genes in *Arabidopsis* (Challis et al., 2013; Danisman et al., 2012; Kieffer et al., 2011).

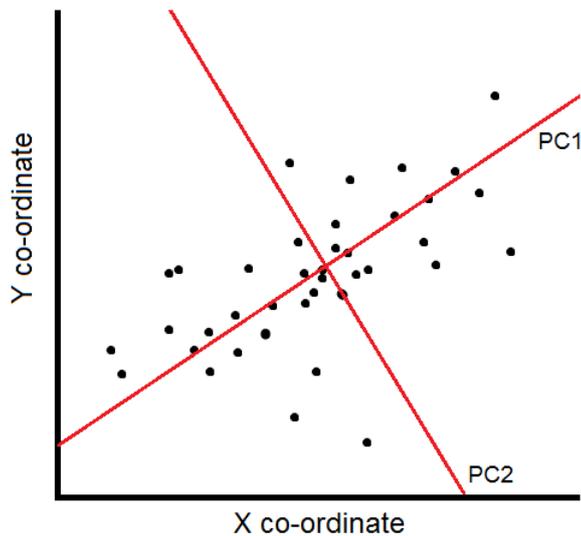


Figure 1-1 Principal Components capture major trends in variation

Figure shows a set of co-ordinates in a two dimensional space. Each point represents a co-ordinate pair, with an x and y value. Each axis describes variation in the x or y co-ordinate. The major variation in the point cloud can be described with principal components, shown as red lines through the point cloud.

1.1.5 Mapping trait variation to natural genetic differences in plants

To investigate naturally occurring leaf shape variation, we would use *Arabidopsis thaliana* as a model, aiming to understand more about the genetic basis of the differences in leaf shape occurring within this species. *Arabidopsis* has been extensively studied as a model plant for biology, resulting in a large body of literature and community resources (Huala et al., 2001).

Much study of the genetics of *Arabidopsis* has used mutant analysis, but natural genetic variation in this species also provides a powerful resource for understanding gene function relating to naturally occurring trait variation (Alonso-Blanco and Koornneef, 2000).

Arabidopsis is predominantly self fertilising and so the many natural accessions collected from the wild are largely homozygous (Bakker et al., 2006; Bergelson et al., 1998) and can be maintained easily as inbred lines, making them a useful subject and tool for analysis of natural genetic variation. Genetic mapping can be used to identify genetic variation responsible for natural trait variation in *Arabidopsis*. This approach uses populations of plants varying in genotype and for a trait of interest. Trait variation can be mapped to genetic loci across the genome by testing for association of variation in allele at a locus with variation in trait in such populations, see Figure 1-2.

Approaches to genetic mapping can be defined by the populations used, and there are two broad categories of experiment. Quantitative Trait Loci (QTL) mapping uses populations created by crossing natural accessions, producing a set of lines in which the alleles of the parent alleles segregate. A more recently developed approach, Genome Wide Association

(GWA), uses collections of hundreds of natural accessions as a mapping population. In both instances, lines are genotyped and scored for the trait of interest. This then allows for testing of possible associations between genetic and trait variation at markers across the genome.

Populations used for QTL mapping, segregating for the alleles of two or more parent accessions, are created by crossing two or more natural accessions to produce a heterozygous F1 plant. The F1 is then self-fertilised to produce F2 offspring, amongst which the different alleles of the parent accessions segregate. By genotyping such a population, and scoring it for a heritable trait, associations between genotype and trait variation can be identified. However, rather than genotyping and scoring a population at the F2 generation, further self fertilisations are often carried out for each plant in the F2, to create a population of F6 lines, known as Recombinant Inbred Lines, or RILs, which are largely homozygous. The advantages of working with homozygous RILs generally outweigh the time required for the further generations of self fertilisation. Plants from F6 RILs can be repeatedly grown and scored for traits with almost no change in genotype in contrast to F2 populations, where each F2 plant can be grown only once. Working with RILs allows multiple plants to be grown per line so that a genotypic average for the trait scored across these plants can be calculated, and also facilitates growing mapping populations in multiple conditions, allowing QTL to be tested for environmental specificity by growing the same genetic line in differing environments.

Although many RIL populations have been created using two natural accessions as parents, populations with greater allelic diversity can be generated by inter-crossing a larger number of parent accessions. The MAGIC and AMPRIL populations were created using crosses between nineteen and eight natural accessions respectively to create RIL populations with a greater number of natural alleles at each locus than in a typical two parent RIL population (Huang et al., 2011; Kover et al., 2009). As well as the increase in the genetic diversity at each locus, the increased number of crosses used to generate multi-parent populations increases the mapping resolution, allowing QTLs to be mapped to narrower regions than typical two parent populations due to greater levels of recombination (Huang et al., 2011; Kover et al., 2009). It is also possible to use combinations of two parent RIL populations which share a common parent accession in complex multi population mapping designs, where scoring such populations for the same trait allows comparisons of QTL between populations (Bentsink et al., 2010; Chardon et al., 2014). Coincidence of QTL can then be examined across multiple related populations, allowing the identification of regions

commonly associated with trait variation in multiple accessions (Chardon et al., 2014), and can lead to identification of new QTL, or narrow the region of identified QTL (Bentsink et al., 2010).

An alternative approach to QTL mapping in F2 and RIL populations is Genome Wide Association mapping (GWA). A Genome Wide Association Study (GWAS) uses natural genetic variation across hundreds of natural accessions to map trait variation (Korte and Farlow, 2013). As natural accessions are largely homozygous, they can be treated similarly to inbred lines such as RILs. Hundreds of these natural accessions have been genotyped as part of the 1001 genomes project, providing marker datasets for genetic mapping (Cao et al., 2011; Long et al., 2013; Ossowski et al., 2008; Schmitz et al., 2013).

The genetic diversity within a collection of natural accessions used for a GWAS is greater than the most diverse multi-parent RIL populations in *Arabidopsis*. The greater genetic differences, built up over generations between natural accessions can allow for greater mapping resolution in GWAS relative to RIL populations (Korte and Farlow, 2013). Also, associated loci identified within a GWAS are more likely to be common variants across *Arabidopsis* as a species, as rare variants are unlikely to be significantly associated with a trait due to low allele frequency (Korte and Farlow, 2013). In contrast RIL populations created from two homozygous parent accessions will have allele frequencies of roughly 50% at loci polymorphic between the parents, as each allele would be expected to segregate evenly. This can lead to the identification of rare variants, QTL mapping in a RIL population has identified an allele found to be unique amongst over 300 accessions (Loudet et al., 2008).

The appropriate steps to further investigate any associated loci identified in a genetic mapping experiment vary with the mapping population used. Once a genetic locus is identified as associated with a trait in a RIL population, the region can be further investigated using remnant heterozygosity within a RIL (Tuinstra et al., 1997). This region of remnant heterozygosity is fixed for either parent allele by self fertilising the RIL, and the resulting near isogenic lines (NILs) can be used to compare the effect of genetic variation at the locus on the trait of interest, as they are genetically identical outside this region. This allows the effect of the associated QTL to be potentially confirmed and physical borders to be established for the locus, by scoring each NIL for the trait of interest. Such lines can also be used as a starting point for fine mapping experiments, to further narrow the variant region between these NILs. In a collection of natural accessions used for a GWAS, many

possible alleles exist at each locus, and so the same approach as in a RIL population for further investigation of an associated locus is not directly applicable. To further investigate the effect of genetic variation at a locus identified in a GWAS, often workers will create separate F2 mapping populations by crossing accessions from the GWAS population with alternate alleles at the locus, and with trait scores matching the estimated direction of effect of that locus (Chao et al., 2012). The effect of each allele on the trait of interest can then be investigated within these mapping populations.

Once a candidate gene responsible for an associated locus in a genetic mapping experiment is identified, natural genetic variation underlying trait variation is sometimes investigated through complementation analysis. In complementation analysis, the effect of variant alleles is examined by comparing trait scores of hybrids of natural accessions and either a mutant line for a candidate gene or the mutant line background. If the natural alleles differ in their ability to complement the candidate gene mutant, this implicates the gene as containing the causative polymorphism responsible for trait variation. This has been used effectively to implicate candidate genes for sodium and copper tolerance in *Arabidopsis* as part of genetic mapping studies in both RIL and natural accession populations (Baxter et al., 2010; Kobayashi et al., 2008).

The genetic architecture of a trait, whether variation is controlled by one or two loci of large effect, or multiple loci of smaller effect, is a factor in the experimental design of further work in genetic mapping (Alonso-Blanco and Koornneef, 2000). Single loci of large effect for example are likely to be easier to identify and study further through independent mapping populations or fine mapping, whereas multiple loci of smaller effect may become difficult to identify in different genetic backgrounds, and so the genetic architecture of traits should inform the approaches taken in follow up work after an initial genetic mapping experiment.

Both QTL mapping in RIL populations, and GWASs with collections of natural accessions have been used to identify loci associated with natural variation in a variety of traits in *Arabidopsis*. In the literature, GWAS have been reported for traits such as hypocotyl length, flowering time, tolerance to salinity and cadmium levels and genome size (Atwell et al., 2010; Baxter et al., 2010; Chao et al., 2012; Filiault and Maloof, 2012; Long et al., 2013). GWASs have been used as a start point to identify and experimentally validate genes responsible for natural trait variation amongst accessions (Baxter et al., 2010; Chao et al., 2012), or to corroborate existing evidence on the effects of allelic variation in genes and to identify candidates for separate experimental investigation (Atwell et al., 2010; Filiault and

Maloof, 2012). QTL mapping has also been used to study the genetic basis of natural trait variation in *Arabidopsis*, and candidate genes have been identified for morphology, disease resistance, metal tolerance and flowering time related traits through QTL analysis (Clarke and Dean, 1994; Hinsch and Staskawicz, 1996; Jiménez-Gómez et al., 2010; Johanson et al., 2000; Kobayashi et al., 2008; Loudet et al., 2008). Although not all QTL studies are continued through to the identification and validation of candidate genes, much has also been learnt about the position and effect size of loci controlling a variety of traits in *Arabidopsis*.

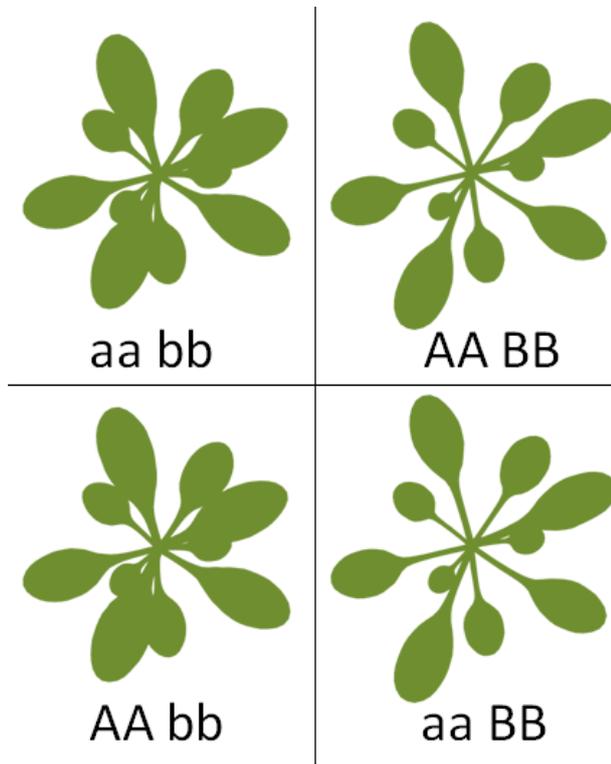


Figure 1-2 Linking genetic variation to trait variation

Figure shows four example plants, each made up of a rosette of varying morphology and two genetic loci. The 'A' locus varies between AA and aa alleles. The 'B' locus varies between BB and bb alleles. Where the B locus has a bb genotype, the rosette is smaller in diameter with more stunted leaves. Where the B locus has a BB genotype, the rosette is wider in diameter, with more elongated leaves. Unlike the 'B' locus, variation in the 'A' locus is not associated with variation in rosette morphology.



Figure 1-3 Natural Arabidopsis accessions

Photo shows natural accessions of *Arabidopsis thaliana*. © Janne Lempe and Detlef Weigel, Max Planck Institute for Developmental Biology

1.1.6 Naturally occurring leaf shape variation between *Arabidopsis* accessions

We chose to use *Arabidopsis* to study the natural genetic variation responsible for leaf shape variation. *Arabidopsis* is a small self-fertilising plant that grows a rosette of simple ovate leaves before developing an inflorescence. Naturally occurring accessions of this species show considerable variation in leaf shape, see Figure 1-3. Accessions of this species grow across a broad geographic range in a variety of environmental conditions (Alonso-Blanco and Koornneef, 2000). Reciprocal transplant experiments and correlations of trait values with accessions latitude of origin suggest some of the differences between the natural accessions are the result of local adaptation (Agren and Schemske, 2012; Hopkins et al., 2008; Stinchcombe et al., 2004). As *Arabidopsis* typically reproduces by self-fertilisation, the accessions are largely homozygous (Bakker et al., 2006; Bergelson et al., 1998) and so can be conveniently maintained as inbred lines for genetic analysis. A small diploid genome, a generation time under two months in some accessions and amenability to controlled cross-pollination also contribute to the suitability of *Arabidopsis* as a model species for study of natural variation.

The community resources available for analysis of natural genetic variation in this species are particularly useful. A number of RIL populations have been produced using natural accessions and are publicly available (Weigel, 2012). These populations are generally available with marker genotype data for each line, and so can be used for QTL mapping experiments without requiring workers to conduct genotyping of a RIL population themselves. A large number of natural accessions now also have been genotyped to produce publicly available marker datasets as part of the 1001 genomes project (Cao et al., 2011; Long et al., 2013; Ossowski et al., 2008; Schmitz et al., 2013). This allows groups to carry out GWASs without a large genotyping effort, and so more resources can be dedicated to scoring collections of accessions obtained from stock centres, allowing small groups to carry out relatively large scale genetic mapping experiments using the publicly available genotype data.

Many genes have been implicated with a function in determining leaf shape in *Arabidopsis* through mutant analysis. A variety of genes have been implicated in leaf development in *Arabidopsis*, through mutant phenotypes such as polarity defects, differences in cell expansion and proliferation and changes in leaf margin serrations (Dkhar and Pareek, 2014).

A reverse genetic screen of T-DNA lines has also found over two hundred T-DNA insertions to be associated with differences in leaf shape (Wilson-Sánchez et al., 2014).

A key process in leaf development is determination of polarity, a major outcome of which is the forming of the adaxial and abaxial layers of the leaf. The boundary between these layers is thought to drive the lateral outgrowth of the leaf, and where it is disrupted, narrow, needle like leaves are formed instead (Waites and Hudson, 1995; Wu et al., 2008; Yamaguchi et al., 2012). In *Arabidopsis* the KANADI1 and ASYMMETRIC LEAVES2 genes are involved in this process (Huang et al., 2014; Wu et al., 2008). Another process key to the development and shape of leaves is cell proliferation and expansion. Cells proliferate early in the development of the leaf, with the majority of cell expansion occurring later in development (Hisanaga et al., 2015). Interestingly, disruption to this process typically leads to compensation, and multiple distinct modes of compensation have been identified (Hisanaga et al., 2015). Mutations in ANGUSTIFOLIA3 for example, can illicit compensation mechanisms in *Arabidopsis*, where cell size increases in similar proportion to a decrease in cell number, minimising disruption to final leaf size (Horiguchi et al., 2006; Kawade et al., 2010). The establishment of polarity and balance between cell proliferation and expansion likely occur similarly for each rosette leaf produced by an *Arabidopsis* plant. However, there is a clear difference in shape over the leaf series. This difference is in part controlled by microRNAs, which have been found to regulate maturity and developmental timing in *Arabidopsis* (Poethig, 2009). Variation in genes responsible for these processes key to leaf development could potentially result in natural differences in leaf shape.

We do not necessarily expect natural leaf shape variation in *Arabidopsis* to be solely explained by polymorphisms in leaf development genes however. For example, variation in a gene characterised with a function in Aluminium tolerance in *Arabidopsis* does not explain a major nearby QTL for this trait in a Landsberg *erecta* x Columbia-0 RIL population (Hoekenga et al., 2003, 2006), and leaf shape QTL in a mapping population of *C. hirsuta* do not map to regions containing known leaf developmental candidate genes (Cartolano et al., 2015). Although leaf development genes have a clear function in leaf shape and striking mutant phenotypes, natural genetic variation may occur in genes with a variety of functions and result leaf shape variation. For example it seems reasonable to expect that any gene that may play a role in vegetative growth, for example growth rate, development, cell proliferation and expansion, metabolism or light response could affect leaf shape, given that leaves are such a large and complex organ crucial to the growth of the plant. As such we do not begin our work on the genetic basis of leaf shape in *Arabidopsis* with a predefined list of

candidates or gene family that we expect to be associated with variation in leaf shape traits, but will evaluate potential candidates occurring within any associated loci.

1.1.7 Investigating the genetic basis of leaf shape variation in *Arabidopsis*

We aimed to use genetic mapping to identify the genetic basis of natural leaf shape variation amongst accessions of *Arabidopsis*. Natural accessions of *Arabidopsis* vary considerably in leaf shape, and we were interested to learn more about the genetic basis of this variation. We would primarily approach this aim using genetic mapping of leaf shape traits, scored using LeafAnalyser, in a RIL population and collections of *Arabidopsis* accessions, using QTL mapping and GWAS approaches. This would allow us to make use of the advantages of both approaches in the project. Our method of leaf shape analysis would allow for a large number of leaves to be harvested relatively quickly, and so we would harvest and analyse the shape of every rosette leaf grown. Although some previous work has mapped shape differences between *Arabidopsis* accessions in RIL populations, measurements were restricted to traits such as length and width of one to two leaves per plant (Jiménez-Gómez et al., 2010; Juenger et al., 2005). We would use a more comprehensive analysis of leaf shape to compare shape across each leaf of the leaf series of each plant's rosette, calculating averages per plant, differences between individual leaves, and heteroblastic differences using shape models made up of multiple leaves. Our efficient approach to accurately record and quantify leaf shape differences would also allow the possibility of growing mapping populations in multiple conditions, to examine how different loci may be associated with leaf shape between different environments.

We were largely successful in mapping the natural genetic basis of leaf shape variation in *Arabidopsis*. We were able to accurately quantify natural leaf shape variation in a two parent RIL mapping population and a collection of hundreds of natural accessions, mapping leaf shape variation to genetic loci in both populations. Interestingly, the major leaf shape variations in these two mapping populations were similar, both to each other and also the leaf library data collected prior to this project. Although loci associated with flowering time appeared to coincide between the two mapping populations used, loci associated with leaf shape variation were specific to each population. There were multiple loci of small effect identified in each population, suggesting leaf shape is a polygenic trait. As well as identifying loci associated with our morphometric leaf shape traits, we identified a striking QTL associated with margin morphology and leaf size in one of the populations. We further

characterised the effect of this locus on silique number and epidermal cell size. Several of the leaf shape associated loci we identified within the collection of hundreds of natural accessions contained NBLRR genes. As such, we took these genes to be possible candidates for natural leaf shape variation, and characterised T-DNA insertion lines for these genes for possible leaf shape differences. We found several lines had leaf shape differences to the Col-0 control line, and found that these differences may be specific to certain light and temperature conditions. Further work based on the results within this project may involve fine mapping specific loci identified, further characterisation of these NBLRR mutant lines, or investigating the effect of natural genetic variation in NBLRR genes in plant growth and morphology.

Chapter 2. Materials and methods

2.1.1 Plant growth

Plants grown for leaf shape analysis or tissue for DNA or RNA extraction were grown in the University of York greenhouses, unless otherwise stated. Plants grown for crossing were grown in the University of York greenhouses, and transferred to University of York growth room after crossing. Once seeds were sown onto soil or agar plates, they were left in a dark 4°C room for three nights for stratification before transfer to specific growth conditions.

All plants grown for leaf shape analysis were grown in individual wells of P24 plastic inserts (H. Smith Plastics, Ltd). Plants grown for crossing were grown in P24, P40 or P60 inserts (H. Smith Plastics, Ltd). Levington F2 soil was used for all soil grown plants. A pesticide treatment, Levington intercept, was applied to the soil to minimise growth of insect larvae.

Greenhouse conditions were long days of 16 hours light and 8 hours dark. Temperature was held between 20°C and 27°C during the day. Temperature during the day was increased by heating if at 20°C, and decreased by partially open vents at 22°C, fully open vents and fans at 24°C and chilling elements if temperature reached 27°C. During times of the year where natural light was not present for 16 hours a day, supplemental lighting was used from overhead 400 watt sodium lamps. Supplemental lighting was activated when a light meter in the greenhouse less than 55 watts/m², and switched off during the 8 hour night, or once this light meter reaches 88 watts/m².

Growth room conditions were also set to long days, of 16 hours light and 8 hours dark. Temperature was kept between 23°C and 24°C during the day, and 20°C and 21°C during the night. No natural light was present and electric lighting was set at an intensity of 80-100µmol/m⁻²s⁻¹.

A Sanyo MLR 350 growth cabinet was used for temperature controlled experiments, and was set for long days of 16 hours light and 8 hours dark. Temperature was set to either 24°C or 16°C for the entire growth experiment. Relative humidity was set at 65% and light intensity was 150µmol/m⁻²s⁻¹.

For different light treatments, plants were grown in a darkened box in a 20°C temperature controlled lab. Light was provided by four LED panels that could emit red or blue light at varying intensity. In these conditions, plants were grown as seedlings on sealed agar plates and subjected to a red, blue or no light treatment in the box for four nights. Light intensity at

the level of the agar plates in the red light treatment was $1.1\mu\text{mol}/\text{m}^{-2}\text{s}^{-1}$ and $0.33\mu\text{mol}/\text{m}^{-2}\text{s}^{-1}$ in the blue light treatment. Different light intensities were used between treatments so that Col-0 control lines would have roughly similar hypocotyl lengths between blue and red treatments to avoid large differences in the effect of measurement accuracy between treatments.

2.1.2 Seed sterilisation and growth on agar plates

Agar for plant growth on plates was produced by autoclaving;

400ml dH₂O

3.2g Oxoid Agar Technical (Thermo Fisher Scientific)

Once autoclaved, 25ml of this 0.8% agar solution was then poured into Sterilin™ petri dishes (Thermo Fisher Scientific) and left to set, before placing in a dark, 4°C room for storage. For plants to be grown on agar plates, seeds were sterilised before sowing. Sterilisation of seed was carried out in a laminar flow hood. Sterilisation first used a bleach solution;

85% autoclaved dH₂O

10% bleach (Haychlor Industrial Sodium Hypochlorite)

5% Tween 20 (Sigma-Aldrich)

Seeds were left in this solution for 15 minutes, then sterilised with 70% ethanol for less than 30 seconds, after which five washes of autoclaved dH₂O were applied. Seeds were first suspended in an agar solution, made as described above, but at a concentration of 0.05%. Seeds were sown onto circular plates filled with set agar.

2.1.3 Crossing Arabidopsis

Plants were selected for crossing, where possible choosing plants that had begun flowering recently and had five to ten open flowers. Crossing was carried out using fine point tweezers, and flowers were observed for emasculation and pollination through a dissection microscope (Nikon Stereoscopic Zoom Microscope 10A). Tweezers were cleaned between touching stamen using ethanol.

To prepare flowers to receive pollen on the plant serving as a female parent for a cross, flower buds were identified that had not yet opened, but were large and with a hint of white at the meeting point of the sepals. For each stem on which possible recipient buds were identified, all other flowers or buds not at this stage were removed. The bud was then

carefully opened by pulling sepals and petals backwards, and then removing anthers separately, leaving the pistil remaining. Stamen were expected to be short and unripe at this stage, with little risk of self pollination of the flower having already occurred. However, if anthers appeared ripe, suggesting possibility of self pollination, the bud in question was abandoned. Stamen were selected from male plants serving as pollen donors by removing anthers from recently opened flowers, using tweezers to separate them from the flower. The anther from the male parent was then rubbed against a stigma of the female parent, and presence of transferred pollen was checked under a dissection microscope.

Plants were left for a day to see if the pistil had grown or withered, indicating pollination or damage to pistil respectively. Mature siliques were collected by removal with tweezers. In the event a silique may dehisce before removal, siliques were cased in a small paper envelope to collect seed.

2.1.4 Epidermal cell size and shape analysis

To measure epidermal cell number and size between lines in HIF118 and HIF102, a rosette leaf was harvested from each plant, the median node leaf from HIF118 plants, and the fourth leaf from HIF102 plants. This leaf was then stuck to a stripe of tape, adaxial side down, as flat as possible. To make a mould of the leaf, clear nail varnish was applied to the abaxial side of the leaf. This was left to dry for 20 minutes. After the varnish had dried, this was removed from the leaf surface with tweezers and placed over a droplet of water on a microscope slide. The imprint of epidermal cells can then be seen and imaged with a microscope (Nikon Optiphot-2) and attached camera (Carl Zeiss Axiocam MRc5).

Size of epidermal cells was then measured by tracing the outline of the cell with the magnetic lasso tool in Photoshop CS4 (Adobe). Cells were measured from the top left, top right, bottom left and bottom right positions on the leaf and on average 43 epidermal cells were measured in total. Epidermal cells were chosen for measurement by picking a point in the centre of the image and measuring the selected cell and those adjacent. As well as measurement of area, highlighted and measured epidermal cells were also saved in a separate image file, and measured for the shape traits circularity and compactness with ImageJ (Schneider et al., 2012).

2.1.5 PCR with genomic DNA and cDNA

PCR from genomic samples of DNA used a standard protocol across the project. The volume of each reaction was 10µl, made up of the following;

- 1 µl genomic DNA
- 1 µl 1x ThermoPol® buffer (New England Biolabs)
- 0.2 µl for each primer (10mM)
- 0.2 µl 10mM dNTPs (Fermentas, Thermo Fisher Scientific)
- 0.05 µl *Taq* DNA polymerase 5000 Units/ml (New England Biolabs)
- Made up to 10 µl with dH₂O

Reactions were then placed in a MJ research DYAD dual block PCR machine for amplification of DNA. The PCR programme used in each experiment followed a standard template;

1. 95°C for 2 minutes
2. 95°C for 30 seconds
3. T_m for 30 seconds
4. 72°C for 5 minutes
5. Go to step 2 (x30)
6. 72°C for 5 minutes

For RTPCR analysis a similar approach was taken. The standard 10 µl reaction used between RTPCR experiments was made up of

- 1 µl cDNA preparation
- 1 µl 1x ThermoPol® buffer (New England Biolabs)
- 0.2 µl for each primer (10mM)
- 0.2 µl 10mM dNTPs (Fermentas, Thermo Fisher Scientific)
- 0.05 µl *Taq* DNA polymerase 5000 Units/ml (New England Biolabs)
- Made up to 10 µl with dH₂O

After preliminary RTPCR experiments with different cycle numbers, the following PCR programme was used for cDNA samples in our RTPCR work.

1. 95°C for 1 minutes

2. 95°C for 30 seconds
3. T_m for 30 seconds
4. 72°C for 5 minutes
5. Go to step 2 (x40)
6. 72°C for 5 minutes

2.1.6 Preparation of DNA

Throughout the project, a standard approach was taken to DNA preparation, following a published protocol (Edwards et al., 1991). First plant tissue was placed into a 2ml centrifuge tube with a 5mm diameter ball bearing (Simply Bearings Ltd, UK) and macerated using a TissueLyser (Qiagen). After the ball bearing was removed, 400µl of DNA extraction buffer was added, and each tube was vortexed. Tubes were then centrifuged (International Equipment Company micromax centrifuge) at 10,000 rpm for 5 minutes. The supernatant was removed, and added to a 1.5ml centrifuge tube containing 300µl isopropanol. This solution was mixed by pipetting and then left for 2 to 30 minutes. After this time, the solution was centrifuged at 10,000 rpm for 5 minutes, the supernatant was discarded, and tubes left to dry. Once dry, 100µl dH₂O was added and the tubes were vortexed before storage in a -20°C freezer.

DNA extraction buffer was made up in 200ml volumes of;

- 10ml 10% SDS
- 50ml 1M NaCl
- 10ml 0.5 EDTA
- 26.7ml 1.5M TRIS pH 7.5
- 103.3ml dH₂O

2.1.7 Preparation of RNA

To extract RNA from plants, tissue was collected from leaves at the middle of the leaf series from bolting plants. Tissue was flash frozen with liquid nitrogen and ground with a ceramic mortar and pestle. The RNA was extracted using RNeasy kit (Qiagen), following the plant tissue protocol provided with the kit, using the RLT buffer and including the optional DNAase step. The concentration of RNA samples prepared was checked using a NanoDrop® ND-1000 spectrophotometer. RNA samples used in the project had concentrations of over 350 µg/ul,

A260/A280 absorption ratios between 2.15-2.16 and A260/A230 absorption ratios of 2.16-2.44.

2.1.8 Preparation of cDNA

To synthesise cDNA from prepared RNA, SuperScript® II Reverse Transcriptase (Invitrogen) was used. cDNA was synthesised with the following protocol;

First a 12µl reaction was prepared of;

- 1µl Oligo(dT)₁₂₋₁₈ (Invitrogen)
- 1µl 10mM dNTPs (Fermentas, Thermo Fisher Scientific)
- 1.5µg prepared RNA
- dH₂O upto 12µl

This reaction was then incubated for 65°C for 5 minutes and then placed on ice. Next, 4µl of 5x First-Strand buffer and 2µl of 0.1M DTT were added. The solution was mixed, and incubated at 42°C for 2 minutes. After this, 1µl of SuperScript II Reverse Transcriptase was added, the solution was mixed, and then incubated at 42°C for 50 minutes, then 70°C for a 15 minutes. Prepared cDNA was then stored at -20°C.

2.1.9 Restriction digest for genotyping of PhyB-9 mutation

To identify the mutation in PhyB-9 line, a region of the PhyB gene was amplified and then subjected to a restriction digest, using the following 20µl reaction;

- 2µl CutSmart® Buffer (New England Biolabs)
- 0.2µl BSA (Promega)
- 5µl PCR product
- 0.5µl MnlI (New England Biolabs, 5000U/ml)
- 12.3 dH₂O

Each sample was then incubated at 37°C for one hour.

2.1.10 Gel electrophoresis

PCR samples were run on agarose gels with DNA stain to visualise amplified fragment sizes. Gels were made with 0.8% electrophoresis grade ultra pure agarose (Invitrogen) in TBE buffer. TBE was kept as a 5x stock and made up from adding the following to 900ml of dH₂O;

54g Tris
 27.5g Boric acid
 2ml 0.5M EDTA pH 8
 dH₂O upto 1L total solution

Gels were run in a gel tank filled with 1% TBE at 80-100 volts. To compare the size of sample bands in reactions run on agarose gels, 2-log ladder (New England Biolabs) was used. See below for a diagram showing band sizes of this ladder.

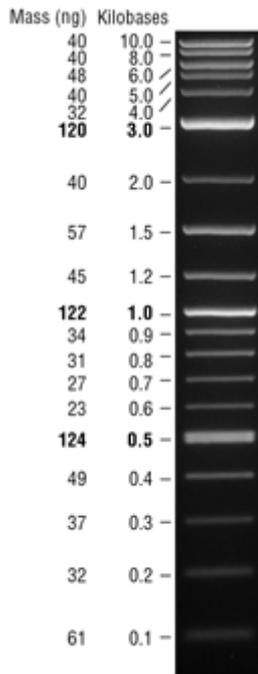


Diagram showing band sizes of 2-log NEB ladder. Image was obtained from <https://www.neb.com/products/n3200-2-log-dna-ladder-01-100-kb> and produced with ethidium bromide staining on a 1.0% TBE agarose gel. Mass values are for 1 µg/lane.

2.1.11 LeafAnalyser leaf shape analysis

For leaf shape analysis the rosette leaf of each plant was harvested at bolting, unless otherwise stated, and placed between two acetate sheets and pressed overnight between magazines. To record an image of the leaves, these acetate sheets were placed on a Hewlett Packard Scanjet 4370 scanner and scanned at 300 dpi, then saved as high quality jpegs. Shape was recorded using LeafAnalyser (Weight et al., 2008) to record the leaf margin with semi-landmarks. The positions of these semi-landmarks were then exported as co-ordinates in text files for analysis. Text files were compiled to a single file using the command line.

To align and scale leaf shape models, the Procrustes fit in MorphoJ (Klingenberg, 2011) was used and shapes were aligned by principal axes and scaled to unit size. Size scaling works by calculating the mean co-ordinate for each shape model, known as the centroid. The absolute

distance of each co-ordinate point along the margin from the centroid is then calculated and the total of these distances describes the size of the shape model, referred to as centroid size. By dividing the co-ordinates of each shape model by that shape models centroid size, all variation in size is removed from the shape models, which are now all scaled to unit size. The procrustes fitted co-ordinates, and centroid size values for each leaf shape model were then exported from MorphoJ.

Shape variation was scored amongst the shape models using the eigensystem created from a PCA on a dataset of leaves in LeafAnalyser, typically either the leaf library reference dataset, or the dataset the leaves to be analysed belonged to.

For the heteroblasty three leaf shape models an alternative approach was required due to the differences in arrangement of shape model data, having been compiled from three mean leaves per plant in R. We ran a PCA on the heteroblasty shape models using the *prcomp* function in R, and used this function to score the heteroblastic shape differences, exporting the PC scores for each shape model from R.

2.1.12 Genetic mapping

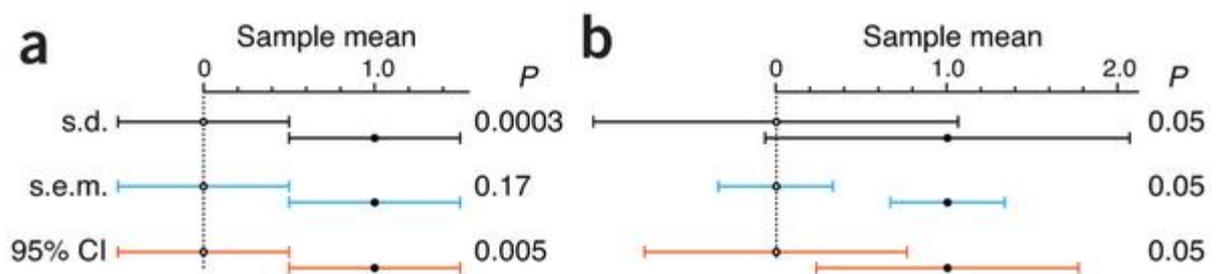
In this project both QTL mapping and GWA were used to identify loci associated with plant traits. QTL mapping was carried out using the R/qtl package (Arends et al., 2010) in R (R core team, 2015). Trait data for the 165 Bay x Shahdara RILs was compiled with genotype data for the population, obtained from <http://publiclines.versailles.inra.fr/page/33>, and loaded into R. QTLs were then detected using the Multiple QTL Mapping (MQM) function, *mqmscan*. Loci were identified using graphical outputs from this package and the *lodint* function to estimate an interval for the QTL region. For GWA work, trait data was also compiled into text files, and then uploaded to GWAPP (<http://gwapp.gmi.oeaw.ac.at/>) (Seren et al., 2012). Both the KW and AMM approaches to mapping available on GWAPP were used. The KW approach is a Wilcox test of association for each SNP. The AMM approach is based on the EMMA method used in Kang et al., (2008) and decreases the chance of identifying false positive associations due to relatedness between accessions using a kinship matrix to reduce the effect of population structure in the association tests. To identify and further investigate loci associated with these approaches, the results files for mapping were downloaded from GWAPP and used to produce graphs in R using base graphics commands.

2.1.13 Hypocotyl measurement

Hypocotyl length was measured from plants grown on agar plates by removing each seedling with tweezers and placing them on an acetate sheet. This was then scanned at 300 dpi on a Hewlett Packard Scanjet 4370 scanner and images were saved as jpegs. Hypocotyl length of seedlings was measured using the freehand draw tool of ImageJ to trace hypocotyl length and measurements were exported to text files for further analysis.

2.1.14 Error bars

The vast majority of graphs in this thesis use 95% confidence intervals for error bars, and are labelled as such. Much of the data in this thesis concerns quantified differences in leaf shape between different *Arabidopsis* lines. Rather than statistically test each leaf of the leaf series and mark those showing a significant difference between lines, 95% error bars are shown along with the mean score at that node for interpretation of the extent of the difference in leaf shape across the leaf series. These error bars, when not overlapping, typically represent a strongly significant difference, and also increase in size in response to low sample size. As such they provide a useful visual guide to the extent of leaf shape differences over the leaf series, as clear differences can be noted where error bars do not overlap, and small sample size, typical at the end of the leaf series, is indicated by large error bars. See Krzywinski & Altman (2013) for more detail. A figure from this publication, showing error bars from example datasets, is reproduced below.



(a,b) Example graphs are based on sample means of 0 and 1 ($n = 10$). (a) When bars are scaled to the same size and abut, P values span a wide range. When s.e.m. bars touch, P is large ($P = 0.17$). (b) Bar size and relative position vary greatly at the conventional P value significance cutoff of 0.05, at which bars may overlap or have a gap.

2.1.15 Primer design

Primers were designed using primer3 (Untergasser et al., 2012); <http://primer3.ut.ee/>. Gene sequence was obtained using the Sequence Bulk Download and Analysis tool from TAIR (Berardini et al., 2015); <https://www.arabidopsis.org/tools/bulk/sequences/index.jsp>.

Primers for Actin 2 were designed by Dr. Michael Schultze. Primers for confirmation of SALK T-DNA insertions were designed using the SALK T-DNA express primer design tool (<http://signal.salk.edu/tdnaprimers.2.html>). Primers for markers BSAT2.010, MSAT2.18, MSAT2.26 and UPSC25794 were obtained from <http://www7.inra.fr/vast/msat.php>.

See Table 2-2 for a list of all primers used during the project. Primers were ordered from Integrated DNA Technologies. Primers were kept as 100mM stock solutions and 10mM working stock solutions, by mixing the delivered desalted primers with amounts of dH₂O to reach the required concentrations.

2.1.16 Plant material

The Col-0 line used throughout this project was donated from Dr. Martin Kieffer. The Bergelson collection of 96 Joy Bergelson accessions and the Detlef Weigel collection of 80 accessions (Cao et al., 2011; Nordborg et al., 2005) were donated by the Ottoline Leyser group. The PhyB-9 mutant line, and Bay-0 and Shahdara accessions were donated by the Seth Davis group. The HapMap collection of natural accessions and SALK T-DNA insertion lines (Alonso et al., 2003) were obtained from NASC (<http://arabidopsis.info>), for NASC IDs see Table 2-1. All Bay-0 x Shahdara HIFs used in the project were obtained from INRA (<http://publiclines.versailles.inra.fr/>). The rHIF47-2 and rHIF47-5 fine mapped lines were donated by Oliver Loudet.

Name	NASC ID
HapMap/Borevitz collection	N76309
Bay-0 x Shahdara 165 core RIL population	N57921
At5g45240-1 SALK T-DNA line	N672671
At5g45240-2 SALK T-DNA line	N655431
At5g45240-3 SALK T-DNA line	N670435
At5g45240-4 SALK T-DNA line	N680733
At1g72840-1 SALK T-DNA line	N658536
At1g72840-2 SALK T-DNA line	N671145
At1g72850-1 SALK T-DNA line	N682142
At4g09420-1 SALK T-DNA line	N678533
At2g17050-1 SALK T-DNA line	N669130

Table 2-1 Arabidopsis lines

Table shows Arabidopsis lines and collections used in the project that were ordered from NASC with the NASC ID for each line.

Name	Pairing	Sequence	Experiment
act2	FP	ACGAGCAGGAGATGGAAACC	RTPCR ACT1 control
act2	RP	ACCCCAGCTTTTAAAGCCTTTG	RTPCR ACT1 control
le12	FP	GCTGTGACGACAAGAAGAGC	At5g45240 RTPCR
le12	RP	ACACAAGAAATGAGCAGGCA	At5g45240 RTPCR
le23	FP	AGAAGTGCCATCTCAACGAG	At5g45240 RTPCR
le23	RP	ATAGACTCAGCGACGTGGTT	At5g45240 RTPCR
le34	FP	GAGCACCCGGATTGGATC	At5g45240 RTPCR
le34	RP	TCCCACTGCAATTAGAGAGATCA	At5g45240 RTPCR
le45	FP	TCTTGGGCACTGTGAGAAAC	At5g45240 RTPCR
le45	RP	CTCCATCCCACAGTTCTCA	At5g45240 RTPCR
le56	FP	TGAGAAGCTGTGGGATGGAG	At5g45240 RTPCR
le56	RP	CCAATAGAGGGAGGAAGCAAC	At5g45240 RTPCR
le67	FP	TGATGTTGCTTCCCTCCTCT	At5g45240 RTPCR
le67	RP	TCGAAAGGTCAGATAGCCCA	At5g45240 RTPCR
le78	FP	GGCTATCTGACCTTTCGATGAG	At5g45240 RTPCR
le78	RP	AGAGAGCGTTGCAGGGAG	At5g45240 RTPCR
le89	FP	AGACTCCAGACACATGGACT	At5g45240 RTPCR
le89	RP	AGGCCAGGCTTGATTCAA	At5g45240 RTPCR
le910	FP	TCTAACCGTCTTTACCTCGT	At5g45240 RTPCR
le910	RP	ACTCTTCGTCGTTCTCTGCA	At5g45240 RTPCR
PhyB-9	FP	CAATGTAGCTAGTGGAAGAAGCTCGATGTGG	PhyB-9 mutant genotyping
PhyB-9	RP	ACATAACAGTGTCTGCGTTCTCAAACGC	PhyB-9 mutant genotyping
At5g45240-1	FP	GAGTCAGCCGTGGTATCTCAG	Genotyping T-DNA insertion
At5g45240-1	RP	GAACCACCTGAGGAGATCTC	Genotyping T-DNA insertion
At5g45240-2	FP	TTTAGTGTACCCAATCGTCC	Genotyping T-DNA insertion
At5g45240-2	RP	TCTCCATGAATTAACCAAACAG	Genotyping T-DNA insertion
At5g45240-3	FP	TGGAAGAGATCGTGAAGATG	Genotyping T-DNA insertion
At5g45240-3	RP	TTCCCATACCTTATATCCTTCCC	Genotyping T-DNA insertion
At5g45240-4	FP	AAATCAATCTTGGGCACTGTG	Genotyping T-DNA insertion
At5g45240-4	RP	GGATCTGGGATTTCTCTCAGG	Genotyping T-DNA insertion
At1g72850-1	FP	GAAGCCACGCTGAAATGATAC	Genotyping T-DNA insertion
At1g72850-1	RP	AATCCCACATCTAAACCCTG	Genotyping T-DNA insertion
At1g72840-1	FP	CGACCGCAGTAAATCTTGAAG	Genotyping T-DNA insertion
At1g72840-1	RP	TGGCTAAAGTCCAAGTGGTTG	Genotyping T-DNA insertion
LBI1.3	RP	ATTTTGCCGATTTCCGGAAC	Genotyping T-DNA insertion
BSAT2.010	FP	CAGGATGACGATATGCTTCG	HIF102 and HIF118 markers
BSAT2.010	RP	CAGGAATTACGAATAAACTTGACG	HIF102 and HIF118 markers
MSAT2.18	FP	TAGTCTCTTTTGGTGCGATA	HIF102 and HIF118 markers
MSAT2.18	RP	AGCCTCTCCAAGCTTAGGTCT	HIF102 and HIF118 markers
MSAT2.26	FP	TCTCCGATTGAGCCCCAAAG	HIF102 and HIF118 markers
MSAT2.26	RP	CGGGGAAAGATGGGTTTTGA	HIF102 and HIF118 markers
UPSC25794	FP	TCATGCGGAAGTGAGTGTTT	HIF102 and HIF118 markers
UPSC25794	RP	TGCTTGAGTTTGGTTTTTGC	HIF102 and HIF118 markers

Table 2-2 Primer sequences

FP/RP column describes whether primer was forward (FP) or reverse (RP) in each pairing.

Chapter 3. Quantitative Trait Loci mapping in the Bay-0 x Shahdara population

3.1 Introduction

3.1.1 Quantitative Trait Loci mapping in the Bay-0 x Shahdara recombinant inbred line population to identify loci associated with natural leaf shape variation

Quantitative Trait Loci (QTL) mapping can be used to identify the naturally occurring genetic variation underlying phenotypic differences in accessions of *Arabidopsis* (Alonso-Blanco et al., 2009; Koornneef et al., 2004). In *Arabidopsis*, a self compatible species, QTL mapping is typically carried out using Recombinant Inbred Line (RIL) populations. Accessions of *Arabidopsis* are largely homozygous (Bakker et al., 2006; Bergelson et al., 1998). To create a RIL population accessions are crossed, producing a heterozygous F1 which is then self fertilised to create an F2 population of plants with a variety of combinations of the parent's alleles. Although an F2 population can be used for genetic mapping, self fertilising F2 plants for 4 to 5 generations will fix the majority of the genotype of each line to homozygosity. This creates an F6 - F7 population that need only be genotyped once, but can be repeatedly regrown with minimal genetic variation, as each line is now almost entirely homozygous. Many RIL populations have been created, genotyped and made available for use by the *Arabidopsis* community (Alonso-Blanco et al., 1998; Clarke et al., 1995; Loudet et al., 2002; Magliano et al., 2005; Wilson et al., 2001).

We chose to use the Bay-0 x Shahdara RIL population to for QTL mapping leaf shape traits. This population is a set of F7 RILs created from two natural accessions, Bay-0 and Shahdara, Bay-0 was the female parent, and Shahdara was the male parent (Loudet et al., 2002). Bay-0 was collected from an area of fallow land near Bayreuth, Germany and Shahdara was collected from the Pamiro-Alay mountains in Tadjikistan (Loudet et al., 2002). The population has been genotyped using 69 markers across the genome and has been used to successfully map QTLs for a variety of traits (Botto and Coluccio, 2007; Jiménez-Gómez et al., 2010; Loudet et al., 2008; Wingler et al., 2010). The difference in origin between Bay-0 and Shahdara suggests these accessions are likely to be adapted to considerably different environments, and vary for a variety of traits. Though they differ in leaf shape, the difference in leaf shape between these two accessions is not striking, see Figure 3-1. We

expect this will allow us to map quantitative variation in shape, rather than a single unigenic shape difference. For example, the *Arabidopsis* line *Ler* has a distinct leaf shape phenotype due to a mutation in the *ERECTA* gene (Tisné et al., 2011) and has been used to generate RIL populations, scoring leaf shape in such populations will identify a large QTL over the *ERECTA* gene (personal observation). Although it is reasonable to assume the association at this locus due to variation in the *ERECTA* gene and factor this into the analysis, the large *ERECTA* QTL may obscure the effect of other nearby polymorphisms. The Bay-0 x Shahdara RILs are a population without a striking difference in leaf shape between the parents of the population. As alleles from either parent will be present in RILs in new combinations, the RILs within this population will likely have a greater range of trait variation than between the parents, and potentially allow us to identify these alleles through association with leaf shape traits.

We would grow the Bay-0 Shahdara RIL core population and measure leaf shape and size across these lines. By creating a leaf shape dataset for this population, we would be able to link variation in leaf shape traits to the natural genetic variation in the population. This would allow us to identify loci associated with natural leaf shape variation in *Arabidopsis*. Both accessions are early flowering and do not require vernalisation before bolting. As such we could use the bolting date of each plant as a point at which to harvest each rosette leaf.

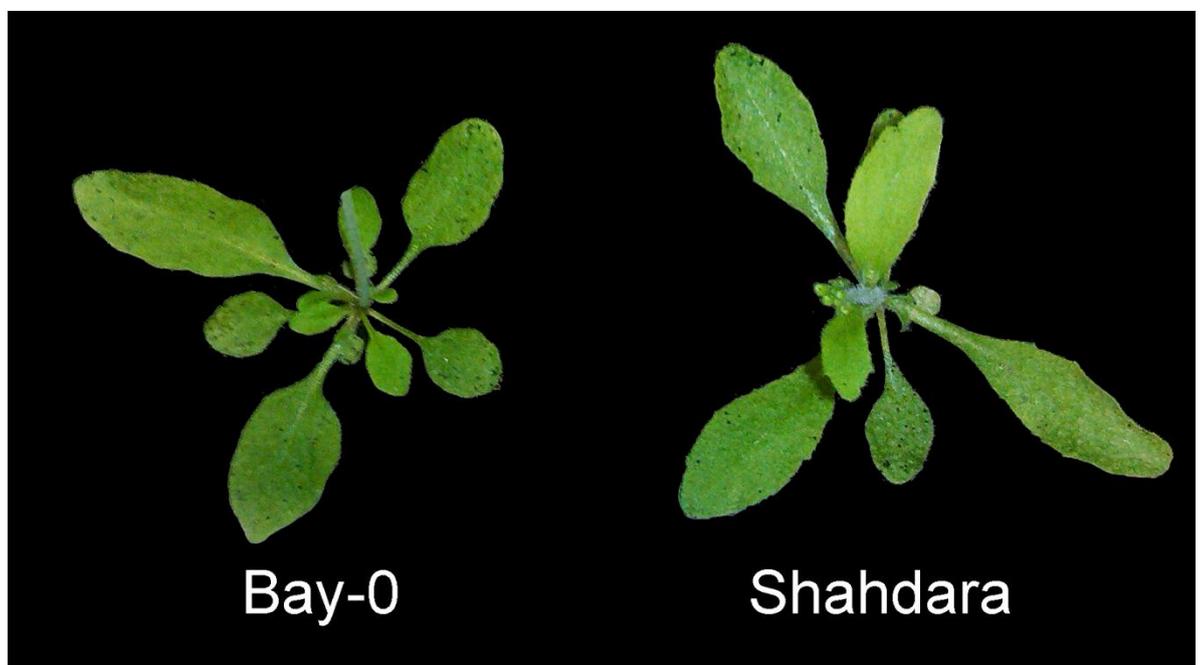


Figure 3-1 Rosettes of Bay-0 and Shahdara accessions

Photo shows the rosettes at bolting of Bay-0 and Shahdara accessions.

3.1.2 Choosing growth conditions for the Bay-0 x Shahdara RILs

We first grew the Bay-0 x Shahdara population in summer 2012. These plants flowered much earlier than we expected given earlier observations of the parent accessions. We suspected this was due to unseasonably warm weather shortly after the plants had germinated. Though we harvested these plants for preliminary data regardless, we felt that the low leaf number of many of the RILs would inhibit our ability to explore the genetic basis of shape changes over the leaf series and decided to regrow the plants. Work done recently within the lab documented the difference in leaf shape amongst a set of accessions grown in light and shade treatments, using white muslin cloth to block and diffuse the light. When regrowing the Bay-0 x Shahdara lines we took the opportunity to include another treatment alongside our standard greenhouse conditions, and also grew the plants in a shade treatment. Previous work suggested plants would vary in leaf shape between the light and shade treatments, but we wondered if the extent of this difference would vary between the RILs, and whether we could map the extent of difference, or plasticity, to QTL in this population. By regrowing the RILs we aimed to increase the number of rosette leaves grown per plant, and potentially map QTL for the plasticity of our leaf shape and size traits between the light and shade treatments, as well QTL for leaf shape and size traits in either treatment respectively.

3.1.3 Software for QTL analysis

Although specific QTL mapping software, such as mapQTL (Wang et al., 2011) and QTLcartographer (Van Ooijen, 2004) exists, we opted to use R/qtl (Arends et al., 2010), a free and open source package for the R statistical environment (R core team, 2015). We opted to work within R for the QTL analysis so that the input and output of data was straightforward to integrate with other work within this project. As we collect a large amount of leaf shape data per population, and analyse this in many ways, it was likely that we would want to test for QTLs with a variety of trait data. For this reason it was important to use an approach to QTL mapping that would facilitate scripting of data input and QTL analysis, allowing large amounts of trait data to be analysed quickly. R/qtl was ideal for this purpose.

3.1.4 A comprehensive approach to leaf shape phenotyping

After harvesting each plant at bolting, we would record the shape of every rosette leaf grown per plant by co-ordinate shape models, using LeafAnalyser (Weight et al., 2008). Principal Component analysis (PCA) can then be used to capture the major variations in shape amongst co-ordinate models with Principal Components (PCs). The shape of individual leaf co-ordinate models can be scored with PCs, and in this way we would create a dataset of leaf shape traits for the Bay-0 x Shahdara RIL population.

There are many possible approaches to quantifying leaf shape variation in *Arabidopsis*, as each plant grows a series of rosette of leaves that vary in shape and total leaf number. By analysing every rosette leaf of each plant we grew, we would be able to investigate changes in leaf shape between lines in depth. Creating a complete dataset of leaf shape at each leaf grown for these lines would allow us to investigate the genetic basis of leaf shape differences across the leaf series, rather than being restricted to differences at a single node of the leaf series. As leaf shape varies over the leaf series, we were interested to see how these changes were controlled genetically. Our expectation was that changes in leaf shape common to each leaf of a rosette would be associated with coinciding QTL when comparing each leaf individually, and the shape of individual leaves would also be modulated by their position in the leaf series. We thought this more likely than separate QTL responsible for specifying the shape of each leaf within the rosette of a plant. If there was a modulating effect of leaf series position on leaf shape, it may be that this also varies in within the population, and so perhaps could also be mapped to QTL.

Changes in shape over the leaf series of plants associated with developmental age are often referred to as heteroblasty (Zotz et al., 2011), and have been mapped to genetic loci in *Antirrhinum* and tomato species (Cartolano et al., 2015; Costa et al., 2012). *Arabidopsis* is a useful system for studying leaf series changes (Poethig, 2010; Telfer et al., 1997; Yang et al., 2011), as each plant grows a defined number of rosette leaves, then bolts and transitions to flowering. By measuring each rosette leaf per plant, we can investigate changes in shape across the leaf series as well as the overall leaf shape differences between plants.

We would also score the plants for bolting date and leaf number. These traits are highly correlated with flowering time in *Arabidopsis*, which is a widely studied trait. As such as well as allowing us to compare leaf shape and size relative to the transition from vegetative to reproductive growth in *Arabidopsis*, these traits would also serve as a point of comparison to work in the literature.

3.2 Results

3.2.1 Growth conditions for the Bay-0 x Shahdara population

We grew the Bay-0 x Shahdara core population of 165 F7 Recombinant Inbred Lines (RILs) (Loudet et al., 2008) in the greenhouses at the University of York. In this experiment we grew two batches of the 165 RILs in different conditions. The first batch was grown in standard greenhouse conditions, see Materials and methods, referred to as the light treatment. The second batch was grown under the shade from a muslin cloth, referred to as the shade treatment, see Materials and methods. In each treatment, three plants were grown for each of the 165 RILs. Once each plant bolted to produce an inflorescence shoot we harvested the rosette leaves of each plant. The shape of each leaf was recorded with LeafAnalyser following standard laboratory protocols, see Materials and methods and (Weight et al., 2008). We also recorded the date each plant bolted and the number of rosette leaves each plant grew.

We had two major aims for this experiment. Firstly, we aimed to map QTLs for the natural leaf shape variation in this population. Secondly we aimed to map loci that may determine variation in the extent of leaf shape differences between RILs when grown in light and shade treatments.

3.2.2 Scoring leaf shape variation within the Bay-0 x Shahdara RIL population

Once the shape of each leaf was recorded with Leaf Analyser we scored differences in leaf shape using the leaf library Principal Components (llpPCs). These PCs represent a standard reference set of leaf shape traits identified with LeafAnalyser that have been used previously to score leaf shape in *Arabidopsis* (Challis et al., 2013; Danisman et al., 2012; Kieffer et al., 2011). Before scoring the leaves with the leaf library PCs, we scaled each leaf co-ordinate model to unit size through Procrustes fitting. This allowed us to separately record the size and shape of each individual rosette leaf. By harvesting and scoring every rosette leaf per plant, we were able to create a comprehensive dataset of differences in shape and size across the leaf series of every plant in the Bay-0 x Shahdara RIL population. This would allow us to compare the leaf shape at each leaf of the leaf series, and to calculate an average leaf shape score for each RIL in the light and shade treatments.

In *Arabidopsis* leaf shape and size varies across the leaf series, see Figure 3-2, Figure 3-3 Difference in lIpPC3 between Bay-0 and Shahdara and Figure 3-4. This is described as heteroblasty. To record and analyse changes in shape across the leaf series, we developed a method to compare the shape of leaves at selected points in the leaf series. From the co-ordinate model dataset we created, which contained a model for each rosette leaf per plant harvested, we selected the co-ordinates for three rosette leaves per plant. We used the co-ordinate shape model of the third leaf, median node leaf and last leaf across the leaf series of each plant harvested to create a new, three leaf co-ordinate model for each plant, see Figure 3-5 and Materials and methods. The median node leaf was the leaf at the halfway point of the leaf series of each plant. We did not Procrustes fit these shape models so that size differences within the leaf series could be compared between lines. We used PCA to identify the major shape variations in these three leaf shape models, creating a set of heteroblasty PCs capturing the major shape variations in the dataset, see Figure 3-5. Heteroblasty PC3 and PC4 describe similar changes in leaf shape at each of the three leaves. Heteroblasty PC1 and PC2 show changes in shape and size that vary between the three leaves. For example, in PC1 the median node leaf appears to vary in size relative to the other two leaves, and PC2 shows a change in which of the three leaves is the largest, varying between the last leaf or the median node leaf.

3.2.3 Summary and Aims

For the population of 165 recombinant inbred lines we built a dataset of 6336 rosette leaves from a total of 938 plants grown in either light or shade conditions. Leaf shape was recorded using LeafAnalyser software and analysed by PCA.

Our aim was to compare leaf shape variation within the Bay-0 x Shahdara RIL population, using our dataset of leaf shape scores to compare shape at individual nodes, for whole plant averages, and differences in heteroblasty between lines. We would compare shape between RILs in both the light and shade treatments, and also examine possible variation in the plasticity of these leaf shape traits between treatments by calculating the extent of the difference in leaf shape between the light and shade conditions for each RIL.

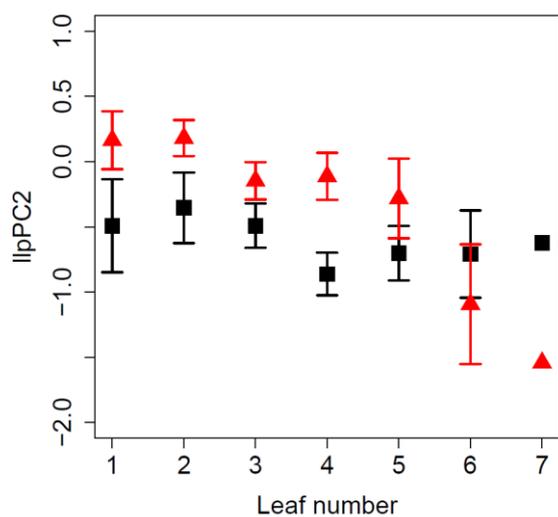


Figure 3-2 Difference in IlpPC2 between Bay-0 and Shahdara

Mean scores per node of the leaf series are shown as black squares for Bay-0 plants and red triangles for Shahdara plants. Error bars show 95% confidence intervals. For Bay-0 and Shahdara, 13 and 14 plants were harvested respectively.

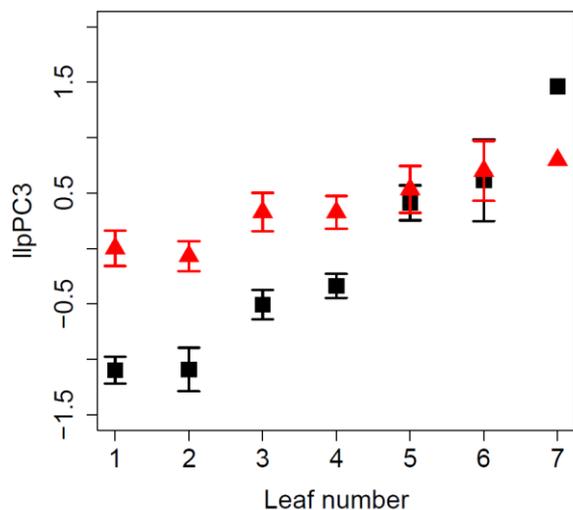


Figure 3-3 Difference in IlpPC3 between Bay-0 and Shahdara

Mean scores per node of the leaf series are shown as black squares for Bay-0 plants and red triangles for Shahdara plants. Error bars show 95% confidence intervals. For Bay-0 and Shahdara, 13 and 14 plants were harvested respectively.

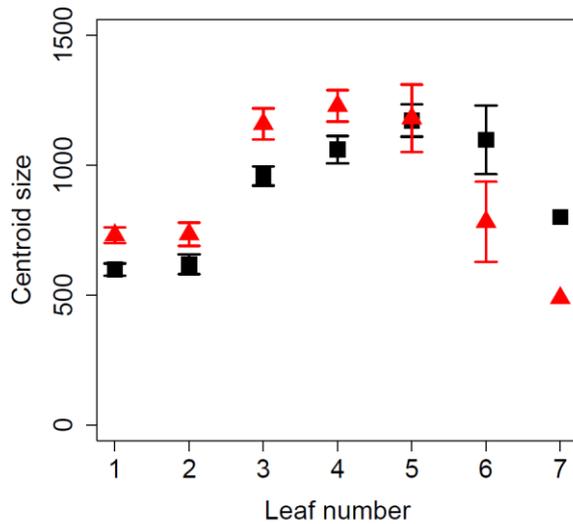


Figure 3-4 Difference in size between Bay-0 and Shahdara

Difference in centroid size across the leaf series between the parents of the Bay-0 x Shahdara RILs. Mean scores per node are shown as black squares for Bay-0 plants and red triangles for Shahdara plants. Error bars show 95% confidence intervals. For Bay-0 and Shahdara, 13 and 14 plants were harvested respectively.

Trait	Treatment	Heritability
days to bolt	light	0.91
days to bolt	shade	0.92
llpPC1	light	0.49
llpPC1	shade	0.59
llpPC2	light	0.8
llpPC2	shade	0.89
llpPC3	light	0.67
llpPC3	shade	0.83
llpPC4	light	0.59
llpPC4	shade	0.588
leaf number	light	0.936
leaf number	shade	0.91
phyllotaxy	light	0.51
phyllotaxy	shade	0.55

Table 3-1 Estimated heriability

This table shows the estimated heriability of each trait in light and shade treatments for the Bay-0 x Shahdara RIL population. Heriability above 0.5 indicates a higher proportion of trait variation is due to between line variation than within line variation, suggesting a genetic basis for the trait.

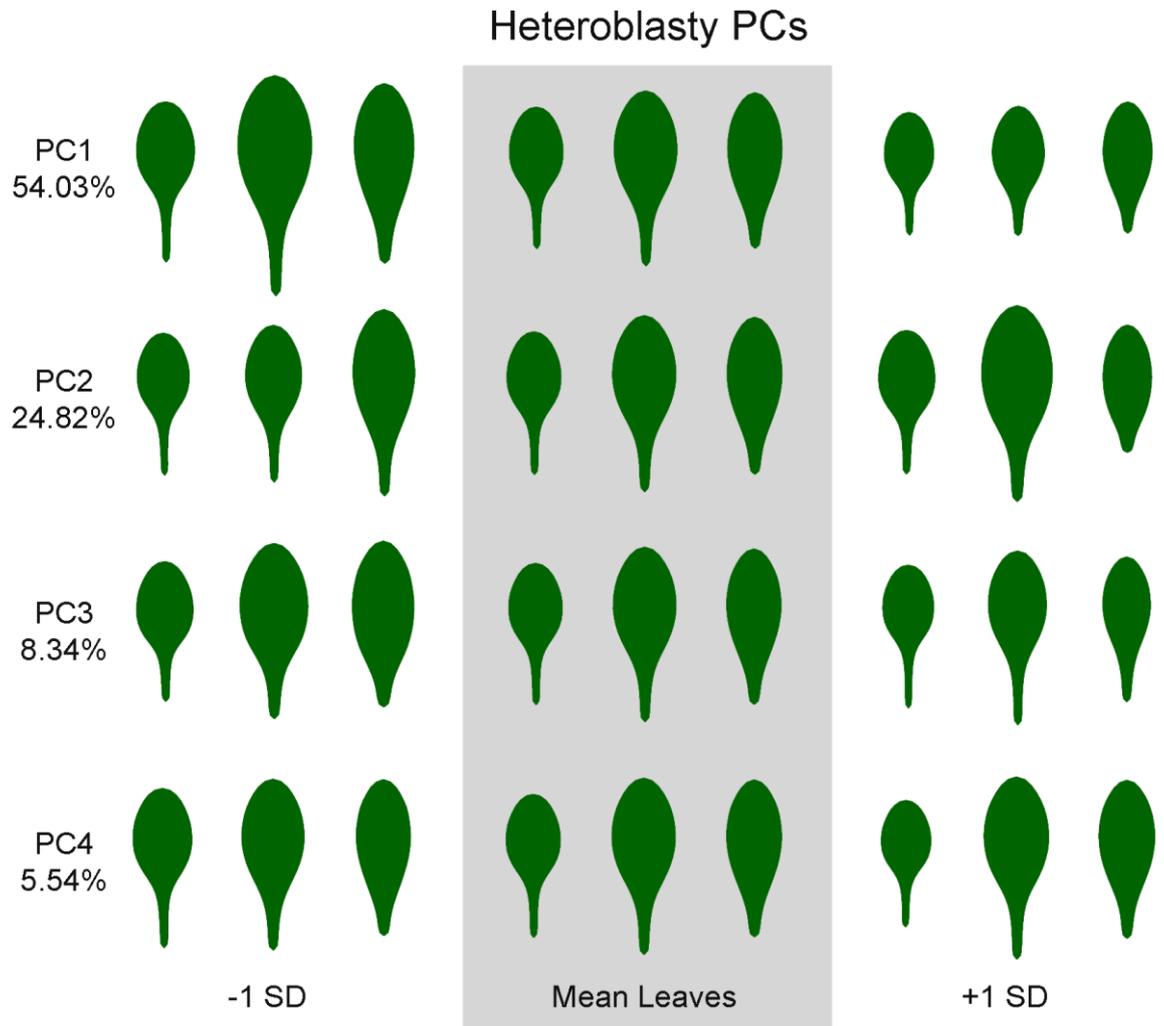


Figure 3-5 Heteroblasty PCA on Bay-0 x Shahdara population

Heteroblasty PCs created using a PCA on three leaf shape models for the Bay-0 x Shahdara population. Each line was represented using three leaves per plant, the third leaf, median node leaf and last leaf. The shape changes captured by these PCs show differences across these three leaves. The mean three leaf shape models are shown in the centre of the figure with a grey background. The shape changes captured by each PC are shown as +/- 1 standard deviation for each PC shown either side of the mean shape models. Some of the changes in shape are common amongst all three leaves, for example PC3 shows a similar change in shape across all three leaves. Others show changes affecting specific leaves, PC2 shows a change in the relative sizes of each leaf, a lower PC2 score meaning that the last leaf is the largest, and a higher PC2 score meaning the median leaf is the largest. The first four PCs of the heteroblasty PCA account for 92.73% of the total variation within the dataset.

3.2.4 High heritability of leaf shape suggests a genetic basis for this trait

We estimated the heritability of our leaf library PCs within the Bay-0 x Shahdara population by comparing between and within line variation. A heritability ratio above 0.5 indicates more variation between than within lines. Heritability was not estimated for the heteroblasty PCs as these were based on shape models using three mean leaves from each line, and so it was not possible to calculate within line variation. However we would expect scores for the heteroblasty PCs to be relatively heritable, given that shape changes scored for individual leaves appear highly heritable.

Traits such as days to bolt and leaf number have a high estimated heritability value of above 0.9, see Table 3-1. This was expected, as much natural genetic variation in flowering time has been identified in *Arabidopsis*, suggesting this trait has a strong genetic basis. We were encouraged to find our llpPC2 trait had a heritability value above 0.8 in both treatments. The score for our llpPC3 trait was lower, 0.67 in the light treatment and 0.83 in the shade, though this still suggested greater variation between than within lines for this trait.

Heritability was close to 0.5 in scores for traits llpPC1 and llpPC4, indicating as much variation for these traits existed within lines as between. This suggests there is no strong genetic basis for variation in these traits within the Bay-0 x Shahdara lines. It is likely variation in scores for llpPC1 and llpPC4 is the result of stochastic effects on plant growth. This is somewhat expected, given the shape variations these PCs describe. llpPC1 and llpPC4 describe changes in the curvature of the leaf or petiole angle, see Figure 3-6. It is likely variation in these traits is due to the position and of individual leaves within the physical environment.

Whilst it is possible that larger between line variation could be the result of methodological differences in the storage or generation of seed for each RIL, it is more likely that high heritability scores indicate variation for some of our leaf shape traits has a genetic basis in this population.



Figure 3-6 Leaf library PCs

The major leaf shape variations identified in the leaf library dataset with a PCA. Each PC is shown as -1 and +1 standard deviation alongside the mean leaf for the dataset. The first 4 PCs account for 85.44% of the variation in the dataset.

3.2.5 Identifying loci significantly associated to variation in leaf shape

Once leaf shape data was collected, R/qtl and Multiple QTL Mapping (MQM) packages (Arends et al., 2010) were used to detect QTLs. A threshold for QTL significance was estimated using the *mqmpermutation* function. This uses the random redistribution of trait scores amongst RILs to estimate the likelihood of identifying a QTL by chance. An LOD threshold, above which a QTL is considered significant, can be calculated by choosing an acceptable probability at which a QTL may occur through chance. We opted to use an LOD threshold based on a 1% probability that a QTL would be identified through chance. This gave us a LOD score of 2.79, above which any QTLs would be considered significant associations of genotype and phenotype.

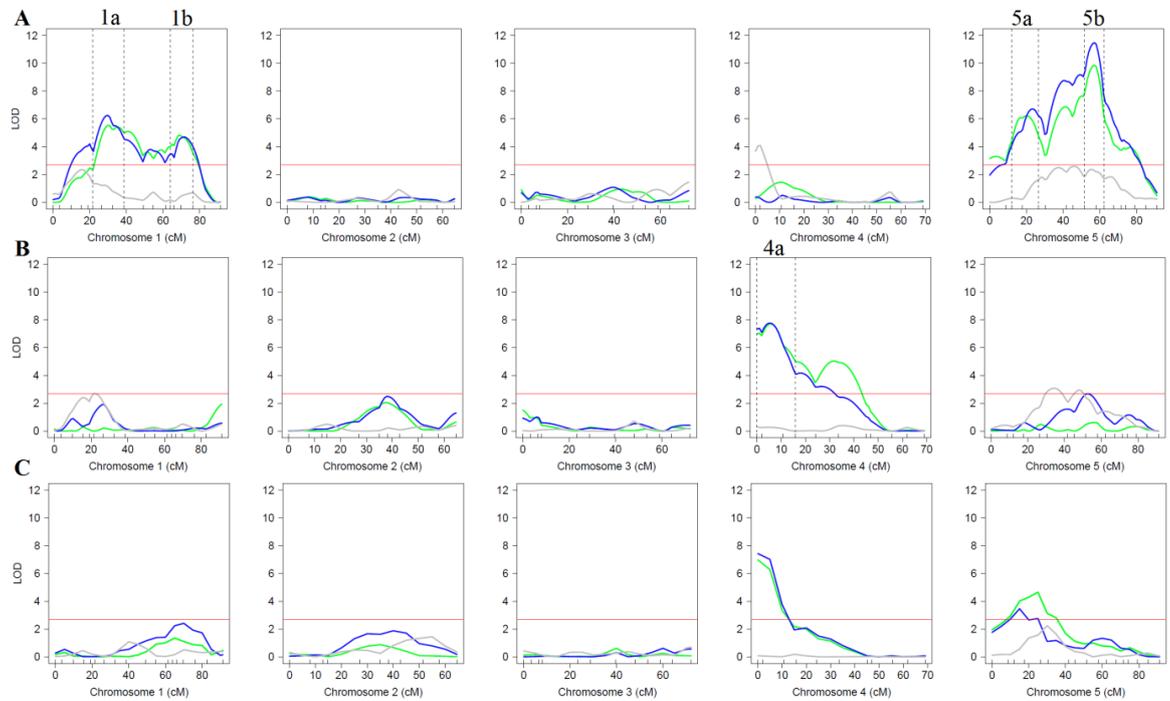


Figure 3-7 QTLs for leaf shape and number traits in the Bay-0 x Shahdara RILs

The QTL plots are shown for IlpPC2; A, for IlpPC3; B, and for leaf number; C. Green, blue and grey lines indicate the likelihood of a QTL existing across the chromosome, for the light, shade and plasticity phenotypes respectively. The red line indicates a LOD score above which QTL are considered significant. Vertical dotted lines show the marker intervals estimated to contain QTL. The major QTL associated with variation in IlpPC2 and IlpPC3 and referred to in the text are labelled as 1a, 1b, 5a, 5b and 4a.

QTL	Peak marker	Estimated effect trait
1a	NGA248	0.28 IlpPC2
1b	F5I14	0.28 IlpPC2
5a	NGA151	0.34 IlpPC2
5b	MSAT520037	0.26 IlpPC2
4a	MSAT4.39	4 leaves

Table 3-2 QTLs identified

This table contains information on the five QTLs shown in Figure 3-7. The marker closest to the QTL peak is shown. The estimated effect on each trait associated with the QTL is shown, calculated as the difference in means between RILs grouped by genotype at the peak marker for the QTL.

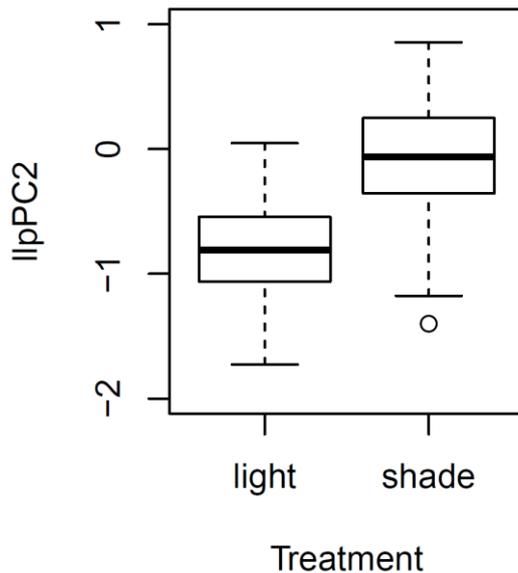


Figure 3-8 Average lIpPC2 score in light and shade treatment RILs

Boxplot shows the difference in lIpPC2 score between the light and shade treatments for the Bay-0 x Shahdara RILs.

3.2.6 Several QTLs are associated with the average leaf shape score for each RIL

We harvested each rosette leaf from plants grown from each RIL line in the light and shade treatments, creating a set of leaf models for the Bay-0 x Shahdara population. We scored the shape variation in these leaf models with our leaf library PCs and recorded the size of leaves as centroid size. We calculated plasticity between our light and shade treatments for each trait using the difference in trait score for each RIL between treatments.

Using the average score for each RIL, we identified several QTLs for variation in lIpPC2 and lIpPC3 score between the RILs. These can be seen in Figure 3-7 and Table 3-2. Shade treatment plants had a higher lIpPC2 score than the same RILs grown in the light treatment (paired t-test, $df=156$, $p<0.001$), see Figure 3-8. In both light and shade treatments we identified four QTLs for lIpPC2, two on both chromosome 1 and 5. No QTLs were identified for plasticity between light and shade treatments. A single large QTL was identified at the beginning of chromosome 4 in both light and shade treatments for average lIpPC3 score in the RILs. No QTL for plasticity were identified for lIpPC3.

We also mapped QTL for variation in leaf number, identifying a strongly associated QTL at the beginning of chromosome 4 in both light and shade treatments, in a similar position to the QTL associated with lIpPC3 score. A second QTL was identified for leaf number in a similar position to one of the lIpPC2 score QTL on the first half of chromosome 5.

The llpPC3 score of each rosette leaf varies over the leaf series in the Bay-0 and Shahdara accessions; leaves at later nodes of the leaf series have a higher llpPC3 score see Figure 3-3 Difference in llpPC3 between Bay-0 and Shahdara . As a result, although average leaf shape score for a RIL describes differences in shape common to each leaf in the leaf series of a plant, average score may also be partly determined by the number of leaves grown. We thought this observation may explain why a QTL associated with variation in llpPC3 coincided with the chromosome 4 QTL identified for leaf number variation.

To examine this we grouped the RILs by genotype at the peak marker for this QTL on chromosome 4, and plotted the mean llpPC3 scores for each node of the leaf series, see Figure 3-9. In both light and shade treatments, there was no difference in llpPC3 score at nodes over the leaf series between the two groups of RILs. This indicated that the difference in average llpPC3 score associated with this region was likely to be the result of later flowering plants producing more leaves with a higher llpPC3 score.

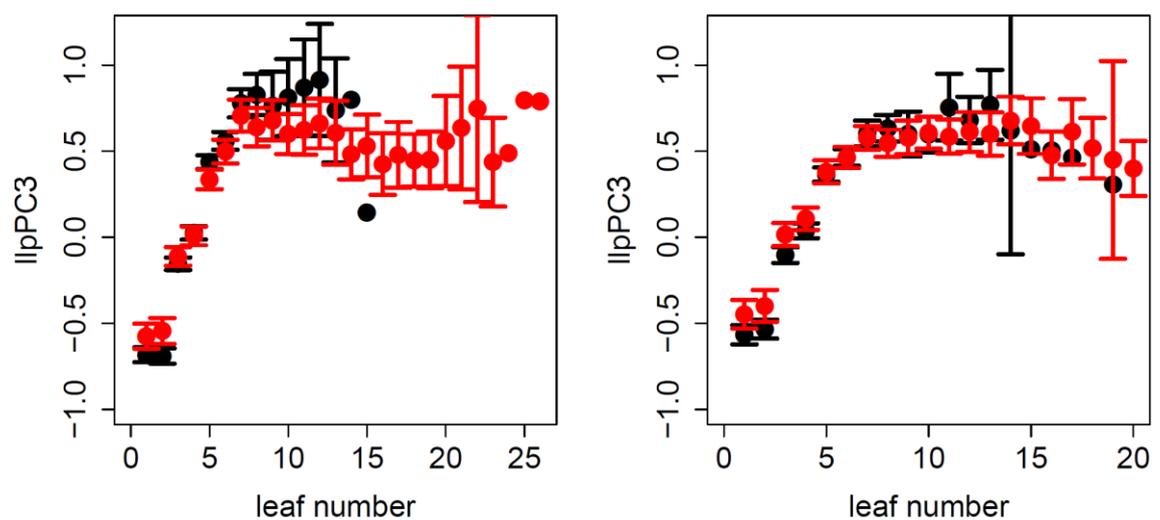


Figure 3-9 Difference in llpPC3 score between RILs

The mean llpPC3 score of light treatment grown RILs grouped by genotype at marker MSAT4.39 is shown in the left figure and for shade grown RILs in the right figure. There is little difference in llpPC3 score across the nodes of the leaf series in both treatments. Black circles show Bay-0 genotype, red circles show Shahdara genotype. Error bars show 95% confidence intervals.

3.2.7 Identifying QTLs for heteroblasty variation

We scored heteroblasty changes between lines by first combining the mean shapes of three leaves from set nodes across the leaf series for each RIL. Shape changes between each three

leaves shape models were identified with a PCA which captured the major changes in leaf shape as heteroblasty PCs (hetPCs), see Figure 3-5. We scored heteroblasty variation for each RIL using these hetPCs, and mapped QTL for variation in the Bay-0 x Shahdara RIL population for these traits see Figure 3-10.

Using each of our four heteroblasty PCs, we identified QTL in similar positions to those identified for the *llpPC2* and leaf number traits we described previously. The regions on chromosome 1 and 5 associated with variation in average *llpPC2* per RIL were also associated with the hetPC3 trait, and other combinations of these QTLs and the chromosome 4 QTL for leaf number variation were also associated with the other heteroblasty PCs, see Figure 3-10.

Interestingly we also found a QTL in a region not previously associated with variation in any of our leaf shape traits. This region was associated with variation in two of the hetPCs; hetPC1 and hetPC3, though LOD score of association was greater for hetPC1. That this region was associated for hetPC1 and hetPC3 indicated this QTL may affect leaf size and the relative size of the median leaf, and also shape across the leaf series in the Bay-0 x Shahdara RILs.

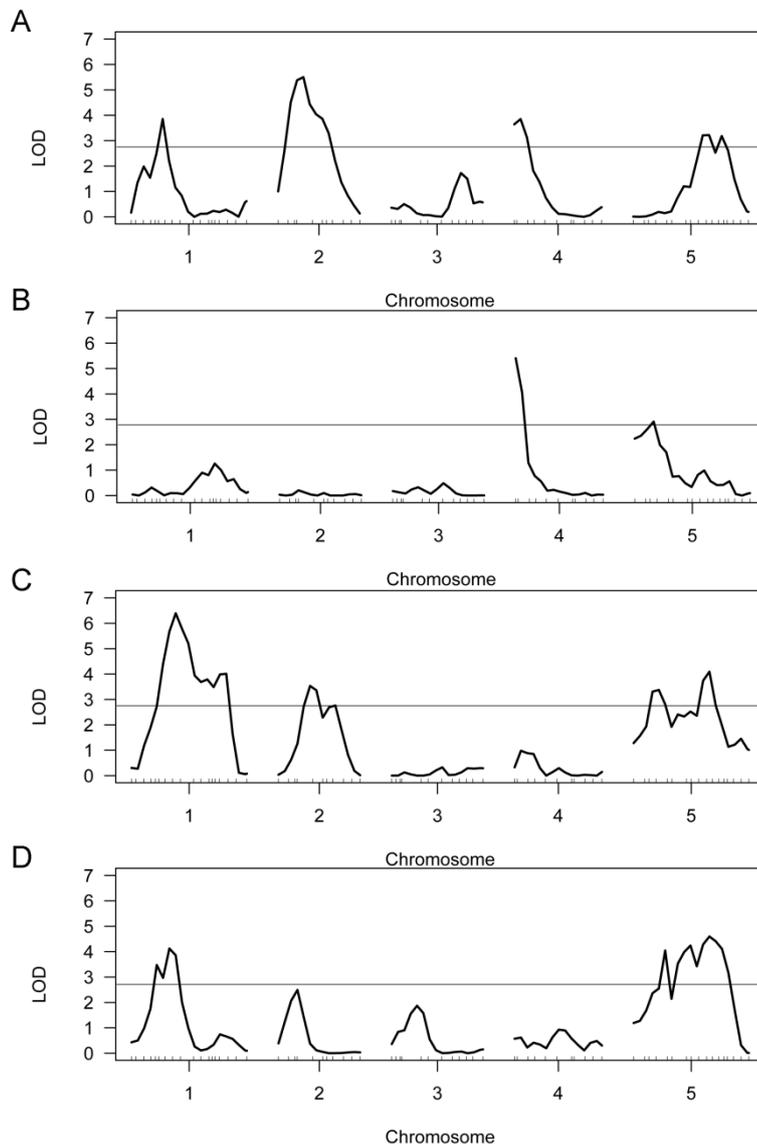


Figure 3-10 QTL plots for the first four heteroblasty PCs

Figure A shows association for hetPC1, figure B shows association for hetPC2, figure C shows association for hetPC3 and figure D shows association for hetPC4. LOD score describes the strength of association and is shown on the Y axis. The higher each peak, the more strongly associated variation in that trait is associated to the region of the genome. A horizontal black line indicates LOD score of 2.79.



Figure 3-11 Margin morphology variation

Photo shows some of the Bay-0 x Shahdara RILs. In the centre is a plant with the wavy margin trait. The leaves of surrounding plants have a flat margin.

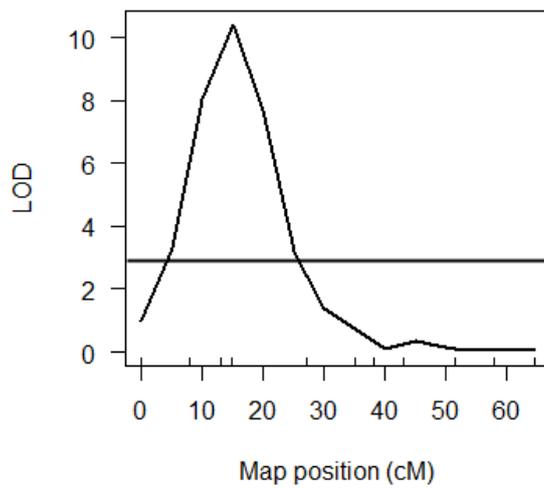


Figure 3-12 QTL for margin morphology

Using the wavy margin phenotype a single QTL was identified. QTL plot shows the position of this QTL on chromosome 2. LOD score is shown on the Y axis, and a horizontal black line shows a significance threshold of 2.79.

3.2.8 Margin morphology varies within the Bay-0 x Shahdara RILs

When growing the Bay-0 x Shahdara RILs we noticed differences in margin morphology that might be segregating within this population. The leaves of some plants had an uneven margin, giving the leaf a crumpled aspect. Other plants had contrastingly flat leaves that were quite distinct in comparison to those with an uneven margin, see Figure 3-11. To test whether this margin morphology variation mapped to QTL, we categorised each RIL for these two traits. 130 of the RIL population were categorised as having leaves with either an uneven margin, a particularly flat margin, or neither. We treated the presence of an uneven margin or a flat margin as separate traits, but found both were associated with a single QTL in almost exactly the same position on chromosome 2, with the peak of each QTL around 17cM, see Figure 3-12 QTL . This suggested variation in both these traits might be controlled by the same loci and that the presence of an uneven or flat margin is determined by alleles of the same polymorphism. Though the same region was identified as a QTL with variation of both uneven and flat margins, the QTL was much more strongly associated using the uneven margin trait, with an LOD score of 10, compared to 5.4 for the flat margin trait. This suggested the uneven margin trait better described the phenotype of this locus, and so we use the uneven margin description in further work on this locus.

It is worthwhile to note this QTL coincides with the chromosome 2 QTL associated with variation in hetPC1 and hetPC3. It may be that these variations in leaf size, shape, and margin morphology are the result of the same underlying polymorphism, though this cannot be confirmed without identification of the underlying polymorphism.

3.2.9 Comparing leaf shape at individual nodes of the leaf series

We had harvested all rosette leaves from each plant to create a leaf shape dataset for the Bay-0 x Shahdara RIL population. We were interested to determine whether leaf shape differences at individual nodes across the leaf series were controlled by the same loci, or whether different loci determined leaf shape at different nodes in the leaf series. We compared leaf shape within the Bay-0 x Shahdara population across the first nine rosette leaves of each plant in our light and shade treatments, and mapped QTLs for shape differences at each node. As this resulted in many individual QTL scans, we presented data as a heat plot, see Figure 3-13. A red region on the heatplot corresponds to a high LOD score, and indicates presence of a QTL. We found regions associated with shape variation at nodes across the leaf series on chromosome 1 and 5. These associations were in similar positions to the four QTLs across chromosome 1 and 5 that were associated with variation in the hetPCs and average llpPC2 score per RIL. Interestingly, there is a difference in node specific QTLs between light and shade treatment plants. For variation in llpPC2 for leaves at nodes four to nine in the light treatment, there is a strongly associated QTL on chromosome 4, in a similar position to the QTL on chromosome 4 associated with leaf number previously. Although a similar region is also associated in shade treatment plants, the strength of association is much weaker, see Figure 3-13.

To investigate further we plotted the llpPC2 score across the leaf series for the Bay-0 x Shahdara RILs, grouping the plants by genotype at the peak marker of this chromosome 4 QTL for light and shade treatments, see Figure 3-14. There was a considerable difference in llpPC2 score between the genotypes in the light treatment. Plants of a genotype associated with growing fewer leaves also had a lower llpPC2 score when comparing shape at nodes three and later. In the shade treatment plants, there appeared to be a similar, but reduced effect.

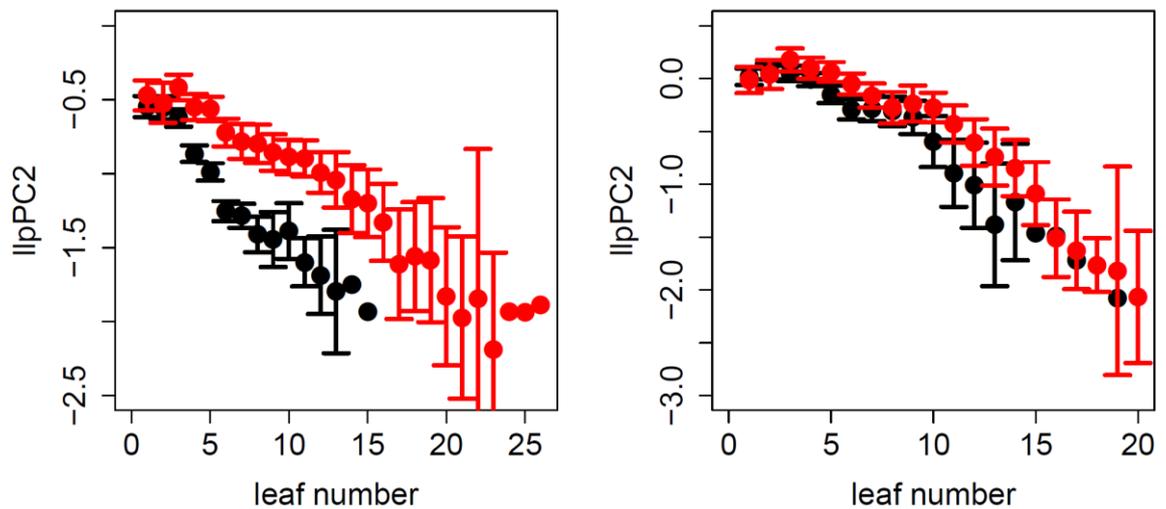


Figure 3-14 llpPC2 score between RILs grouped by genotype at a QTL

This figure shows the llpPC2 score across the leaf series for RILs grouped by genotype at marker MSAT4.39. The Shahdara genotype plants are shown as red circles, and Bay-0 genotype plants as black circles. This marker is at the peak of a leaf number QTL on chromosome 4. A clear difference in llpPC2 can be seen across the leaf series in the light grown plants, shown in the left graph. In shade grown plants, shown in right graph, there is no difference in leaf shape between plants grouped by MSAT4.39 genotype. Error bars show 95% confidence intervals.

3.2.10 Differences in leaf shape and number between light and shade conditions

We had grown the Bay-0 x Shahdara RIL population in light and shade treatments. A QTL on chromosome 4, associated with variation in llpPC2 score across the leaf series, had a stronger effect on llpPC2 in the light treatment compared to the shade treatment.

To investigate the effect of our treatments on plant growth, we compared trait values for RILs between treatments using paired t tests. Interestingly despite a significant difference in the number of days until bolting between light and shade treatments, see Figure 3-16, there was no difference in leaf number between RILs in the light and shade treatments. This meant plants grown in the light treatment typically bolted earlier, but grew the same number of leaves as those in the shade treatment, see Figure 3-15. We found a significant difference in llpPC2 score between RILs grown in light and shade conditions. When grown in shade, RILs had a significantly higher llpPC2 score than when grown in light, with a mean difference of roughly 0.7 llpPC2.

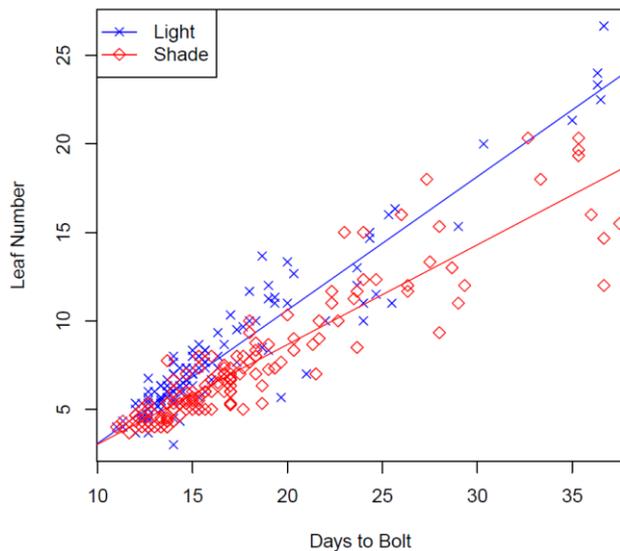


Figure 3-15 Relationship between leaf number and days to bolt

Figure shows the leaf number of each plant in the shade and light treatments, against the number of days until each plant bolted. The lines show a linear line of best fit for each treatment, with an r^2 value of 0.90 and 0.84 for the light and shade treatments respectively. In both cases, each treatment grows more leaves the longer the time till a plant bolts. However for shade grown plants, fewer leaves are grown per day before bolting.

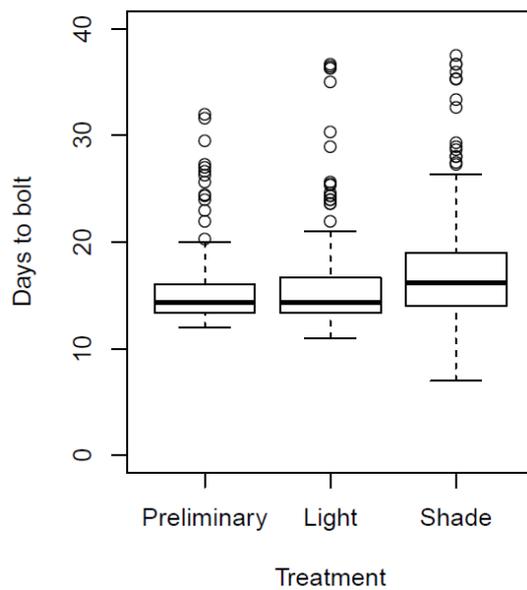


Figure 3-16 Days to bolt in Bay-0 Shahdara RILs between growth experiments

Difference in days to bolt for the Bay-0 x Shahdara population between the preliminary growth experiment, and the light and shade conditions of the main growth experiment discussed in this chapter. Paired t-tests identified significant differences ($p < 0.001$) when comparing RILs grown in light and shade conditions. The mean difference between RILs grown in these two conditions was 1.67 days to bolt less in light conditions.

3.2.11 Leaf shape analysis of the Bay-0 x Shahdara RIL population identifies several QTL

We had scored the Bay-0 x Shahdara RILs for our leaf library and heteroblasty PCs, and identified several QTL for these traits. We found six QTL controlling leaf shape, size and number across chromosomes 1, 2, 4 and 5. There was little difference between QTL

associated with these traits between the light and shade treatments, and we did not identify any QTL for plasticity of these traits between treatments. However, a region at the start of chromosome 4 contained QTL for both leaf number and node by node llpPC2 score, and the effect of genotype at this region on node by node llpPC2 score was considerably stronger in light grown plants than for those grown in shade. We also found that similar QTL were associated with shape variation across the leaf series. We next aimed to investigate further and confirm the effects of the QTL we identified on leaf shape traits.

3.2.12 Epistatic interactions between leaf shape QTL

We tested for the presence of epistatic interactions between the QTL we had identified. Epistasis occurs where the effect of a QTL is dependent on the genotype at a region outside the QTL. As well as suggesting a mechanism linking the causative polymorphisms of the two QTLs, it is also useful to know whether there is an epistatic effect between markers before investigating a QTL further with near isogenic lines (NILs). If a NIL is created to examine the effect of a small region of the genome, and other regions in epistasis are not fixed to a facilitating allele, the QTL effect cannot be detected within the NILs.

We tested for interaction between the four QTLs on chromosome 1 and 5 associated with variation in average llpPC2 score, see Figure 3-7 and Table 3-2. RILs were grouped by genotype at the peak markers for two QTL in turn and examined for an interaction between the markers, using the *effectplot* function in R/qtl.

After testing for epistasis across all combinations of our llpPC2 QTL, we found a mild epistatic effect of QTL5a, around 5Mb on chromosome 5, on QTL1a, around 10Mb on chromosome 1, see Figure 3-17. Genotype at QTL1a only has an effect on llpPC2 score when the peak marker for QTL5a, NGA151, has a Bay-0 genotype. We found this epistatic effect was only present in light grown plants, see Figure 3-17.

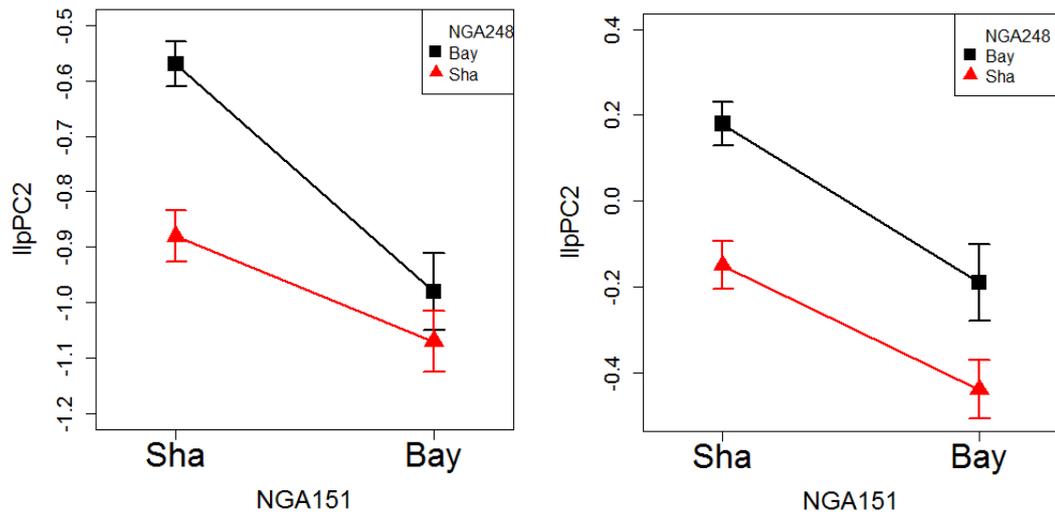


Figure 3-17 Epistatic relationship between QTL 1a and QTL 5a

Plots showing epistatic effect between QTL 1a and 5a. This effect is only present in the light treatment, shown in the left graph. Genotype at peak markers for QTL1a and 5a, NGA248 and NGA151 respectively, are used to group the plants. In the light treatment, shown left, a difference in genotype at marker NGA248 only affects the lIpPC2 score of the plants when marker NGA151 is the Shahdara genotype. This does not appear to be the case in the shade treatment RILs, shown right. Error bars show standard error.

3.2.13 Confirming effect of QTLs and defining causative region with HIFs

We used a Heterozygous Inbred Family (HIF) approach to test for the effect of QTL regions associated with variation in leaf shape traits between Near Isogenic Lines (NILs). HIFs are created from a RIL with a region of remnant heterozygosity (Tuinstra et al., 1997). This region is then fixed for either parent allele by self fertilisation to create lines that are genetically identical outside of this region. A set of these NILs, differing in genotype at one region, is referred to as a HIF, and can be used to examine the trait effect of genotypic variation at a specific region. Identifying a predicted QTL effect within a HIF for a region allows defined limits to be established for the QTL, using the known variant region between the NILs in a HIF. Multiple HIFs can be used to narrow the causative region for a QTL, for example if a QTL effect found within two HIFs for which the variant region overlaps, the polymorphism must lie within the overlap, and so the causative region for the QTL effect can be narrowed to the region of overlap.

We obtained HIFs from The French National Institute for Agricultural Research, (INRA), which covered the regions of the five QTLs we identified across chromosome 1, 2 and 5, see Figure 3-19. We grew these HIFs in our standard greenhouse conditions, referred to as the light treatment when growing the 165 Bay-0 x Shahdara RILs. The leaves of each plant grown per HIF were harvested once the plant had bolted and leaf shape was scored with our leaf library PCs. At least ten plants per HIF genotype were analysed for leaf shape. To increase the likelihood we were identifying the same variant regions for the initial QTL, any leaf shape effect identified within a HIF was compared to the direction of effect of each parent allele in the initial QTL.

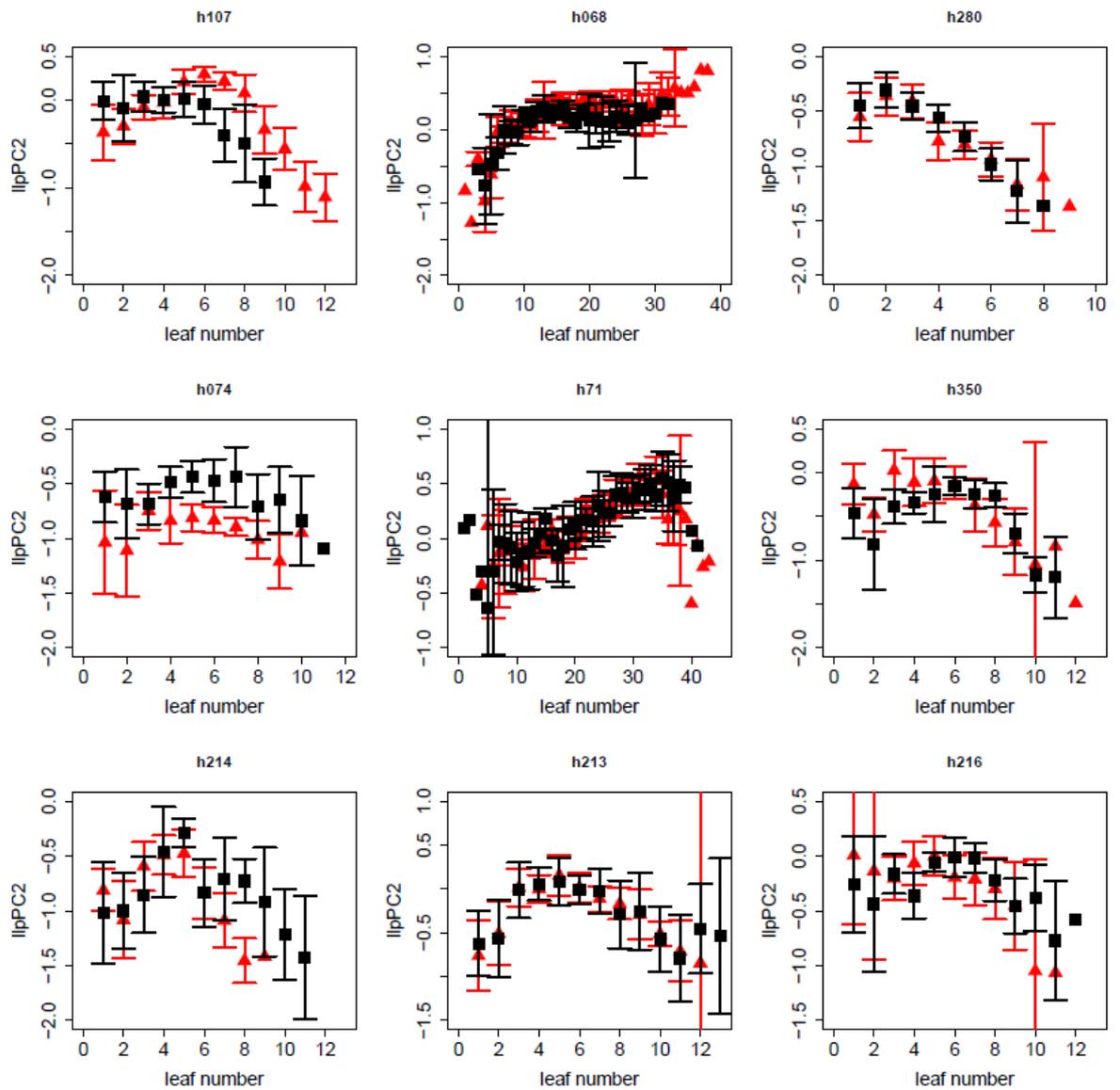


Figure 3-18 Differences in llpPC2 within HIFs

HIFs covering each llpPC2 QTL were analysed for differences in leaf shape. Each plot shows the mean llpPC2 score for plants grouped by their genotype at the variant region within the HIF. The mean score at each node for plants with a Bay-0 allele is shown as black squares and the mean for those with a Shahdara allele are shown by red triangles. Error bars show 95% confidence intervals.

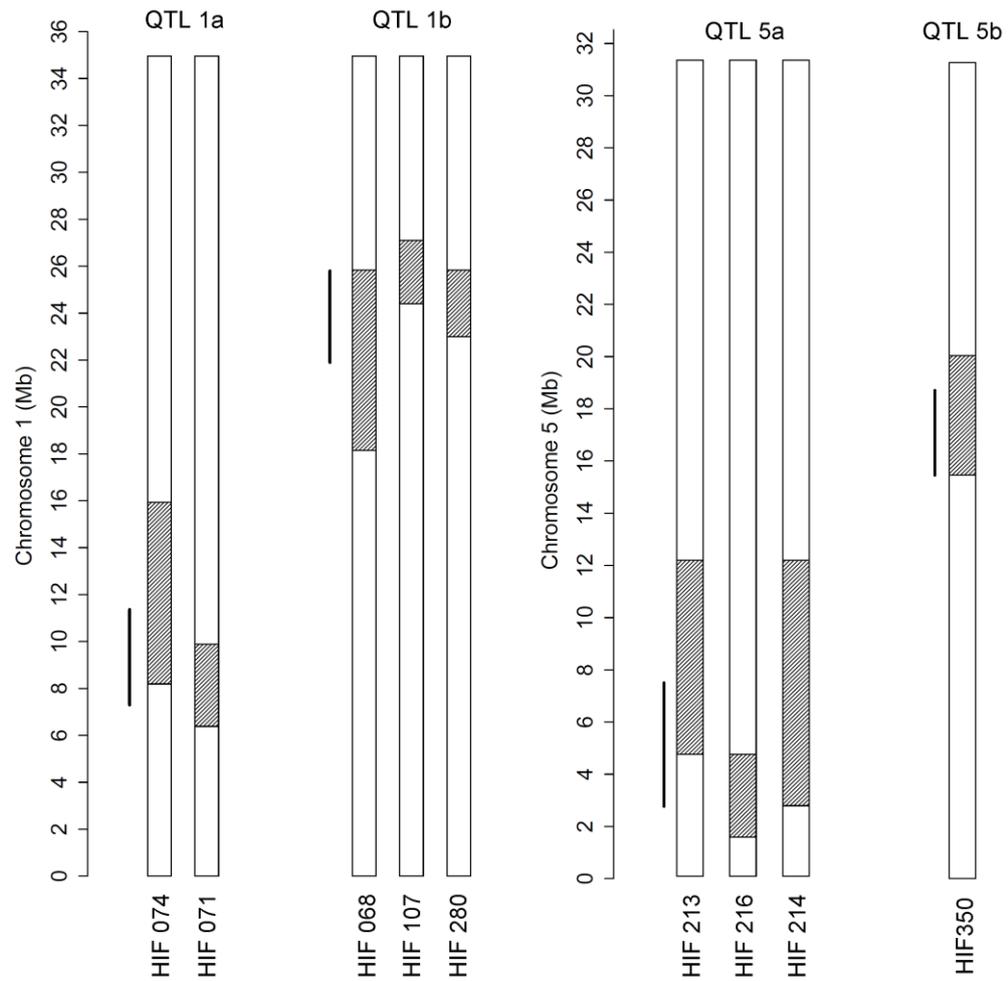


Figure 3-19 Genetically varying regions within HIFs

This figure shows the variant regions within HIFs used for validating each QTL. The white regions of the chromosome represent the area in which each does not differ within a HIF. The shaded regions show the region which varying in genotype within each HIF. The estimated position of each QTL is indicated against the relevant HIF genotypes by a vertical black line. Leaf shape effects were identified within HIF074, HIF107 and HIF350.

3.2.13.1 A similar leaf shape effect for QTL1a is identified within HIF074

QTL1a is a region of chromosome 1 centred around 10Mb associated with variation in average *llpPC2* score across the RILs. The effect of this QTL is dependent on the genotype of an epistatic QTL, QTL5a at marker NGA151. Of the two HIFs obtained to confirm the effect of QTL1a, HIF074 and HIF071, HIF074 had the Shahdara genotype at NGA151 and HIF071 had the Bay genotype. There was an *llpPC2* effect found within HIF074; lines with a Bay-0 allele had a higher *llpPC2* score as predicted for this region in the Bay-0 x Shahdara RIL population, see Figure 3-18 and Figure 3-20. This confirmed the QTL1a effect, and defined the loci as within roughly 8 to 16Mb on chromosome 1. HIF074 also had an effect on leaf number, lines with the Bay genotype at this region had significantly more leaves (Wilcox test, $W = 164$, $p < 0.025$). HIF071 has a Bay-0 genotype at NGA151. As such, although no effect similar to QTL1a was found within this HIF, we could not use HIF071 to inform us further about the location of QTL1a without also phenotyping this HIF in shade conditions, as a Bay-0 allele for the epistatic QTL at marker NGA151 is predicted to negate the effect of this QTL in light treatment conditions.

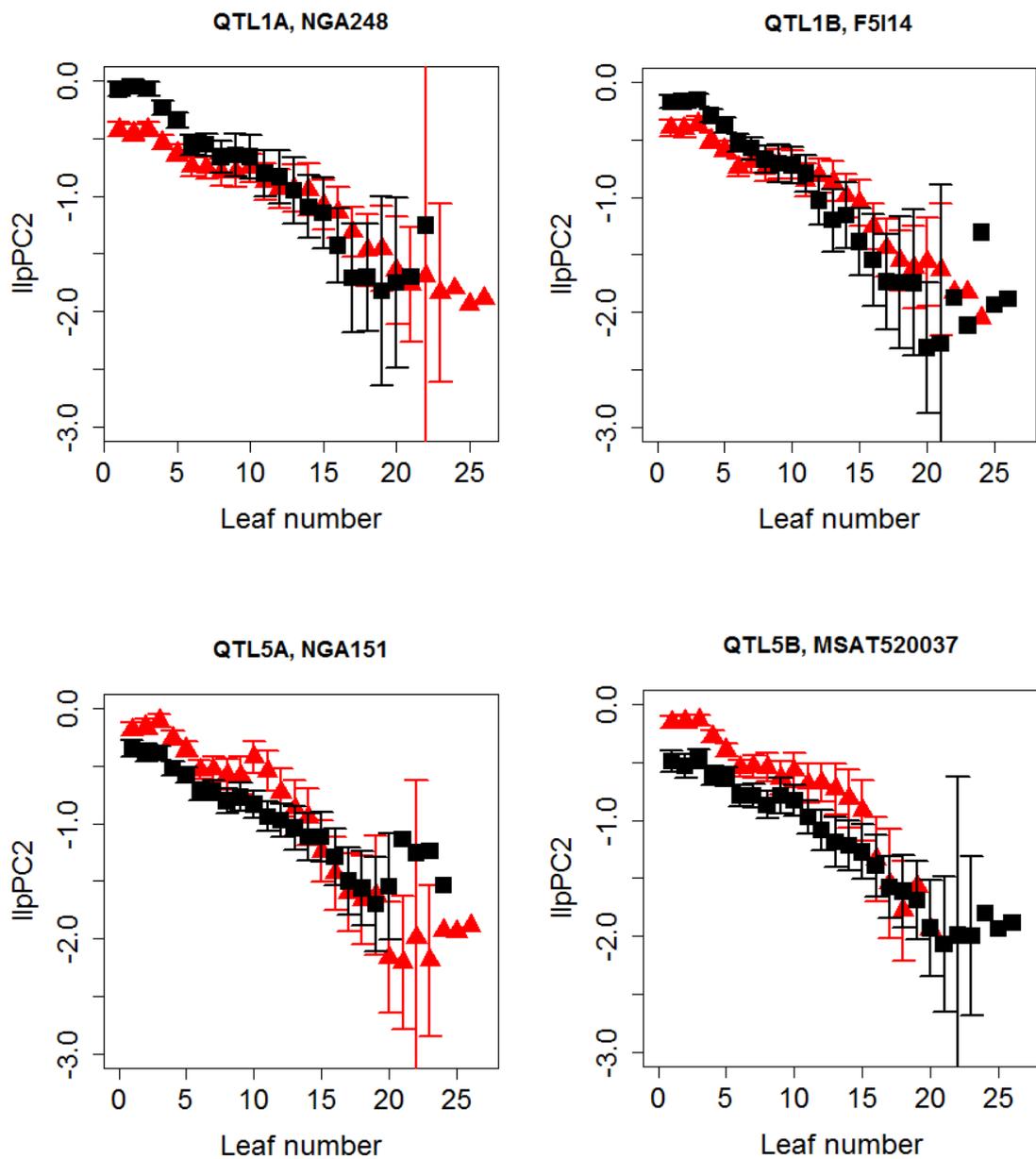


Figure 3-20 lIpPC2 effect of QTLs across RILs

This figure shows the effect of each of the QTLs associated with variation in mean lIpPC2 score per line. The Bay-0 x Shahdara population was grouped by genotype at the peak marker for each QTL to create the plots, using plants grown in both light and shade treatments. The mean lIpPC2 score across the leaf series is plotted. The mean for plants with a Bay-0 genotype at the peak marker of the QTL are shown as black squares, and the mean for plants with a Shahdara genotype are shown as red triangles. Error bars show 95% confidence intervals.

3.2.13.2 The variant region in HIF107 has a leaf shape effect similar to QTL1b

For QTL1b, a QTL associated with variation in average llpPC2 score in the RILs on chromosome 1 centred at 24Mb, We obtained three HIFs; HIF068, HIF107 and HIF280. We measured the leaf shape within thesis HIFs and found that llpPC2 score varied between plants with different alleles at the variant region within HIF107, suggesting the causative region for QTL1b lies within the variant region of HIF107, but not HIF068 or HIF280. HIF107 also had an effect on leaf number; lines with the Shahdara genotype at this region had significantly more leaves (Wilcox test, $W = 10.5$, $p < 0.001$).

3.2.13.3 The effect of QTL5a is not observed within HIF213, HIF214 or HIF216

We measured leaf shape in HIF213, HIF216 and HIF214, however, there was no effect on llpPC2 score similar to that for QTL5a, a QTL on chromosome 5 centred at 5Mb, and associated with average llpPC2 score variation in the RILs.

3.2.13.4 The effect of QTL5a is identified within HIF350

We obtained HIF350, which contained a variant region over QTL5b, a QTL on chromosome 5 centred at 17Mb. We found that genotype at the variant region within HIF350 was associated with a difference in leaf shape similar to the QTL5b effect. HIF350 also varied in leaf number, lines with the Bay genotype at this region had significantly more leaves (Wilcox test, $W = 92.5$, $p < 0.03$).

3.2.14 HIF102 and HIF118 vary for leaf size and margin morphology

Two HIFs, HIF102 and HIF118 were used to investigate the margin morphology QTL we identified on chromosome 2, centred at 5Mb. The region of this QTL was also associated with variation in leaf size identified with the hetPCs, see Figure 3-10.

Leaf size and margin morphology clearly differed with each allele at the variant region of both HIFs, see figures Figure 3-21, Figure 3-22, Figure 3-23 and Figure 3-24. As such the causative region is likely to be within the overlapping region between these HIFs, between MSAT200897 and IND628, see Figure 3-25.

The markers used to genotype the variant regions in each HIF at INRA are quite broadly spaced along the chromosome, see Figure 3-25. We hoped to narrow the potential causative

region for this QTL by genotyping this region with more markers, to identify the limits of each variant region in HIF118 and HIF102, see Figure 3-25.

We were unable to narrow the variant region of HIF102 further, as each additional marker we used within the annotated variant region for this HIF confirmed the region to vary between the NILs of the HIF. We found that NILs within HIF118 did not vary at the MSAT2.26 additional marker however, and so were able to identify a new border for the known variant region in this HIF, see Figure 3-25. The overlap of the newly defined variant region in HIF118, and that in HIF102 allowed us to identify a 4.33Mb region likely causative for the margin morphology and leaf size effects shared between these HIFs, between markers MSAT2.26 and IND628 on chromosome 2.



Figure 3-21 HIF118 plants

Line 118-12 has the Bay-0 allele at the variant region for this HIF, and smaller leaves with a wavy margin. Line 118-13 has the Shahdara allele at the variant region within this HIF, and has larger leaves with a flatter margin.



Figure 3-22 HIF102 plants

Line 102-10 has the Bay-0 allele at the variant region for this HIF, and smaller leaves with a wavy margin. Line 102-14 has the Shahdara allele at the variant region within this HIF, and has larger leaves with a flatter margin.

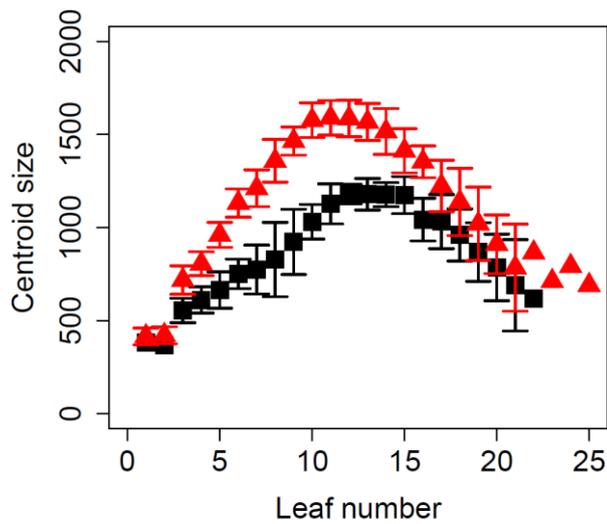


Figure 3-23 Difference in leaf size between HIF118 plants

The mean centroid size for HIF118-12 and HIF118-13 is plotted. HIF118-12 has a Bay-0 allele at the variant region in this HIF, and values for this line are plotted as black squares. HIF118-13 has a Shahdara allele at the variant region, and values are plotted as red triangles. For each line, 10 plants were harvested.

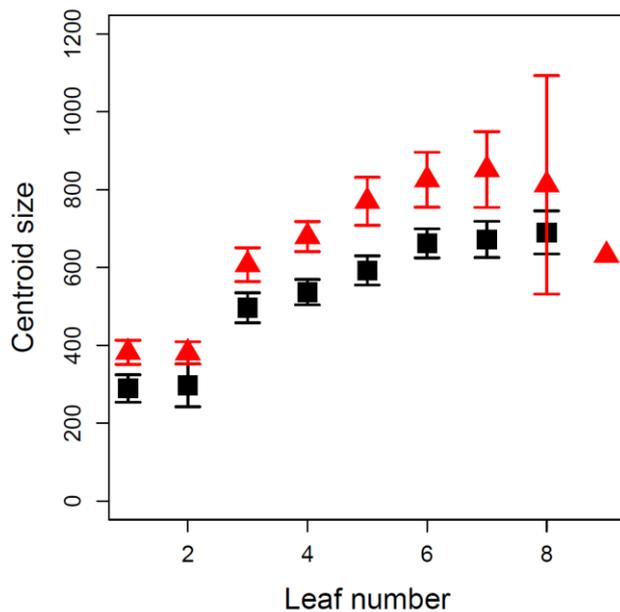
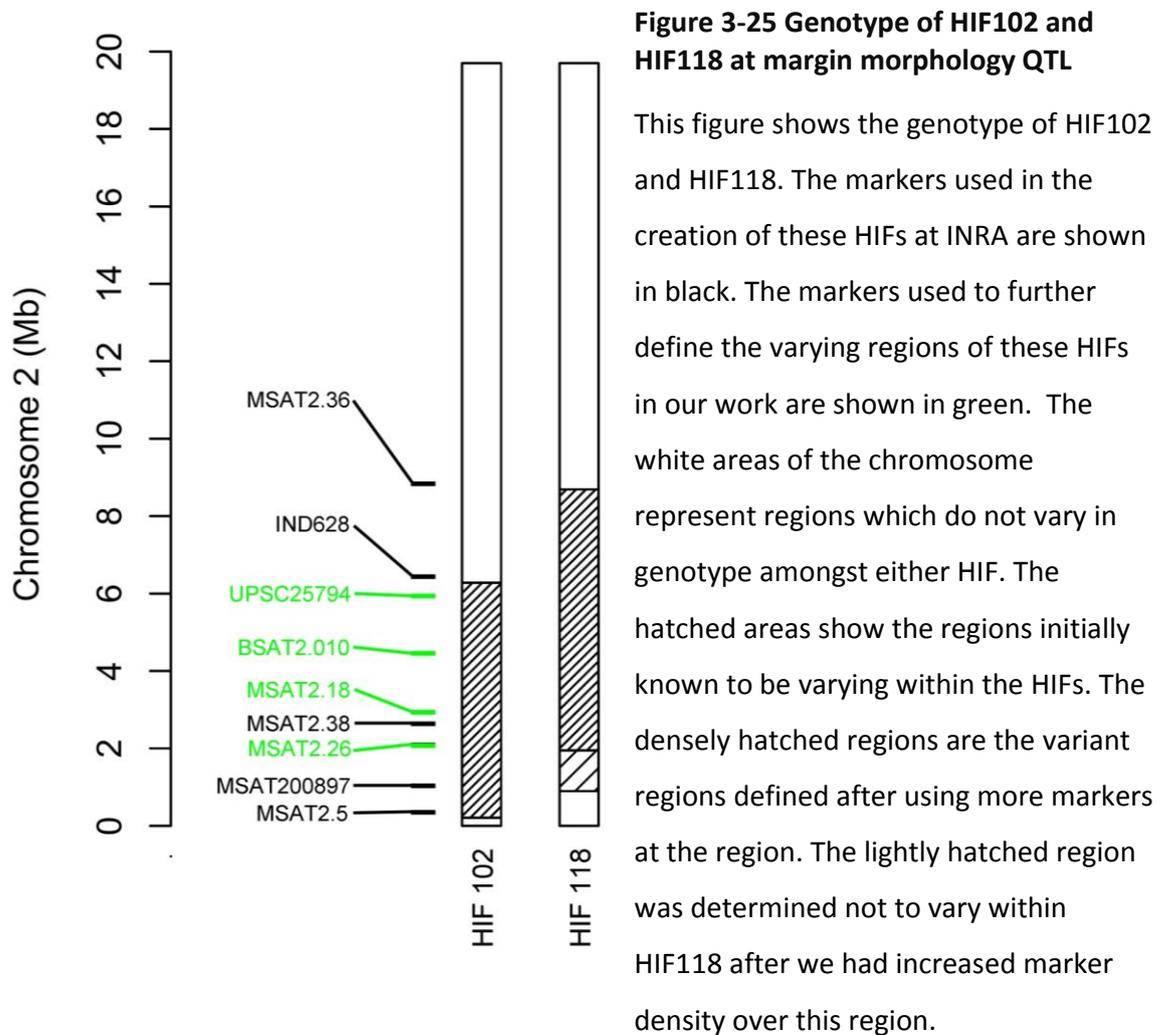


Figure 3-24 Difference between HIF102 plants

The mean centroid size for HIF102-10 and HIF102-14 is plotted. HIF102-10 has a Bay-0 allele at the variant region in this HIF, and values for this line are plotted as black squares. HIF102-14 has a Shahdara allele at the variant region, and values are plotted as red triangles. For each line, 10 plants were harvested.



3.2.15 Variation in life history and silique number in HIF102 and HIF118

We used HIF118 and HIF102, containing the causative region for a margin morphology and leaf size QTL to further investigate potentially pleiotropic effects on other aspects of plant growth. As the segregating phenotype within HIF118 and HIF102 has a considerable effect on leaf size, see Figure 3-23 and Figure 3-24, we wondered if this impacted the growth or life history of the plant. To explore this possibility we recorded days to bolt, silique number and length, and leaf number between the HIFs. There was no significant difference in days until bolting, within either HIF102 or HIF118, see Figure 3-27. Leaf number did not vary significantly within HIF102 or HIF118 either. Silique length did not vary significantly within HIF118, however, in HIF102, plants of NILs with the Shahdara allele produced more siliques per plant than those with the Bay-0 allele (Kruskal-Wallis, $df=1$, $\chi^2=5.2851$, $p<0.025$), see Figure 3-26. There was no significant difference in silique number within HIF118, see Figure 3-26 .

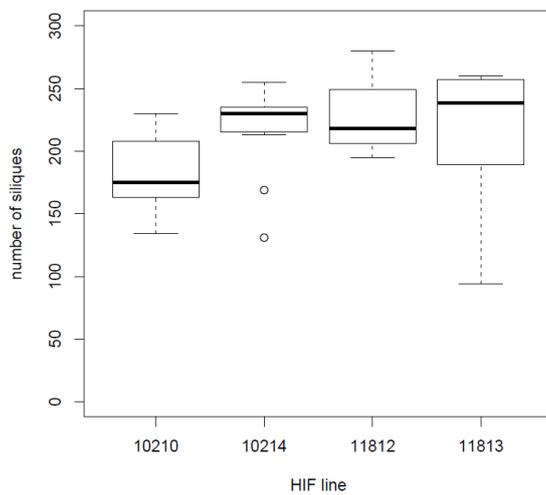


Figure 3-26 Difference in silique number within HIF102 and HIF118

Y axis shows the number of siliques whilst the X axis shows the HIF and genotype. Between genotypes of HIF102 there is a difference in silique number. Silique number does not differ between genotypes in HIF118.

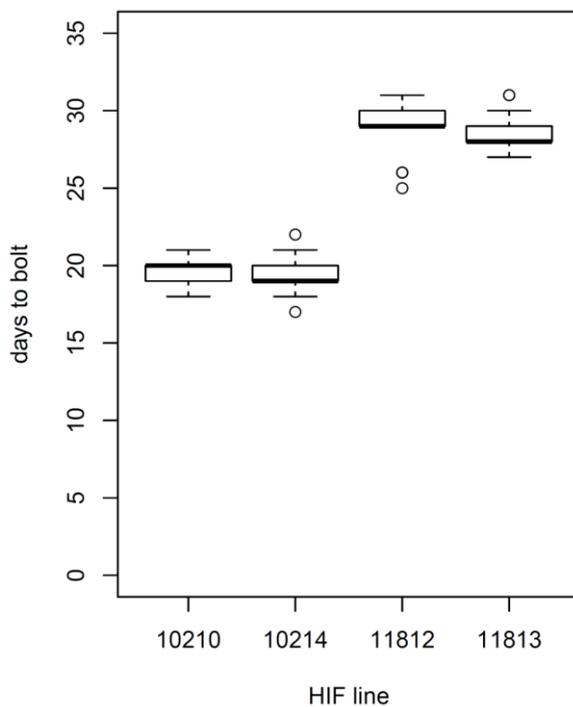


Figure 3-27 Difference in days to bolt within HIF102 and HIF118

Y axis shows the number of days before bolting whilst the x axis shows the HIF and genotype. There was no difference within HIF102 or HIF118 in number of days before bolting.

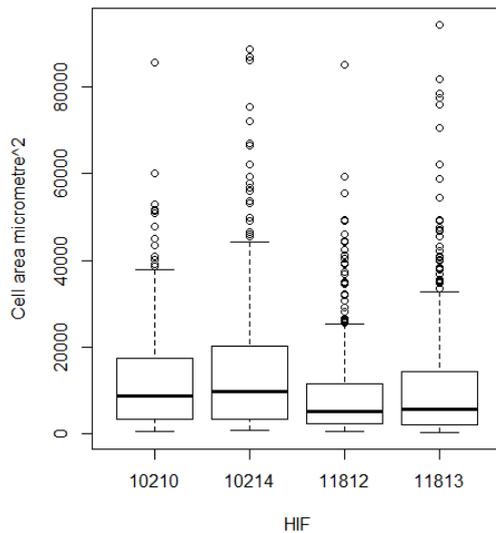


Figure 3-28 Epidermal cell area between HIFs

Graph shows the epidermal cell area in μm^2 per HIF line. There is no significant difference in cell area within either HIF, although within both HIFs, the Shahdara allele appears to confer a higher upper range and median of cell areas.

3.2.16 Variation in epidermal cell size and shape in HIF118 and HIF102

We tested whether the difference in leaf shape identified within HIF118 and HIF102 coincided with a difference in abaxial epidermal cell size and shape. For HIF118 we measured the size of epidermal cells of the median node leaf from 9 to 10 plants of 11812 and 11813. For HIF102 we measured epidermal cells from the fourth leaf of 8 plants of lines 10210 and 10214. Cells were measured from four positions on the abaxial side of the leaf and on average 43 cells were measured from each leaf. There was a similar trend for larger cell size in plants with a Shahdara allele at the variant region in both HIF102 and HIF118 plants, however the difference in cell size was not significant in either HIF, see Figure 3-28. This suggests that epidermal cell number, rather than cell size is largely responsible for the difference in leaf area observed within HIF102 and HIF118.

As a measure of cell shape, we tested perimeter, solidity, circularity and aspect ratio of the cells within the HIFs, see Figure 3-29 and Figure 3-30. Solidity, the amount of area of the convex hull of a shape the shape itself occupies, differed between the HIFs (HIF102; Kruskal-Wallis- $X^2=6.6$, $df=1$, $p<0.015$, HIF118; Kruskal-Wallis- $X^2=4.3$, $df=1$, $p<0.04$). Circularity, a measure of shape similarity to a circle, also differed between the parent alleles in both HIFs (HIF102; Kruskal-Wallis $X^2=6.1$, $df=1$, $p<0.015$, HIF118; Kruskal-Wallis $X^2=11.8$, $df=1$, $p<0.001$). This suggests there is a difference in the shape of the abaxial epidermal cells between Bay-0 and Shahdara alleles of the variant regions in HIF102 and HIF118.

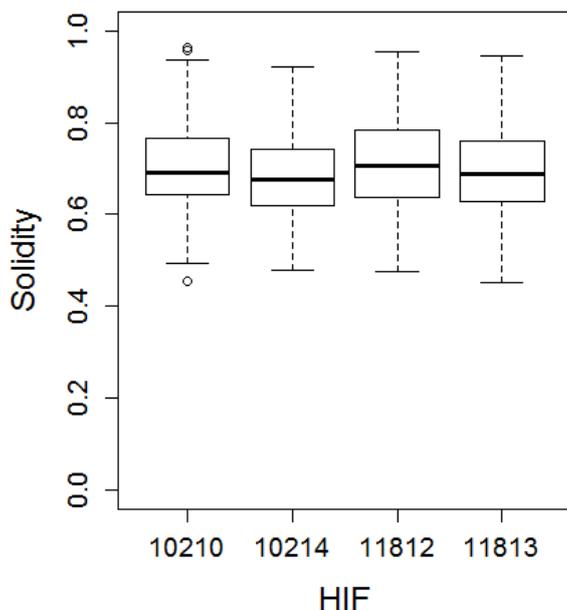


Figure 3-29 Shape differences between HIF abaxial cells

This graph shows the values for the solidity of abaxial epidermal cells between the HIFs. Within each HIF, the line with the Bay-0 allele, 10210 or 11812, has a significantly higher solidity value than the line with the Shahdara allele. Solidity is the proportion of a convex hull for a shape that shape takes up.

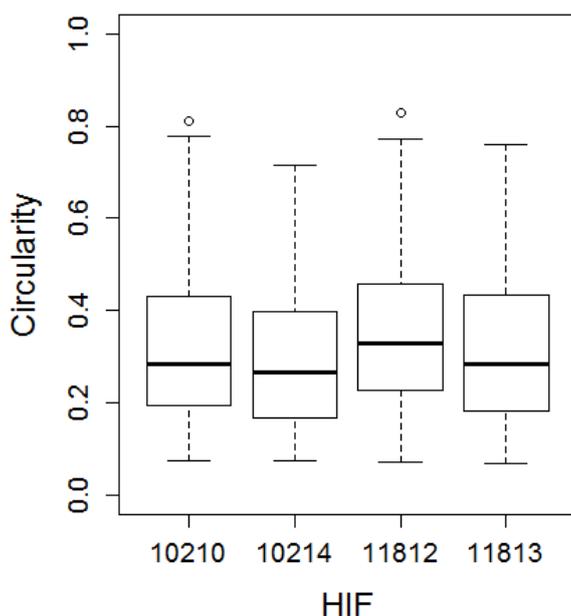


Figure 3-30 Shape differences between HIF abaxial cells

This figure shows the circularity values for each HIF. Within each HIF, the line with a Bay-0 allele in the HIF variant region, 118-12 or 10210, is associated with a higher circularity value. Circularity is calculated as $4\pi \cdot \text{area} / \text{perimeter}^2$. The higher the circularity value, the closer the shape is to a circle.

3.2.17 Network analysis to narrow list of candidate genes in the variant region between HIF102 and HIF118

The causative region for the margin morphology QTL was narrowed to 4.33Mb using the overlap of HIF102 and HIF118 and additional markers. This region contains 1069 gene models in TAIR9. A network analysis has been previously used in work on the Bay-0 x Shahdara population to identify a candidate gene for the shade avoidance response (Jiménez-Gómez et al., 2010). These authors identified and then experimentally validated a

candidate gene, ELF3, from a list of 363 genes within their causative region. As the number of genes within our 4.33Mb region was too large to work through using solely the published research on gene function in the literature to choose candidate genes, we opted to apply a network analysis to narrow our list of candidates.

Following the methods of (Jiménez-Gómez et al., 2010), the network analysis filters gene lists and links candidates in networks based on gene expression and regulation. Each gene within the 4.33Mb region was first connected to coexpressed genes, identified with the ATTED-JP dataset (Obayashi et al., 2009). Then these coexpressed genes were filtered so that only coexpressed genes with an expression QTL (eQTL) in the region of the networked candidate gene were left. This was done using gene expression data for the Bay-0 x Shahdara population (West et al., 2006, 2007). As well as the size of networks, Gene Ontology (GO) terms (Berardini et al., 2004) were used to identify networks made up of similarly functioning genes, more likely to be meaningful associations.

When collating coexpressed genes from the ATTEDJP dataset, we found coexpression data was available for only 292 of the 1069 candidate genes within the 4.33Mb region. This may be due to the position on chromosome 2 of this causative region, which contains part of the centromere. Centromeric regions often contain a high proportion of transposons and genes are typically repressed in expression (Copenhaver et al., 1999; Fransz et al., 2000). We were unsure as to whether the missing data was a result of biological reasons relating to the nearby centromere or due to genes in this region being excluded from microarray design. We proceeded with the network analysis and identified 10 genes with 8 or more direct network connections. These can be seen in Figure 3-31 and Table 3-3.

To test for possible effects on margin morphology and leaf size in these candidate genes, we obtained SALK T-DNA lines annotated with a T-DNA insertion in these genes. We grew plants for these lines in our standard greenhouse conditions, alongside the Col-0 background, and the HIF102 and HIF118 lines. We compared plants of the T-DNA lines to the Col-0 background, looking for an effect on margin morphology and leaf size similar to that observed within the HIF102 and HIF118 lines. However, no similar effect was observed within the 6 plants grown for each of the T-DNA lines during the plants development or at bolting, by which point the difference in leaf morphology and size was clear between HIF102 and HIF118 lines grown alongside these plants.

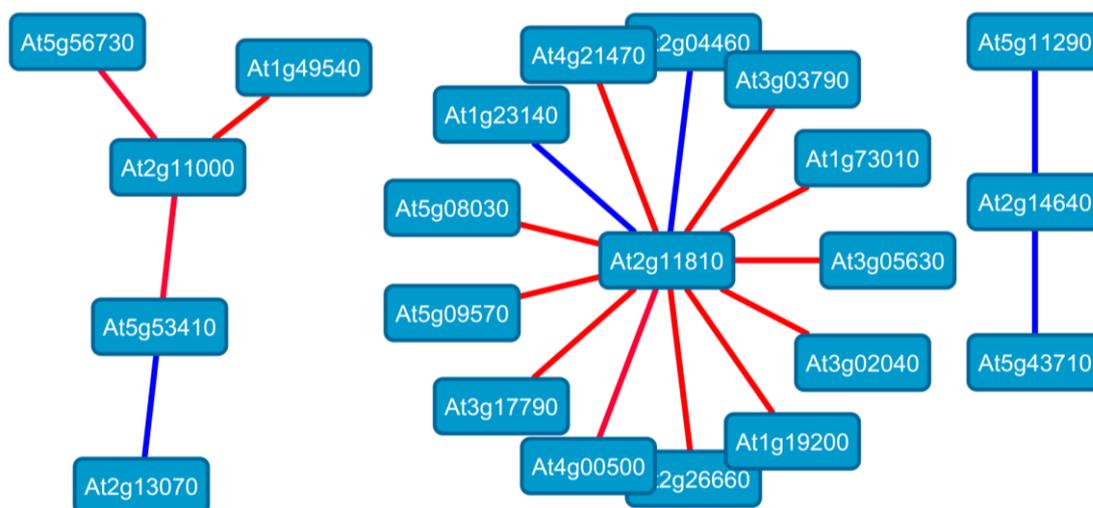


Figure 3-31 Example of network analysis results

An example of networks created in order to narrow the list of candidate genes from the wrinkled leaf causative region. Each candidate gene is connected to coexpressed genes that have an eQTL within the region of the candidate gene. A red connecting edge indicates at least one shared gene ontology term, a blue edge indicates none.

Gene	Edges	Shared GO terms	T-DNA lines scored
At2g10940	34	yes	1
At2g10950	25	yes	1
At2g10740	24	no	0
At2g11520	20	yes	2
At2g12550	13	no	2
At2g11910	12	yes	1
At2g10340	9	yes	1
At2g12400	9	yes	3
At2g11810	8	yes	2
At2g07690	8	yes	1

Table 3-3 Candidate genes identified with network analysis

This table shows a list of all candidate genes directly connected to 8 or more other genes in the network analysis, and the number of T-DNA lines scored for these genes

3.2.18 A polymorphism in At5g43630 may be responsible for llpPC2 variation in the Bay-0 x Shahdara RILs

QTL5b, centred around 18Mb on chromosome 5 and associated with variation in average llpPC2 amongst the RILs, lies within a similar region to a QTL detected for hypocotyl length variation in the literature (Loudet et al., 2008). These authors measured hypocotyl variation in the Bay-0 x Shahdara RIL population across light and temperature treatments (Loudet et al., 2008). Using HIF and fine mapping approaches, they mapped a QTL associated with hypocotyl variation on chromosome 5 to a 17.545–17.552Mb interval, containing one complete gene and portions of two others. The authors used transgenic lines to investigate the effects of the Bay-0 and Shahdara alleles of these three genes and noted that lines overexpressing the Shahdara allele of At5g43630 in a background with the Bay-0 allele had longer hypocotyls than in the untransformed background (Loudet et al., 2008). They also noted that ‘All linearly elongating organs were more extended, including petioles, internodes, and peduncles and the main floral stem’ (Loudet et al., 2008). They did not find any effect on hypocotyl length in transgenic lines expressing the other two candidate genes from the interval, indicating At5g43630 as the likely candidate gene for their chromosome 5 hypocotyl length QTL.

We wondered if the variant region identified by Loudet et al., (2008) was also responsible for llpPC2 associated QTL we identified in the same region. To test this, we obtained seeds for the fine mapped rHIF lines created by Loudet et al., (2008), rHIF47, containing either a Bay-0 or Shahdara allele at the 17.545Mb – 17.552Mb interval, but identical across the rest of the genome.

We grew plants for this rHIF line alongside plants of HIF350. This HIF had a variant genotype at the chromosome 5 llpPC2 QTL centred at 18Mb and showed a similar difference in llpPC2 to the initial QTL, and so we wondered if there would be a similar difference in leaf shape between plants of the rHIF as in HIF350. We measured the shape of each rosette leaf of these plants after harvesting at bolting and found the rHIF47 plants varied in leaf shape in a very similar way to HIF350, see Figure 3-33. In both HIF350 and rHIF74, the Shahdara allele results have a higher llpPC2 score compared to the line with a Bay-0 allele. This suggested the 17.545Mb - 17.552Mb interval contained the causative polymorphism for the QTL we found associated with llpPC2 variation at this region. Given the work of (Loudet et al., 2008) on the At5g43630 gene, we thought it very likely that a polymorphism in this gene was also responsible for association of variation in llpPC2 to this region in our work.

We further explored the relationship of hypocotyl length and leaf shape in the Bay-0 x Shahdara population by obtaining the hypocotyl phenotype data from the supplementary materials of (Loudet et al., 2008), and testing for correlations with llpPC2. We found there was a significant correlation between hypocotyl length published by these authors and the llpPC2 trait, see Figure 3-32.

We wondered if other QTLs we had found associated with variation in leaf shape traits were also associated with hypocotyl variation, and so tested for QTLs using the hypocotyl data of Loudet et al., (2008). We also created a new trait, using the residuals from the correlation of llpPC2 and hypocotyl length. This trait corresponds to variation in llpPC2 score not explained by a linear model with hypocotyl length. Comparing the QTLs from these two traits, see Figure 3-34, showed that QTL1a, centred at 9Mb on chromosome 1 also appeared to be associated with hypocotyl length variation. Using the residual llpPC2 to map QTLs identified associations with similar regions to QTL1b and QTL5a, on chromosome 1 and 5, centred at 24Mb and 5Mb respectively. This indicated QTL1b and QTL5a in our work were probably not also associated with hypocotyl length variation as phenotyped by Loudet et al., (2008).

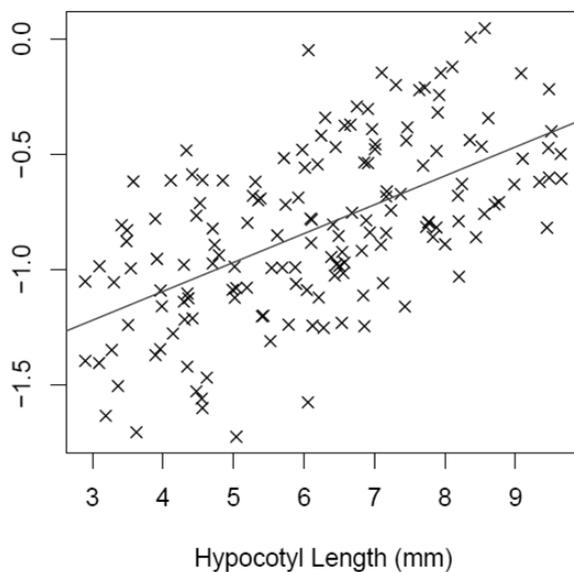


Figure 3-32 Correlation of llpPC2 and hypocotyl data

This figure shows the relationship between the leaf shape trait llpPC2, and the Loudet et al., (2008) hypocotyl data. There is a positive correlation between the traits. A line of best fit has been plotted, showing a linear model of the relationship between these traits, with an r^2 value of 0.29. Residual llpPC2 scores were calculated using this model.

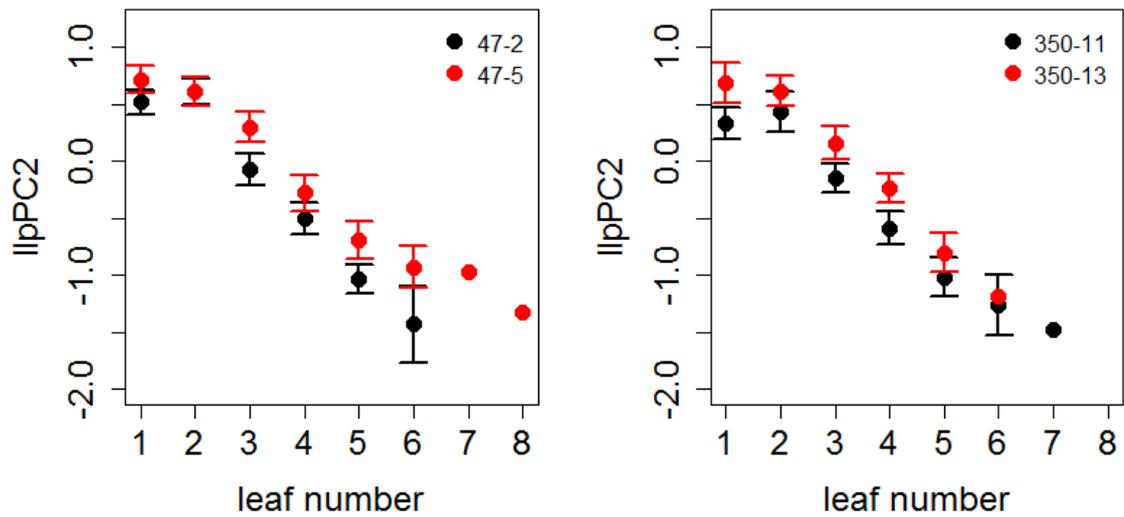


Figure 3-33 Difference in lIpPC2 score between HIF350 and rHIF47

This figure shows the mean lIpPC2 across the leaf series scores for rHIF47 and HIF350 lines shown on the left and right graphs respectively. In both HIFs the Shahdara allele lines, shown in red, have a higher lIpPC2 score than the Bay-0 allele lines, shown in black. Error bars show 95% confidence intervals.

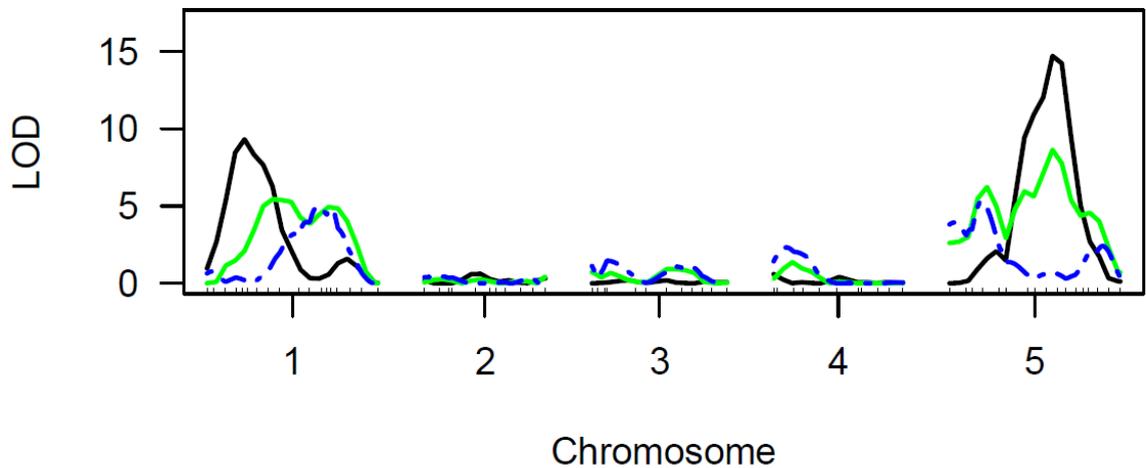


Figure 3-34 Coincidence of QTLs for lIpPC2 and hypocotyl traits

QTL plot showing QTLs for the Loudet et al., (2008) hypocotyl data in black, alongside the QTLs for our lIpPC2 phenotype in light conditions, shown in green. The blue dashed line shows the residuals from the correlation of light treatment lIpPC2 and hypocotyl length.

3.3 Conclusions

3.3.1 Multiple QTLs detected for leaf shape traits within the Bay-0 x Shahdara RIL population

We detected five QTL for variation in average *llpPC2* score amongst the Bay-0 x Shahdara RILs. QTLs were on chromosome 1, 4 and 5. That we detected several QTL of small to medium effect is consistent with leaf shape variation as a quantitative trait. As expected by the relatively similar leaf shape of the Bay-0 and Shahdara parent accessions, the direction of effect on *llpPC2* score amongst these five QTLs varied for each parent allele. A Bay-0 genotype at two of the five QTLs increased *llpPC2* score, as did a Shahdara genotype for the other three QTL. The chromosome 4 QTL for *llpPC2* variation coincided with another for leaf number variation. We also detected a chromosome 2 QTL affecting leaf size and margin morphology. Variation for the margin morphology trait appeared to be controlled by a single locus in the Bay-0 x Shahdara population, as only one QTL of large effect was identified.

3.3.2 Some QTL effects are confirmed with a HIF approach

We used a HIF approach to try to confirm the effect of the QTLs associated with leaf shape traits, and identified similar effects within HIFs for four of the five QTLs. However, for QTL5a, a chromosome 5 QTL centred around 5Mb associated with variation in average *llpPC2* score, no similar effect could be identified within a HIF. There are several possible explanations for this. One is that this QTL was sensitive to small differences in greenhouse conditions between experiments, and so was not detectable during the HIF growth experiments. The specific genetic background of the HIFs used to try and confirm this QTL might also have obscured the effect of the QTL. There is also a small QTL at this region associated with the leaf number and days to bolt traits amongst the light condition grown RILs. We detected a leaf number difference within one of the HIFs at this region; HIF214. There was also a small difference in leaf shape identified for HIF214, however, the effect on *llpPC2* in this HIF was dissimilar to that for the QTL5a in the Bay-0 x Shahdara RILs. It may be we were unable to detect a similar leaf shape effect to QTL5a within HIFs for this region due to a confounding effect on leaf number within this region.

3.3.3 No QTLs were identified for leaf shape response to shade conditions

We grew the Bay-0 x Shahdara RILs in a light treatment and a shade treatment. By doing this we intended to identify QTL involved in trait plasticity between the two treatments.

Although we identified an effect of this treatment on the leaf shape of the plants, there did not appear to be genetic variation in the extent of this effect on leaf shape between the RILs, and so we were unable to map QTLs for plasticity.

This was somewhat surprising as studies of the shade response in hypocotyls and petioles had identified QTLs within the Bay-0 x Shahdara RILs (Jiménez-Gómez et al., 2010). A key difference in methodology may explain why we did not find the similar plasticity QTLs to these authors. In their work, Jiménez-Gómez et al., (2010) used white light supplemented with far red light as a shade treatment. They kept photosynthetically active radiation (PAR) constant across simulated shade and sun treatments, altering red to far red ratio using supplementary far red light. A decrease in the red to far red ratio, even under non shade levels of PAR can activate a specific type of shade response phenotype in plants (Morgan and Smith, 1976). The shade treatment in our work on the Bay-0 x Shahdara RILs was created by growing plants underneath white muslin cloth. The effect of this treatment on light would be most likely a decrease in PAR, with relatively little alteration of wavelength ratios. Response to changes in levels of PAR can be separate from the response to change in red to far red ratio, which is monitored by phytochrome states within the plant (Smith and Whitelam, 1997). This potentially explains why we did not detect a QTL for plasticity between treatments in a similar region to (Jiménez-Gómez et al., 2010).

Changes in red to far red ratio are associated with the presence of overshadowing neighbouring plants (Schmitt, 1997). It is interesting that the Bay-0 x Shahdara population varies in response to red to far red ratio, but not to a general reduction in PAR. It could be that these two *Arabidopsis* accessions are differently adapted in tolerance of plant neighbour shading, resulting in variation in response to red:far red ratios. PAR may not vary consistently between the native environments of the Bay-0 and Shahdara accessions. The absence of QTLs for this trait suggests that not only do the Bay-0 and Shahdara accessions respond similarly to a decrease in PAR, but that there is no variation between the parents at the individual loci controlling this response. There is some evidence supporting this lack of variation in response to changes in PAR in this population. (Loudet et al., 2008) found when measuring hypocotyl length at $10 \mu\text{mol}/\text{m}^2\text{s}^{-1}$ and $17 \mu\text{mol}/\text{m}^2\text{s}^{-1}$, that although hypocotyl

length amongst the Bay-0 x Shahdara RILs changed between treatments, the extent of this difference did not vary significantly amongst the RILs (Loudet et al., 2008).

3.3.4 Light condition may affect the effect size and interaction between leaf shape QTL

Although we did not identify any QTL for plasticity in our leaf shape traits between our light and shade treatments, we found that in some cases light treatment altered the effect size or interaction of our QTLs. The chromosome 4 QTL associated with leaf number and *llpPC2* had a much stronger effect on *llpPC2* score in light treatment than in shade, and an epistatic interaction between QTL1a and QTL5a was only present in light grown plants.

One possible explanation for the larger effect of the chromosome 4 QTL on *llpPC2* score in light conditions is that the difference in *llpPC2* score associated with this region is a result of leaf number variation also linked to this locus. Plants with a Shahdara allele at this locus have more leaves and, in the light treatment, a higher *llpPC2* score at each node than plants with a Bay-0 allele. As plants responding to shade conditions typically have a higher *llpPC2* score, we wondered if the increased number of leaves within the rosette for Shahdara allele plants led to greater shading within the rosette, causing a higher *llpPC2* score. This effect is then perhaps much reduced in shade conditions, where plants are already responding to a high level of shade regardless of leaf density within the rosette. It is possible the light treatment specific epistatic interaction between QTL1a and QTL5a is the result of a similar effect, as the region containing the *llpPC2* associate QTL5a is also associated with a small leaf number QTL. Work in the literature has identified leaf morphology effects of a gene associated with differences in leaf number in *Cardamine Hirsuta* (Cartolano et al., 2015) and a tomato species (Shalit et al., 2009), however as these studies did not examine these effects across light and shade conditions it is unclear how similar such effects may be to the coinciding *llpPC2* and leaf number QTLs we identified in the Bay-0 x Shahdara population.

3.3.5 Major trends in leaf size variation involve relative changes across the leaf series

To identify changes in shape across the leaf series, we used the mean third, median node and last leaf of each line to build a three leaf shape model for each RIL. The median node and last leaf were chosen as these are set positions in the leaf series common to every plant grown. The third leaf was chosen rather than the first or second to represent shape early in the leaf series as the first two leaves of *Arabidopsis* typically appear somewhat distinct in

shape and size from the rest of the leaf series. The first two leaves also grow 180 degrees from one another, indicating a different pattern of phyllotaxy more similar to cotyledons than the majority of rosette leaves. The third leaf in *Arabidopsis* grows at an angle closer to the average 137.5 degrees of *Arabidopsis* spiral phyllotaxy, and also appears more similar in shape and size to the other leaves within the leaf series. Using this approach we created a set of hetPCs to score heteroblastic variation in the Bay-0 x Shahdara RILs. We found QTLs in regions previously associated with variation in llpPC2 score were also associated with heteroblasty PCs. We also found a QTL region on chromosome 2 associated with hetPC1 and hetPC3 traits, which describe variation in size and shape.

There is previous work on genetic mapping of heteroblastic leaf shape changes (Costa et al., 2012). These authors mapped differences in heteroblasty in *Antirrhinum*. In contrast to *Arabidopsis*, *Antirrhinum* grows the majority of its leaves along a central stem which later becomes the inflorescence. Although this is a distinct system to the rosette leaves of *Arabidopsis*, their work provides an interesting comparison. Costa et al., (2012) show in their PCA on *Antirrhinum* heteroblasty shape changes that the PC explaining the largest amount of co-ordinate variation is one controlling size, similar to our results. Interestingly, the changes in size across their leaf series captured in their PC1 vary, so that some leaves change in size more than others. This is very similar to our hetPC1, perhaps suggesting that increases in leaf size are generally coupled with relative differences in size in the leaf series, or that after a point, a minimum size constraint is reached for leaves, and so relative differences in size are lost in lines within plants with small leaves.

3.3.6 Identification and characterisation of a margin morphology QTL

We identified a leaf margin morphology phenotype segregating within the Bay-0 x Shahdara RILs and were able to identify a single large QTL responsible for variation in this phenotype. The QTL associated with margin morphology provided an interesting comparison to those associated with variation in the llpPC2 trait. In comparison to quantitative traits such as llpPC2, which was associated with multiple loci of small to medium effect, the leaf margin morphology trait was associated with one large distinct peak, suggesting a single polymorphism was responsible for variation in this trait.

We used HIFs to confirm the effect of this margin morphology QTL. Although both HIFs showed a similar difference in leaf size and margin morphology to the QTL in the Bay-0 x Shahdara RILs, the difference in leaf size was clearer within HIF118 than HIF102, likely due to

the different genetic background of either HIF. Given that this QTL has a considerable effect on the size and morphology of the leaves, the lines with a Bay-0 allele have a much smaller rosette than Shahdara allele, we wondered whether this trait would affect other aspects of plant growth, and measured days to bolt, leaf number and silique number and length within HIF102 and HIF118. There were no significant differences in these traits across both HIFs, suggesting the difference in leaf size did not alter wider aspects of plant growth, although silique number did differ within HIF102.

Similar results have been reported elsewhere, such as in (Costa et al., 2012), where QTLs controlling the size and shape of leaves did not coincide with a difference in flowering time. It was unclear why silique number would only vary within HIF102. It may be that because HIF102 lines flower considerably earlier than the HIF118 plants the change in leaf size has a more severe effect on the plants growth and ability to produce seeds. One way to test whether this region affects silique number in different genetic backgrounds would be to measure silique number in the Bay-0 x Shahdara RILs. If the QTL for leaf size and margin morphology affects silique number, the same chromosome 2 QTL should be identified for variation in silique number, and if the effect on siliques is background dependent, epistatic loci may be identified too. Given the possibility that the later flowering of HIF118 could obscure a difference in silique number, we may expect such epistatic loci to have life history related phenotypes.

To determine whether the difference in leaf size within HIF118 and HIF102 coincided with differences in epidermal cell size or number, we measured the size of epidermal cells in leaves from the HIF102 and HIF118 lines. There was no difference in cell size between the NILs within HIF102 and HIF118, suggesting the difference in leaf area coincided with a difference in cell number rather than cell size. Analysis of cell shape showed there was a difference in solidity, the area of the shape's convex hull that is taken up by the shape, and circularity, the similarity of the shape to a circle between NILs in each HIF. The Bay-0 and Shahdara alleles were associated with the same direction of effect in HIF102 and HIF118, suggesting there may be a difference in cell shape associated with this leaf size and margin morphology trait. Although slight, these differences suggest this QTL has some effect on the formation of the epidermal cells in these HIFs.

3.3.7 Loci controlling leaf shape in the Bay-0 x Shahdara RILs may have wider effects on plant morphology

Two of the QTLs we identified for *llpPC2* in the Bay-0 x Shahdara RILs coincide with two QTL identified for hypocotyl length in this population (Loudet et al., 2008). We found our *llpPC2* QTL resided in the same 7kb interval on chromosome 5 as these authors hypocotyl QTL using a fine mapped rHIF. Loudet et al., (2008) overexpressed the Shahdara allele of each of the three genes within this 7kb region in a line containing the Bay-0 allele and found only the expression of *At5g43630* caused an increase in hypocotyl length (Loudet et al., 2008). These authors found overexpression of this gene led to an extension of all linearly elongating organs, such as petioles, internodes and the main floral stem (Loudet et al., 2008). There were 14 amino acid changes, three deletions and an 8 bp insertion varying between the Bay-0 and Shahdara allele of this 7 kb region, but Loudet et al., 2008 identify the 8bp insertion in *At5g43630* as causative of the phenotype through comparison of an alternate Bay-0 line without the insertion. It is possible that any one of these polymorphisms or others in the two other genes partly within the fine mapped 7 kb interval could cause the *llpPC2* effect observed in the fine mapped rHIF we scored for leaf shape, however, given the phenotype described for lines overexpressing *At5g43630*, it seems likely to be a polymorphism within this gene that is responsible for the *llpPC2* effect we found when growing the rHIF. The *llpPC2* effect we identified at this region is probably the result of the same 8 bp insertion that is causing the hypocotyl phenotype, but without comparing the leaf shape of the alternate Bay-0 line lacking the insertion to the Bay-0 parent of the Bay-0 x Shahdara RILs, we are unable to differentiate between any of the polymorphisms within *At5g43630* as the probable cause of the association of *llpPC2* variation to this region.

Other work on the genetic mapping of leaf shape traits has also identified loci that control multiple aspects of plant morphology. Costa et al., (2012) identify QTLs controlling the leaf shape of *Antirrhinum*, and some of these QTL also have an effect on the height of plants when first flowering, and the internode length of the plants. QTL controlling leaf size in *Antirrhinum* have also been found to control petal size too (Feng et al., 2009). Our results have shown that it is very likely at least one of the QTLs we identified for leaf shape has a hypocotyl phenotype. We also found a locus associated with *llpPC2* coincided with a leaf number QTL. It seems reasonable to conclude that the major variations in leaf shape within a plant may be genetically controlled by loci with that also affect other aspects of plant morphology.

3.3.8 Using a network analysis to identify candidate genes

To create a list of candidate genes from the 1069 genes in the 4.33Mb causative region we identified for the margin morphology QTL, we used a network analysis approach. This strategy had been used previously to identify ELF3 as a candidate for controlling the shade response in the Bay-0 x Shahdara RIL population (Jimenez-Gomez et al., 2011). The method uses coexpressed gene databases and eQTL data to filter a list of candidate genes within the QTL region (Obayashi et al., 2009; West et al., 2006).

We repeated the analysis in (Jiménez-Gómez et al., 2010) using the same datasets, successfully identifying the candidates identified in these authors work. We then used our set of 1069 genes from the causative region of the QTL as the input, to create a shortlist of highly networked candidates for further investigation. Since the publication of (Jiménez-Gómez et al., 2010) there had been an update to the ATTEDJP database. The ELF3 paper used a gene coexpression dataset of 1388 gene chips (v4.1) (Obayashi et al., 2009). We had the option to use a later dataset, c5.0, a set of 11171 gene chips. As the v4.1 dataset was successful in identifying the causative gene for those Jiménez-Gómez et al., (2010), we decided to use both coexpression datasets in our analysis, and take the top genes with 8 or more direct network links for candidates. This left us with a shortlist of 10 candidate genes.

Unfortunately there was an issue of data availability when carrying out the data analysis for our chromosome 2 QTL. Of the 1069 candidate genes within the causative region, coexpression data was only available for 292 genes. The centromere of chromosome 2 begins around 6Mb (The *Arabidopsis* Genome Initiative, 2000), this is within the causative region we identified for this QTL, which spans from 1.95Mb to 6.28Mb. It is possible that the lack of available coexpression data for this region reflects the presence of the centromeric region as such regions are associated with a high density of transposons and low expression in *Arabidopsis* (Copenhaver et al., 1999; Schmid et al., 2005; Yamada et al., 2003). If this was the case, then the impact on the results of the network analysis may be limited, as these genes would be less likely to be responsible for the QTL. We proceeded with the network analysis, reasoning that if we were to find a likely candidate gene for this QTL, we would be able to validate this independently of the network analysis, and so it was worth continuing with this approach despite the problem of missing data.

We used SALK T-DNA insertion lines annotated with insertions in these candidate genes to test for phenotypes similar to that of the margin morphology QTL. However, none of the lines showed a morphological difference similar to that observed for the uneven margin QTL

within HIF102 or HIF118. There are a number of possible explanations for this result. The first is that the network analysis did not pick out the candidate gene amongst the highest connected genes. The next is that for this list of candidates, the T-DNA lines were not obtained for the causative gene, as was the case for 1 of the 10 candidates shortlisted. The gene with no T-DNA insertion lines available was At2g10950, and had 25 direct network connections. This gene is identified on TAIR as a gypsy-like retrotransposon, expressed during two and six leaf visible stage, and petal differentiation and expansion stage. It is possible that this gene contains the causative polymorphism for the leaf margin QTL, although given that it is not expressed during more stages of plant growth this seem unlikely given the clear difference in leaf size and margin morphology at this QTL. A third possibility is that we did identify the gene in which a polymorphism resulted in the QTL, but the mutations caused by the T-DNA insertions in these lines were not equivalent to this polymorphism, and so we did not see a similar effect on leaf size and margin morphology. SALK T-DNA insertion lines typically produce null mutations (Wang, 2008), and it is not necessarily the case that the polymorphism between Bay-0 and Shahdara is the result of a null mutation or similar polymorphism. For a lowly expressed gene, the overexpression of a truncated gene can have a much more dramatic effect than a null mutation (Faigón-Soverna et al., 2006). Given the number of genes within the causative interval for QTL2a, it seems most likely that the network analysis simply did not identify the gene containing the causative polymorphism.

Chapter 4. Genome Wide Association mapping for leaf shape traits

4.1 Introduction

In our previous work to map the genetic basis of leaf shape traits we measured leaf shape across the Bay-0 x Shahdara Recombinant Inbred Line (RIL) population and mapped 5 QTLs for leaf morphology. We were able to identify a likely candidate gene for one of our QTL. After our work with the Bay-0 and Shahdara accessions, we were interested to understand how leaf shape was controlled more widely across a greater number of *Arabidopsis* accessions.

An alternative approach to QTL mapping in a RIL population is Genome Wide Association (GWA). In a Genome Wide Association Study (GWAS), a population is genotyped and scored for a trait of interest, and variation in traits can be linked to regions of the genome. GWA mapping differs from a QTL approach in the population used. The RIL or F2 populations used in QTL mapping are created from the self fertilised offspring of a cross between two or more *Arabidopsis* accessions. This results in a population segregating for alleles of the parents. Although RIL populations can be created using multiple parents to increase the number of alleles segregating, at most 19 accessions have been used to create these multi-parent populations in *Arabidopsis* (Huang et al., 2011; Kover et al., 2009). In a GWAS, hundreds of natural *Arabidopsis* accessions are used as the mapping population. *Arabidopsis* is predominantly self fertilising and so natural accessions of the species are highly homozygous (Bakker et al., 2006; Bergelson et al., 1998). As such these lines need only be genotyped once, but can be regrown and phenotyped repeatedly, similarly to a RIL population.

In contrast to a RIL population, where alleles segregate due to recombination during the crosses that create the population, GWA relies on genetic differences built up over generations between accessions. Although this can result in a high level of genetic variation between lines useful for mapping, the balanced segregation of alleles typical in RIL populations cannot be assumed. For example, accessions collected from the same geographic area are more likely to be genetically similar, and allele frequencies amongst natural accessions will be much more variable than in a RIL population (Nordborg et al., 2005). These effects can cause spurious associations when testing for linked loci; regions may be linked to a trait simply because an allele correlates with presence of another, physically distant, allele with a causative polymorphism (Korte and Farlow, 2013; Zhao et al.,

2007). This has led to development of complex association analysis to facilitate mapping in GWAS (Kang et al., 2008; Segura et al., 2012).

Hundreds of *Arabidopsis* accessions have been genotyped as part of the 1001 genomes project (Cao et al., 2011; Long et al., 2013; Ossowski et al., 2008; Schmitz et al., 2013). This data has been used to provide markers for GWAS; single nucleotide polymorphisms (SNPs) are used in GWA mapping to compare genotype across the accessions. The marker density is often several orders of magnitude greater than used in RIL populations. One tool for testing for associations in a GWAS, GWAPP, currently uses a dataset of over 200,000 SNPs (Seren et al., 2012). In combination with the increased mixture of alleles typical amongst accessions used in GWASs, this increase in genetic markers can allow traits to be mapped to loci much narrower than equivalent work in RIL populations (Chao et al., 2012).

We were interested in the possibilities of a GWAS for leaf shape. The potentially greater resolution for mapping relative to our Bay-0 x Shahdara work was exciting, and we hoped it could allow us to relatively quickly identify a list of candidate genes. We also expected that any loci identified using a large panel of natural accessions would therefore be relatively common in the species, and so likely represent widespread loci controlling leaf shape in *Arabidopsis*. We obtained a collection of natural accessions to grow and phenotype for our GWAS. We chose to use the HapMap core population of 360 accessions, as used in (Baxter et al., 2010) and (Chao et al., 2012) and genotyped at 213,497 SNPs. We opted for this collection as we knew it had been used successfully to map traits by other groups, had readily available genotype data and contained accessions over a broad geographic range.

We used the online association analysis tool GWAPP (Seren et al., 2012) for our analysis. This was a convenient tool for testing for associations in our GWAS, as the calculations would be carried out remotely on hardware more powerful than the available desktop PCs. As such testing large amounts of trait data was relatively quick, particularly useful for our work given the number of combinations of data subsets and traits we were interested in. Within GWAPP there are two main approaches to the association analysis, the non parametric Wilcoxon rank sum test (KW) and the Accelerated Mixed Model (AMM) method. AMM is based on a variation of the Efficient Mixed-Model Association (EMMA) method (Kang et al., 2008, 2010) which uses a kinship matrix to control for relatedness between the accessions. The differences between each method are important to interpreting the results from a GWAS. The KW approach includes no correction for population structure and so p values for the association of each SNP are inflated, and falsely associated loci may occur

through linkage with true positives. The AMM method may overcorrect true associations, particularly if phenotypes vary geographically similarly to genetic relatedness across the accession used. For these reasons we considered associations detected by both methods as most credible, but also reported loci exclusive to either method, given the advantages of each, as in (Filiault and Maloof, 2012). We used a cut off of $5.5 -\log(p)$ for classifying loci as significant, based on the Benjamini-Hochberg-Yekutieli correction for multiple testing used within GWAPP (Benjamini and Yekutieli, 2001; Seren et al., 2012). Although it is possible that variation at a single isolated SNP could be associated with a trait, it seems much more likely that a region of associated SNPs would be identified, due to linkage between physically nearby SNPs (Atwell et al., 2010). As such we used criteria of loci made up of a tower of associated SNPs that met our significance threshold to identify associations of interest.

We created datasets of leaf shape variation for the accessions used in our GWAS by harvesting each rosette leaf per plant, and scoring these leaves with our leaf shape traits. With these datasets we would be able to examine differences in the shape of individual leaves as well as calculate an average leaf shape score for each accession. By harvesting each leaf of the rosette, we would also be able to capture changes in shape across the leaf series of an individual plant, and compare these leaf series shape changes between lines. We would record the leaf number of each plant, a trait closely related to life history in *Arabidopsis*. Life history is commonly studied as flowering time in *Arabidopsis*, and so by scoring a related trait we would also be able use our leaf number trait as a point of comparison with others work.

To score the leaf shape differences in these datasets we would apply the lIpPCs, created with the 10 accession leaf library. As we would be scoring shape variation in leaves from more than 300 accessions with considerable variation in leaf shape, we would also use dataset specific PCs, created from a PCA on the leaves in these GWAS datasets. Although the lIpPCs had been effective in mapping genetic variation in leaf shape in the Bay-0 x Shahdara RILs, due to the potentially greater variation in leaf shape and leaf number in the HapMap population, it would be worthwhile to investigate leaf shape variation in this collection of accessions with a new PCA also. This way, if the major variations in leaf shape across the HapMap collection of accessions differed to those in the 10 accessions used to create the leaf library PCs, we would still be able to effectively quantify variation in leaf shape in the HapMap collection of accessions.

4.2 Results

4.2.1 Leaf shape varies considerably between natural accessions of *Arabidopsis*

Genome wide association can be a powerful approach to genetic mapping and we were interested in the possibility of a GWAS for leaf shape. To further explore the idea and see whether this was something we could incorporate into the project, we obtained small sets of *Arabidopsis* accessions; the Joy Bergelson set of 96 and the Detlef Weigel set of 80 (Cao et al., 2011; Nordborg et al., 2005). We grew these in the University of York greenhouses and saw there were large differences in leaf shape between the lines. That there were visually clear between line differences in leaf shape amongst the natural accessions suggested a strong genetic basis to these differences in leaf shape, and encouraged us towards a GWAS. Relative to the Bay-0 x Shahdara RILs, these accessions also varied more in flowering time and leaf number. Many accessions of *Arabidopsis* are winter annuals and can require a period of vernalisation before transitioning to flowering (Michaels and Amasino, 1999; Shindo et al., 2005). As such when grown in greenhouse conditions many of these accessions grew a large number of leaves before bolting, and some did not bolt even after two months growth. Although we had previously used bolting date as a harvest point in our work so far, we would require new harvest points in our GWAS work, so that these late flowering plants could be harvested and analysed for leaf shape. We decided to carry out a GWAS for leaf shape, and obtained a larger mapping population of accessions, the core collection of 360 accessions from the HapMap project (<http://www.naturalvariation.org/hapmap>).

4.2.2 Creating leaf shape datasets at new harvest points for GWA mapping

Our work using GWA to map loci for leaf shape is made up of three separate growth experiments. For each growth experiment we grew the HapMap accessions in the University of York greenhouses and harvested every rosette leaf from each plant. We used different growth experiments so that we could record leaf shape at different harvest points. Our first GWAS was GWAS60, where plants were harvested at bolting, or after 60 days, whichever came first. By harvesting the majority of plants at bolting we were able to create a dataset using a similar approach to our other work in the project, and collect leaf shape data for each leaf grown by late flowering plants. The second GWAS was GWAS10, where plants were harvested after growing more than 11 leaves longer than 5mm in length. This allowed

us to compare leaf shape between plants with relatively little variation in leaf number. The third is GWAS09, harvested at a similar point to GWAS10, for which we also collected hypocotyl and cotyledon data, to investigate whether changes in rosette leaf shape were correlated with other aspects of plant morphology. Our separate GWA growth experiments would allow us to create datasets at both late and early harvest points and make use of the advantages of each. We harvested every rosette leaf of each plant. This allowed us to compare leaf shape between the accessions across a variety of subsets of the data. We used comparisons of leaf shape at individual leaves and whole plant averages. We were also interested in the shape changes across the leaf series, and so used median node leaf comparisons and measures of heteroblasty to identify changes in shape at the across the leaf series too. The results of each growth experiment will be reported in turn.

4.2.3 Mapping leaf shape differences between accessions at bolting to create the GWAS60 dataset

In the first of our three growth experiments we harvested the HapMap accessions once they had bolted or after 60 days, whichever came first. Two plants were grown for each accession in May 2013. In total we harvested 468 plants for 306 lines, recording the shape of 8309 leaves. Any leaves that were severely damaged during plant growth were not harvested. All plants that did not flower were harvested after 60 days of growth. The early leaves of such long lived plants were often wilted and decayed, and so unsuitable for harvesting. The number of leaves collected for each node across the leaf series can be seen in Figure 4-1. Some of the 360 HapMap accessions were never received from the stock centre and during the growth experiment some plants were lost due to greenhouse pests or poor growth and germination. Though we were unable to phenotype each line of the HapMap collection, we were confident this would not limit our GWAS, given that the lines grown showed much variation in leaf shape, and successful GWASs have been carried out using less than 200 accessions (Atwell et al., 2010; Filiault and Maloof, 2012). We used LeafAnalyser to capture the shape and size of each rosette leaf to produce the leaf shape dataset GWAS60.

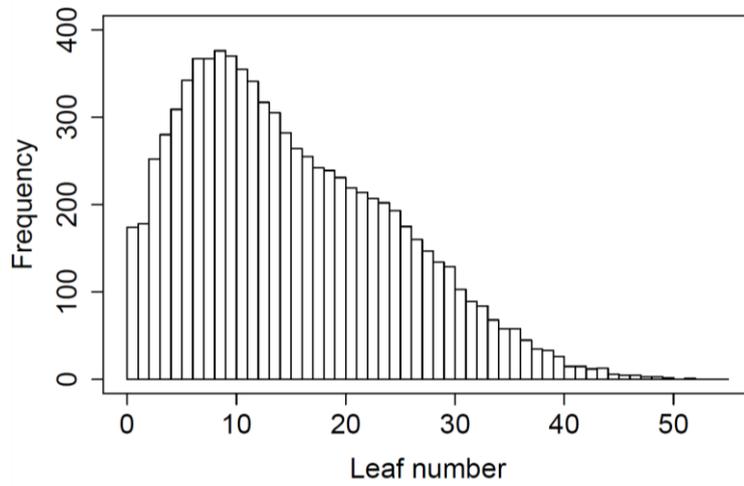


Figure 4-1 Leaves collected for GWAS60 dataset

Histogram showing the number of leaves collected for each node of the leaf series to create the GWAS60 dataset. The most common leaf in the dataset is the 9th leaf, with 375 leaves harvested and analysed. Relatively few leaves were harvested for the first four nodes, indicating many early leaves were lost from the plants harvested.

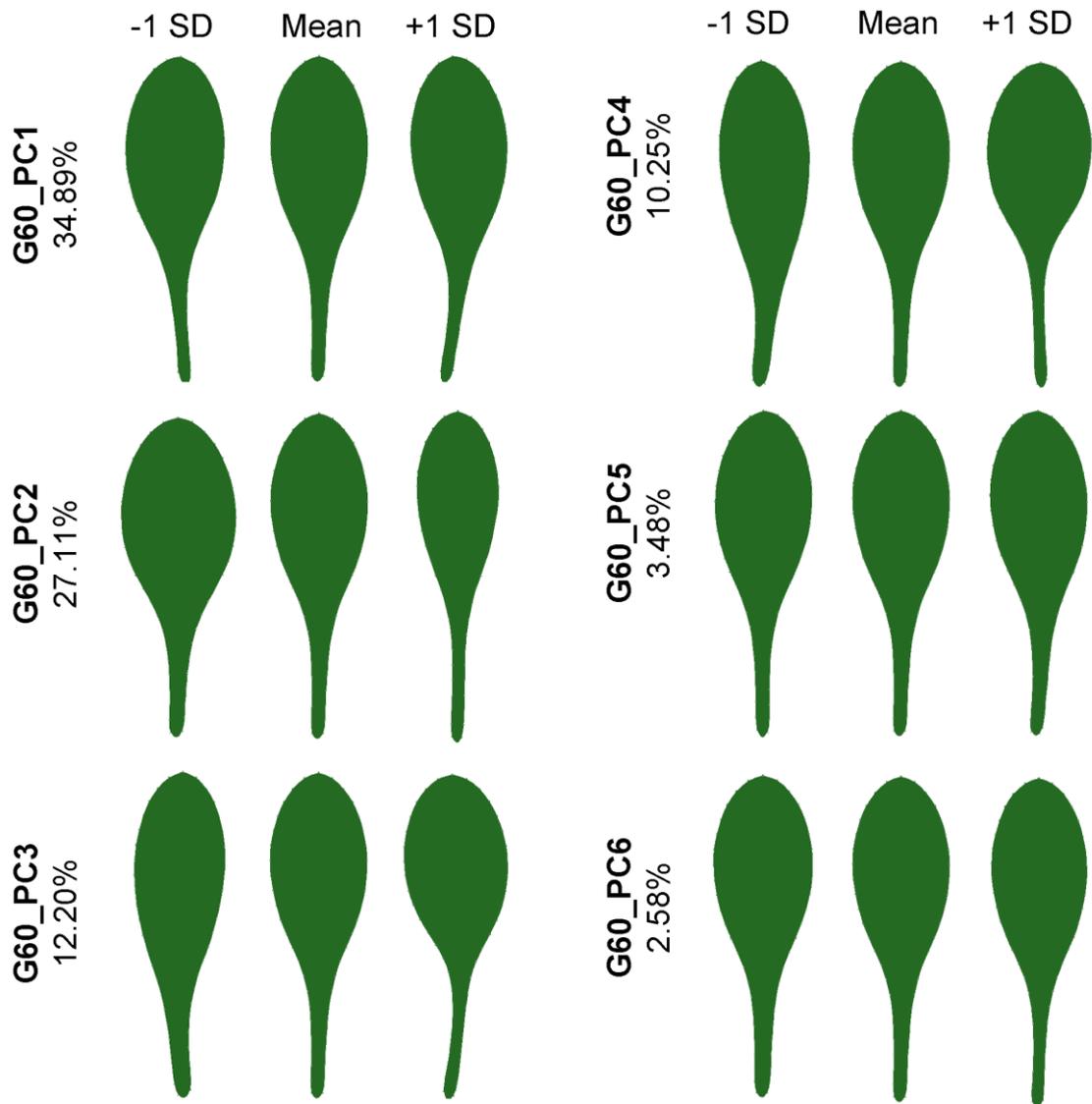


Figure 4-2 Leaf shape changes in GWAS60 leaf shape dataset

Figure shows the major leaf shape changes captured in a PCA on the GWAS60 leaves. The change in shape captured by each G60_PC is shown by as -1 and +1 Standard Deviation (SD) alongside the mean leaf. The first 6 PCs account for 90.51% of the variation in the dataset.

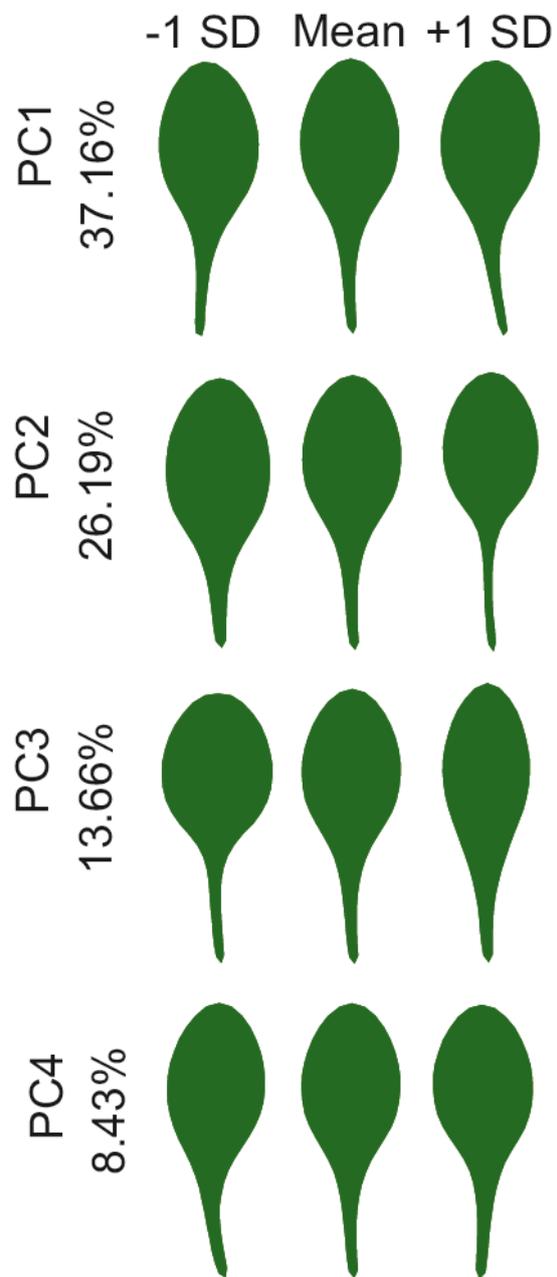


Figure 4-3 Leaf library PCs

The major leaf shape variations identified with a PCA on the leaf library dataset. Each PC is shown as -1 and +1 Standard Deviation alongside the mean leaf. The first 4 PCs account for 85.44% of the variation in the dataset.

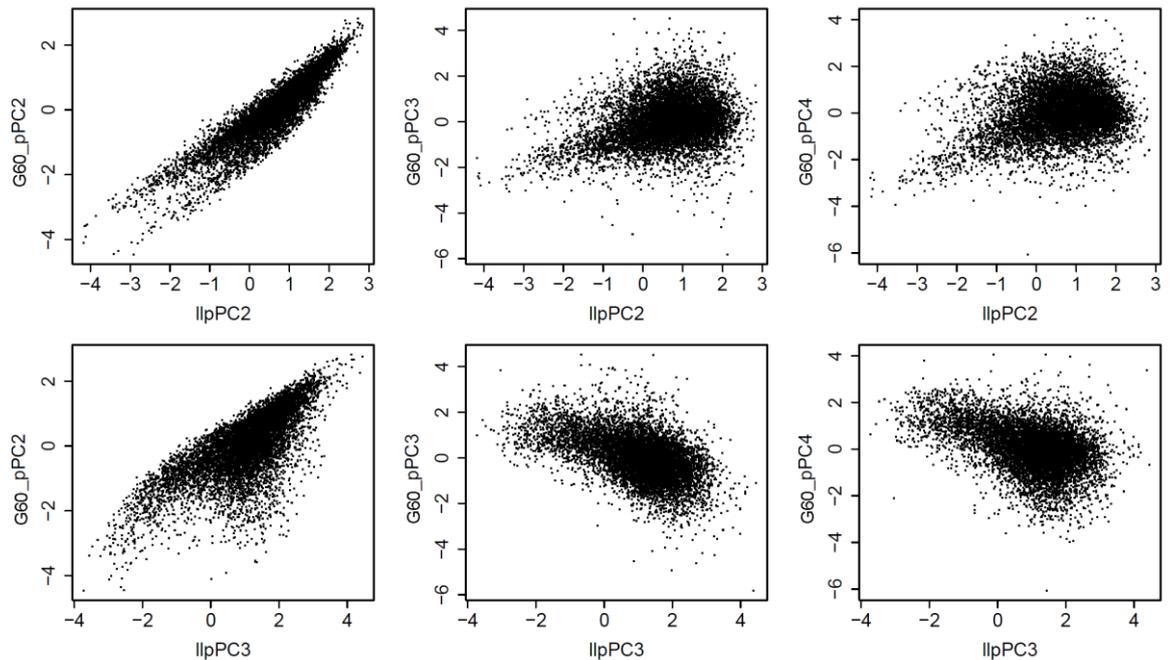


Figure 4-4 Correlations between leaf shape traits in the GWAS60 dataset

Scatterplots showing the correlations between the most heritable llpPC traits, llpPC2 and llpPC3, and the G60_pPCs within the GWAS60 dataset. Though there are some correlations between G60_pPCs and llpPCs, none of the G60_pPCs appear directly equivalent for the llpPCs.

4.2.4 Scoring shape variation in late flowering accessions

Once we had recorded the shape of each leaf harvested during the GWAS60 growth experiment, we scored each of these leaves for leaf shape. We used our llpPCs and GWAS60 dataset pPCs from a PCA on the procrustes fitted GWAS60 leaves. The shape changes captured in the GWAS60 pPCs (G60_pPCs) can be seen in Figure 4-2 Leaf shape changes in GWAS60 leaf shape dataset . By taking this approach we could score leaf shape with the same traits we used in the Bay-0 x Shahdara lines and throughout the project, but also account for the potentially different variations in leaf shape variation in the GWAS60 dataset.

To evaluate whether the major trends in leaf shape captured by the dataset specific G60_pPCs were distinct from those captured with the llpPCs, we looked at correlations within the leaves harvested for the GWAS60 dataset between the two sets of PCs. We were most interested in correlations with llpPC2 and llpPC3, shown in Figure 4-3 . These llpPCs describe changes in leaf shape such as ratio of leaf blade area to petiole length and represent heritable differences in leaf shape between genetically different lines. Although

there were correlations between the llpPC2 and llpPC3 and some of the G60_pPCs, see Figure 4-4 Correlations between leaf shape traits in the GWAS60 dataset , the trends did not show any of the PCs between the datasets to be directly equivalent, suggesting some differences in the major trends in leaf shape variation within the GWAS60 dataset captured by the G60_pPCs, and those described by the leaf library pPCs.

We were most interested in scores for the G60_pPCs defining differences in overall leaf shape such as G60_pPC2 and G60_pPC3, rather than those scoring curvature or petiole angle. The changes in leaf shape such as those described by G60_pPC2 and G60_pPC3 are typically highly heritable and we expected variation in these G60_pPCs were likely to have a genetic basis. Although we were unable to estimate heritability for these G60_pPCs in the absence of a large number of repeats per line, that these G60_pPCs showed most correlation with llpPC2 and llpPC3, the most heritable of the llpPCs in the Bay-0 x Shahdara population, supports this assumption.

4.2.5 Leaf number variation is associated with a commonly identified flowering time locus

There is typically strong correlation between leaf number and flowering time in *Arabidopsis* (Koornneef et al., 1991) and so we used our leaf number trait as a point of comparison with studies on natural variation of flowering time. After carrying out a GWAS with the traits we scored during the GWAS60 growth experiment, we identified a locus at the start of chromosome 4, between the start of the chromosome and 0.5Mb, associated with variation in our leaf number trait, see Figure 4-5. We had also found this region to be associated to our leaf number trait in the Bay-0 x Shahdara population, and it is commonly associated with flowering time traits in QTL populations and GWASs grown across different greenhouses and growth chambers (Atwell et al., 2010; Clarke et al., 1995). Polymorphisms within the gene *FRIGIDA* have been identified as responsible for this leaf number and flowering time variation associated with locus in work in the literature (Johanson et al., 2000; Shindo et al., 2005), making this gene the likely candidate for this locus in our GWAS. We also found a region of associated SNPs for variation in leaf number over the gene *Delay Of Germination 1* (*DOG1*), see Figure 4-6. Although characterised for its role in the natural variation of seed germination (Alonso-Blanco et al., 2003; Chiang et al., 2011), this gene has also been found associated with flowering time phenotypes in *Arabidopsis* GWAS (Atwell et al., 2010). That we identified these commonly reported flowering time loci for our leaf number trait

confirms that our growth conditions are within a similar range as others work, and that data reproduces some commonly reported loci.

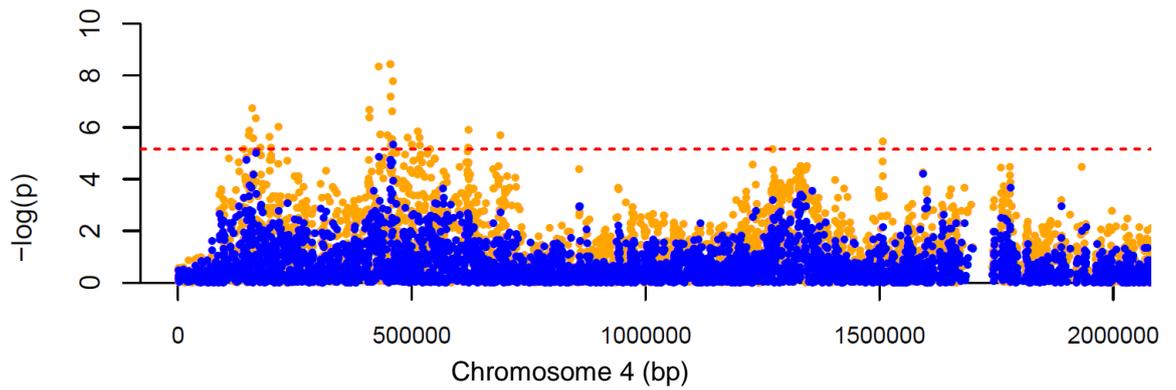


Figure 4-5 Locus associated with leaf number variation on chromosome 4

Locus at the beginning of chromosome 4 associated with leaf number variation in the GWAS60 dataset. Associated SNPs with the KW and AMM methods are shown as orange or blue dots respectively. Each dot represents a SNP marker, and the higher on the Y axis a SNP is, the more strongly associated the SNP is to variation in the trait. The position of each SNP on chromosome 4 is shown along the X axis. The gene *FRIGIDA* begins at 269025bp on chromosome 4.

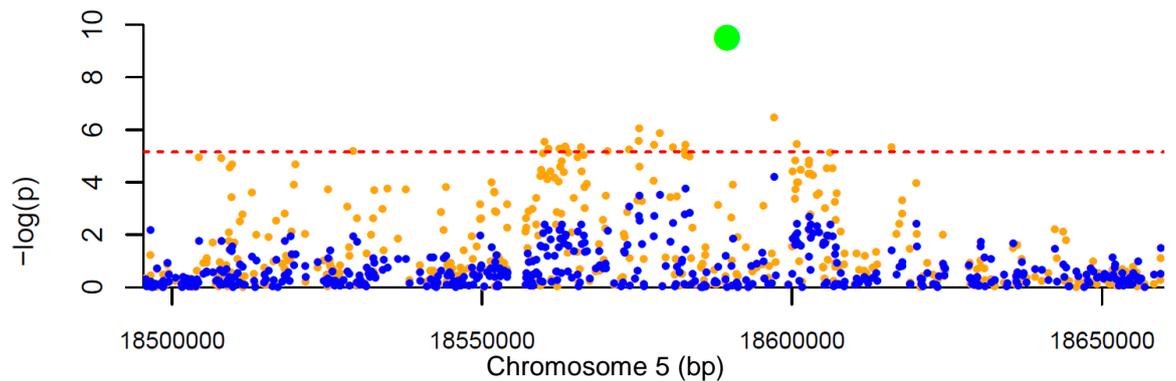


Figure 4-6 Locus associated with leaf number variation on chromosome 5

A locus associated with variation in leaf number in the middle of chromosome 5. Associated SNPs with the KW and AMM methods are shown as orange or blue dots respectively. Each dot represents a SNP marker, and the higher on the Y axis a SNP is, the more strongly associated the SNP is to variation in the trait. The position of each SNP on chromosome 5 is shown along the X axis. The gene *DOG1* is at 18,589,482bp and marked with a green circle.

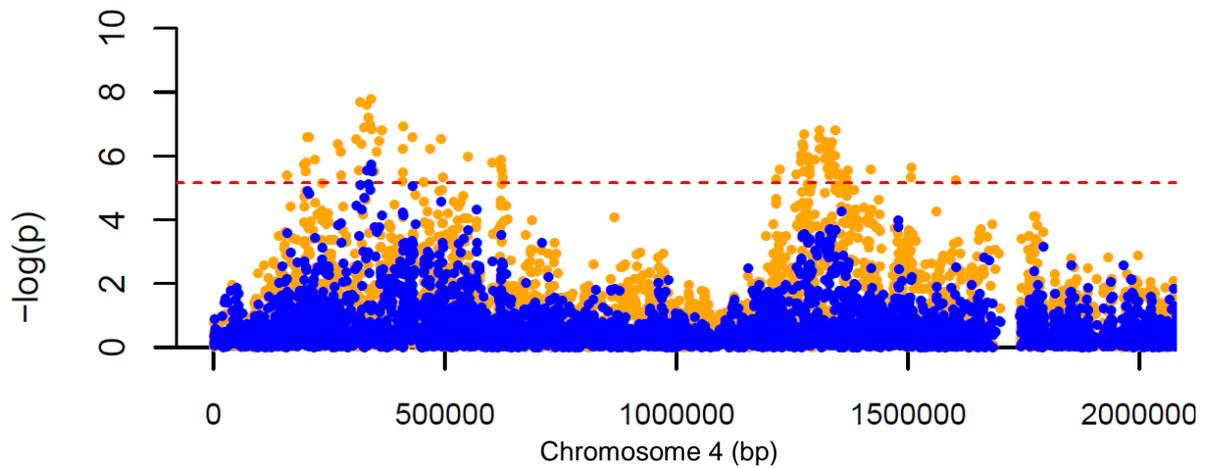


Figure 4-7 Locus associated with G60_pPC2 on chromosome 4

A locus associated with variation in trait G60_pPC2 at the start of chromosome 4. Associated SNPs with the KW and AMM methods are shown as orange or blue dots respectively. Each dot represents a SNP marker, and the higher on the Y axis a SNP is, the more strongly associated the SNP is to variation in the trait. The position of each SNP on chromosome 4 is shown along the X axis.

4.2.6 A locus associated with leaf number variation is also associated with differences in leaf shape

We scored our GWAS60 dataset with the llpPCs and G60_pPCs then calculated the average leaf shape score per line for each of these traits. Using these average trait scores for GWA mapping identified associated loci. The most strongly associated of these was within 0Mb to 0.5Mb on chromosome 4, associated with variation in llpPC2 and G60_pPC2, see Figure 4-7. We had previously found this region associated with variation in leaf number see Figure 4-5.

As leaf shape changes over the leaf series of a plant, and leaf number varies greatly in the GWAS60 dataset, we wondered if the association of this locus with some of our leaf shape traits was the result of some lines having more leaves at later nodes of the leaf series.

To examine the effect of this locus on leaf shape more closely, we grouped the accessions by their genotype at highly associated SNPs in this locus. We then plotted the leaf shape scores for G60_pPC2 across the leaf series for accessions in either group, see Figure 4-8 Difference in G60_pPC2 across the leaf series. The accessions with an allele associated with increased leaf number had a higher G60_pPC2 score at all leaves of the leaf series. This confirmed that this locus had an effect on G60_pPC2 as well as leaf number.

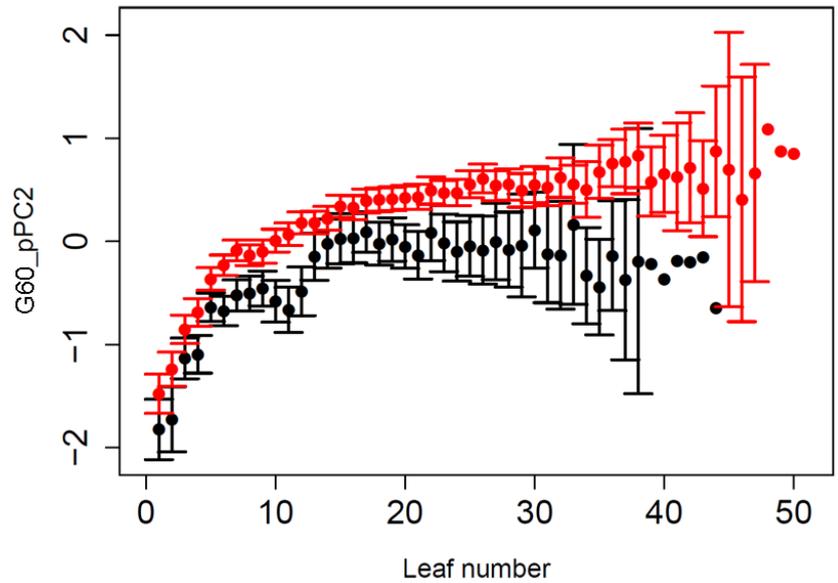


Figure 4-8 Difference in G60_pPC2 across the leaf series

This figure shows the difference in G60_pPC2 score across the leaf series between the accessions when grouped by genotype at a SNP on chromosome 4 at 340503bp within a locus associated with leaf number variation. Values for accessions with an allele at this SNP associated with greater leaf number are shown in red, those with an allele associated with lower leaf number are shown in black.

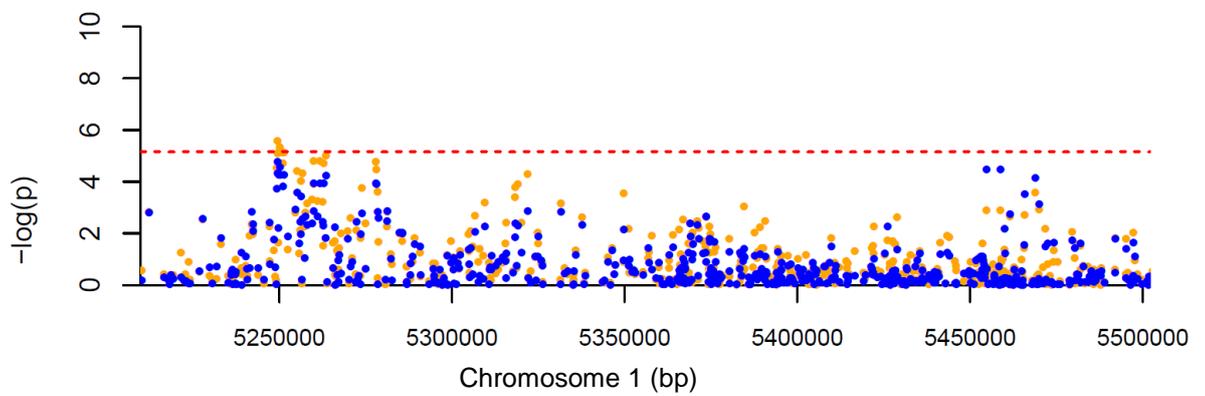


Figure 4-9 Locus associated with lIpPC2 variation on chromosome 1

A locus associated with variation in the trait lIpPC2 amongst median node GWAS60 leaves on chromosome 1. Associated SNPs with the KW and AMM methods are shown as orange or blue dots respectively. Each dot represents a SNP marker, and the higher on the Y axis a SNP is, the more strongly associated the SNP is to variation in the trait. The position of each SNP on chromosome 1 is shown along the X axis.

4.2.7 Variation in shape between accessions at individual nodes of the leaf series

As we had harvested every rosette leaf per plant in the GWAS60 growth experiment, we were able to make subsets of the leaf shape dataset we created to compare shape differences at individual nodes of the leaf series.

We used subsets of our GWAS60 leaf shape data of the third, sixth and ninth node leaves from each plant, and tested for loci associated with shape differences at each of these nodes. No loci were associated with shape variation in these subsets. It may be that low sample size restricted the effectiveness of these subsets, although whilst node three had relatively few leaves sampled, over 300 leaves were harvested at the sixth and ninth nodes. We next created a new subset of the GWAS60 dataset collecting the median node leaf from each plant in the total dataset, and compared leaf shape at this point in the leaf series between the accessions.

When comparing shape between median node leaves of accessions in the GWAS60 dataset, we found a locus on chromosome 1 associated with variation in the traits *llpPC2* and *G60_pPC2*, see Figure 4-9, and a locus on chromosome 5 associated with variation in *llpPC3*, see Figure 4-10. These loci were identified in both KW and AMM methods, see Table 4-1. A whole genome plot for the *llpPC2* median leaf trait can be seen in the appendices.

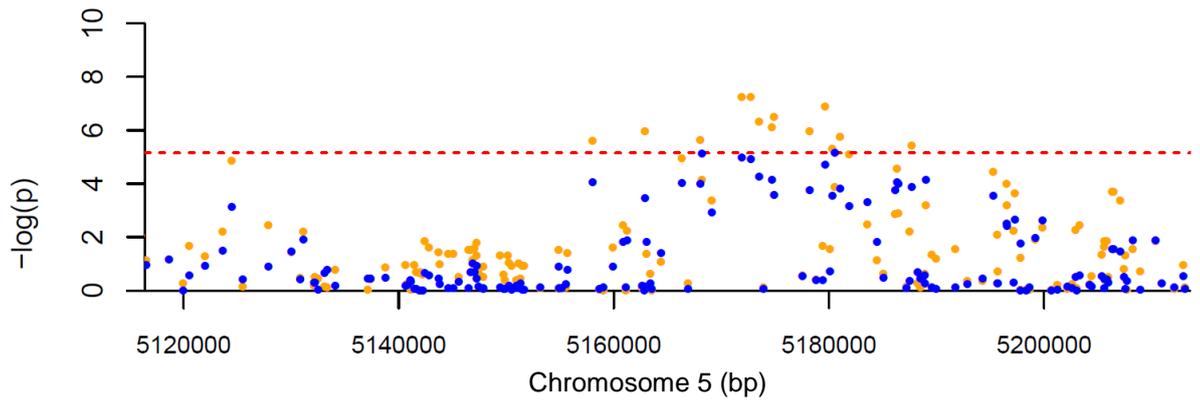


Figure 4-10 Locus associated with llpPC3 variation on chromosome 5

A locus associated with variation in the trait llpPC3 amongst median node GWAS60 leaves on chromosome 5. Associated SNPs with the KW and AMM methods are shown as orange or blue dots respectively. Each dot represents a SNP marker, and the higher on the Y axis a SNP is, the more strongly associated the SNP is to variation in the trait. The position of each SNP on chromosome 5 is shown along the X axis.

Trait	GrowthExperiment	Position	KW Peak $-\log(p)$	AMM Peak $-\log(p)$
N6PCA_PC2	GWAS10	Chr1:26Mb	7.35	5.5
N6PCA_PC5	GWAS10	Chr4:5Mb	7.08	6.21
hetPC4	GWAS10	Chr1:26Mb	5.8	5.6
hetPC8	GWAS10	Chr2:4Mb	6.02	5.6
hetPC8	GWAS10	Chr1:12Mb	8.6	5.12
mn_llpPC2	GWAS60	Chr1:5.24Mb	5.56	4.76
mn_llpPC3	GWAS60	Chr5:5.17Mb	7.24	5.15
avg_llpPC2	GWAS60	Chr4:3.31Mb	7	4.81

Table 4-1 Loci identified in leaf shape GWAS

Table shows the loci identified in both KW and AMM association tests across our leaf shape GWASs. 'mn' indicates traits in a median leaf subset and 'Avg' indicates a trait calculated by average across all leaves for a plant. For lists of the genes within each of these loci see appendices.

4.2.8 Mapping leaf shape variation in leaves harvested at bolting with the GWAS60 dataset

Using our GWAS60 trait data we had found a commonly reported locus for flowering time linked to our leaf number trait, suggesting we had identified similar variation for this trait as had other groups. We found this locus was also associated with a difference in leaf shape traits *llpPC2* and *G60_pPC2*. We also mapped loci responsible for shape variation at the median node leaf. We were unable to apply a heteroblasty measure to the GWAS60 leaf shape dataset to measure shape variation at multiple points of the leaf series, as data for the early leaves was missing for a number of lines due to the wilting of early leaves in late flowering plants. We were encouraged by our GWAS60 results, having identified several loci linked to changes leaf shape. We planned to grow the HapMap accessions again and harvest at an earlier time point to reduce the extent of leaf number variation in the plants harvested, and potentially map more loci associated with leaf shape variation.

4.2.9 A leaf shape GWAS with a new harvest point

In our second growth experiment, GWAS10, we used an earlier harvest point than for the GWAS60 dataset. Rather than using a harvest point after the majority of plants per accession had bolted, we harvested the lines for our GWAS10 dataset once plants had grown 11 leaves longer than 5mm in length, or had bolted, whichever came first. This minimised the loss of early leaf series leaves due to wilting and so resulted in a more complete dataset. We also expected that lower variation in leaf number would make the dataset more appropriate to use for comparisons of leaf shape across the leaf series.

We grew the core HapMap accessions in mid June 2013, sowing two plants for each line, though some of these were lost through low germination or during the growth of the plants. 652 plants were harvested for 345 lines and the shape of 6142 leaves was recorded. The total number of leaves harvested is lower than in GWAS60 as in this growth experiment the majority of plants were harvested before growing more than 12 leaves because of the chosen harvest point. In comparison to the GWAS60 dataset, leaf shape data was collected for a greater number of accessions, with a higher number of repeats and fewer leaves lost to wilting. This resulted in a more complete dataset, as can be seen in the distribution of leaves harvested across the leaf series, shown in Figure 4-11. At least 600 leaves were analysed for the first six nodes of the leaf series.

We scored the GWAS10 leaves with our lIpPCs. We also performed a PCA on the leaves to create PCs specific to the major shape variations in the GWAS10 dataset. We were interested in capturing the major trends in shape within the GWAS10 dataset, as we had not previously scored leaf shape in plants using this harvest point. Creating a new set of PCs specific to the changes in leaf shape in GWAS10 dataset would allow us to score potentially different variations in leaf shape that may not be captured by the lIpPCs.

We were interested as to whether the GWAS10 pPCs (G10_pPCs) captured leaf shape changes distinct from those captured by the lIpPCs. To evaluate this we looked at correlations between values for the G10_pPCs and the lIpPCs in the GWAS10 dataset. We found that two of the G10_pPCs, G10_pPC1 and G10_pPC3 were strongly correlated with lIpPC2 and lIpPC3 respectively. This suggested that although the leaf library and GWAS10 leaf shape datasets used for each PCA differed in harvest point and accessions harvested, the major shape changes captured by both PCAs were similar. When testing for association of loci to variation in both sets of PCs, we found no loci were unique to either set of PC traits. As such for convenience we report the associated loci using the lIpPCs, as these are used more widely throughout the project.

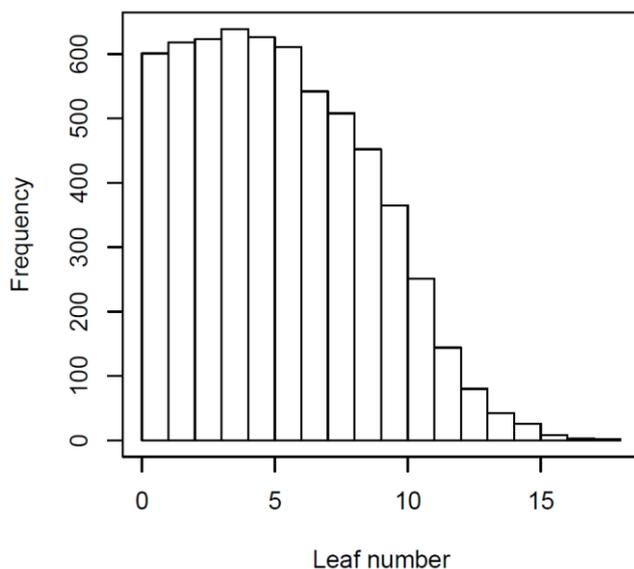


Figure 4-11 Number of leaves collected for GWAS10 dataset

Histogram showing the number of leaves collected per node of the leaf series in the GWAS10 dataset. There are over 600 leaves in the dataset for nodes one to six.

4.2.10 Capturing leaf shape variation early in the leaf series with the GWAS10 dataset

The earlier harvest point chosen for the GWAS10 growth experiment meant that a larger number of leaves were sampled early in the leaf series relative to the GWAS60 work, see Figure 4-1 and Figure 4-11. We were interested in whether shape variation at the first two leaves of the leaf series could be mapped. We tested for associations with our llpPCs, however we found no associated loci. As the first two leaves are relatively small and distinctly shaped in comparison to the majority of the leaf series, we also used a new PCA on a subset of the first two leaves in the GWAS10 dataset, reasoning that dataset specific PCs may more accurately capture the variation at these nodes. However, variation in these PCs specific to this subset of the data were not associated with any loci either.

4.2.11 NBLRR genes underlie some loci associated with node specific leaf shape changes

We next compared leaf shape in leaves across the leaf series. We tested for loci associated with shape differences between leaves at the 4th, 5th, 6th, and 7th nodes of the leaf series in turn using llpPC scores for these leaves. There were between 542 and 639 leaves harvested and scored at each of these nodes.

For node six leaves we identified several loci associated with variation in llpPC2. Using the KW test of association, there was a locus on chromosome 1 at 27.4Mb with a tower of SNPs above a group of genes within the Nucleotide Binding Leucine Rich Repeat (NBLRR) family. This was one of the most strongly associated loci we found within our GWAS work with a –log(p) score of 7.9 using the KW method, see Figure 4-12. A whole genome plot for this trait can be seen in the appendices. A region at the start of chromosome 4, very similar to that for leaf number and shape in GWAS60, was also associated with this trait. Interestingly two other clusters of NBLRR genes, on chromosome 1 at 21.7Mb and chromosome 5 at 14.5Mb, were within loci associated for llpPC2 variation in node six leaves, see Table 4-2. These loci were associated with this trait with only the KW method.

Combinations of these loci were also identified when testing for associations with llpPC2 variation for leaves four, five and seven. The NBLRR loci on chromosome 1 at 27.4Mb and 21.7Mb were associated with our llpPC2 trait in subsets for the 4th and 7th leaves, and llpPC2 score for the 5th leaf was associated with the 14.5Mb locus on chromosome 5, see Table 4-2.

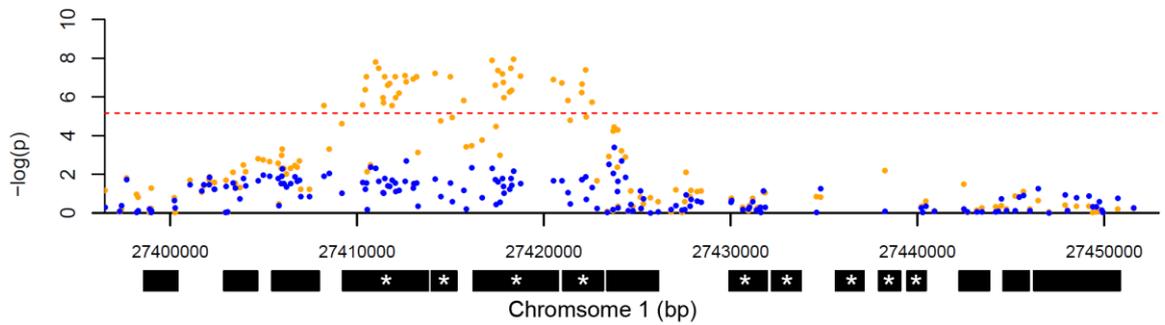


Figure 4-12 Locus associated with *lIpPC2* variation on chromosome 1

A locus associated with variation in the trait *lIpPC2* amongst node 6 GWAS10 leaves on chromosome 1. Associated SNPs with the KW and AMM methods are shown as orange or blue dots respectively. Each dot represents a SNP marker, and the higher on the Y axis a SNP is, the more strongly associated the SNP is to variation in the trait. The position of each SNP on chromosome 1 is shown along the X axis. Beneath the SNPs the gene models present in the TAIR10 genome annotation are shown as black boxes. NBLRR genes are marked with a white asterisk.

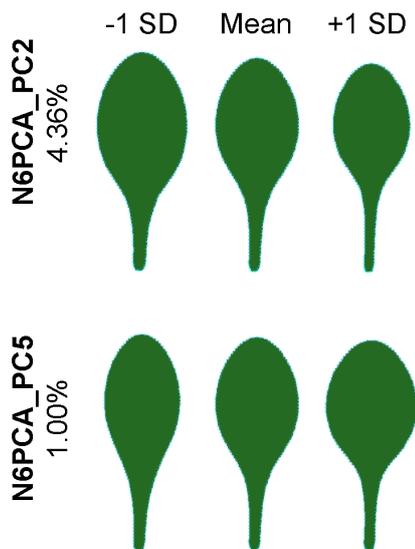


Figure 4-13 Node 6 specific leaf shape variation

Figure showing the node 6 specific PCs (N6PCA_PC_s). These were created using a PCA on node 6 leaves of the GWAS10 dataset. Using these subset specific PCs identified loci associated with variation in these traits that were not associated with the leaf shape changes described by the *lIpPCs*.

Trait	GrowthExperiment	Position	-log(p) (KW)
N6_IIpPC2	GWAS10	Chr1:21Mb	6.04
N4_IIpPC2	GWAS10	Chr1:21Mb	5.76
N7_IIpPC2	GWAS10	Chr1:21Mb	6.96
hetPC1	GWAS10	Chr1:21Mb	7.31
N6_IIpPC2	GWAS10	Chr1:27Mb	7.93
N4_IIpPC2	GWAS10	Chr1:27Mb	6.15
N6_IIpPC3	GWAS09	Chr1:27Mb	6.21
hetPC8	GWAS09	Chr1:27Mb	6.47
G09_L2_pPC4	GWAS09	Chr1:27Mb	5.51
N7_IIpPC2	GWAS10	Chr1:27Mb	6.15
hetPC1	GWAS10	Chr1:27Mb	6.23
N6_IIpPC2	GWAS10	Chr5:14Mb	5.42
N5_IIpPC2	GWAS10	Chr5:14Mb	5.86
hetPC8	GWAS09	Chr5:14Mb	5.59
hetPC1	GWAS10	Chr5:14Mb	6.1
hetPC1	GWAS10	Chr5:18Mb	6.24
hetPC8	GWAS09	Chr5:18Mb	6.58
hetPC7	GWAS09	Chr1:27Mb	5.69
hetPC7	GWAS09	Chr5:18Mb	6.84

Table 4-2 Loci associated with leaf shape traits in GWAS10 and GWAS09

Table showing the loci associated with leaf shape traits in GWAS10 and GWAS09 growth experiments that contain NBLRR genes. Traits scored within subsets of nodes are indicated by prefix 'N4' for node 4, 'N5' for node 5, 'N6' for node 6 and 'N7' for node 7. Traits scored within a subset of the first two leaves are indicated with prefix 'L2'. -log(p) KW shows the association score of the most associated SNPs within each locus. For a list of the genes within each of these loci see appendices.

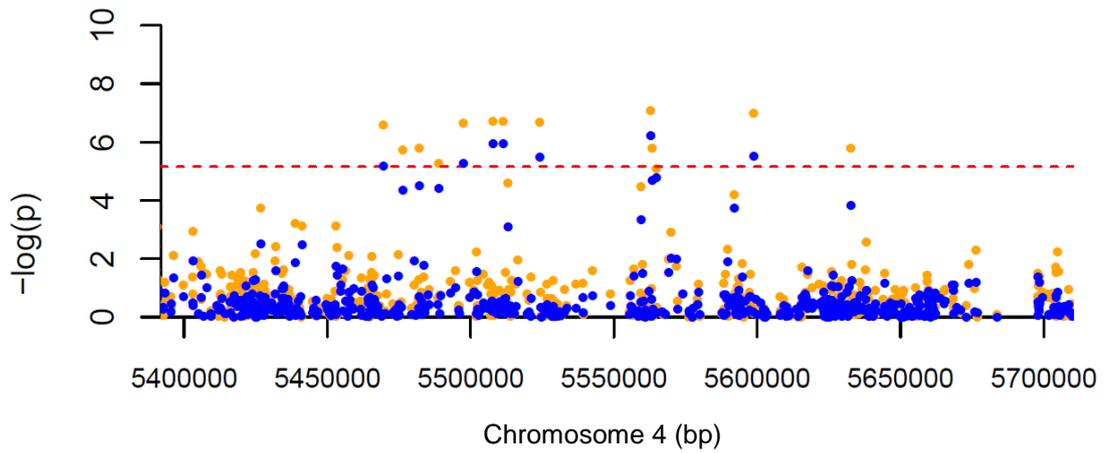


Figure 4-14 Locus associated with variation in node 6 shape variation on chromosome 4

A locus associated with variation in the trait N6PCA_PC5 amongst node 6 GWAS10 leaves on chromosome 4. Associated SNPs with the KW and AMM methods are shown as orange or blue dots respectively. Each dot represents a SNP marker, and the higher on the Y axis a SNP is, the more strongly associated the SNP is to variation in the trait. The position of each SNP on chromosome 4 is shown along the X axis.

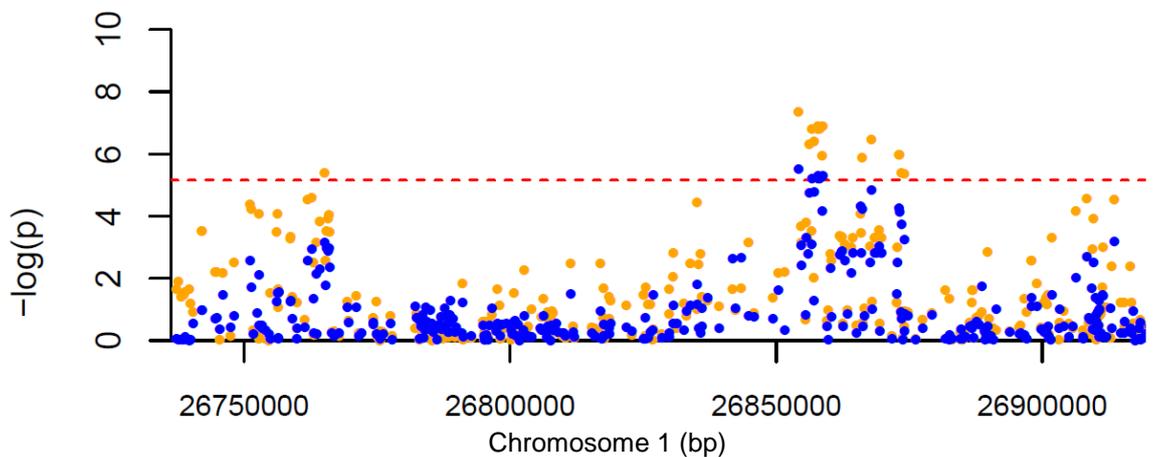


Figure 4-15 Locus associated with node 6 shape variation on chromosome 1

A locus associated with variation in the trait N6PCA_PC2 amongst node 6 GWAS10 leaves on chromosome 1. Associated SNPs with the KW and AMM methods are shown as orange or blue dots respectively. Each dot represents a SNP marker, and the higher on the Y axis a SNP is, the more strongly associated the SNP is to variation in the trait. The position of each SNP on chromosome 1 is shown along the X axis.

4.2.12 Further investigation of shape variation at node six leaves

We had identified several loci associated with the *llpPC2* variation in node six leaves. As the *llpPCs* were created with a PCA on a dataset of leaves across the leaf series, we wondered if a PCA on node 6 leaves would capture more accurately variations in shape at this node, potentially enabling us to map more loci for shape variation at this node. We ran a PCA on the node six subset of the GWAS10 leaves and captured shape variation with a set of node six specific PCs (*N6PCA_PCs*).

Using these *N6PCA_PCs* we identified the two *NBLRR* gene containing chromosome 1 loci and the locus at the start of chromosome 4 that we had found associated with the *llpPC2* trait for node six leaves previously. Interestingly we found new loci associated with traits *N6PCA_PC2* and *N6PCA_PC5*. Variation in *N6PCA_PC2* was associated with a locus at 26.8Mb on chromosome 1, with a $-\log(p)$ score above 5.5 in both KW and AMM approaches, see Figure 4-15. The *N6PCA_PC5* trait was linked to a locus on chromosome 4 at 5.5Mb with a $-\log(p)$ score above 6.1 for both KW and AMM, see Figure 4-14.

4.2.13 Investigating changes in shape across the leaf series with a heteroblasty PCA

An alternative approach to comparing shape between accessions at individual leaves, or using the average score for all leaves grown per accession, is to use several points across the leaf series at once for a measurement of shape. This can be done through a heteroblasty PCA, where each plant is represented by three leaves from set points in the leaf series. The shape of rosette leaves change across the leaf series of individual plants. Using a heteroblasty approach allows us to capture changes in shape common across the leaf series, but also to identify the differences in the leaf series shape changes between plants.

We used the third leaf, median node leaf, and leaf three quarters through the leaf series to represent each plant. As these three leaves are at set points in the leaf series, they provide an indication of shape across the leaf series for each plant. We captured the major variations in shape and size across these three leaf shape models with a PCA. The changes in shape and size captured by these heteroblasty PCs (*hetPCs*) can be seen in Figure 4-16. As we were interested to see how size varied over the leaf series we did not Procrustes fit the leaves we used for our heteroblasty PCA. Some PCs captured changes common to each of the three leaves, for example *hetPC4* captures changes in petiole length and leaf blade size

ratio that affect the three leaves in a similar way. Other hetPCs capture relative differences in size and shape between the three leaves in the shape model, for example in hetPC2 captures variation in relative size amongst the leaves, a high hetPC2 score corresponds to a larger last leaf, and a lower HetPC2 score is associated with a larger median leaf, see Figure 4-16.

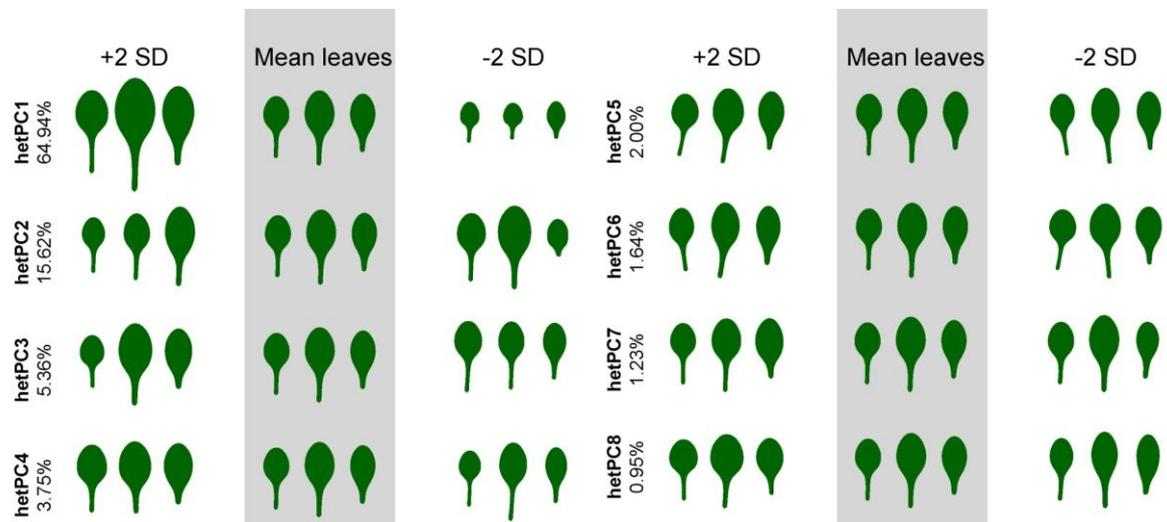


Figure 4-16 Heteroblasty PCs

Figure shows the major leaf shape changes from the GWAS10 heteroblasty PCA. The change in shape captured by each PC is shown as -2 and +2 standard deviation alongside the mean leaves. The first 8 PCs account for 95.49% of the variation in the three leaf heteroblasty dataset.

Though our method to create the three leaf shape models in our GWAS10 work differed slightly from our approach in the Bay-0 x Shahdara population, similar changes in leaf shape and size were captured with both approaches. Rather than use the last leaf of each plant for the third leaf of the shape model, as we did when working with the Bay-0 x Shahdara population, we used the leaf three quarters of the way through the leaf series. This was a result of the different harvest point between the Bay-0 x Shahdara population and the GWAS10 growth experiment. It was not possible to harvest each of the 652 GWAS10 plants at the exact point the 11th leaf was longer than 5mm, and so variability was increased in the size of the last leaf of each plant. This extra non genetic variability deterred us from using the last leaf and so instead we opted to use the leaf three quarters of the way through the leaf series. This would likely minimize the effect of variation in harvest point whilst allowing us to still record the leaf shape of plants near the end of the leaf series.

We found several loci associated for variation in these hetPCs. Although when using the KW method with trait hetPC1 we found a large number of isolated SNPs with a high $-\log(p)$ values suggestive of false positives, there were also towers of associated SNPs. Several of these loci associated with variation in HetPC1 were above NBLRR genes, including loci previously identified on chromosome 1 at 27.4Mb and 21.7Mb, and on chromosome 5 at 14.5Mb. There was a new NBLRR cluster picked out on chromosome 5 at 18.3Mb with a $-\log(p)$ value of 6.2, identified using KW, see Figure 4-17. For variation in Het_PC4 we identified a locus on chromosome 1 at around 26.8Mb, this had a $-\log(p)$ score of 5.8 using the KW method and 5.6 with the AMM method. This region was identified previously, associated with variation in the N6PCA_PC2 trait, see Table 4-1.

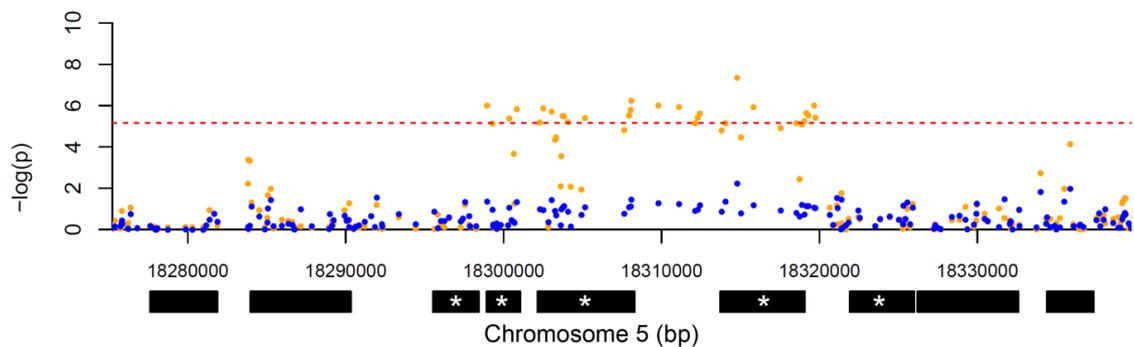


Figure 4-17 A chromosome 5 locus associated with HetPC1 variation

A locus associated with variation in the trait HetPC1 in GWAS10 leaves on chromosome 5. Associated SNPs with the KW and AMM methods are shown as orange or blue dots respectively. Each dot represents a SNP marker, and the higher on the Y axis a SNP is, the more strongly associated the SNP is to variation in the trait. The position of each SNP on chromosome 1 is shown along the X axis. Beneath the SNPs the gene models present in the TAIR10 genome annotation are shown as black boxes. NBLRR genes are marked with a white asterisk.

For variation in the het_PC8 trait, a locus on chromosome 2 around 4.1Mb was identified using both KW and AMM methods with a score of roughly $-\log(p)$ 7, see Table 4-1. There was also another locus associated with this trait, at 12Mb on chromosome 1 with a KW $-\log(p)$ value of 8.6, and 5.1 for the AMM method. This locus had a NBLRR gene, *Activated Disease Resistance 1 (ADR1)* within 30Kb from the peak SNP. We did not include this locus within our list of associated NBLRR loci throughout our GWA work because of the distance between ADR1 and the nearest significantly associated SNP. Although within seven genes of the most associated SNP, *ADR1* may be outside linkage with the nearby associated SNPs, as

linkage disequilibrium is estimated to reduce to 50% within 5Kb in *Arabidopsis* (Gan et al., 2011). We still felt it worthwhile to acknowledge the proximity of *ADR1* however. *FRIGIDA* is often named as a likely candidate gene for the 0 – 0.5Mb chromosome 4 locus associated with flowering time traits in the literature (Atwell et al., 2010), though the landscape of associated SNPs over this region can be complex, without a clear peak centred on *FRIGIDA* itself (Atwell et al., 2010; Johanson et al., 2000; Shindo et al., 2005). In the chromosome 4 locus at this region associated with leaf number variation in the GWAS60 dataset, *FRIGIDA* is further than 5kb from any of the most associated SNPs. As such we mention the proximity of *ADR1* in the event this is a similarly complex association. We also tested for loci associated with variation in hetPC2, hetPC3, hetPC5, hetPC6 and hetPC7, however these had no significantly associated loci.

4.2.14 NBLRR loci are associated with several phenotypes in GWAS10

Using an earlier harvest point in our second growth experiment, GWAS10, allowed us to create a dataset of leaf shape with less variability in leaf number and less missing data at early nodes than in the GWAS60 dataset. This allowed us to compare leaf shape at individual nodes across the leaf series and also to conduct a heteroblasty PCA. Although we did not find any loci associated with shape variation at the first two leaves of the leaf series, or using average leaf shape per accession, we did identify multiple loci associated with changes in our llPC traits across the 4th to 7th leaves, and also for our heteroblasty PCs. We found four of the loci associated with leaf shape variation in the GWAS10 dataset contained NBLRR genes, suggesting this family may be associated with shape differences between natural accessions of *Arabidopsis*.

4.2.15 A third growth experiment to further explore early leaf series shape variation

We had found two of five leaf shape QTLs detected in our work on the Bay-0 x Shahdara population coincided with QTLs in the literature for hypocotyl length. We were curious as to whether a similar correlation of hypocotyl length and leaf shape may exist amongst the accessions within the HapMap population, and whether some of the loci we had found associated with leaf shape traits in our GWA work so far were also associated with variation in hypocotyl length. In the GWAS10 work we had observed variation in cotyledon shape, and wondered if variation in cotyledon shape could be mapped to genetic loci, or if this correlated with the leaf shape of the plants. As both cotyledon shape and hypocotyl length

can be recorded early in a plants growth, finding correlations between these traits and leaf shape may allow us to use these traits as useful proxies or indicators of rosette leaf shape in other areas of the project.

We grew the HapMap accessions in late September 2013 to measure hypocotyl length and cotyledon shape in this population. We harvested the plants at the point the first two true leaves were a similar size as the cotyledons and measured hypocotyl length. After collecting our hypocotyl data we allowed some of the remaining seedlings to continue growing so that we could record the leaf shape in these lines at a similar stage to our GWAS10 work. This would allow us to test for correlations of hypocotyl length and cotyledon shape to leaf shape in the rosette leaves within a group of plants grown in the same growth experiment. It also provided an opportunity to repeat our GWAS10 work. We had identified a variety of loci associated with leaf shape variation within GWAS10, but we were limited in the number of replicates we could harvest concurrently. Although this third growth experiment would not be an exact replicate, it gave us the chance to test whether loci identified in the GWAS10 work would be associated with shape variation in a separate growth experiment.

We aimed to use the same harvest point as the GWAS10 work; harvesting plants once they had grown 11 leaves longer than 5mm or bolted. Unfortunately, due to ongoing work in other areas of the project we were unable to harvest at exactly this point. Instead we allowed the unbolted plants to grow as much as possible up to the GWAS10 harvest point and then harvested all unbolted plants within a week. We harvested 646 plants for 368 lines, collecting a total of 5031 leaves for analysis for in third growth experiment, GWAS09.

Due to the difference in harvest point between this dataset and the GWAS10 work, there is some difference in distribution of leaves harvested over the leaf series, though in both datasets around 600 leaves were harvested for each of the first six leaves of the leaf series, see Figure 4-18.

We recorded the leaf shape of every rosette leaf harvested per plant, and scored these leaves with the llpPCs. We also conducted a PCA on the GWAS09 leaves themselves, to identify the major shape variations within the GWAS09 dataset, creating the G09_pPCs. Similarly to our work with the GWAS10 dataset, we found that within the GWAS09 leaves scored for both sets of PCs, scores for the G09_pPCs were correlated with scores for llpPC2 and llpPC3. When testing for loci associated with variation in the G09_pPCs we found any loci identified were also associated with the leaf library pPCs, and so for convenience we report the loci using the leaf library llpPCs.

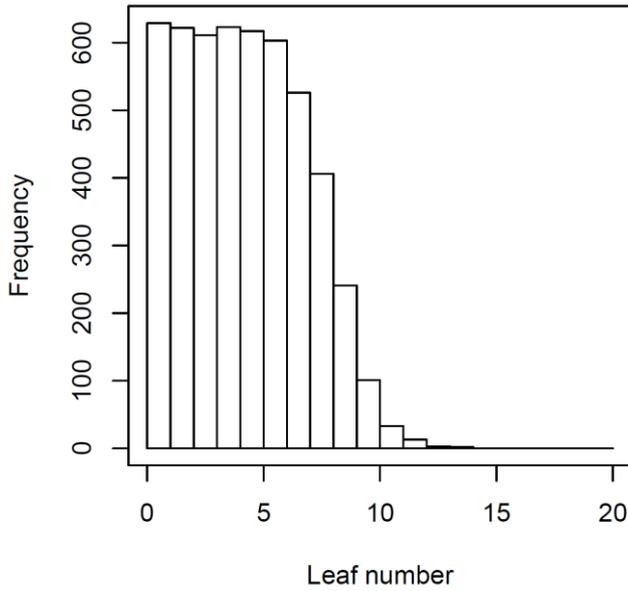


Figure 4-18 Leaves harvested for GWAS09 dataset

Histogram showing the number of leaves harvest at each node of the leaf series in the GWAS09 growth experiment. There are over 600 leaves sampled for the first 6 leaves of the leaf series in this dataset.

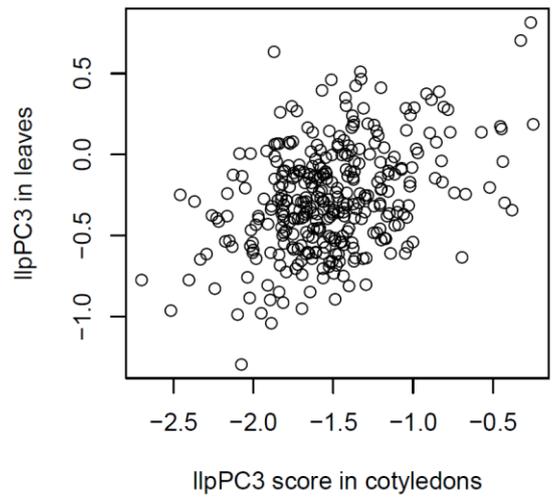
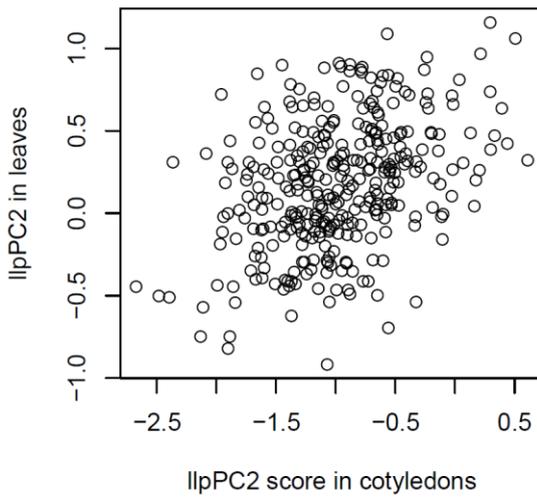


Figure 4-19 Correlation of shape traits between leaves and cotyledons

Correlation of average lIpPC2 and lIpPC3 score between the cotyledons and rosette leaves of each plant in the GWAS09 dataset. Linear models for the correlations between lIpPC2 and lIpPC3 score in leaves and cotyledons both an r^2 value of 0.15.

4.2.16 Variation in cotyledon shape and hypocotyl length in the natural accessions

We harvested the cotyledons of each plant and scored them for our llpPCs. As the cotyledons are distinct in size and shape from the rosette leaves, we also created a set of cotyledon specific PCs, using a PCA on the harvest cotyledons. When using either cotyledon dataset specific PCs, or the llpPCs to score shape in cotyledons in a GWAS, no loci were associated to either set of traits. We were interested in whether variation in cotyledon shape was correlated with variation in rosette leaf shape, and looked for trends between the llpPC2 and llpPC3 score of cotyledons and rosette leaves in the GWAS09 dataset, see Figure 4-19 . There was a positive correlation for both llpPC2 and llpPC3 between cotyledon and rosette leaves, with a Spearman's Rho of 0.36 for llpPC2, and 0.33 for llpPC3. That there was a significant correlation between cotyledon and rosette leaf shape is interesting, suggesting that although cotyledons are relatively distinct in size and shape from the rosette leaves, there are similarities between the shapes of these lateral organs. We tested for loci associated with variation in hypocotyl length, but found no significantly associated loci. We next looked at correlations between hypocotyl length and our llpPC shape traits within the GWAS09 dataset, however we found no trends between the traits in the GWAS09 leaves. This was in contrast to our work in the Bay-0 x Shahdara population, where we found the llpPC2 trait to be positively correlated with hypocotyl length.

That we did not map any loci for our cotyledon and hypocotyl phenotypes does not suggest there is no genetic basis for these traits. Given the successful work on mapping loci for hypocotyl length in *Arabidopsis* (Coluccio et al., 2010; Filiault and Maloof, 2012; Jiménez-Gómez et al., 2010) it is more likely that our methods resulted in too much non genetic variance in the data collected. It is not possible to estimate heritability of variation in cotyledon shape in our dataset due to the low number of replicates, however that cotyledon and rosette leaf shape are correlated suggests there may be a genetic component to this trait, given we had found previously that a large proportion of variation in rosette leaf shape is heritable.

As hypocotyls and cotyledons develop and grow when the plant is relatively small, it may be that finescale variation in environmental conditions can have a larger effect on variation in these traits, resulting in a lower proportion of genetic variation for these traits in the GWAS09 dataset. The heterogeneous nature of the soil, or variation in watering and temperature, are factors that may be averaged out over a plants life or once a root system

becomes more developed, minimising the effect on rosette leaf traits due to these factors. For hypocotyl and cotyledon traits, these factors may have resulted in higher non genetic variability in our scores for these traits.

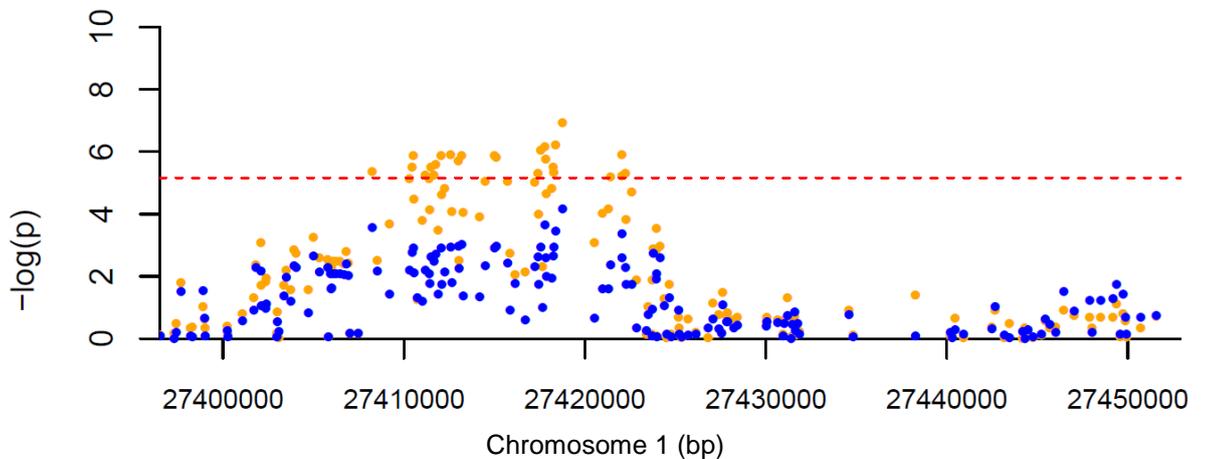


Figure 4-20 Locus associated with *lIpPC3* variation in node 6 leaves

A locus associated with variation in the trait *lIpPC3* amongst node 6 GWAS09 leaves on chromosome 1. Associated SNPs with the KW and AMM methods are shown as orange or blue dots respectively. Each dot represents a SNP marker, and the higher on the Y axis a SNP is, the more strongly associated the SNP is to variation in the trait. The position of each SNP on chromosome 1 is shown along the X axis.

4.2.17 NBLRR loci are associated with leaf shape traits in a separate growth experiment

In GWAS10 we found NBLRR clusters within loci associated with variation in the *lIpPC2* trait when comparing leaf shape between the accessions in the 4th to 7th leaves of the leaf series. To investigate whether similar loci would also be associated to shape variation in our third growth experiment, we tested for associations with the *lIpPCs* using the same node subsets of data with the GWAS09 leaves.

When comparing shape across at nodes of the leaf series in the GWAS09 dataset, we identified a chromosome 1 locus at 27.4Mb associated with variation in *lIpPC3* for node six leaves, using the KW method, see Figure 4-20. This locus contained a cluster of NBLRR genes and was associated with variation in *lIpPC2* for node six leaves within the GWAS10 growth experiment, see Table 4-2.

We compared leaf shape across the first two leaves of the leaf series using a subset of the GWAS09 leaves. We used the *lIpPCs* to score shape variation for these leaves and also

created new PCs for shape variation in this subset using a PCA on the first two leaves of the GWAS09 data, see Figure 4-21, as the first two leaves of the leaf series are distinct in shape and size from the majority of the leaf series. We found that variation in one of these L2_pPCs was associated with a locus at 27.4Mb on chromosome 1, containing a cluster of NBLRR genes, see Table 4-2.

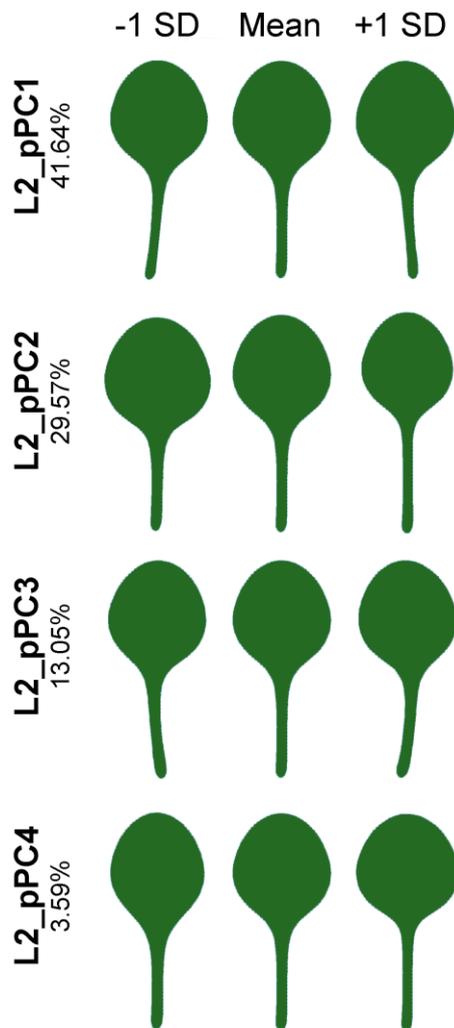


Figure 4-21 Major shape variations in leaves at nodes 1 and 2

The major PCs created with a PCA on the first two leaves of GWAS09 dataset, describing shape variation specific to the first two leaves of the leaf series. These capture 87.85% of the variation in shape amongst the first two leaves.

4.2.18 Loci associated with variation in heteroblasty traits in GWAS09 contain NBLRR clusters

We scored the leaves in the GWAS09 dataset using the heteroblasty PCs we had created previously to score variation with the GWAS10 leaves, see Figure 4-16. We tested for associations using these hetPCs and identified two loci associated with variation in the het_PC7 trait which contained clusters of NBLRR genes, one on chromosome 1 at 27.4Mb, and the other at 18.3Mb on chromosome 5, both associated using the KW method, see Table 4-2. We had previously found these loci associated with variation in the hetPCs with

the GWAS10 dataset, although interestingly these loci were associated with variation in hetPC1 in the GWAS10 dataset, as rather than het_PC7 as in the GWAS09 dataset.

4.2.19 Multiple loci associated with variation in leaf shape across two separate growth experiments contain clusters of NBLRR genes

Shape variation in the GWAS09 dataset was associated with several loci containing NBLRR genes that were also associated with leaf shape variation in the GWAS10 dataset. For a diagram of all NBLRR clusters in the *Arabidopsis thaliana* genome, with those within our leaf loci marked, see the appendices. Of the four loci containing NBLRR genes we identified associated with shape variation in the GWAS09 and GWAS10 experiments, only the locus on chromosome 1 at 21.4Mb, found in GWAS10, was exclusive to data from one growth experiment.

Interestingly the loci were not always associated with the same shape traits between GWAS09 and GWAS10. Only in GWAS09 was shape variation at the first two leaves associated with a locus at chromosome 1 at 27.4Mb containing an NBLRR gene cluster, and although shape variation at node six leaves in both growth experiments was associated with this chromosome 1 locus at 27.4Mb, in GWAS09 leaves this was for trait llpPC3, whereas in GWAS10 the association was for differences in llpPC2. We also found that although some of the same loci were associated with shape variation in hetPCs in both GWAS09 and GWAS10, the same loci were not necessarily associated with the same hetPC between growth experiments.

The differences in the shape traits these NBLRR containing loci were associated with may be the result of the difference in harvest point and growth conditions between the GWAS10 and GWAS09 datasets. We had limited time to harvest the GWAS09 plants before continuing with ongoing work in another area of the project. This resulted in a harvest point based on time, rather than the growth stage harvest point used in the GWAS10 growth experiment. We suspect this resulted in more non genetic variation in the GWAS09 dataset, as variation in growth due to environmental variability was perhaps able to impact the morphology of plants at harvest point more so than in the GWAS10 dataset. For example, plants with identical growth rates that germinated on different days would be harvested at different times in the GWAS10 dataset, once they had grown 11 leaves longer than 5mm. In the GWAS09 growth experiment, these plants would be harvested at the same time due to the time based harvest point.

It is interesting that variation in leaf shape of the first two leaves was associated with loci only in the GWAS09 dataset. One explanation for this is that the level of variation in growing conditions itself varied at different times between the GWAS09 and GWAS10 growth experiments. For example, greater variability of growth conditions early in the growth of GWAS10 plants could result in a greater proportion of non genetic variability in the leaf shape of early leaves in the GWAS10 dataset relative to the GWAS09 dataset.

4.2.20 GWAS for leaf shape in *Arabidopsis* identifies loci for a variety of traits

We had conducted three separate GWASs, using leaf shape data from three independent growth experiments. We were able to map loci associated with variation in a variety of traits, scoring leaf shape for accessions at the level of individual leaves, whole plant averages, and for heteroblastic changes in shape, and found loci associated with variation in traits at each of these levels. We had found the effect size of our loci to be small, and for multiple loci to be associated with many of the traits. This suggests a polygenic genetic architecture for leaf shape variation amongst the HapMap accessions, with many loci of small effect contributing to the natural leaf shape variation in this collection of accessions.

4.3 Conclusions

4.3.1 Practical considerations for the scale of phenotyping experiments

We had run several GWASs to map loci responsible for natural leaf shape variation in *Arabidopsis*, creating leaf shape datasets for three separate growth experiments and identifying multiple loci associated with our leaf shape traits. An alternative approach would have been to consolidate the effort in carrying out each growth experiment to one larger experiment with an increased number of repeats. In each of our three growth experiments we aimed to harvest two plants per accession. As such an alternative approach, using similar effort over one single growth experiment, would allow us to harvest six plants per accession. The advantage of this approach would be decreased non genetic variability for our leaf shape traits, and the ability to calculate within line variation to make an estimate of the heritability of our traits.

Although worthwhile to consider, the disadvantages of growing many plants in one large experiment meant this approach was less suitable to our project. Practically speaking, the effort of one worker over a period of time cannot be condensed to a shorter period without more personnel, and so our resources determined our approach to an extent. We were also unsure of how successful our GWA work would be. Though we had shown in work to identify QTLs in the Bay-0 x Shahdara RIL population that variation in our leaf shape traits could be mapped to genetic loci, it was possible that mapping leaf shape traits within a panel of natural accessions would not be as successful if leaf shape is determined by many loci of small effect. GWA mapping works best with traits with a strong genetic basis across few loci (Korte and Farlow, 2013), and so variation in such traits can be mapped to a small number of strongly associated loci (Baxter et al., 2010; Chao et al., 2012). In contrast, a trait controlled by many loci of small effect, or rare polymorphisms, may be much trickier to map (Korte and Farlow, 2013), as such we were wary of immediately committing a large number of resources to a single GWAS.

We did not wish to compromise on the number of accessions we used as a decrease in the number of unique genetic lines used would reduce our mapping resolution, and this was one of the key benefits to carrying out the GWAS during the project. We were also wary of growing batches of the accessions at different times, then collating the data obtained from each into a single dataset. Although we could include control lines between growth

experiments in an attempt to normalise data across seasonal differences, this would rely on an assumption that each line responded in the same way to environmental variables such as temperature and day length. It is also likely that batches grown at different times of the year would vary in leaf number and flowering time. Differences in leaf number between batches due to variation in greenhouse conditions would greatly hinder our investigation of heteroblastic changes in shape across the leaf series. As such we ruled out growing the accession collection in smaller batches throughout the year.

Our approach using multiple growth experiments for separate GWASs meant each experiment could be informed by the previous analysis. We were able to choose a new harvest point in response to our results from first GWAS growth experiment; GWAS60. With the GWAS60 dataset we examined leaf shape at two different time points and mapped loci for leaf shape unique to each dataset. We were encouraged to find variation in leaf number in this dataset was associated with two loci, containing the genes *FRIGIDA* and *DOG1*, that had been reported as associated with flowering time in a GWAS in the literature (Atwell et al., 2010). This suggested our approach was capable of reproducing association of candidate genes in commonly studied traits, despite the relatively low number of repeats per accession we had used. It is also worthwhile noting that although a low number of plants were harvested for each accession, as each rosette leaf per plant was harvested, we recorded multiple data points for each plant harvested. On average for the GWAS60, GWAS10 and GWAS09 growth experiment datasets, 17.8, 9.4 and 7.8 leaves respectively were harvested from every plant. This made a variety of leaf shape comparisons possible within the datasets we created.

We used a similar harvest point for the GWAS09 and GWAS10 growth experiments, and it was promising that for leaf shape traits in both these experiments, some of the associated loci were common to data from both growth experiments.

4.3.2 Scoring leaf shape variation effectively in large datasets

We scored leaf shape differences within the shape datasets of the GWAS growth experiments using PCA to identify major variations in leaf shape, capturing these variations in a set of PCs, and scoring the shape of individual leaves with these PCs. Leaves can be scored using PCs created from the dataset to which they belong, or for PCs created with a PCA on an external dataset, such as the ten accession leaf library used to create the leaf library PCs (llpPCs) (Matser, 2014). For each of the leaf shape datasets created with our three GWAS growth experiments, we scored leaf shape variation using PCs created from

each individual dataset respectively, and also the llpPCs. This allowed us to capture the major shape trends in this new collection of accessions, which may differ from those within the 10 accession dataset used to create the leaf library PCs.

Within each of our three GWAS growth experiments, we compared values for the dataset specific PCs and the llpPCs. Within the GWAS10 and GWAS09 datasets, scores for dataset specific PCs correlated strongly with the leaf library PCs. PCs from a PCA on the GWAS60 growth experiment dataset were less strongly correlated with the llpPCs. However, across our three GWASs, any loci associated with changes in dataset specific PCs were also associated with variation in the llpPCs. This suggested that the major shape variations in the leaves harvested for our GWAS growth experiments were similar to those captured by the leaf library PCs, despite the difference in the number of accessions used to create each set of PCs.

Throughout our GWAS work we investigated shape changes between accessions at individual nodes of the plant, or across set points of the leaf series. As the major shape variations in these subsets may differ from those in the total dataset, we used PCA on these subsets of leaves to create subset specific PCs. In some instances variation in these subset specific PCs was associated with loci not identified when using the llpPCs, for example for a subset of the first two leaves in the GWAS09 dataset, or the 6th leaf in the GWAS10 dataset. On the whole however, using subset specific PCs did not identify any new loci not also associated with the llpPCs, for example for the median node leaf subset of the GWAS60 dataset, or the node 4 to 7 subsets of the GWAS10 and GWAS09 datasets.

Our series of GWAS analyses is perhaps the first detailed assessment of natural leaf shape variation across a collection of over 300 *Arabidopsis* accessions. We analysed this dataset in different ways in order to best identify loci associated with the natural differences in leaf shape between the accessions. Using a combination of our leaf library PCs, and new PCs capturing variation in subsets of the growth experiment datasets, we mapped a variety of loci linked to natural leaf shape variation in this collection of accessions.

4.3.3 A locus associated with variation in leaf shape and number

A locus associated with leaf number variation in GWAS60 plants was also associated with variation in average llpPC2 score in this growth experiment. Leaf number is a trait closely related to bolting and flowering time in *Arabidopsis*. During the transition from vegetative to reproductive growth, plants cease growing rosette leaves, bolt, and produce an

inflorescence. Mutants in flowering time are also known to have a variety of other phenotypes, including leaf morphology (Pouteau et al., 2004). It is known that the *FRIGIDA* gene has an effect on photosynthetic rate and water use efficiency (Lovell et al., 2013). Variation in genes with flowering time phenotypes also affect leaf morphology in *Cardamine Hirsuta* and a tomato species (Cartolano et al., 2015; Shalit et al., 2009). It seemed possible to us therefore that the polymorphism underlying a locus controlling leaf number was also affecting leaf shape in these lines. The later flowering lines grew more leaves, which had a higher *llpPC2* score compared to the earlier flowering lines with fewer leaves. We noted whilst carrying out QTL mapping in the Bay-0 x Shahdara RIL population that *llpPC2* score increases in plants under shade conditions. It is possible the increase in *llpPC2* in later flowering plants is due to greater self shading within rosettes due to the presence of more leaves that overlap existing ones. Growing groups of late and early flowering plants alongside one another, harvesting them at a stage where both had similar leaf number, and again once both had bolted, would reveal if the later flowering plants had a higher *llpPC2* score at leaves throughout the plants growth or if this developed later, perhaps as a response to a denser rosette in later flowering plants.

4.3.4 Incorporating our GWAS results into the wider project

Further experimental work is required to identify a causative gene within a locus associated with trait variation in a genetic mapping experiment. Alternatively, genes with functions or structures likely to be related to the trait within the locus can be identified as candidate genes and subjects for further investigation. Within this work, several of the loci associated with changes in leaf shape contained clusters of genes in the NBLRR family. This suggested that genes from this family might be responsible for some of the differences in leaf shape amongst these accessions. In the wider context of the project, this was interesting because it may implicate this gene family in contributing to the natural differences of leaf shape in *Arabidopsis*. We chose to focus on the genes within this family at these loci in our further work. The alternative strategy of attempting to identify the causative polymorphism responsible for each association appeared less likely to be successful within the time available.

We considered the option of using a statistical measure to test whether there was an enrichment of NBLRR family members across all loci associated with our leaf shape traits. This would be relatively simple to implement using a variation of previous scripts we had created for a network analysis carried out in the Bay-0 x Shahdara RIL population. Each SNP

above a threshold of significance would be listed, and from this, a list of genes within 5kb either side of each SNP could be collected. This would provide a list of all genes within estimated linkage disequilibrium with every significantly associated SNP. This list could be tested for enrichment of NBLRR family members relative to the genome. A problem with this approach however is that the NBLRR genes often reside within clusters in the genome (Meyers et al., 2003). There are upto nine NBLRR genes within clusters underneath some of our GWAS loci, and so even if only one gene amongst the cluster contained associated SNPs and harboured the causative polymorphism, the other genes within the cluster would greatly enrich the results of the analysis, as they lie within linkage. If we attempted to counter this by considering each cluster or isolated gene as an equal 'instance' of NBLRR genes, then we would likely find a lack of NBLRR genes amongst all our peaks, because potentially causative NBLRR genes would be combined into one instance. As the NBLRR family has evidence of undergoing extensive duplication (Meyers et al., 2003), it is quite possible NBLRR genes within the same cluster have a similar function and so multiple polymorphisms within different NBLRR genes in a cluster could potentially be associated with variation in leaf shape. To identify a significant enrichment, an enrichment analysis of loci associated to a trait in a genetic mapping experiment has to overcome a low ratio of causative to non causative genes in loci (Filiault and Maloof, 2012). Assuming only one gene is responsible for each associated loci, the ratio of causative to non causative genes in the enrichment analysis depends on the size of the linkage disequilibrium window chosen and number of loci associated for the trait. We decided the outcome of an enrichment analysis would be too dependent on how we designed the analysis; the size of linkage window we chose and assumptions we made on whether a single NBLRR gene or several within a locus were causative, and so decided not to apply an enrichment analysis.

We also identified other loci associated to our leaf shape traits that did not contain NBLRR genes. For these non NBLRR loci we reported the loci associated in both KW and AMM methods, aiming to identify the strongest associations for potential further experimental work.

In contrast to some studies which identify a single clear peak for their phenotypes (Baxter et al., 2010; Chao et al., 2012), we had identified multiple peaks, making it difficult to choose any as a focal point for further investigation. If our GWAS had presented a small number of highly significant loci associated with a large phenotypic effect, we would have been able to focus the project on one or two of these major loci, creating F2 populations to find the gene responsible for the difference in phenotype. This requires a considerable amount of work,

scoring the shape of and genotyping hundreds to thousands of plants to narrow the causative region to the width of a single gene (Loudet et al., 2008). We were concerned the small to medium effect sizes of our loci would make this approach difficult, as identifying the effect of such a polymorphism in an F2 population may be trickier to resolve against the genetic background of the chosen parents. Small regions of the genome can have complex effects on a trait (Kroymann and Mitchell-Olds, 2005), and so even after fine mapping, identifying the responsible polymorphism within a region can be difficult. We were wary of committing such a large effort to one of our loci, given the relatively small effect of each.

One option for further investigating peaks of association is to use a candidate gene approach. When QTL mapping in RIL populations, a large number of genes generally fall within the confidence intervals of the QTL, and so compiling a list of candidate genes with known functions similar to the trait the QTL is associated with can be relatively ineffective due to the number of possible candidate genes. Within a GWAS however, associated loci are much narrower, in part due to the decay of linkage disequilibrium amongst natural accessions, estimated to be to 50% within 5kb (Gan et al., 2011). This allows for considerably smaller windows in which to collect candidate genes. We did not think this would be an appropriate approach for our results however. Although there are a number of genes implicated in leaf development in *Arabidopsis* that represent candidate genes (Berná et al., 1999; Dkhar and Pareek, 2014; Gonzalez and Inzé, 2015; Hepworth and Lenhard, 2014), it is very possible that genes not involved in leaf development were responsible for the natural variation in leaf shape we observed. The causative polymorphism for one of our leaf shape QTL identified in the Bay-0 x Shahdara RIL population was likely within such a gene, *TZP*, characterised for a role in morning specific growth (Loudet et al., 2008). Given the number of genes in which a polymorphism could potentially result in a difference in leaf shape, we felt creating a list of candidates for each loci would be unlikely to produce a particularly short list, and would be at risk of overlooking a potentially causative gene.

It is worthwhile to note that the loci associated to our leaf shape traits containing NBLRR genes in the GWASs were only identified using the KW approach. This left open the possibility of these loci being false positives potentially the result of population structure within *Arabidopsis*. There is evidence that leaf shape varies geographically in *Arabidopsis* accessions (Hopkins et al., 2008) and this has been suggested to result in overcorrection of loci for known natural variation in candidate genes in GWAS (Filiault and Maloof, 2012). Although we identified several loci with the both the KW and AMM methods, and so less at risk of being false positive loci, none of these had a large effect size. Though we could

perhaps have more confidence in these loci as they were significantly associated in both association tests, we were wary of committing to the resource intensive route of confirming these loci experimentally without a strong allelic effect. It was possible that pursuing these associations further would not result in the identification of the causative polymorphism, or likely candidate gene, within the time available. Although we created numerous F2 populations based around all the associated loci we identified as a starting point for further investigation, we decided to take the opportunity the NBLRR loci presented, and to focus on the possible role of members of this gene family identified in our GWAS results in leaf shape variation for the next stage of the project. This would allow us to answer specific questions about the role of NBLRR genes in leaf shape in *Arabidopsis*.

Chapter 5. Analysis of T-DNA insertion lines for NBLRR genes associated with natural leaf shape variation

5.1 Introduction

We identified potential associations between variation in our leaf shape traits and natural genetic variation at loci using natural accessions of *Arabidopsis* in our GWAS work. We found clusters of Nucleotide Binding Leucine Rich Repeat (NBLRR) genes within four loci associated with leaf shape traits in our GWAS work. We decided to investigate these genes further, and identify whether they might have a role in leaf shape variation in *Arabidopsis*. Estimates of the number of NBLRR genes in *Arabidopsis* range from 149 to 159 (Guo et al., 2011; Meyers et al., 2003). NBLRR family genes can be divided by the type of N terminus domain. TIR-NBLRR genes are defined by a N terminal TIR domain resembling *Drosophila* Toll and mammalian IL-1 receptors, and CC-NBLRR genes feature a coiled-coil (CC) domain (Guo et al., 2011). Of the 14 NBLRR genes within loci associated to leaf shape traits in our GWAS work, only one was a CC type NBLRR. The remaining 13 genes associated to leaf shape traits were not restricted to a distinct subgroup in the TIR-NBLRR clade, see appendices for a phylogenetic tree for this family.

We focused on NBLRR genes within loci associated to leaf shape traits in our GWAS and their possible role in determining leaf shape variation. These 14 NBLRR genes were located across four loci associated to leaf shape traits identified on chromosome 1 at 27Mb and 21Mb, and chromosome 5 at 14Mb and 18Mb. We would characterise mutant lines for these 14 genes to investigate possible gene function. We opted to use T-DNA insertion lines to test whether mutations in any of these genes resulted in changes in leaf shape. SALK T-DNA insertion lines are a convenient tool for investigating gene function and are available for a large number of genes in *Arabidopsis* (Alonso et al., 2003).

NBLRR gene phenotypes have been characterised in the literature using both natural genetic variation and mutant lines. Natural variation in disease resistance between the accessions Ws-0 and RLD was used to map the *RPS4* NBLRR gene (Gassmann et al., 1999; Hinsch and Staskawicz, 1996) and naturally occurring NBLRR loci have been found to create incompatible F1 plants between different *Arabidopsis* accessions (Alcázar et al., 2008; Bomblies et al., 2007). Mutations in NBLRR genes have also been studied for their disease

resistance phenotypes (Shirano et al., 2002; Sohn et al., 2014). Although we had first identified these NBLRR genes through work on natural genetic variation in a GWAS, we were confident we could continue to investigate a potential leaf shape function of these genes using mutant analysis.

Mutant analysis can be used in combination with studies of natural genetic variation through complementation. Kobayashi et al., (2008) use complementation to investigate the *HMA5* gene as a candidate for a root length QTL in a RIL mapping population created from *Cvi* and *Ler Arabidopsis* accessions. These authors find that in F1 plants hybrid between a T-DNA insertion line for gene *HMA5* with a recessive short root phenotype, and two RIL lines with variant alleles at a root length QTL, the RIL with a *Ler* QTL allele is able to rescue the short root phenotype of the *HMA5* T-DNA line, whereas the RIL with a *Cvi* allele cannot. Crossing the *Cvi* QTL allele RIL to Col-0, the T-DNA line background, rescued the short root phenotype, indicating *HMA5* as the gene responsible for this QTL (Kobayashi et al., 2008). It is possible we could use a similar approach to compare any leaf shape difference we identify within the T-DNA lines against those of natural alleles amongst the *Arabidopsis* accessions.

The phenotypes of NBLRR genes studied in the literature are often specific to environmental conditions. NBLRR genes are known to have phenotypes that vary with temperature, often becoming severe at lower temperatures. *RPP4* for example has a stronger phenotype at 4°C than at 22°C (Huang et al., 2010). The *SNC1* NBLRR gene has a dramatic effect on plant morphology at 24°C but not 28°C (Zhu et al., 2010). Naturally occurring NBLRR loci causing hybrid defects between *Arabidopsis* accessions have also been found to be temperature sensitive, with severe defects occurring at lower temperatures (Alcázar et al., 2008; Bomblies et al., 2007). The phenotypes of caused by mutations in NBLRR genes have also been known to be sensitive to humidity (Noutoshi et al., 2005). Interestingly there is work reporting a NBLRR mutation with a light specific phenotype. A mutation in the NBLRR gene *Constitutive Shade Avoidance1 (CSA1)* causes a dramatic leaf shape phenotype similar to that observed as a result of the shade response in *Arabidopsis*, and shows a hypocotyl elongation phenotype in only red light (Faigón-Soverna et al., 2006).

We obtained T-DNA insertion lines for the NBLRR genes within our leaf shape associated GWAS loci to test for leaf shape differences between these lines and the Col-0 background. We would harvest each rosette leaf of the T-DNA lines at bolting and score shape variation in these leaves using the leaf library PCs. The SALK T-DNA lines are created in a Col-0 background (Alonso et al., 2003), which is early flowering. Harvesting these plants at bolting

will allow us to identify any variation in the number of leaves grown in the T-DNA lines. For our GWAS work, we had used the greenhouses at the University of York to grow the natural accessions and so would use this same environment to test for leaf shape differences amongst the NBLRR T-DNA lines. Given the effects of light and temperature on NBLRR gene phenotypes in the literature, we were also interested to examine the response of any NBLRR T-DNA lines with differences in leaf shape to variation in temperature and light conditions.

5.2 Results

5.2.1 Leaf shape varies amongst T-DNA insertion lines for candidate genes within GWAS loci

We used SALK T-DNA insertion lines to further investigate NBLRR genes within loci associated to leaf shape traits in our GWAS work. We obtained 18 SALK T-DNA insertion lines annotated for insertions in 14 genes, see Table 5-1. As we would be unable to carry out a detailed characterisation of each mutation due to time constraints, we instead screened the set of T-DNA insertion lines for differences in leaf shape. This allowed us to identify T-DNA insertion lines that varied in leaf shape, potentially implicating candidate NBLRR genes with a leaf shape function, which we could focus on in further work. In February 2014 we grew six plants for each of these T-DNA lines and harvested the leaves once plants had bolted. We scored leaf shape and size variation using the leaf library PCs and centroid size. The leaf shape of each T-DNA insertion line was compared to the background line, Col-0, grown alongside these lines. In this growth experiment we identified two T-DNA lines, At5g45240-1 and At1g72840-1, as having a lower *llpPC2* score than Col-0, see Figure 5-1.

NASC ID	Gene	Insert location
N655775	At1g72890	Exon 2
N655912	At5g36930	Exon 7
N658536	At1g72840	Exon 4
N658563	At1g72860	Exon 3
N658582	At1g72860	Intron 2
N659957	At5g45230	Exon 4
N660465	At5g45230	Exon 3
N662567	At5g45250	Exon 2
N670661	At1g58602	1000-Promotor
N672671	At5g45240	Exon 9
N673326	At1g72910	300-UTR3
N676125	At1g59218	300-UTR5
N676263	At5g45200	Intron 2
N681211	At5g45250	300-UTR5
N682142	At1g72850	Exon 1
N683557	At5g45260	Exon 5
N685562	At1g72940	Intron 1
N686636	At1g72940	Exon 2

Table 5-1 T-DNA insertion lines annotated for NBLRR genes

Table listing the 18 T-DNA insertion lines for 14 GWAS associated NBLRRs that we analysed for leaf shape. Where possible we obtained lines annotated as having a T-DNA insertion within an exon

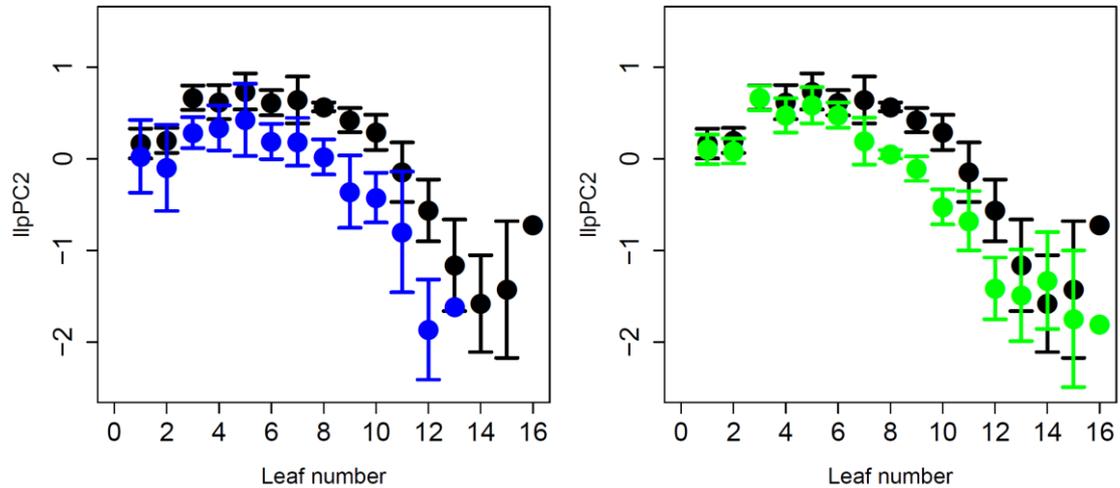


Figure 5-1 Difference in llpPC2 between Col-0 and T-DNA lines At5g45240-1 and At1g72840-1

Left figure shows the difference in mean llpPC2 score across the leaf series between Col-0, shown in black, and At5g45240-1, shown in blue. Right figure shows difference in mean llpPC2 score across the leaf series between Col-0, shown in black and At1g72840-1, shown in green. Error bars show 95% confidence intervals.

5.2.2 The T-DNA insertion line At1g72850-1 may have a condition specific phenotype

In repeated growth experiments with the NBLRR T-DNA insertion lines we found a third T-DNA insertion line, At1g72850-1, had a difference in llpPC2 score relative to Col-0. An image of rosettes of Col-0 and At1g72850-1 plants can be seen in Figure 5-4. However, the difference in shape observed for this line was not present in all growth experiments with these lines. Variability in the phenotype of this line may be the result of differences in greenhouse conditions. Figure 5-2 shows the difference in llpPC2 score between the At1g72850-1 line and Col-0 in a growth experiment in January. Growth conditions in the greenhouse may vary throughout the year. Comparison of llpPC2 score of the Col-0 control lines between January and February growth experiments showed the control line had a small difference in leaf shape between the January and February growth experiments. Interestingly, for line At1g72850-1, the difference in llpPC2 score between January and February growth experiments was much greater, see Figure 5-3. This suggested the leaf shape of At1g72850-1 may vary more than Col-0 in response to differences in growth conditions.

We aimed to examine how the leaf shape of the T-DNA lines varied in response to differences in growth conditions. Our work growing the Bay-0 x Shahdara RILs in light and shade treatments suggested that shade grown lines will have a higher llpPC2 score. We grew

the T-DNA lines At1g72840-1, At5g45240-1 and At1g72850-1 in light and shade conditions to see how the leaf shape of these lines would respond to the difference in conditions. The average llpPC2 score for each line can be seen in Figure 5-5. Each line had a higher llpPC2 score in the shade, and the difference in llpPC2 score between each T-DNA line and Col-0 was also greater in shade conditions. This result suggested that the extent of the llpPC2 difference between line At1g72580-1 and Col-0 may vary with environmental conditions, and is perhaps greater in environments resulting in higher llpPC2 score.

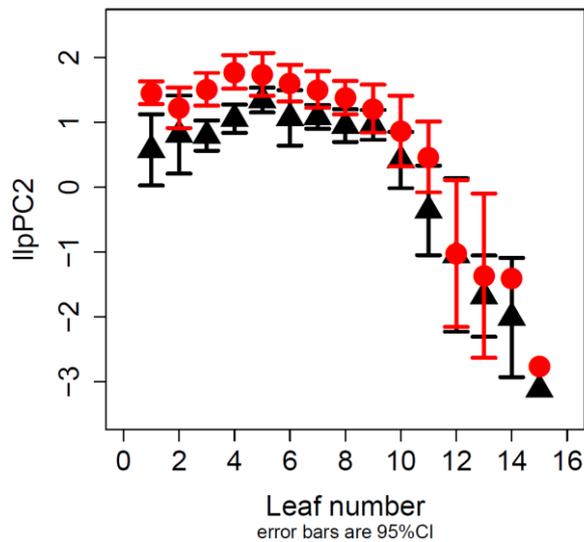


Figure 5-2 Difference in llpPC2 score between Col-0 and At1g72850-1

This figure shows the difference in mean llpPC2 score across the leaf series between Col-0, shown in black, and At1g72850-1, shown in red. Error bars show 95% confidence intervals.

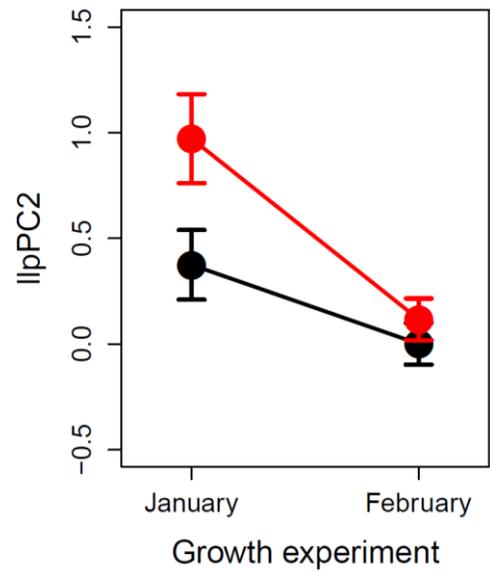


Figure 5-3 llpPC2 score of Col-0 and At1g72850-1 in two growth experiments

Figure shows the whole plant llpPC2 scores of Col-0 and At1g72850-1 plants grown in two growth experiments during January and February. Error bars show 95% confidence intervals.



Figure 5-4 Rosettes of Col-0 and T-DNA line At1g72850-1
Image shows a representative plant from the T-DNA line At1g72850-1 and Col-0.

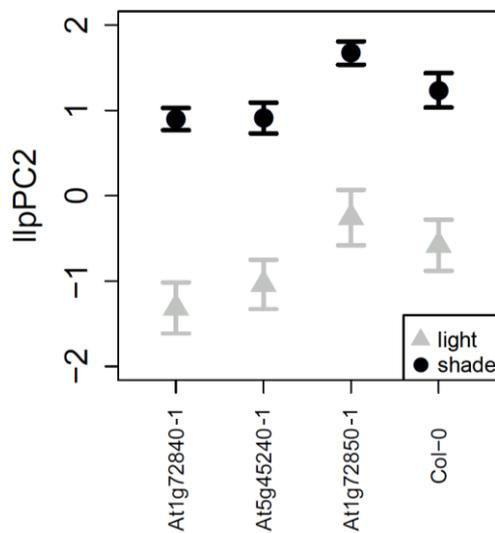


Figure 5-5 lIpPC2 score for three T-DNA lines and Col-0 in two light conditions

Figure shows the whole plant lIpPC2 scores for three T-DNA lines, At1g72840-1, At5g45240-1 and At1g72850-1, grown in either light conditions, shown by grey triangles, or shade conditions, shown by black circles. Eight plants were grown for each line. Error bars show 95% confidence intervals.

5.2.3 Characterisation of further T-DNA insertion lines for genes At5g45240 and At1g72840

We identified leaf shape differences in three SALK T-DNA insertion lines annotated as having insertions in the genes At1g72840, At1g72850 and At5g45240 after measuring leaf shape across 18 T-DNA insertion lines for 14 NBLRR genes. We wondered if the leaf shape differences in these lines would also be found in other available T-DNA insertion lines for these genes. We were able to obtain one other T-DNA insertion line for At1g72840, At1g72840-2, and three more for At5g45240, At5g45240-2, At5g45240-3 and At5g45240-4. We grew these lines in the University of York greenhouses, harvesting each rosette leaf per plant at bolting and scoring leaf shape with our leaf library PCs.

The second T-DNA insertion line we obtained for At1g72840, At1g72840-2, had a similar leaf shape to At1g72840-1. Both At1g72840-1 and At1g72840-2 had a lower llpPC2 score than Col-0 across leaves at nodes six to ten of the leaf series, see Figure 5-6. That the two T-DNA lines we obtained for At1g72840 had a similar leaf shape suggests mutant alleles of this gene may have similar effects on leaf shape. The At1g72840-2 line also differed in leaf size compared to Col-0, although no difference in leaf size was found between leaves of Col-0 and the line At1g72840-1, see Figure 5-6. Although both lines have similar llpPC2 scores over the leaf series, the size difference between leaves of the At1g72840-2 and At1g72840-1 lines suggests the effect of each T-DNA insertion per line on leaf shape and size may not be entirely identical. An image of the plants from these lines and a Col-0 plant grown alongside them can be seen in Figure 5-7.

Of the four T-DNA lines annotated with insertions in At5g45240 that we characterised, two of these, At5g45240-3 and At5g45240-4, showed no difference in leaf shape and size compared to Col-0. The At5g45240-2 line had a similar llpPC2 score to the At5g45240-1 line; both lines had a lower llpPC2 score than Col-0 across the leaf series, see Figure 5-9. Both At5g45240-1 and At5g45240-2 had a smaller leaves than Col-0, but the leaves of At5g45240-2 plants were smaller still than At5g45240-1, see Figure 5-10. Images of representative plants of the At5g45240-1 and At5g45240-2 lines can be seen in Figure 5-11. At5g45240-1 and At5g45240-2 are described as having insertions within the 9th and 1st exons of the gene respectively, whereas At5g45240-3 and At5g45240-4, are as described as having insertions in the introns of the gene, see Figure 5-8. We wondered if the variation in leaf shape and size amongst these four T-DNA lines for At5g45240 related to the individual insertion described for each line.

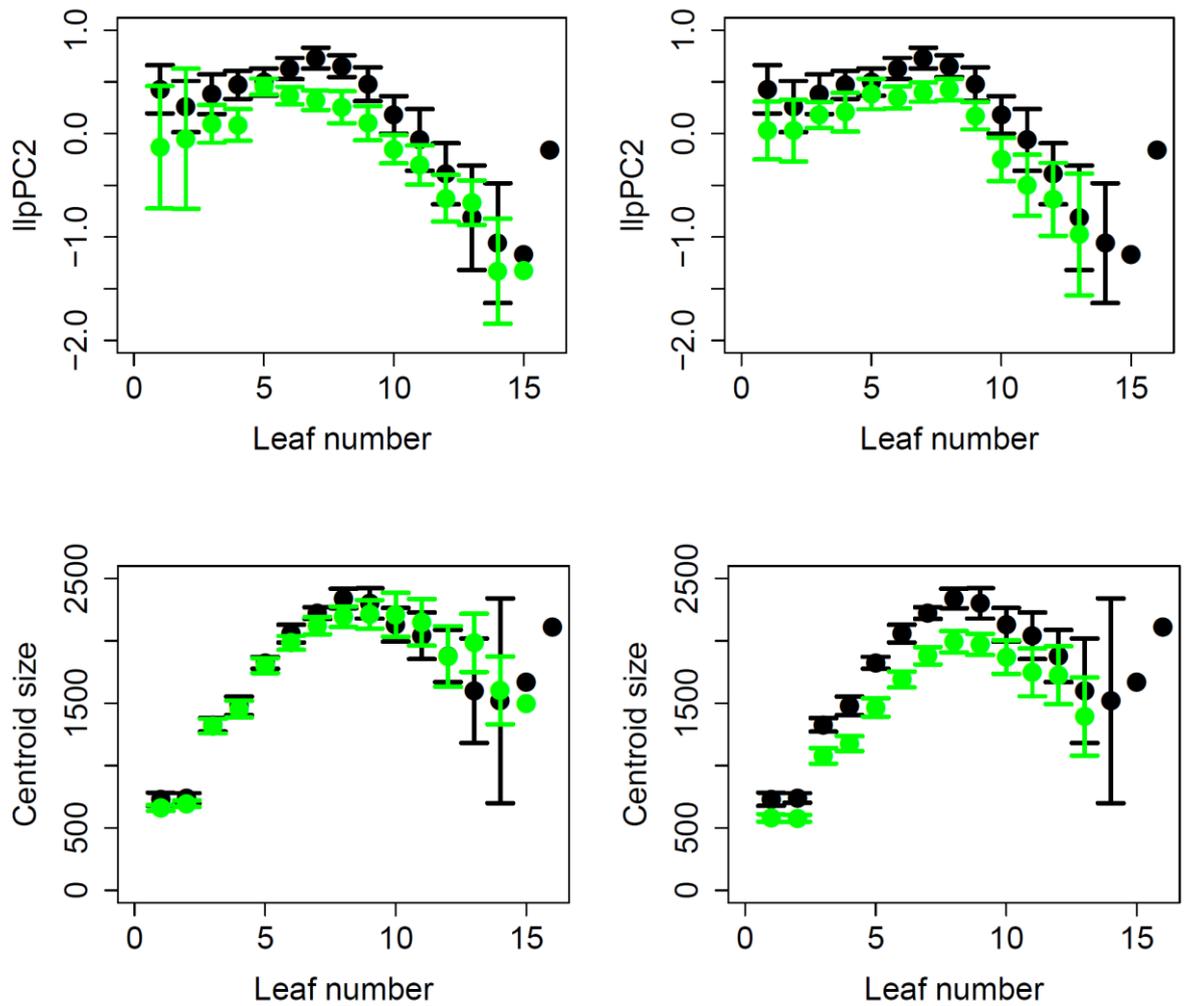


Figure 5-6 lIpPC2 score and size differences between Col-0 and T-DNA lines annotated for At1g72840

Of the two upper figures; left figure shows the mean lIpPC2 values across the leaf series for Col-0, shown in black, and At1g72840-1, in green. Right shows the mean lIpPC2 values for At1g72840-2, shown in green, and Col-0 shown again in black. For the two lower figures; left figure shows the mean centroid size values across the leaf series for Col-0, shown in black, and At1g72840-1, in green. Right shows the centroid size for At1g72840-2, shown in green, and Col-0 shown in black. Error bars show 95% confidence intervals.



Figure 5-7 Rosettes of Col-0 and T-DNA lines At1g72840-1 and At1g72840-2
 Photograph shows a representative plant from the lines Col-0, At1g72840-1 and At1g72840-2.

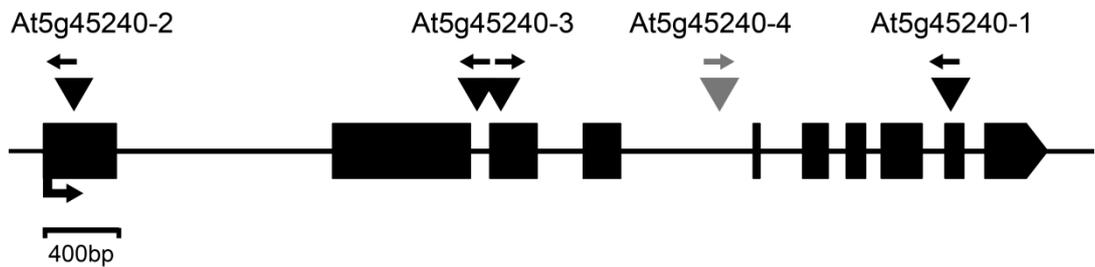


Figure 5-8 At5g45240 gene model
 Diagram of At5g45240 gene model, using TAIR10 annotation, from 5' to 3'. Exons of the gene are indicated by black squares and introns by a thin black line. The position of T-DNA insertions are indicated by triangles. PCR was used to confirm the presence of these insertions, and for At5g45240-1, -2 and -3, a band was produced supporting the annotated insertion site. For At5g45240-3 PCR results suggested a second nearby insertion in the opposite orientation. A grey triangle marks the annotated insertion site for At5g45240-4, as although PCR results confirm the presence of at least one T-DNA insertion within this region, further work is required to understand the likely integration of this insertion.

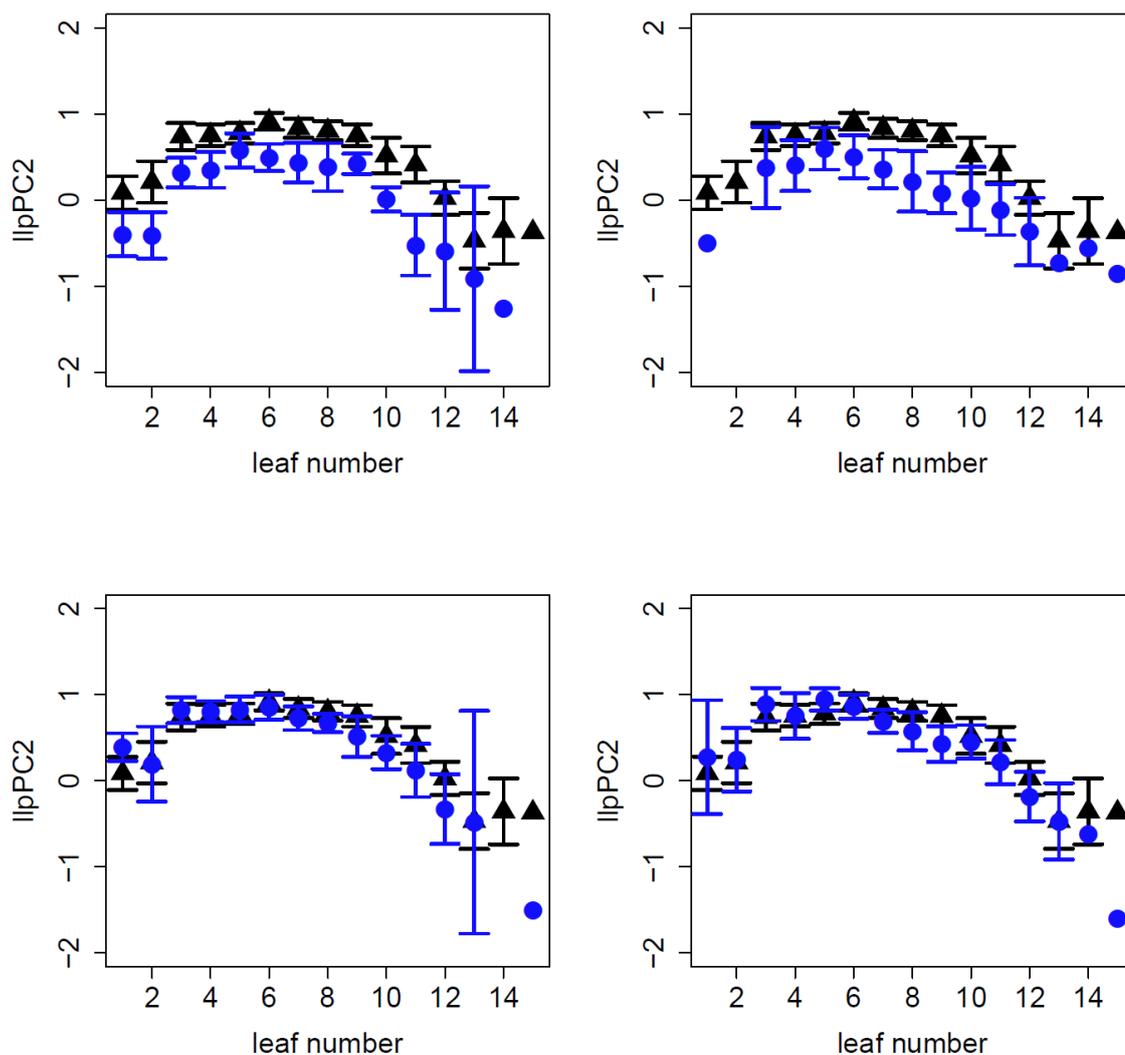


Figure 5-9 Difference in lIpPC2 score between Col-0 and T-DNA lines for At5g45240

Figures show the mean lIpPC2 score for the four T-DNA lines of At5g45240 and Col-0. Values for Col-0 are shown in black and values for each T-DNA insertion line are shown in blue. The T-DNA lines are shown compared to Col-0 as follows; top left; At5g45240-1, top right; At5g45240-2, bottom left; At5g45240-3, and bottom right; At5g45240-4. 12 plants were harvested for each line.

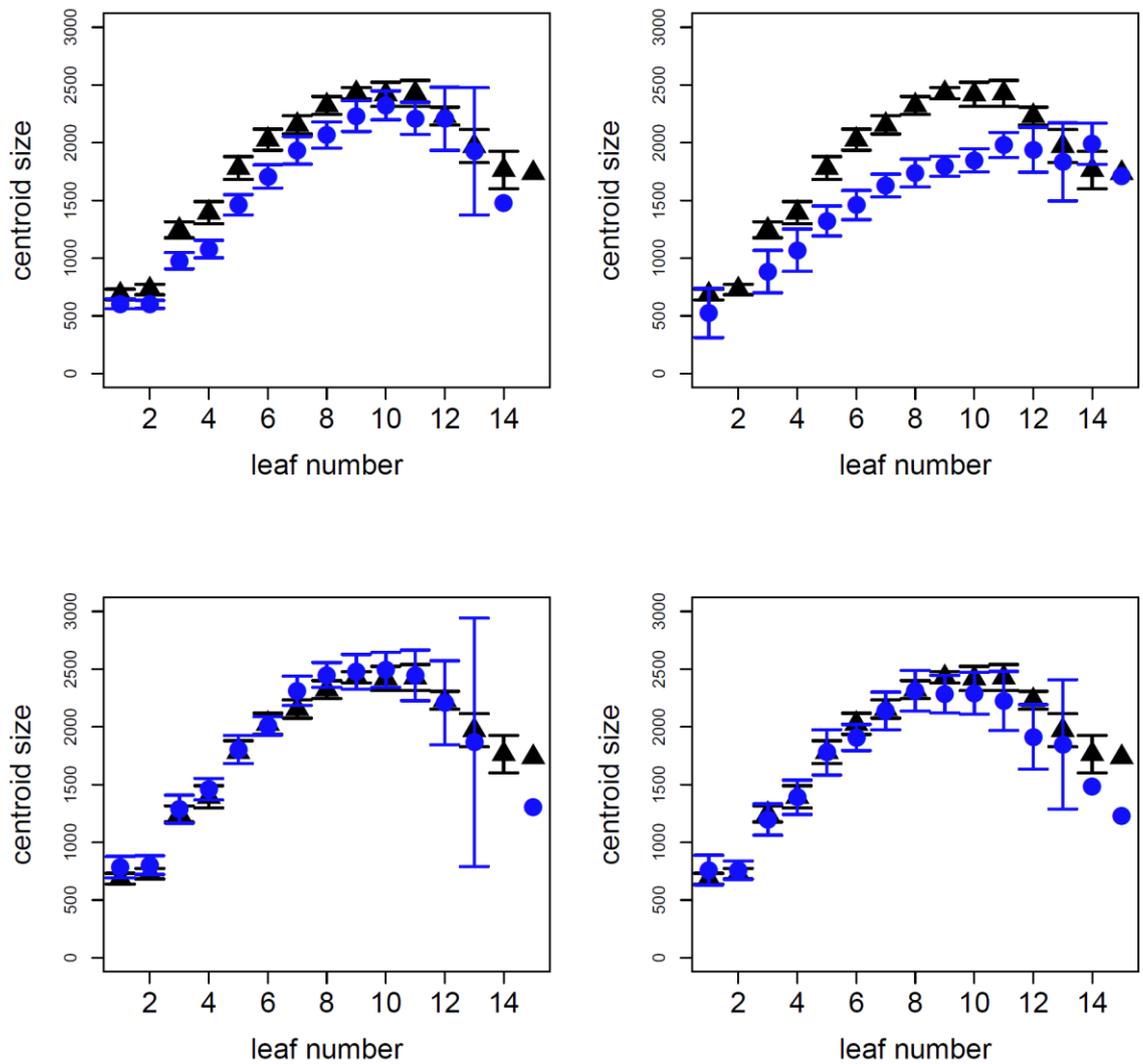


Figure 5-10 Difference in size between Col-0 and T-DNA lines for At5g45240

Figures show the mean centroid size for four T-DNA insertion lines for At5g45240 and Col-0.

The values for Col-0 are shown in black and values for each T-DNA line are shown in blue.

The T-DNA lines are shown compared to Col-0 as follows; top left; At5g45240-1, top right;

At5g45240-2, bottom left; At5g45240-3, and bottom right; At5g45240-4. To calculate mean

trait scores for each line, 12 plants were harvested.



Figure 5-11 Image of Col-0 and T-DNA lines At5g45240-1 and At5g45240-2
Image shows a representative plant from the lines Col-0, At5g45240-1 and At5g45240-2.

5.2.4 PCR genotyping to confirm presence of T-DNA insertion in lines

We aimed to verify the annotated location of the T-DNA insertions in the SALK T-DNA lines we had obtained from the Nottingham Arabidopsis Stock Centre (NASC). To do this we used PCR, using primers created with the online SALK T-DNA primer tool to genotype each line (<http://signal.salk.edu/tdnaprimers.2.html>).

Three primers are used to verify the presence of the T-DNA insertion; a pair that will amplify a region of the host genome, and a third primer complementary to the left border of the T-DNA insertion. The pair annealing to the host genome are designed to amplify the region otherwise containing the T-DNA insertion and so will only produce a band for a wild type allele. When using all three primers at once, the presence of a large T-DNA insertion between this primer pair will prevent amplification across the region between host genome primers, and if a T-DNA insertion is present a smaller band than for the two host genome primers will be produced, typically between the T-DNA left border primer and the right host genome primer, though T-DNA insertions may insert in a reverse orientation. As each outcome produces a different size band, plants heterozygous for an insertion can be expected to have two bands of different sizes, a larger host genome band, and a smaller band from between the T-DNA border and a host genome primer.

Using the SALK primer design tool to design primers, we verified the presence of T-DNA insertions in the expected regions for the insertion lines At1g72840-1, At1g72850-1, At5g45240-1 and At5g45240-2. Each set of primers produced a wild type and single mutant

band of the expected size for DNA of Col-0 and the corresponding T-DNA line, see Figure 5-12, Figure 5-13, and Figure 5-14.

When genotyping for the insertions in At5g45240-3 and At5g45240-4 with the respective set of three primers we found that two bands were produced for each of these lines. Col-0 DNA with the respective primers for each of these lines produced a wild type band of the expected size, but for DNA of either of these two T-DNA lines, there was one band within the size range expected for the insertion, and another larger band, a different size to that found with each primer set and Col-0 DNA, see Figure 5-15. This suggested a more complex arrangement than a single T-DNA insertion. Given the variation in leaf shape across the four T-DNA insertion lines annotated for the At5g45240 line and the unexpected second bands when genotyping lines At5g45240-3 and At5g45240-4, we decided to further investigate the arrangement of the T-DNA insertions in these lines.

So far we had used both host genome primers in combination along with the T-DNA left border primer, assuming that if the band produced with T-DNA line DNA was within the expected size range, that this was produced with the T-DNA left border primer, LBI1.3, and the right host genome primer. As At5g45240-3 and At5g45240-4 produced two bands for the three primer combination, we wanted to test this assumption, and used the left and right host genome primers individually with LBI1.3.

Using the right host genome primer and LBI1.3 with DNA for lines At5g45240-1 and At5g45240-2 produced the same band identified when using the three primer combination for these lines, see Figure 5-17 and no bands were identified when using the left host genome primer and LBI1.3 with these two lines, see Figure 5-16. This suggested the bands produced for At5g45240-1 and At5g45240-2 were the result of amplification of the region between the left T-DNA border primer and the host genome right primer, as assumed previously.

For At5g45240-3, we found that using the right host genome primer and LBI1.3 produced the band of the expected size for this line we had seen previously, see Figure 5-15 and Figure 5-17. Using the left host genome primer and LBI1.3 also produced a band for At5g45240-3, see Figure 5-16. This band was a similar size to the larger band produced when using both host genome primers with LBI1.3 for this line earlier, see Figure 5-15. This suggested there may be two T-DNA insertions in opposite orientations in the At5g45240-3 line; one which produces a band between LBI1.3 and the host genome right primer, and

another insertion in the reverse orientation producing a band between the host genome left primer and LBI1.3.

For At5g45240-4, only the left host genome primer in combination with LBI1.3 produced a band, and this matched the size of the larger of the two bands identified with the left and right host genome primers initially for this line. This suggested the presence of a T-DNA insertion in the reverse orientation near the left host genome primer. Given that the second band produced with DNA from this line using both host genome primers and LBI1.3 earlier was of the expected size for the right host genome and LBI1.3, it was surprising we did not find such a band with At5g45240-4 DNA using solely the right host primer and LBI1.3, however repeating the experiment with this combination also did not show a band.

That the smaller band for produced with At5g45240-4 DNA requires all three primers is curious. A worthwhile follow up experiment would be to use just the two host genome primers with DNA from this line to verify that no band is produced, as one possible explanation for the results observed is that the host genome has been rearranged, resulting in a amplifiable region between the two host genome primers, but that is shorter than found in wild type.

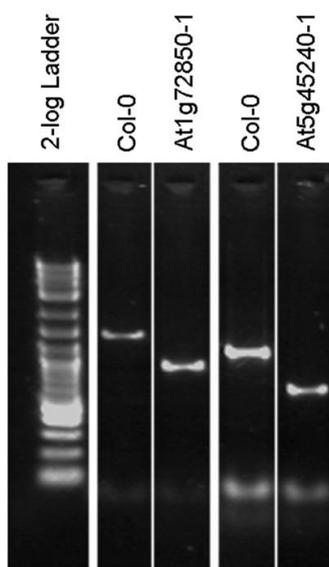


Figure 5-12 Testing At1g72850-1 and At5g45240-1 insertion position

In the first lane is a 2-log ladder. The second and third lanes use left (LP) and right primers (RP) for the host genome across the region of the At1g72850-1 insertion and an LBI1.3 primer complementary for the left border of the T-DNA insertion. The fourth and fifth lanes contain LP and RP for the host genome across the region containing the At5g45240-1 insertion, and the LBI1.3 primer. The DNA sample used for the reaction run in each lane is listed above each lane respectively. DNA for both insertion lines produces a single band different to Col-0, suggesting each has a homozygous insertion at this point.

Expected size for wild type band with At1g72850-1 LP and RP primers is 1219bp. If there is a T-DNA insertion in this interval, expect a band of 580-880bp. For At5g45240-1 LP and RP in a wild type background, a band of 991bp is expected. If there is an insertion in this interval, a band between 464-764bp is expected.

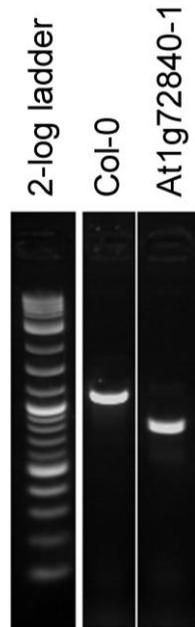


Figure 5-13 Testing At1g72840-1 insertion position

In the first lane is a 2-log ladder. The second and third lanes use left (LP) and right primers (RP) for the host genome across the region of the At1g72840-1 insertion and an LBI1.3 primer complementary for the left border of the T-DNA insertion. The DNA sample used for the reaction run in each lane is listed above each lane respectively.

Expected size for wild type band with At1g72840-1 LP and RP primers is 1105bp. If there is a T-DNA insertion in this interval, expect a band of 585-885bp.

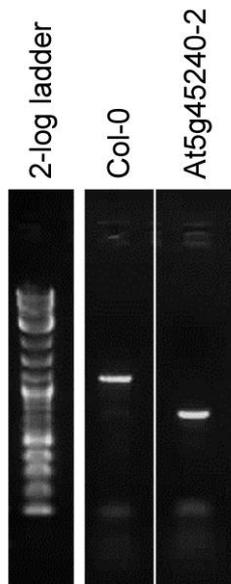


Figure 5-14 Testing At5g45240-2 insertion position

In the first lane is a 2-log ladder. The second and third lanes use left (LP) and right primers (RP) for the host genome across the region of the At5g45240-2 insertion and an LBI1.3 primer complementary for the left border of the T-DNA insertion. The DNA sample used for the reaction run in each lane is listed above each lane respectively.

Expected size for wild type band with At5g45240-2 LP and RP primers is 990bp. If there is a T-DNA insertion in this interval, expect a band of 438-738bp.

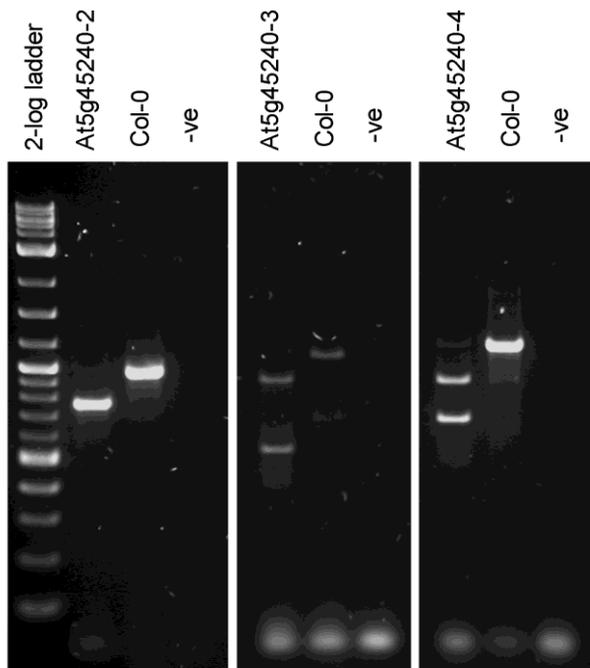


Figure 5-15 Testing insertion position in T-DNA lines At5g45240-2, At5g45240-3 and At5g45240-4

In the first lane is a 2-log ladder. Lanes 2 to 4 use left (LP) and right primers (RP) for the host genome over a region of the At5g45240-2 insertion and an LBI1.3 primer complementary for the left border of the T-DNA insertion. The expected band for the presence of the At5g45240-2 insertion is 438-738bp, the wild type expected band is 990bp. Lanes 2 to 4 use left (LP) and right primers (RP) for the host genome over a region of the At5g45240-3 insertion and a LBI1.3 primer complementary for the left border of the T-DNA insertion. The expected band for the presence of the At5g45240-3 insertion is 450-750bp, the wild type expected band is 1111bp. Lanes 2 to 4 use left (LP) and right primers (RP) for the host genome over a region of the At5g45240-4 insertion and a LBI1.3 primer complementary for the left border of the T-DNA insertion. The expected band for the presence of the At5g45240-4 insertion is 598-898bp, the wild type expected band is 1222bp.

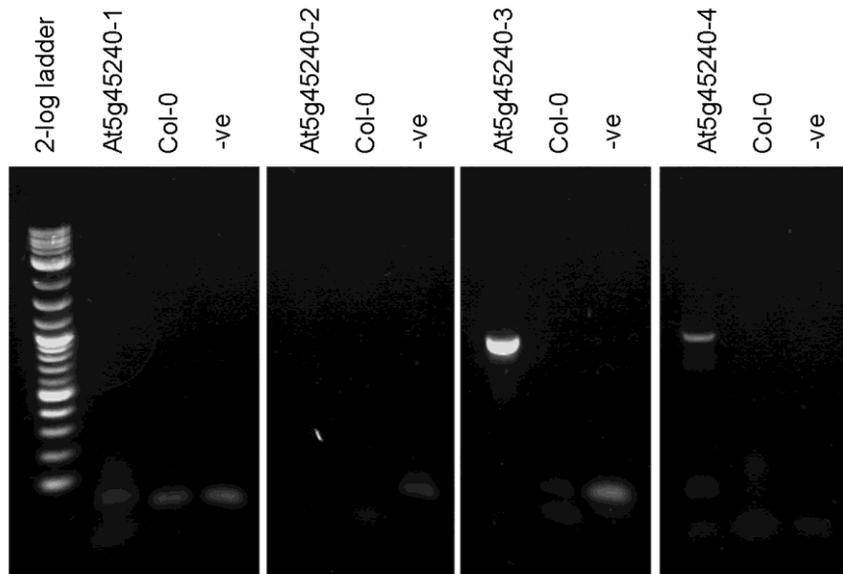


Figure 5-16 Testing insertion position in T-DNA lines annotated for At5g45240

For PCR reactions with each of the T-DNA lines annotated for At5g45240, we used the left primer for each T-DNA insertion and LBL1.3. Along with DNA from each T-DNA insertion line of At5g45240, we used Col-0 and a negative control with the reactions. T-DNA insertions in the expected orientation would not produce a band with the left primer and LBL1.3 alone. Bands for At5g45240-3 and At5g45240-4 may indicate the presence of T-DNA insertions in reverse orientation.

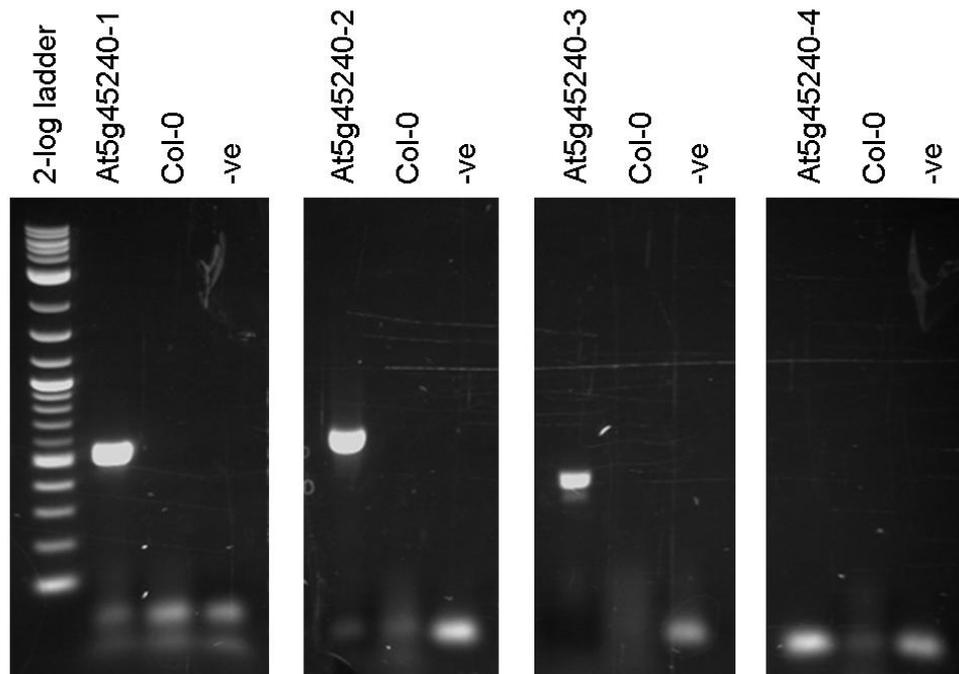


Figure 5-17 Testing insertion position in T-DNA lines annotated for At5g45240

For PCR reactions with each of the T-DNA lines annotated for At5g45240, we used the right primer for each line and LBL1.3. Along with DNA from each T-DNA insertion line of At5g45240, we used Col-0 and a negative control for the reaction. T-DNA insertions in the expected orientation will produce a band with the right primer and LBL1.3. Bands for At5g45240-1, At5g45240-2 and At5g45240-3 indicate the potential presence of T-DNA insertions in expected orientation.

T-DNA line	At5g45240-1	At5g45240-2	At5g45240-3	At5g45240-4
Size with LP (bp)	-	-	900	900
Size with RP (bp)	600	700	500	-
Size with LP + RP (bp)	600	700	900, 500	900, 700
Expected size (bp)	464-764	438-738	450-750	598-898
Wild type size (bp)	991	990	1111	1222

Table 5-2 Confirming presence of insertions in At5g45240 T-DNA lines

Table shows the band sizes produced when using the LBI1.3 primer with the left, right or both left and right host genome primers across the annotated insertion site for T-DNA lines for At5g45240. Dashes mark combinations where no band was produced. The expected insertion and wild type band sizes for each T-DNA line are also shown. Both At5g45240-1 and At5g45240-2 show a band of the expected size using the right host genome primer and LBI1.3, suggesting a T-DNA insertion in the expected annotated orientation and position. For At5g45240-3, a band of the expected size is produced when using the right host genome primer and LBI1.3, but another band, larger than the expected size, is produced when using the left host genome primer, suggesting there may be two T-DNA insertions of opposite orientation at this position. A band of the expected size is produced for At5g45240-4 only when using both the left and right host genome primers with LBI1.3, as well as a larger band which is also produced using the left host genome primer, suggesting a more complex pattern of insertion the annotated insertion for this T-DNA line.

5.2.5 Association of T-DNA line leaf shape differences to T-DNA insertions in F2 populations

The SALK T-DNA insertion lines we obtained from NASC were annotated as having a single homozygous insertion in one of the 14 NBLRR candidate genes. However, as these lines are produced and analysed on large scale as one of many thousand, it is useful to verify that the phenotype observed for a line is not the result of another unidentified insertion or other mutation elsewhere in the genome.

To confirm that each T-DNA insertion segregated with the difference in leaf shape we had identified for that line, we created F2 populations by crossing each T-DNA line to Col-0 and self fertilising the offspring. This produced an F2 population segregating for the presence of the T-DNA insertion. We would be able to test whether the differences in leaf shape we observed for the T-DNA lines were observed in F2 plants with the T-DNA insertion. We would also be able to analyse the leaf shape of plants heterozygous for these insertions. We created F2 populations for the T-DNA lines with differences in leaf shape; At1g72850-1, At1g72840-1, At5g45240-1 and At5g45240-2.

For At5g45240-1 we genotyped 76 F2 plants and compared leaf shape between the three genotypes, Col-0/Col-0, At5g45240-1/At5g45240-1 and At5g45240-1/Col-0. The frequency of these alleles did not differ significantly from expected values using a Chi² test. We found that lines homozygous for the At5g45240-1 insertion had leaf shape very similar to the At5g45240-1 control line plants, see Figure 5-18 and Figure 5-19. This suggested the At5g45240-1 insertion was responsible for the leaf shape difference we observed for the At5g45240-1 line. The effect of the At5g45240-1 insertion is recessive, as heterozygous F2 plants did not differ in leaf shape from F2 plants without an insertion, see Figure 5-19.

We carried out a similar experiment for the At5g45240-2 insertion. Between the At5g45240-2 and Col-0 lines grown as controls alongside the F2 population there was a difference in leaf size across the 3rd to 9th leaves and the llpPC3 score over the 5th to 9th leaves of these plants, though no difference in llpPC2 score, see Figure 5-20 and Figure 5-21. The At5g45240-2 line had previously been identified with a different in leaf size and llpPC2 score when compared to Col-0. This suggested that leaf size was the most consistent phenotype for this line between experiments, and that although there was also a difference in leaf shape, the precise difference in shape may vary between growth experiments.

We genotyped 46 F2 plants and compared the effect of the At5g45240-2 insertion on leaf size and llpPC3, as these differed between the Col-0 and At5g45240-2 control lines grown alongside the F2 plants. Allele frequencies were not significantly different from expected values, tested by Chi² test. There was a difference in llpPC3 score across the 5th to 9th leaves and a difference leaf size across the leaf series between lines homozygous for the At5g45240-2 insertion and lines without, see Figure 5-23 and Figure 5-22. We found that F2 plants heterozygous for the At5g45240-2 insertion had a wild type phenotype, suggesting the effect of this insertion on leaf shape and size is recessive.

We next looked to confirm the segregation of the At1g72840-1 insertion with a difference in leaf shape. There was a difference in llpPC2 score between the Col-0 and At1g72840-1 lines grown as controls for the F2 population, the At1g72840-1 line had a lower llpPC2 score across the 4th to 10th leaves. We genotyped 81 plants for this F2 population and allele frequencies were not significantly different from expected values. However, when comparing the leaf shape of plants within the F2, there was no effect of the At1g72840-1 insertion on llpPC2. This suggested that the leaf shape difference between Col-0 and At1g72840-1 we observed was not the result of the T-DNA insertion in the At1g72840 gene.

We were surprised that the At1g72840-1 insertion did not segregate with a difference in llpPC2, as another T-DNA line for this gene, At1g72840-2, had a very similar leaf shape to the At1g72840-1 line. This had made us more confident that the similar leaf shape differences of these two T-DNA lines relative to Col-0 were the result of T-DNA insertions in the At1g72840 gene. However, our segregation analysis showed this was not the case for the At1g72840-1 line. We chose to discard the At1g72840-2 line from further analysis, as we suspected the leaf shape difference of the At1g72840-2 line might not segregate with the At1g72840-2 insertion due to the similarity in leaf shape effect between the At1g72840-2 and At1g72840-1 lines. This decision allowed us to focus remaining time on further work with the At5g45240-1 and At5g45240-2 lines, further work would be required to establish whether the leaf shape difference observed for the At1g72840-2 line would segregate with the At1g72840-2 insertion.

We attempted to carry out the test of association between T-DNA insertion and leaf shape effect for the At1g72850-1 line. However, we could not find a difference in leaf shape between the At1g72850-1 and Col-0 control lines we grew alongside the F2 plants. We tried using both our standard greenhouse conditions, and a shade treatment, but no difference in llpPC2 score was identified between the At1g72850-1 and Col-0 lines. This prevented us

from testing for an effect on leaf shape of the At1g72850-1 insertion in the segregating F2 population. Despite this we decided to continue using this line in our further experiments, as we were curious if we could identify environmental conditions in which this line showed a consistent difference in leaf shape to Col-0. If we were able to identify these conditions, further work could be carried out to test for the segregation of the At1g72850-1 insertion with a leaf shape effect using such conditions.

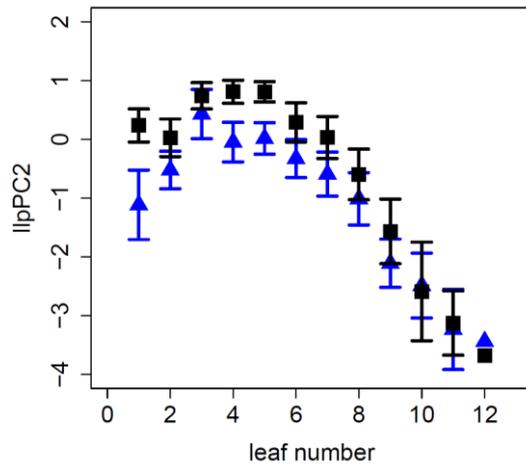


Figure 5-18 Difference in lIpPC2 score between Col-0 and At5g45240-1 control lines

Mean lIpPC2 score for At5g45240-1 and Col-0 control lines across the leaf series. Values for Col-0 are shown by black squares and for At5g45240-1 values are shown as blue triangles. Error bars show 95% confidence intervals. 12 plants were harvested for each line.

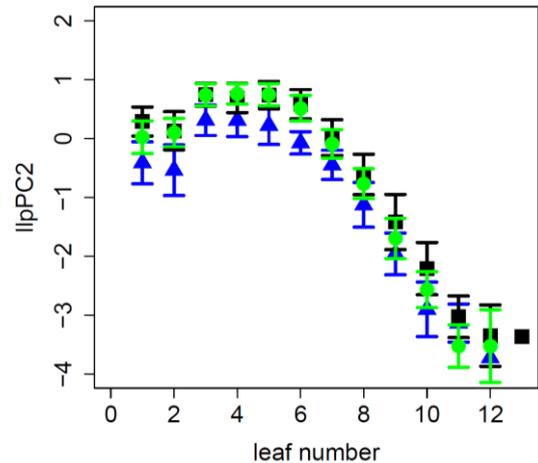


Figure 5-19 lIpPC2 score in F2 plants with or without At5g45240-1 insertion

Figure shows the mean lIpPC2 score across the leaf series for the F2 population segregating for the At5g45240-1 insertion. The mean score for plants without the insertion are shown in black, plants homozygous for the insertion are shown in blue, and plants heterozygous for the insertion are shown in green. Error bars show 95% confidence intervals.

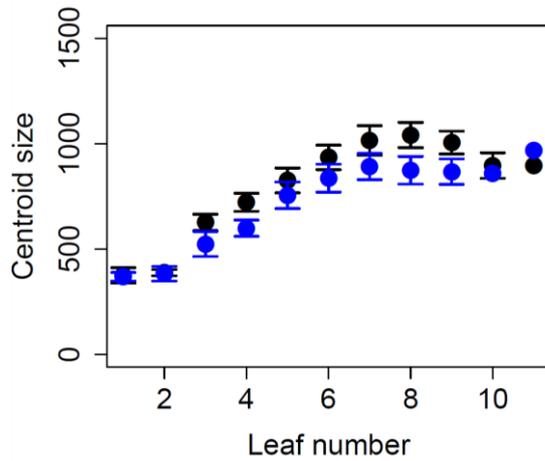


Figure 5-20 Difference in size between Col-0 and At5g45240-2 control lines

Mean centroid size score for At5g45240-2 and Col-0 control lines grown alongside the F2 population segregating for the At5g45240-2 insertion. Values for Col-0 are shown as black circles, values for At5g45240-2 are shown as blue circles. 12 plants were harvested for each line. Error bars show 95% confidence intervals.

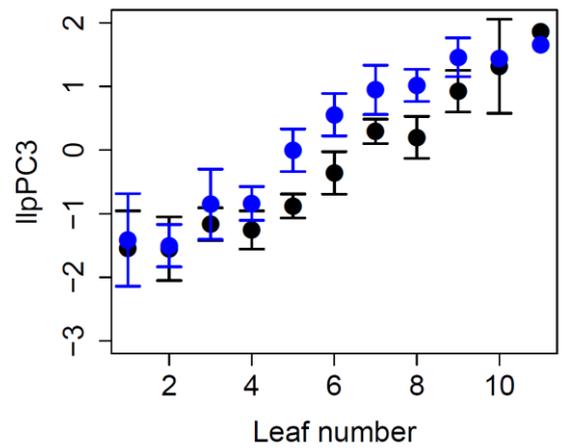


Figure 5-21 Difference in llpPC3 score between Col-0 and At5g45240-2 control lines

Mean llpPC3 score for At5g45240-2 and Col-0 control lines grown alongside the F2 population segregating for the At5g45240-2 insertion. Values for Col-0 are shown as black circles, values for At5g45240-2 are shown as blue circles. 12 plants were harvested for each line. Error bars show 95% confidence intervals.

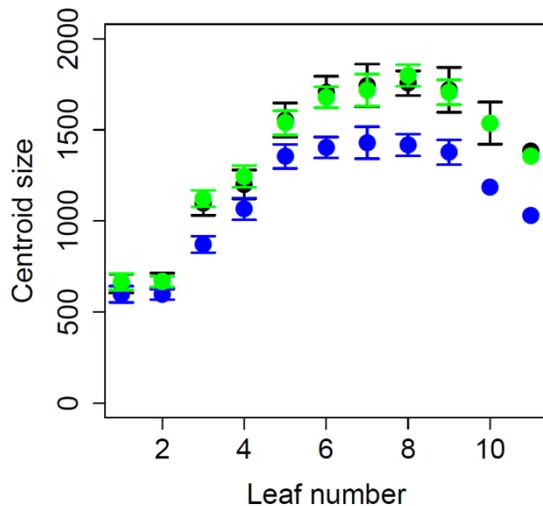


Figure 5-22 Size in F2 plants with or without At5g45240-2 insertion

Figure shows the mean centroid size across the leaf series for the F2 population segregating for the At5g45240-2 insertion. The mean score for plants without the insertion are shown in black, plants homozygous for the insertion are shown in blue, and plants heterozygous for the insertion are shown in green. Error bars show 95% confidence intervals.

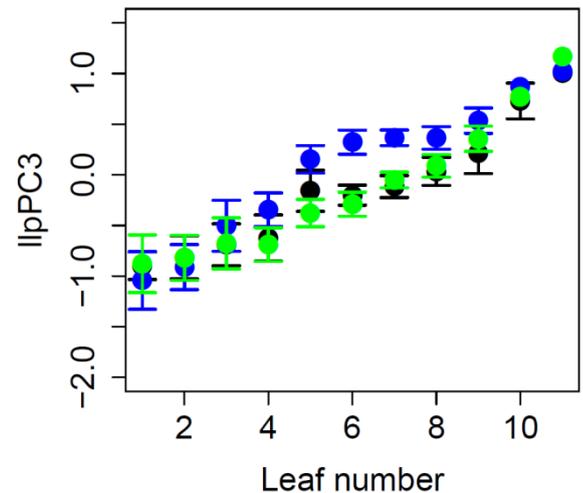


Figure 5-23 llpPC3 score in F2 plants with or without At5g45240-2 insertion

Figure shows the mean llpPC3 score across the leaf series for the F2 population segregating for the At5g45240-2 insertion. The mean score for plants without the insertion are shown in black, plants homozygous for the insertion are shown in blue, and plants heterozygous for the insertion are shown in green. Error bars show 95% confidence intervals.

5.2.6 T-DNA insertion lines for homologues of At1g72850 and At5g45240 vary in leaf shape and size

We were curious as to whether T-DNA insertion lines for genes closely related to At1g72850 and At5g45240 would also show leaf shape phenotypes. We used the BLAST webpage at TAIR (Berardini et al., 2015) to identify genes similar to At1g72850 and At5g45240, using the nucleotide sequence of each gene including introns and untranslated regions. For each similar gene with an E value below 0.002 we obtained SALK T-DNA mutants where possible and scored these lines for leaf shape differences.

We obtained four T-DNA lines for three genes similar to At1g72850, and five lines for four genes similar to At5g45240. We grew 12 plants for each line and harvested the rosette leaves of each plant at bolting, scoring the shape of the leaves with the leaf library PCs. We

identified one line, At4g09420-1, with a difference in leaf shape compared to Col-0, see Figure 5-24. This line was annotated for a T-DNA insertion in At4g09420, a gene related to At1g72850. We also identified a T-DNA line with a lower number of leaves than Col-0, At2g17050-1. This line had significantly fewer leaves than Col-0, (Kruskal-Wallis, $\chi=8.12$, $df=1$, $p<0.005$). The leaves of this line were also smaller than those of Col-0 near the end of the leaf series. This line was annotated for a T-DNA insertion in At2g17050, a gene similar to At5g45240. That the majority of the T-DNA lines for which we scored shape, that were annotated for seven genes most similar to At1g72850 and At5g45240 did not vary in leaf shape may indicate that these genes closely related to At1g72850 and At5g45240 do not also have an effect on leaf shape.

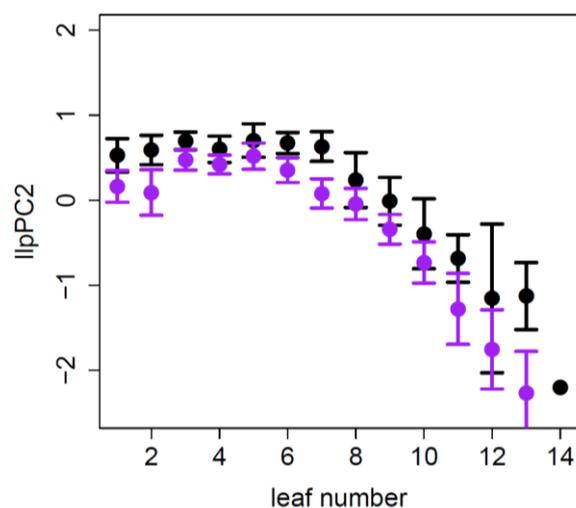


Figure 5-24 Difference in lIpPC2 score between Col-0 and At4g09420-1

Figure shows the mean lIpPC2 for Col-0 and the At4g09420-1 T-DNA line. Values for Col-0 are shown in black and those for At4g09420-1 are shown in purple. 12 plants were harvested for each line. Error bars show 95% confidence intervals.

5.2.7 T-DNA insertions in some NBLRR genes are associated with differences in leaf shape

We had found differences in leaf shape and size between multiple T-DNA insertion lines for NBLRR genes and Col-0. NBLRR genes are often characterised with phenotypes specific to environmental conditions and we were interested to test the response of T-DNA lines we identified with differences in leaf shape to a range of conditions. The At1g72850-1 line had shown a difference in leaf shape in some growth experiments but not others. Our work so far suggested that environments resulting in a high lIpPC2 score for Col-0 leaves, such as shade conditions, would also result in a greater increase in lIpPC2 score for leaves of the T-DNA line At1g72850-1.

We were curious as to whether we could identify an environmental change resulting in a difference in lIpPC2 score between the At1g72850-1 line and Col-0. We were also interested

to test whether differences in light and temperature conditions might affect the leaf shape and size of the four T-DNA lines annotated for At5g45240, given that so far only two of these lines, At5g45240-1 and At5g45240-2, showed differences in leaf shape compared to Col-0.

We wondered as to how T-DNA insertions in different NBLRR genes might interact, and so created an At1g72850-1 At5g45240-1 double mutant to include in further analysis. For further information on the creation of this double mutant, see appendices. We had also created an At5g45240-1 At1g72840-1 double mutant, however this was discarded after a segregation analysis for the At1g72840-1 insertion showed the insertion did not segregate with the leaf shape difference identified in this line. Previously work on a NBLRR gene with a leaf morphology phenotype had found an interaction with phytochrome B (PhyB) (Faigón-Soverna et al., 2006), and so we also opted to create a PhyB-9 At5g45240-1 double mutant using a mutant line for PhyB (PhyB-9) (Reed et al., 1998).

We would continue our investigation of the possible leaf shape functions of NBLRR genes now focused on T-DNA insertions lines for genes that had leaf shape differences relative to Col-0 in our earlier work. We would investigate the leaf shape and size of these lines in varying environmental conditions, test the expression levels of At5g45240 transcripts in the T-DNA lines for this gene, examine complementation of T-DNA lines with natural accessions and explore interactions of T-DNA insertions in these NBLRR genes in double mutants.

5.2.8 The At5g45240-1 insertion does not decrease llpPC2 score in a Phytochrome B mutant background

A mutation in the *CONSTITUTIVE SHADE AVOIDANCE1 (CSA1)* NBLRR gene has been suggested to interfere with phytochrome signalling (Faigón-Soverna et al., 2006). A dominant mutation in *CSA1* causes a shade avoidance response in *Arabidopsis* similar to that observed for *Phytochrome B (PhyB)* mutants, but does not have an effect in a *PhyB* background, suggesting this mutation interferes with phytochrome B signalling (Faigón-Soverna et al., 2006). We created a At5g45240-1 *PhytochromeB-9* (referred to as At5g45240-1 PhyB-9) double mutant to examine the effect on leaf shape of the At5g45240-1 insertion in a PhyB-9 background (Reed et al., 1998). For further information on the creation of this double mutant, see the appendices. We grew Col-0, At5g45240-1, PhyB-9 and the At5g45240-1 PhyB-9 double mutant in the University of York greenhouses in June 2015.

The At5g45240-1 line has a lower llpPC2 score than Col-0, and PhyB-9 plants have a higher llpPC2 score than Col-0, see Figure 5-25. Despite lower llpPC2 score in the At5g42540-1 line,

we found that the double mutant did not have a lower *llpPC2* score than the *PhyB-9* single mutant. Indeed, for some leaves, the *At5g45240-1 PhyB-9* double mutant had a higher *llpPC2* score than the *PhyB-9* single mutant, see figure Figure 5-25. There was no indication that the *At5g45240-1* insertion had any effect on *llpPC3* or centroid size in a *PhyB-9* background, despite causing a decrease in both traits as a single mutant, see Figure 5-26 and Figure 5-27. This result shows the effects of the *At5g45240-1* insertion on leaf shape and size are negated in a *PhyB-9* background.

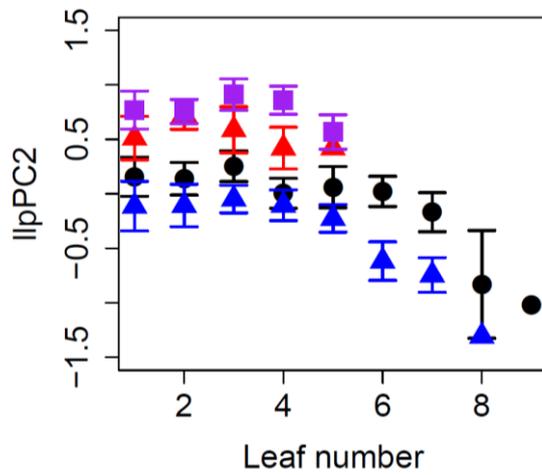


Figure 5-25 lIpPC2 score in At5g45240-1 PhyB-9 double mutant

Figure shows the mean lIpPC2 score for Col-0, shown in black, At5g45240-1, shown by blue triangles, PhyB-9, shown by red triangles, and the At5g45240-1 PhyB-9 double mutant, shown as purple squares. Error bars show 95% confidence intervals. Around twenty plants were harvested for each line.

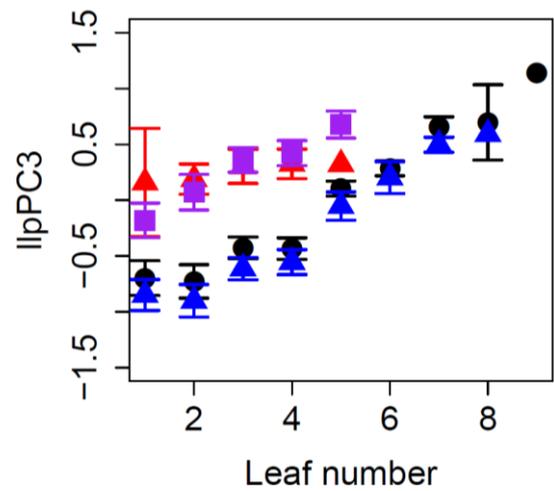


Figure 5-26 lIpPC3 score in At5g45240-1 PhyB-9 double mutant

Figure shows the mean lIpPC3 score for Col-0, shown in black, At5g45240-1, shown by blue triangles, PhyB-9, shown by red triangles, and the At5g45240-1 PhyB-9 double mutant, shown as purple squares. Error bars show 95% confidence intervals. Around twenty plants were harvested for each line.

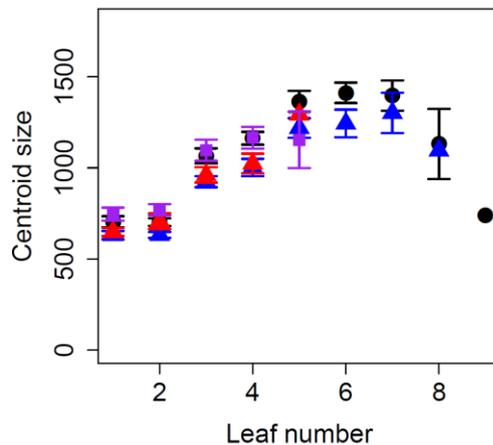


Figure 5-27 Leaf size in At5g45240-1 PhyB-9 double mutant

Figure shows the mean centroid size for Col-0, shown in black, At5g45240-1, shown by blue triangles, PhyB-9, shown by red triangles, and the At5g45240-1 PhyB-9 double mutant, shown as purple squares. Error bars show 95% confidence intervals. Around twenty plants were harvested for each line.

5.2.9 Hypocotyl length of T-DNA lines for some NBLRR genes may vary in response to different wavelengths of light

Mutations in the NBLRR gene *CSA1* have been found to show red light specific hypocotyl phenotypes (Faigón-Soverna et al., 2006). We tested whether any of the T-DNA lines annotated for NBLRR genes with leaf shape differences had a light specific effect on hypocotyl growth by growing them in red and blue light and a dark treatment. We grew the four T-DNA lines annotated for At5g45240, and the T-DNA line for At1g72850 and the double mutants we had created, At5g45240-1 PhyB-9 and At1g72850-1 At5g45240-1 in these conditions. As controls we used Col-0 and PhyB-9. For each line we harvested between 19 and 35 plants in each experiment. Seeds were sown onto agar plates and stored at 4°C for three nights before being transferred to a light treatment for four nights before harvest.

In both red and blue light treatments, hypocotyls of At5g45240-1 line plants were significantly shorter than for Col-0 plants, see Figure 5-28, Figure 5-29, Figure 5-30 and Figure 5-31. In dark conditions, only one of three experimental repeats showed At5g45240-1 plants to have a significantly shorter hypocotyl than Col-0, see Figure 5-32, Figure 5-33 and Figure 5-34. These results suggest that the hypocotyl length of At5g45240-1 line plants may differ from Col-0 in the presence of light.

Hypocotyls of At5g45240-2 plants were not significantly different from Col-0 in blue light or dark conditions, but were significantly shorter in red light, see Figure 5-30 and Figure 5-31, suggesting the At5g45240-2 line has a shorter hypocotyl than Col-0 in red light conditions.

The At5g45240-3 plants did not differ in hypocotyl length from Col-0 when grown in red light, blue light or dark conditions. The At5g45240-4 line had a significantly shorter hypocotyl than Col-0 in red light and one of two blue light experiments. In the dark treatment, this At5g45240-4 line had a significantly shorter hypocotyl in one of three experimental repeats. This suggested that this line may have a shorter hypocotyl than Col-0 in red light, however further repeats would be necessary to confirm this, given the variation in hypocotyl length for this line in blue and dark treatments. Nonetheless, these results suggested that despite so far not identifying a leaf shape effect for the At5g45240-4 line, this T-DNA line appears to have a difference in hypocotyl length from Col-0.

We found the PhyB-9 mutant line had an elongated hypocotyl in only red light, consistent with the phenotypes described for PhyB mutants in the literature (Faigón-Soverna et al., 2006), see Figure 5-30 and Figure 5-31. The At5g45240-1 PhyB-9 double mutant also showed

an elongated hypocotyl in red light, and hypocotyl length of this line was not significantly different from PhyB-9. This showed that in red light, where PhyB-9 has a significantly longer hypocotyl than Col-0, the presence of an At5g45240-1 insertion does not significantly decrease hypocotyl length in a PhyB-9 background, despite a significant decrease in the At5g45240-1 single mutant. In one of two blue light experiments, and one of three dark experiments, the At5g45240-1 PhyB-9 double mutant had a significantly shorter hypocotyl than Col-0, though the PhyB-9 single mutant did not, see Figure 5-29 and Figure 5-32. This suggests there may be an effect of the At5g45240-1 insertion in the At5g45240-1 PhyB-9 double mutant in the light treatments other than red.

Hypocotyl length of At1g72850-1 plants was significantly shorter than Col-0 in one experimental repeat for each of the red, blue and dark treatments. This lack of consistency between experimental repeats prevents us concluding and describing hypocotyl difference for this line across light conditions.

Interestingly, hypocotyl length for the At5g45240-1 At1g72850-1 double mutant was not significantly different from wild type plants across all treatments in all but one experiment; a single dark treatment experiment, in which this double mutant had a significantly higher hypocotyl length than Col-0, At5g45240-1 and At1g72850-1, see Figure 5-34. This was unexpected, as the At5g45240-1 line had a consistently shorter hypocotyl length than Col-0 across blue and red light treatments, and although the hypocotyl length of the At1g72850-1 line was inconsistent across treatments, plants of this line were only either significantly shorter than Col-0, or a similar length. This result suggested that the red and blue treatment short hypocotyl of the At5g45240-1 insertion was negated in the At5g45240-1 At1g72850-1 double mutant.

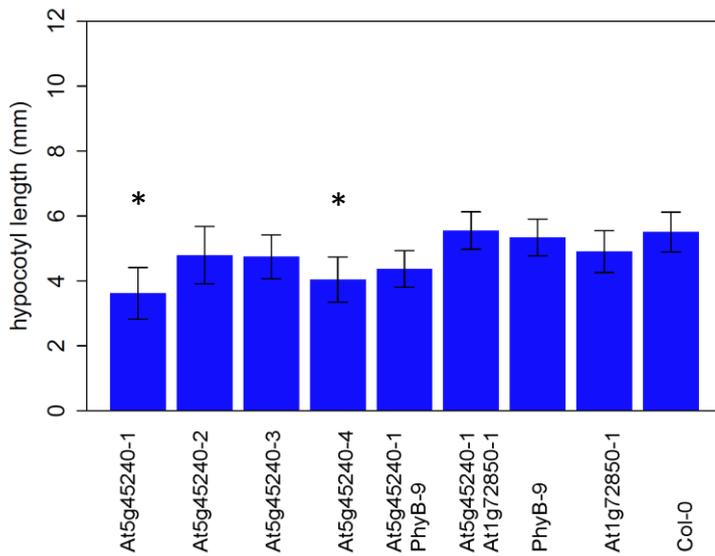


Figure 5-28 Hypocotyl growth in blue light

Figure shows hypocotyl lengths of seedlings grown in blue light. A significant difference was identified between the lines with an ANOVA, and lines significantly different from Col-0 in a Tukey post doc test are marked with an asterisk. Both of these lines had a lower hypocotyl length than Col-0. Error bars show 95% confidence intervals.

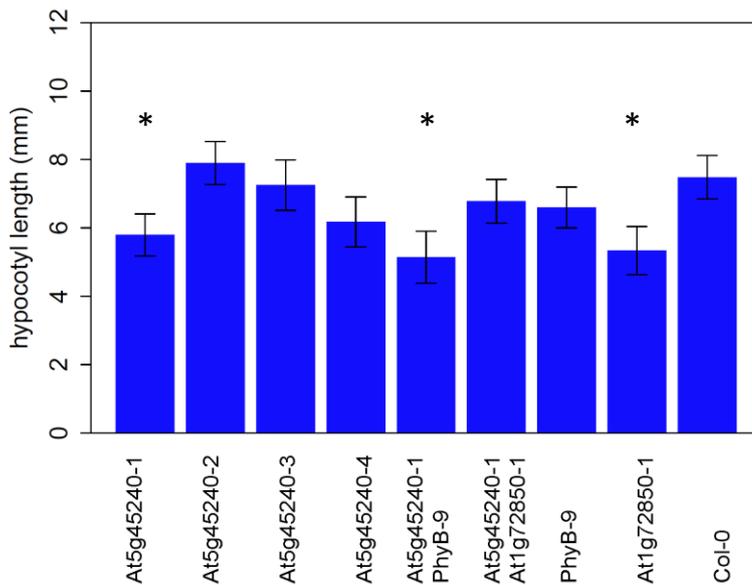


Figure 5-29 Hypocotyl growth in blue light

Figure shows hypocotyl lengths of seedlings grown in blue light. A significant difference was identified between the lines with an ANOVA, and lines significantly different from Col-0 in a Tukey post doc test are marked with an asterisk. Each of the latter three lines had a lower hypocotyl length than Col-0. Error bars show 95% confidence intervals.

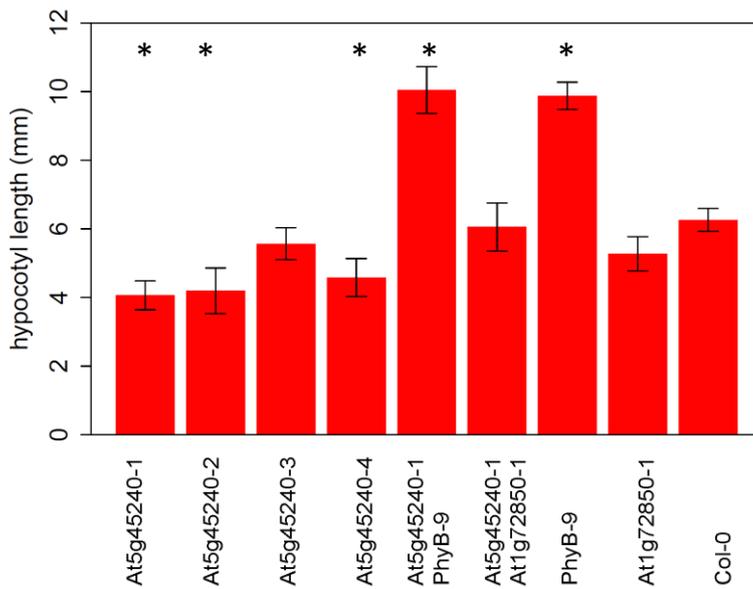


Figure 5-30 Hypocotyl growth in red light

Figure shows hypocotyl lengths of seedlings grown in red light. A significant difference was identified between the lines with an ANOVA, and lines significantly different from Col-0 in a Tukey post doc test are marked with an asterisk. Error bars show 95% confidence intervals.

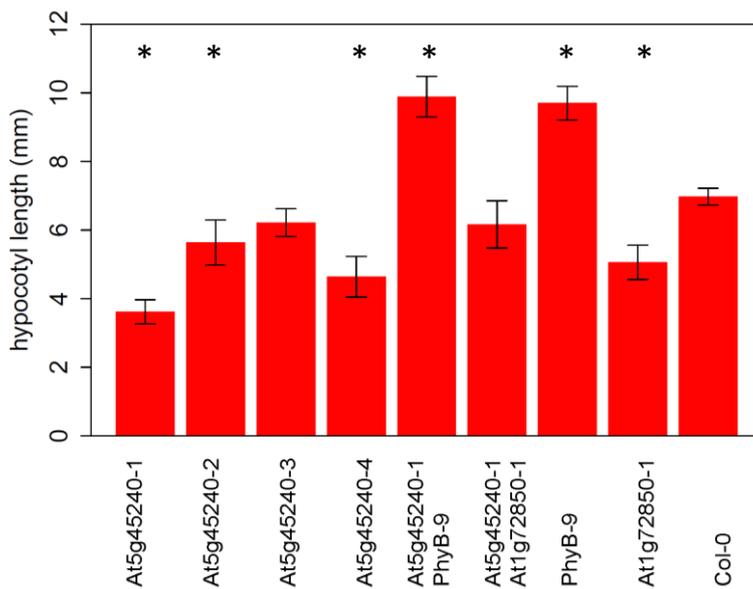


Figure 5-31 Hypocotyl growth in red light

Figure shows hypocotyl lengths of seedlings grown in red light. A significant difference was identified between the lines with an ANOVA, and lines significantly different from Col-0 in a Tukey post doc test are marked with an asterisk. Error bars show 95% confidence intervals.

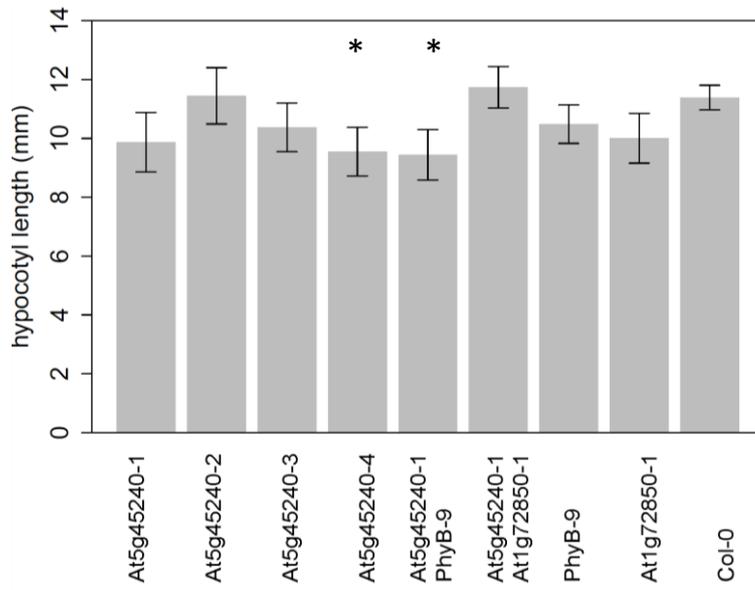


Figure 5-32 Hypocotyl length in dark grown seedlings
 Figure shows hypocotyl lengths of seedlings grown in dark conditions. A significant difference was identified between the lines with an ANOVA, and lines significantly different from Col-0 in a Tukey post doc test are marked with an asterisk. Asterisks indicate lines significantly different from Col-0. Error bars show 95% confidence intervals.

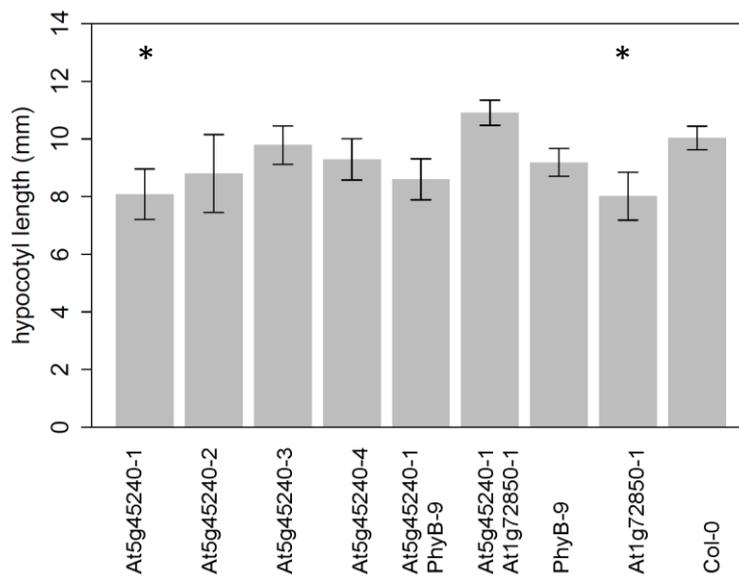


Figure 5-33 Hypocotyl length in dark grown seedlings

Figure shows hypocotyl lengths of seedlings grown in dark conditions. A significant difference was identified between the lines with an ANOVA, and lines significantly different from Col-0 in a Tukey post doc test are marked with an asterisk. Error bars show 95% confidence intervals.

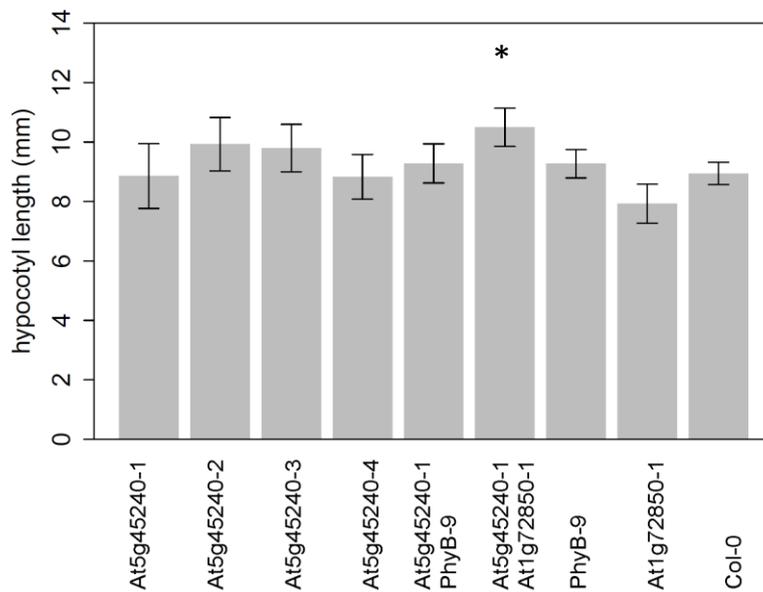


Figure 5-34 Hypocotyl length in dark grown seedlings

Figure shows hypocotyl lengths of seedlings grown in dark conditions. A significant difference was identified between the lines with an ANOVA, and lines significantly different from Col-0 in a Tukey post doc test are marked with an asterisk. Error bars show 95% confidence intervals.

5.2.10 Leaf shape of some T-DNA lines for NBLRR genes vary between temperatures

We were interested to see whether the leaf shape differences we had identified amongst T-DNA lines for some NBLRR genes varied in response to temperature. We also wondered if some of the other 18 T-DNA lines for NBLRR candidate genes, shown in Table 5-1, that had not shown a difference in leaf shape in greenhouse conditions, would differ in shape when grown in cooler or warmer temperatures.

We grew plants in a growth cabinet over July and August 2015 at two temperatures; 16°C and 24°C. In both treatments we grew six plants for each of the 18 T-DNA insertion lines we had previously scored for leaf shape, to see if in any of these lines showed differences in leaf shape when grown in colder conditions. For the lines we had already identified with differences in leaf shape, and the Col-0 background, we grew twelve plants per line. We were interested to see how the leaf shape of these lines might vary between temperature conditions.

We also included the At5g45240-3, At5g45240-4 and the At5g45240-1 At1g72850 double mutant. We were interested as to whether the lines At5g45240-3 and At5g45240-4, despite not showing a difference in leaf shape when grown in greenhouse conditions, would show differences in shape in cooler or warmer temperatures. We had not yet measured leaf shape of the At5g45240-1 At1g72850-1 double mutant, and were curious as to whether this line would respond similarly to differences in temperature as the At5g45240-1 and At1g72850-1 single mutants.

As time was limited at this point in the project, we recorded an image of the rosette of each plant and used visual inspection to compare leaf shape amongst the 18 T-DNA lines we grew for the NBLRR candidate genes that had not previously shown differences in leaf shape. No new differences in leaf shape to Col-0 in either 16°C or 24°C were observed. This suggested these lines did not differ in leaf shape in our standard greenhouse conditions or in these temperature conditions.

For the four T-DNA lines annotated for At5g45240, the T-DNA lines At1g72850-1, the At5g45240-1 At1g72850-1 double mutant, and the T-DNA lines At4g09420-1 and At2g17050-1 lines annotated for homologues of At1g72850 and At5g45240, and Col-0, we harvested each rosette leaf per plant grown at bolting in both 16°C and 24°C treatments. We recorded the shape of each rosette leaf using LeafAnalyser (Weight et al., 2008) and scored differences in shape with the leaf library pPCs.

Of T-DNA lines for the At5g45240 gene, the At5g45240-3 and At5g45240-4 lines did not differ from Col-0 in leaf shape or size when grown in either temperature treatment. Interestingly, although At5g45240-1 and At5g45240-2 showed difference in leaf shape similar to those we had identified previously for these lines in greenhouse conditions, the extent of leaf shape and size differences varied between the temperature treatments.

At5g45240-1 line plants had smaller leaves with a lower *llpPC2* and *llpPC3* score compared to the Col-0 background in only the 24°C temperature treatment, see figures Figure 5-35, Figure 5-36 and Figure 5-37. At 16°C, At5g45240-1 plants did not show any differences in leaf shape and size from Col-0. At5g45240-2 had smaller leaves than Col-0 in both 16°C and 24°C, though the difference was greater in 16°C, see Figure 5-38. The *llpPC3* score for this line was also higher than Col-0 in 16°C but not 24°C. Interestingly, for this line, it appeared that differences in leaf shape and size to Col-0 were stronger in the cooler temperature, whereas the leaf shape differences of the At5g45240-1 line were only present in the warmer temperature treatment.

We were curious as to whether the leaf shape of the At1g72850-1 line would differ from Col-0 in either of the temperature treatments. In some of our growth experiments, this line had a higher *llpPC2* score than Col-0, though this was not the case in all growth experiments. When comparing the leaf shape of this line to Col-0 in our two temperature conditions, we found plants of the A1g72850-1 line had a higher *llpPC2* score than the Col-0 background in only the 24°C treatment, see Figure 5-42. As such it may be that the leaf shape difference for the At1g72850-1 line is specific to higher temperatures.

Plants of the At4g09420-1 line had smaller leaves with a lower *llpPC2* and *llpPC3* score than Col-0 plants when grown at 16°C, see Figure 5-39, Figure 5-40 and Figure 5-41. There were no differences between this line and Col-0 in 24°C, suggesting the leaf shape differences of the At4g09420-1 line are specific to the colder temperature treatment.

Leaf shape and size differences between the T-DNA insertions lines for At5g45240, At1g72850 and At4g09420 did not show a common response to the temperature treatments. For example, lines At5g45240-1 and At1g72850-1 were distinct in leaf shape from Col-0 in the 24°C treatment, but not at 16°C. In contrast, the leaf shape differences observed for the At4g09420-1 line appeared only in the 16°C treatment. The shape and size differences of leaves from At5g45240-2 line plants was more distinct than those of Col-0 in the 16°C treatment also, though there was a difference in size and shape for this line relative to Col-0 at 24°C.

We analysed the leaf shape and size of the At5g45240 At1g72840 double mutant in both 16°C and 24°C temperatures, however we saw no effect on leaf shape in either treatment. This was somewhat surprising given that the lines At5g45240-1 and At1g72850-1 used to create this line had differences in leaf shape and size relative to Col-0 in 24°C conditions; relative to Col-0 the lines At5g45240-1 and At1g72850-1 had a decrease and increase in

llpPC2 respectively. It is possible that as both lines individually had an opposing effect on llpPC2 score the llpPC2 differences of each line were masked in the double mutant.

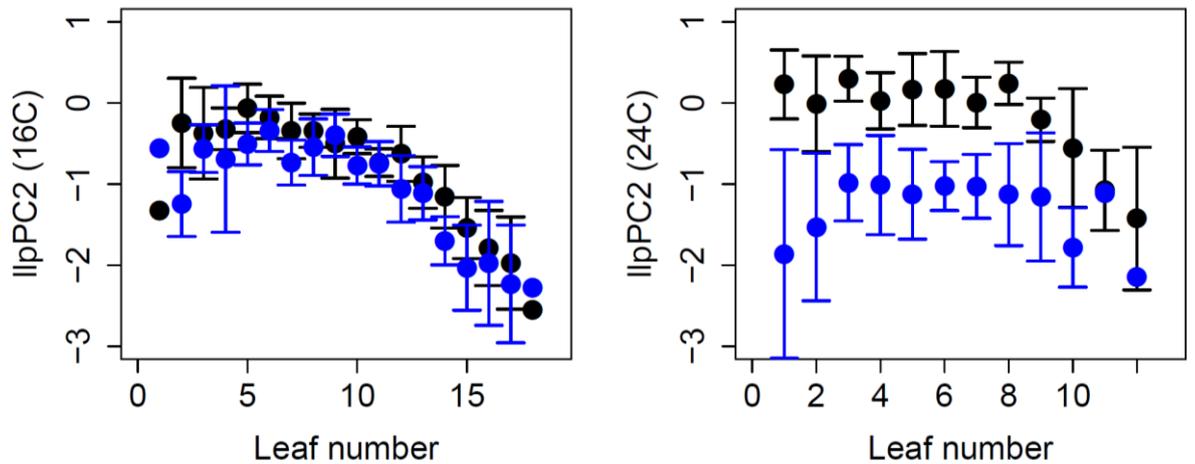


Figure 5-35 Difference in llpPC2 between Col-0 and At5g45240-1 at two temperatures
 Figures showing the mean llpPC2 score for Col-0 and At5g45240-1 across the leaf series in two temperature treatments. Values for Col-0 are shown as black circles, values for At5g45240-1 are shown as blue circles. For the left figure, plants were grown at 16°C. For the right figure, plants were grown at 24°C. 12 plants were harvested for each line. Error bars show 95% confidence intervals.

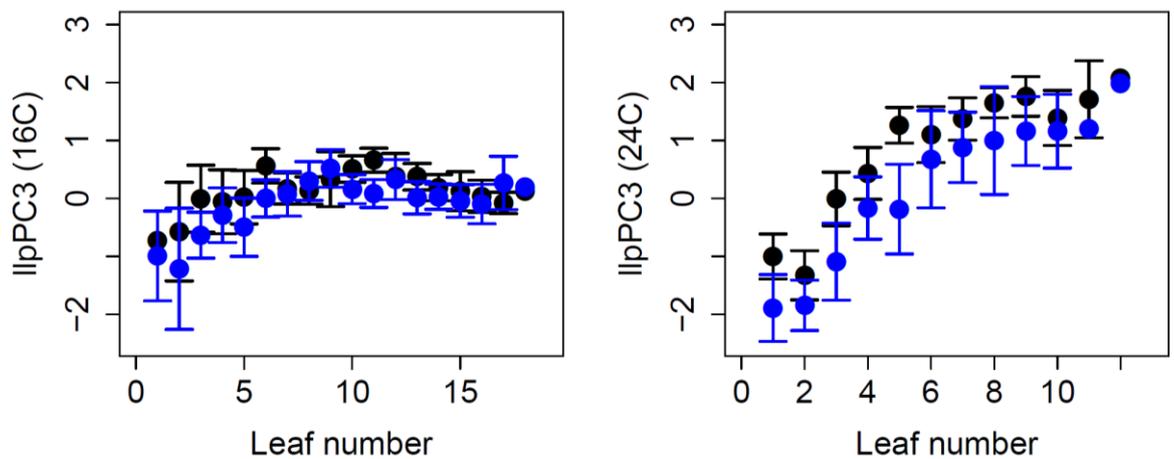


Figure 5-36 Difference in llpPC3 between Col-0 and At5g45240-1 at two temperatures
 Figures showing the mean llpPC3 score for Col-0 and At5g45240-1 across the leaf series in two temperature treatments. Values for Col-0 are shown as black circles, values for At5g45240-1 are shown as blue circles. For the left figure, plants were grown at 16°C. For the right figure, plants were grown at 24°C. 12 plants were harvested for each line. Error bars show 95% confidence intervals.

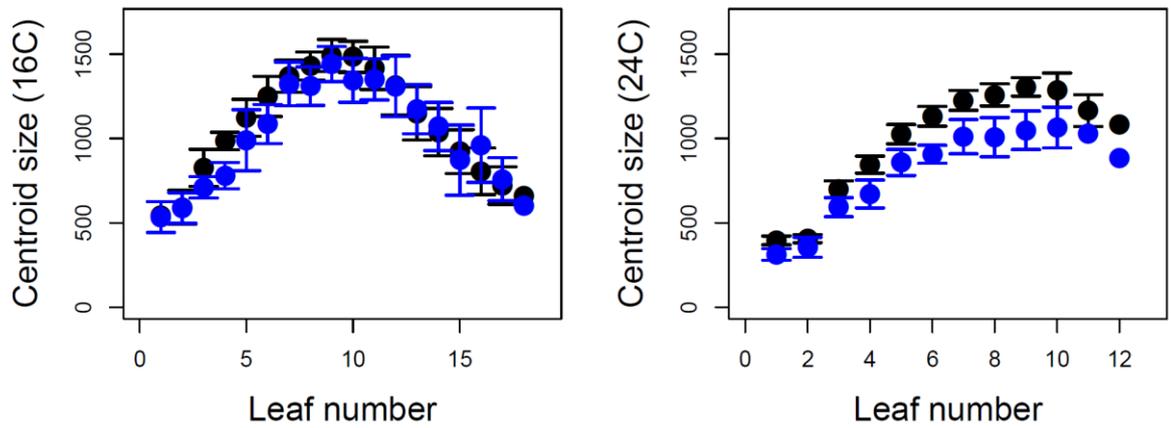


Figure 5-37 Centroid size difference between Col-0 and At5g45240-1 at two temperatures
 Figures showing the mean centroid size for Col-0 and At5g45240-1 across the leaf series in two temperature treatments. Values for Col-0 are shown as black circles, values for At5g45240-1 are shown as blue circles. For the left figure, plants were grown at 16°C. For the right figure, plants were grown at 24°C. 12 plants were harvested for each line. Error bars show 95% confidence intervals.

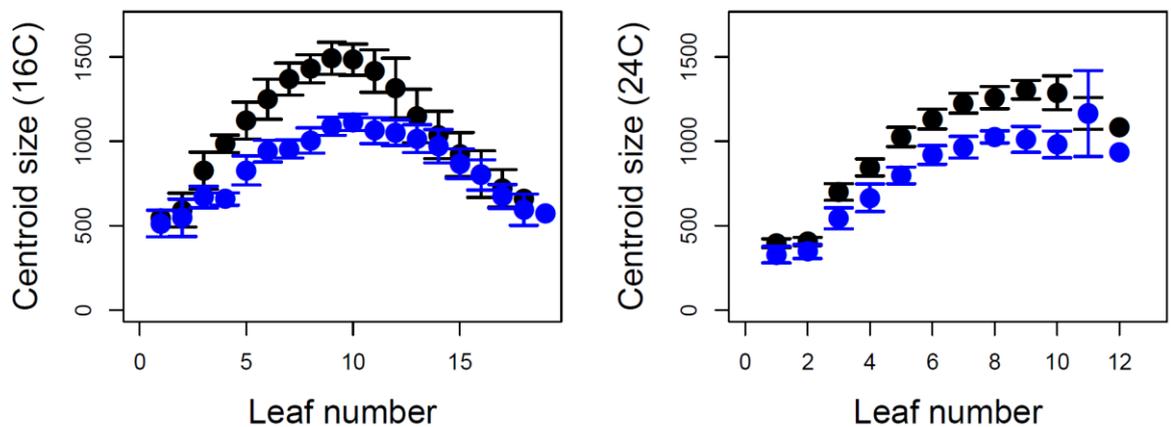


Figure 5-38 Centroid size difference between Col-0 and At5g45240-2 at two temperatures
 Figures showing the mean centroid size for Col-0 and At5g45240-2 across the leaf series in two temperature treatments. Values for Col-0 are shown as black circles, values for At5g45240-2 are shown as blue circles. For the left figure, plants were grown at 16°C. For the right figure, plants were grown at 24°C. 12 plants were harvested for each line. Error bars show 95% confidence intervals.

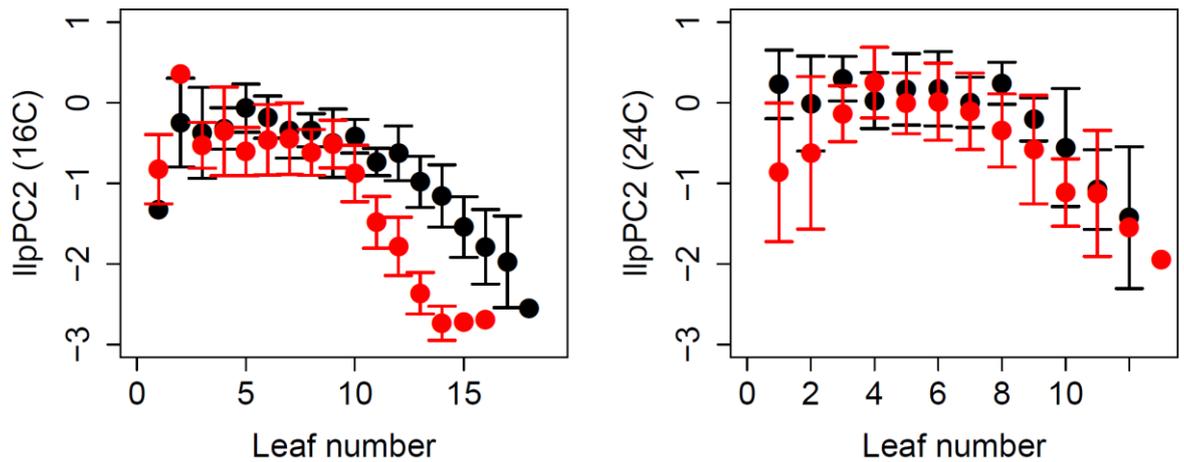


Figure 5-39 Difference in lIpPC2 between Col-0 and At4g09420-1 at two temperatures
 Figures showing the mean lIpPC2 score for Col-0 and At4g09420-1 across the leaf series in two temperature treatments. Values for Col-0 are shown as black circles, values for At4g09420-1 are shown as red circles. For the left figure, plants were grown at 16°C. For the right figure, plants were grown at 24°C. 12 plants were harvested for each line. Error bars show 95% confidence intervals.

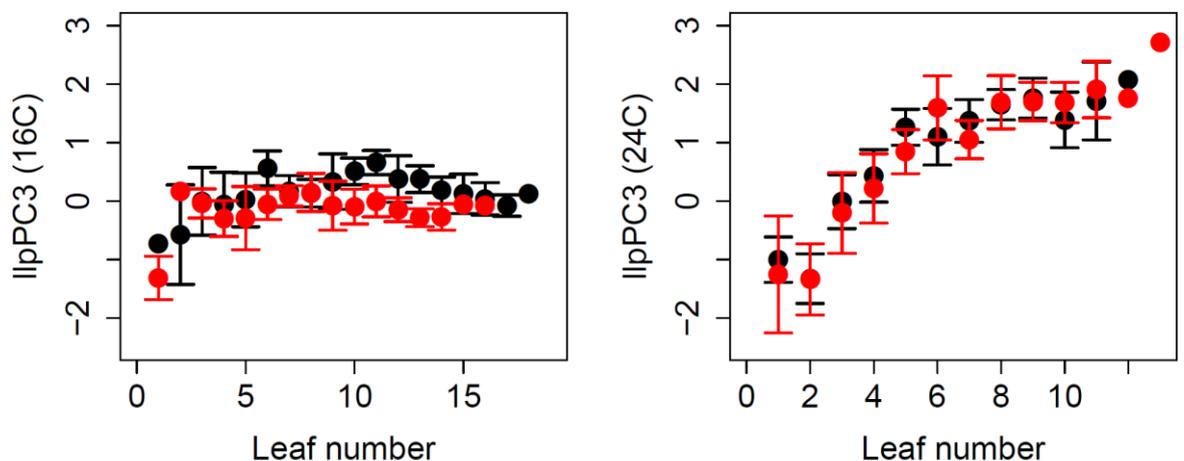


Figure 5-40 Difference in lIpPC3 between Col-0 and At4g09420-1 at two temperatures
 Figures showing the mean lIpPC3 score for Col-0 and At4g09420-1 across the leaf series in two temperature treatments. Values for Col-0 are shown as black circles, values for At4g09420-1 are shown as red circles. For the left figure, plants were grown at 16°C. For the right figure, plants were grown at 24°C. 12 plants were harvested for each line. Error bars show 95% confidence intervals.

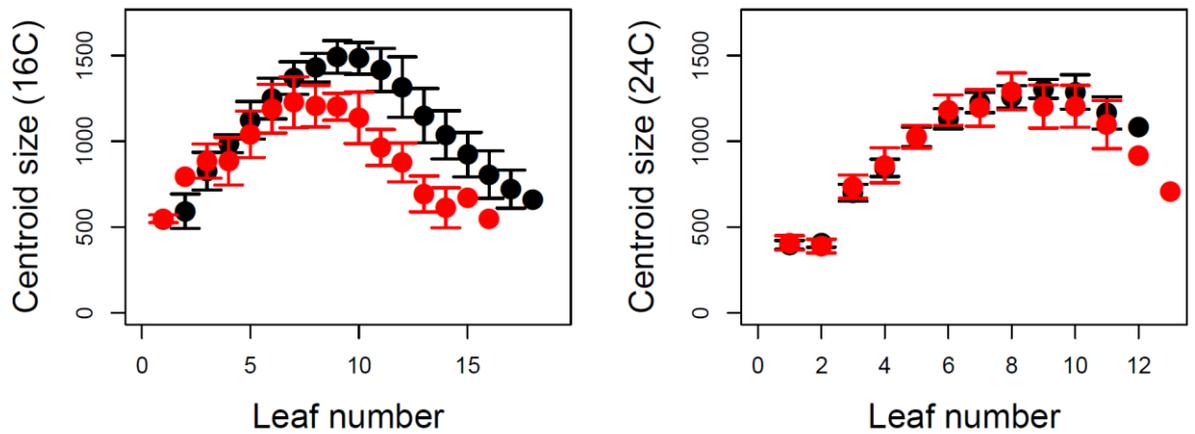


Figure 5-41 Size difference between Col-0 and At4g09420-1 at two temperatures

Figures showing the mean centroid size for Col-0 and At4g09420-1 across the leaf series in two temperature treatments. Values for Col-0 are shown as black circles, values for At4g09420-1 are shown as red circles. For the left figure, plants were grown at 16°C. For the right figure, plants were grown at 24°C. 12 plants were harvested for each line. Error bars show 95% confidence intervals.

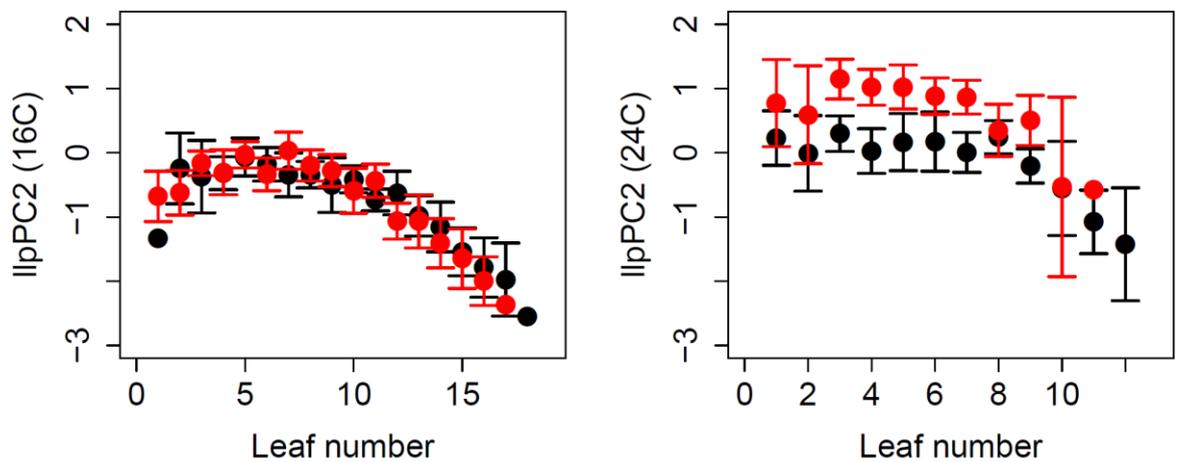


Figure 5-42 Difference in llpPC2 between Col-0 and At1g72850-1

Figures showing the mean llpPC2 score for Col-0 and At1g72850-1 across the leaf series in two temperature treatments. Values for Col-0 are shown as black circles, values for At1g72850-1 are shown as red circles. For the left figure, plants were grown at 16°C. For the right figure, plants were grown at 24°C. 12 plants were harvested for each line. Error bars show 95% confidence intervals.

5.2.11 Expression of At5g45240 in Col-0 and four T-DNA insertion lines

We wondered if the leaf shape differences between the four At5g45240 T-DNA lines related to variation in At5g45240 gene expression between the lines and used Reverse Transcriptase PCR (RT-PCR) to investigate transcript level across this gene. We grew Col-0 and the four T-DNA insertion lines of At5g45240 in the University of York greenhouses and harvested the rosette leaves of bolting plants to prepare cDNA. We visually inspected the leaves of these plants and confirmed the leaf shape and size differences we had characterised the lines for were present in this growth experiment.

We chose to prepare cDNA from bolting plants as the differences in leaf shape and size between lines are most apparent at this stage. However, it is possible that harvesting tissue at an earlier stage in the growth and development of the leaves of these plants would be more informative in relating gene expression to leaf shape differences between lines. Certainly, further work using cDNA prepared from tissue taken at different stages of plant growth and development is required before any relationship between gene expression of At5g45240 and the leaf shape differences between the four T-DNA lines for this gene can be concluded on.

In preliminary work we amplified a region from the 5th to 8th exon, see Figure 5-43. A shorter band was produced using Col-0 cDNA than with genomic DNA, though the cDNA band was larger than predicted from the TAIR10 At5g45240 gene annotation, see Figure 5-43 and Table 5-3. Another band of similar size to that produced for genomic DNA, was also produced with cDNA, despite no indication of genomic contamination in Actin 2 controls. This preliminary result suggested there may be differences in transcript length and splicing pattern to the TAIR10 annotation for this gene.

We were interested as to whether band sizes of other amplified regions of At5g45240 would match predictions with the TAIR10 gene annotation, and whether expression of At5g45240 varied between the T-DNA lines for this gene. We designed consecutive exon spanning primers across At5g45240 to test this, see Figure 5-43. Many of the exon spanning primers produced bands of different sizes to those predicted with the At5g45240 TAIR10 annotation. Only three of the nine exon spanning primer pairs produced a band matching the expected transcript size, see Table 5-3. In seven of these nine primer pairs, a band the same size as the genomic band was identified, despite no indication of genomic contamination in the Actin 2 control for any of the cDNA preparations.

Primers spanning the 1st and 2nd exons just downstream of At5g45240-2 T-DNA insertion annotation, see Figure 5-43, did not produce a transcript of the predicted size with cDNA for At5g45240-2, but did with cDNA of the other three T-DNA lines and Col-0, see Figure 5-44. The At5g45240-3 line is annotated for an insertion beginning in the 2nd intron of the gene, see Figure 5-43. Primers for this region, spanning from the 2nd to 3rd exons were used with cDNA from each line, however only a band for Col-0 was detected, see Figure 5-45. The T-DNA insertion annotation for the At5g45240-4 line begins within the 4th intron of At5g45240. We used two different pairs of primers to try to amplify this region, however neither pair produced bands using cDNA from Col-0 or any of the T-DNA lines. A primer pair amplifying a region from the 5th to 6th exon did produce a band for At5g45240-4, see Figure 5-47. The T-DNA insertion annotation for At5g45240-1 begins in the 9th exon. Primers spanning the 8th to 9th and 9th to 10th exons produce a band for all lines bar At5g45240-1, see Figure 5-48 and Figure 5-49. Of the four T-DNA insertion lines only the At5g45240-1 line showed absence of expression downstream of the annotated T-DNA insertion site.

Our results suggest there may be an effect of the T-DNA insertions in these four lines on the expression of regions of At5g45240, however further work is required to fully understand the exon intron structure of this gene before the effect of these T-DNA insertions can be more fully explained, as this may differ from the TAIR annotation for this gene.

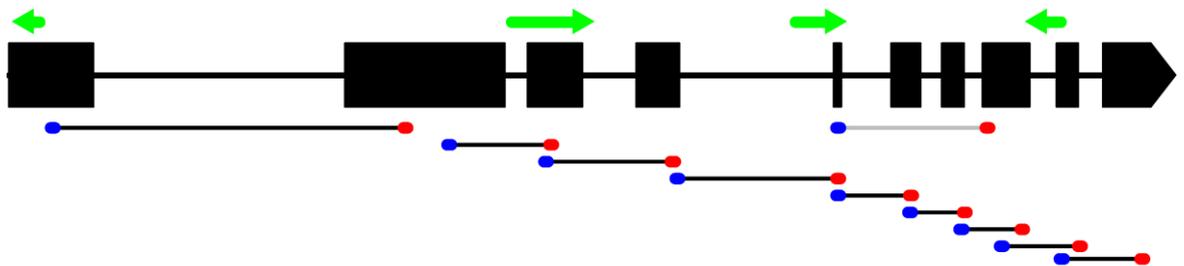


Figure 5-43 At5g45240 gene model

Diagram of At5g45240 gene model, using TAIR10 annotation, from 5' to 3'. Exons of the gene are indicated by black squares and introns by a thin black line. The regions amplified by exon spanning primers are shown by black and grey horizontal lines. The left and right primer positions are shown as blue and red dots respectively. The position of the T-DNA insertion start points as listed on SALK website for the T-DNA lines for At5g45240 are shown as green arrows. The order of insertions for each line, from left to right are At5g45240-2 starting in the 1st exon, At5g45240-3 in the 2nd intron, At5g45240-4 in the 4th intron, and in the 9th exon, At5g45240-1.

Primers	Expected transcript length (bp)	Expected genomic length (bp)	cCol-0	cAt5g45240-1	cAt5g45240-2	cAt5g45240-3	cAt5g45240-4	gCol-0
Exons1-2	497	1703	500	500, 200	200	500, 200	500	1750
Exons2-3	394	503	500	blank	blank	blank	blank	500
Exons3-4	364	622	625	625	625	625, 600, 375	625	625
Exons5-6	129	370	350, 150	blank	150	150	350, 150	350
Exons6-7	159	263	300	300	300	300	300	300
Exons7-8	200	295	300	300	300	300	300	300
Exons8-9	243	374	400	blank	400	400	400	400
Exons9-10	266	384	400	blank	400	400	400	400
Actin 2	494	580	500	500	500	500	500	600
Exons5-8	297	725	500, 650, 700	500, 650, 700	500	500	blank	700
Exons5-8	297	725	blank	500, 650, 700	500, 650, 700	500, 650	500, 650, 700	700

Table 5-3 Summary of RT-PCR results using exon spanning primers over at5g45240

Table lists the exon spanning primers used in RTPCR analysis. The expected band size, estimated using TAIR10 sequence is listed along with the sizes of all observed bands for each set of primers. Columns for samples of cDNA are indicated by a c, for example cCol-0, and for samples of genomic DNA columns are indicated by a g, for example gCol-0.

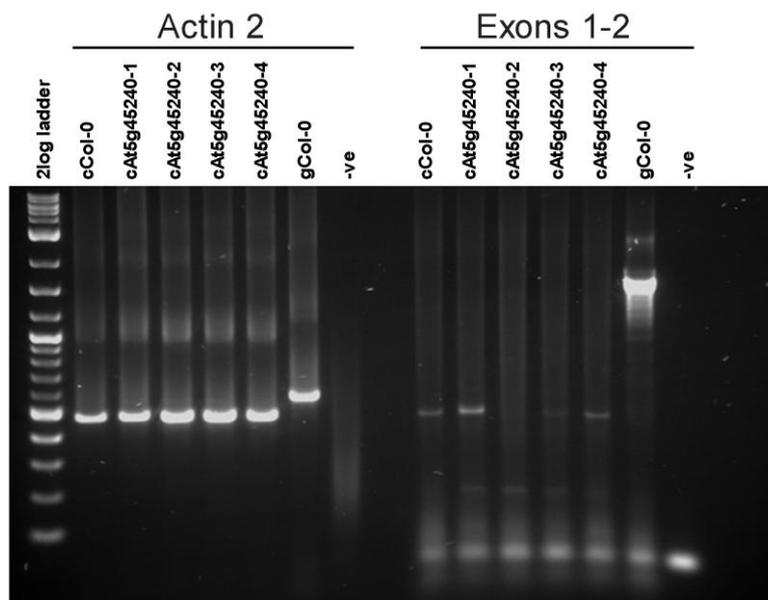


Figure 5-44 Transcript levels between exons 1-2
Figure shows RTPCR results for cDNA samples from T-DNA insertion lines in At5g45240. Presence of transcript was tested with primers covering Actin 2 and exons 1 to 2 of At5g45240. cDNA from Col-0 was used as comparison, and genomic Col-0 DNA and a reaction with no DNA were used as control.

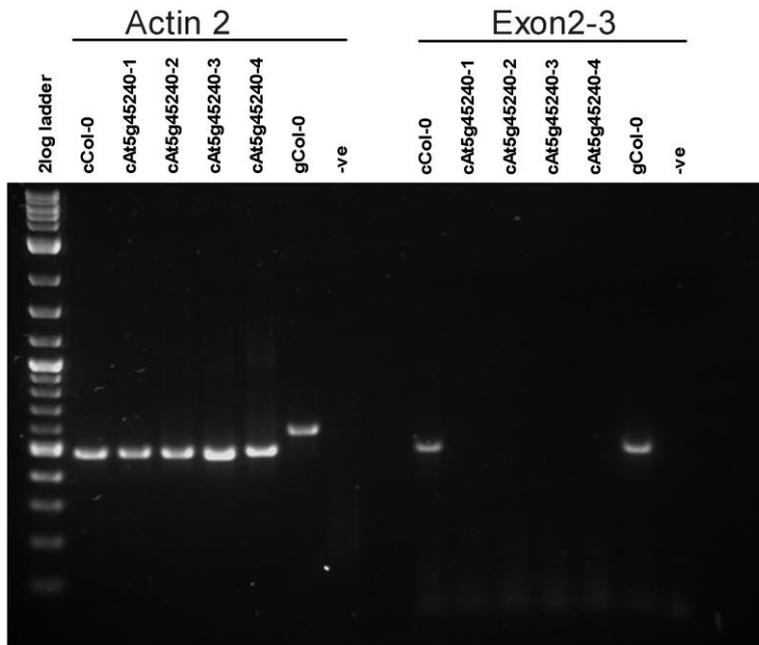


Figure 5-45 Transcript levels between exons 2-3
 Figure shows RTPCR results for cDNA samples from T-DNA insertion lines in At5g45240. Presence of transcript was tested with both primers covering Actin 2 and exons 2 to 3 of At5g45240. cDNA from Col-0 was used as comparison, and genomic Col-0 DNA and a reaction with no DNA were used as control.

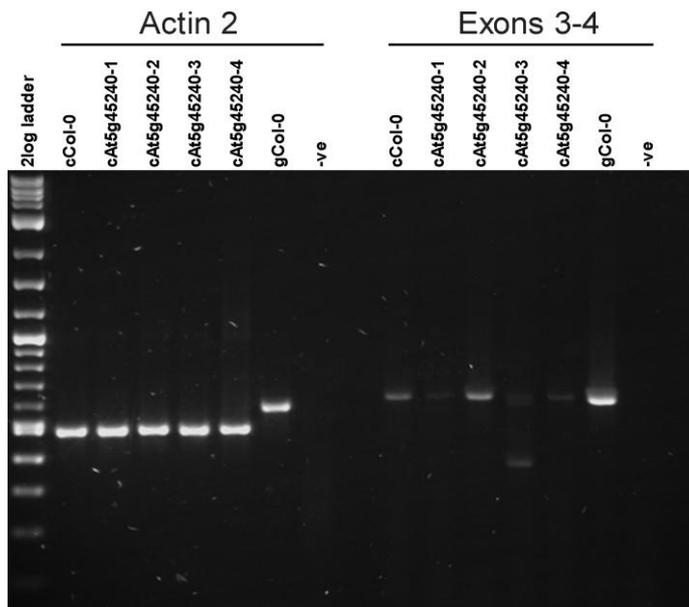


Figure 5-46 Transcript levels between exons 3-4
 Figure shows RTPCR results for cDNA samples from T-DNA insertion lines in At5g45240. Presence of transcript was tested with both primers covering Actin 2 and exons 3 to 4 of At5g45240. cDNA from Col-0 was used as comparison, and genomic Col-0 DNA and a reaction with no DNA were used as control.

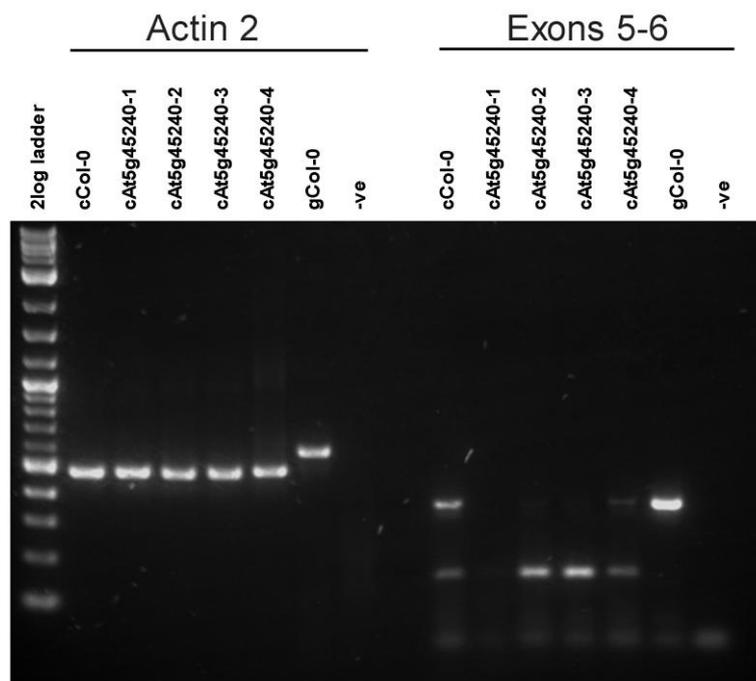


Figure 5-47 Transcript levels between exons 5-6
 Figure shows RTPCR results for cDNA samples from T-DNA insertion lines in At5g45240. Presence of transcript was tested with both primers covering Actin 2 and exons 5 to 6 of At5g45240. cDNA from Col-0 was used as comparison, and genomic Col-0 DNA and a reaction with no DNA were used as control.

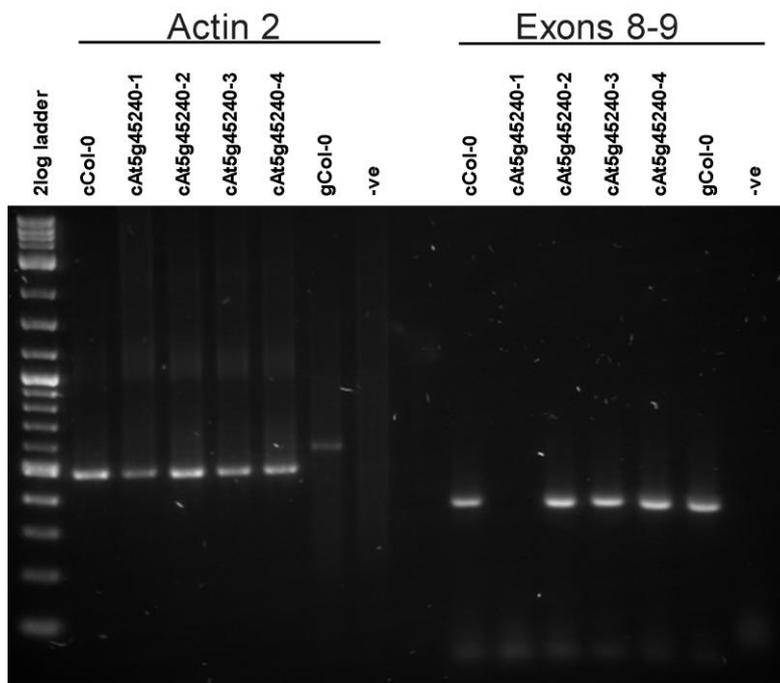


Figure 5-48 Transcript levels between exons 8-9
 Figure shows RTPCR results for cDNA samples from T-DNA insertion lines in At5g45240. Presence of transcript was tested with both primers covering Actin 2 and exons 8 to 9 of At5g45240. cDNA from Col-0 was used as comparison, and genomic Col-0 DNA and a reaction with no DNA were used as control.

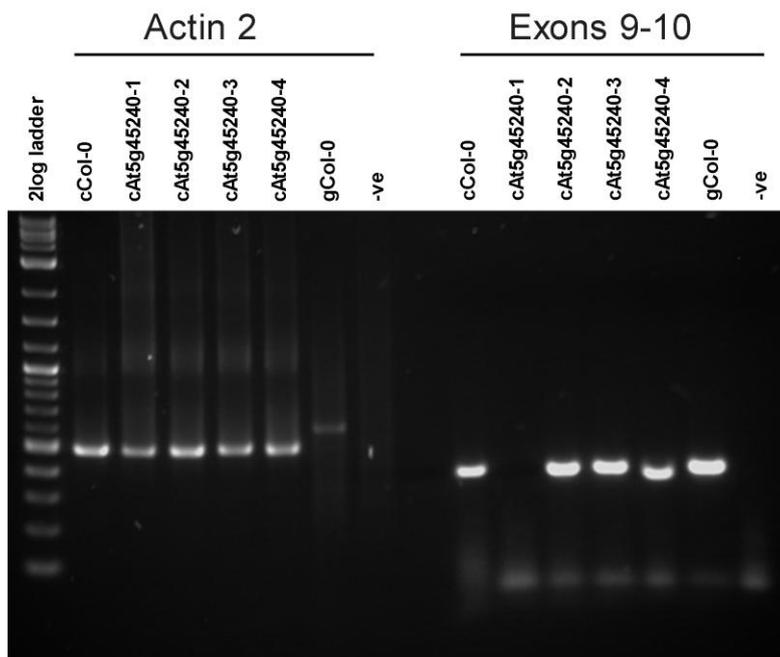


Figure 5-49 Transcript levels between exons 9-10
 Figure shows RTPCR results for cDNA samples from T-DNA insertion lines in At5g45240. Presence of transcript was tested with both primers covering Actin 2 and exons 9 to 10 of At5g45240. cDNA from Col-0 was used as comparison, and genomic Col-0 DNA and a reaction with no DNA were used as control.

5.3 Conclusions

5.3.1 Analysis of T-DNA insertion lines for NBLRR genes associated with leaf shape traits in a GWAS

We measured the leaf shape of a set of T-DNA lines annotated with insertions for the 14 NBLRR genes within four GWAS loci associated to our leaf shape traits. Although the 14 NBLRR genes we obtained T-DNA lines for were within loci associated with leaf shape traits in our GWAS work, we did not necessarily expect each of these genes to have a leaf shape function. As these 14 NBLRR genes lie within four clusters, it may be that some of these genes are associated with leaf shape variation in the natural accessions through linkage to polymorphisms in nearby NBLRR genes within the same cluster. Identifying mutant lines for NBLRR genes with differences in leaf shape would support the hypothesis that some NBLRR genes in these loci are associated with natural leaf shape variation. T-DNA insertion lines for NBLRR genes within two of four clusters associated with our leaf shape traits had differences in leaf shape compared to Col-0, these loci were on chromosome 1, at 27Mb and chromosome 5, at 18Mb. For NBLRR genes in the GWAS associated loci at 21Mb on chromosome 1 and 14Mb on chromosome 5, no T-DNA insertion lines had a leaf shape or size different to Col-0. In total we found leaf shape differences for T-DNA lines for three NBLRR genes when compared to Col-0.

SALK T-DNA insertions in genes often cause a decrease or absence of protein expression for a gene (Wang, 2008). Whilst a low proportion of the 18 T-DNA lines we examined had a difference in leaf shape or size, an alternative approach might reveal more evidence for the function of these genes. Other types of mutation, such as those generating overexpression phenotypes may provide better evidence of gene function. For example, a mutation causing overexpression of a truncated transcript of the *CSA1* NBLRR gene has a considerably stronger morphology phenotype compared to a mutation causing almost complete absence of expression (Faigón-Soverna et al., 2006). Given that NBLRR genes are typically expressed at low levels (Tan et al., 2007) and recessive mutations in NBLRR genes are considered rare (Sohn et al., 2014), it may be worthwhile to assess leaf shape amongst overexpression lines for these genes.

It is also possible that leaf shape differences amongst these T-DNA lines were but were not detectable in our growth conditions. Several studies have reported NBLRR gene phenotypes to be more severe, or present only in cooler temperatures (Alcázar et al., 2008; Bombliet et

al., 2007; Huang et al., 2010; Zhu et al., 2010). Although we initially scored the T-DNA lines for the 14 NBLRR genes in standard greenhouse conditions, visual inspection of these lines when grown at 16°C did not identify any previously undetected differences in leaf shape, suggesting these lines did not have an effect on leaf shape in cooler conditions either.

NBLRR genes often have tissue specific expression (Tan et al., 2007), and may have phenotypes we would not have detected in our leaf shape analysis, such as root growth (Kim et al., 2012). It is possible therefore that T-DNA insertions in NBLRR genes have an effect on the plants growth in traits we did not measure. Indeed the majority of work in the literature has characterised NBLRR genes with disease resistance phenotypes, something we did not measure. As such we do not conclude that T-DNA insertion lines for some of these genes have no difference in plant growth, but that they have no effect on leaf morphology when grown in our greenhouse conditions or in growth cabinets at 16°C and 24°C.

5.3.2 Leaf shape and size vary between different T-DNA insertion lines for At5g45240

We characterised the leaf shape of four T-DNA insertion lines for the gene At5g45240. T-DNA insertions At5g45240-1 and At5g45240-2 were associated with differences in leaf shape and size, however no difference in leaf shape was found for lines At5g45240-3 and At5g45240-4. Variable phenotypes across T-DNA insertion lines for a single gene have been previously reported in the literature (Valentine et al., 2012). Valentine et al., (2012) scored differences in growth between seven and six T-DNA insertion lines for two different genes, finding considerable variability in traits such as fruit number and biomass between the lines, and some evidence of intron or exon specific effects on leaf number and germination rate. We aimed to investigate the T-DNA insertion configuration and effect on gene expression across the four T-DNA insertion lines of At5g45240 to provide a possible link between the specific properties of each mutation and the leaf shape variation between the T-DNA lines.

The annotated positions for the T-DNA insertions in the At5g45240-3 and At5g45240-4 lines are the 2nd and 4th introns respectively and for the At5g45240-1 and At5g45240-2 line insertions the annotated positions are 9th and 1st exons respectively. PCR genotyping of these insertions suggested At5g45240-1 and At5g45240-2 had single insertions in the expected orientation, for At5g45240-3 results suggested two insertions may be in a back to back configuration with one T-DNA in a reverse orientation. For At5g45240-4, results suggested a more complex configuration of insertions. As both lines with differences in leaf shape appeared to have single insertions beginning in exons of the gene, and both lines

without effects on leaf shape had more complex configurations of insertions beginning in introns, it is possible that site and configuration of insertion is causing the presence or absence of effects on leaf shape between the four lines. However more work is required to explore this fully, such as sequencing the bands produced when genotyping these insertions to confirm the exact position of each insertion in these lines. The position of a SALK T-DNA insertion has been found to differ from the annotated position in work in the literature (Rodríguez et al., 2014), and so until the position of the At5g45240 T-DNA insertions is verified, we cannot be sure that each insertion in the T-DNA lines of At5g45240 is within either an exon or intron, though our PCR genotyping of the insertions suggests they are not more than 100bp from the described position.

We used RTPCR to examine transcript level across At5g45240 in each of the four T-DNA lines and the Col-0 background. We found that large amounts of prepared cDNA, 1 µl – 1.5 µl and 35 PCR cycles were required to identify a clear band for cDNA of Col-0, suggesting transcript for At5g45240 was present at low levels. Transcript was amplified for each of the four T-DNA lines across most exons of the gene, suggesting that none of the insertions prevented expression completely. Band size and presence varied differently for each of the T-DNA lines for At5g45240, and so we were unable to draw any clear link between presence of leaf shape difference and transcript. It is possible that the absence of a 500bp band with the At5g45240-2 line for primers spanning the 1st and 2nd exons, and the absence of a band for the At5g45240-1 line with primers spanning the 8th to 9th and 9th to 10th exons relates to the leaf shape of each line, as these transcript profiles are unique to these lines and are likely the result of the position of the T-DNA insertions. However, further work testing the effect of varying levels of transcript at these regions would be required before the leaf shape of these lines can be attributed to absence of transcript between these exons. As such, whilst our analysis of insertion configuration and transcript expression level for the T-DNA alleles of At5g45240 is informative, we are unable to conclude as to how differences in expression between the four T-DNA lines correspond to the differences in leaf shape between these lines. Further work could use transgenic approaches to introduce regions of transcript absent in At5g45240-1 and At5g45240-2 to see if this alleviates the leaf shape differences observed for these lines.

Band sizes produced for many of the primers we used to amplify regions of At5g45240 were different from those predicted using the TAIR10 annotation of At5g45240. This is not unexpected given that alternative splicing has been reported for a number of NBLRR genes (Tan et al., 2007). Indeed, alternative splice variants have been shown to be induced by, and

are necessary for, disease response through the *RPS4* NBLRR gene (Zhang and Gassmann, 2003, 2007). Given the prevalence of bands produced at the same size as the genomic control, and the absence of any genomic contamination identified with the Actin 2 primers, it may be that many of the predicted introns of At5g45240 are not spliced out in the cDNA samples we used. Although quantitative analysis of band strength could be used to inform suggestions as to effect of the T-DNA insertions on transcript abundance, the deviation from predicted band sizes, high number of PCR cycles used and absence of visually clear variation in transcript abundance discouraged us from taking this approach. Instead we focused on the size of and presence or absence of bands detected. Further work is required to identify the alternative transcripts produced for this gene before the effect of the T-DNA insertions in transcript level can be fully understood. In particular the structure and number of exons and introns may be different from the reference genome annotation for At5g45240. Further knowledge of gene structure could provide greater understanding as to the effect of the T-DNA insertions in the four T-DNA lines for At5g45240.

One possibility we did not investigate is that the leaf shape differences observed for the T-DNA lines were caused by an effect of the T-DNA insertion on neighbouring genes, instead of or as well as an effect on the gene insertions were within. Although possible, we thought this relatively unlikely as we had analysed leaf shape of T-DNA insertion lines for neighbouring NBLRR genes within each cluster and not observed differences. This reduces the likelihood that the leaf shape differences observed are the result of effects on neighbouring NBLRR genes, although we did not examine T-DNA lines for any neighbouring non NBLRR genes. Analysis of expression levels of neighbouring genes of At5g45240 across the four T-DNA insertion lines for this gene would be worthwhile to carry out to further establish perturbation of the At5t45240 gene as responsible for the observed leaf shape differences in the At5g45240-1 and At5g45240-2 T-DNA lines.

5.3.3 Effect of the At5g45240-1 T-DNA insertion in a At1g72850-1 and PhyB-9 background

Using the At5g45240-1 insertion we created two double mutants; At5g45240-1 At1g72850-1 and At5g45240-1 PhyB-9. We were interested in the possible interaction between the At5g45240-1 insertion and another mutation in a separate NBLRR gene, and also to see whether the At5g45240-1 insertion would alter leaf shape in a PhyB-9 background. A mutation in the NBLRR gene *CSA1* has no effect in a PhyB mutant background, despite a

strong morphology phenotype as a single mutant, suggesting this mutation interferes with PhyB signalling (Faigón-Soverna et al., 2006).

We measured the leaf shape of the At5g45240-1 PhyB-9 double mutant in greenhouse conditions and found the At5g45240-1 insertion did not decrease llpPC2 score in a PhyB-9 background, despite decreasing llpPC2 score relative to the Col-0 background as a single mutant. Surprisingly, the At5g45240-1 insertion actually increased llpPC2 score of leaves at two nodes in the double mutant. The At5g45240-1 insertion did not decrease hypocotyl length in the At5g45240-1 PhyB-9 double mutant when plants were grown in red light, despite a decrease in hypocotyl length in the At5g45240-1 T-DNA line. Although these results could be interpreted as the At5g45240-1 insertion having no effect in a PhyB-9 background in conditions where PhyB-9 is significantly different from wild type, it is also worth considering the hypocotyl length of At5g45240-1 lines in the dark treatment. In these conditions, At5g45240-1 lines only had a hypocotyl length shorter than Col-0 in one of three experimental repeats. This suggests that the effect of the At5g45240-1 insertion may be negated in strongly etiolating plants, as caused by the PhyB-9 mutation or the dark treatment, rather than the PhyB-9 background specifically. It would be interesting to explore this in further work, to see if the At5g45240-1 insertion had a leaf shape effect in dark grown plants for example.

In an At1g72850-1 background, the At5g45240-1 insertion did not decrease hypocotyl length relative to Col-0 in red or blue light, despite the short hypocotyl of the At5g45240-1 line in these conditions. Interestingly, in some repeats of these experiments, the At1g72850-1 line had a shorter hypocotyl than Col-0 also, and so despite the lines At5g45240-1 and At1g72850-1 having shorter hypocotyls in some experimental repeats, the double mutant had a wild type hypocotyl length. When measuring leaf shape for lines grown at 24°C, we found the At1g72850-1 line had an increased llpPC2 score compared to Col-0, and the At5g45240-1 line had a lower llpPC2 score. The double mutant with both insertions from these lines did not differ from Col-0 at 24°C or 16°C for any of our leaf shape and size traits.

The morphology of the At5g45240-1 At1g72850-1 double mutant cannot be entirely explained by the opposing direction of effects on llpPC2 score found in the At5g45240-1 and At1g72850-1 lines cancelling each other out. Although this fits with results for leaf shape traits in 24°C it does not seem appropriate when considering the hypocotyl data, where the At1g72850-1 line either did not differ in hypocotyl length to Col-0, or had an effect in the same direction as At5g45240-1, as the double mutant had a wild type phenotype. Instead it

appears that regardless of the morphology of the At1g72850-1 line, in combination, the At5g45240-1 and At1g72850-1 insertions produce a wild type morphology as a double mutant.

Epigenetic suppression of a T-DNA insertion mutant by another T-DNA insertion has been documented in Arabidopsis (Gao and Zhao, 2013). Although there are other explanations for the phenotypes we observed in our At5g45240-1 At1g72850-1 and At5g45240-1 PhyB-9 double mutants, we note that as we did not test the phenotype of each insertion after going through an F2 population to create the double mutant, this remains a possibility. As T-DNA insertions remained silenced over subsequent generations (Gao and Zhao, 2013), it may be worthwhile to outcross each double mutant insertion and verify the effect of each mutation individually to rule out epigenetic suppression of single mutant phenotypes.

5.3.4 Further characterisation of NBLRR genes with possible leaf shape function

Through analysis of T-DNA lines we had identified several NBLRR genes that may have a role in leaf shape. Within the time available to the project, we identified temperature and light specific morphologies for these lines, and created double mutants to further explore the leaf shape effects of these lines. However, several aspects of characterisation for these mutants were unable to be completed in the remaining time and are outlined here.

We tested the association of T-DNA insertion and leaf shape effect in a segregating F2 population for At4g45240-1, At4g45240-2 and At1g72840-1. At1g72840-1 was discarded from the project after the At1g72840-1 T-DNA insertion was not associated with the expected leaf shape difference in an F2 population segregating for the insertion. The insertions for At4g45240-1 and At4g45240-2 were associated with leaf shape and size differences in F2 populations. Although we identified differences for leaf shape in the At1g72850-1 and At4g09420-1 lines we did not have time to test for association of the insertions in these lines and leaf shape effect in segregating F2 populations. We found At1g72850-1 line leaves had an increased llpPC2 score when grown at a temperature of 24°C, and so rather than using standard greenhouse conditions, the growth cabinet environment at this temperature should provide suitable conditions for testing the effect of the At1g72850-1 insertion in a segregating F2 population. A difference in leaf shape was identified for the At4g09420-1 line in our standard greenhouse conditions as well as in a 16°C growth cabinet, and so either of these conditions would be suitable for testing association of the leaf shape effect we identified for this line and the At4g09420-1 insertion

in a segregating F2 population. Although the line At5g45240-4 did not show a difference in leaf shape in any of the conditions we tested, it appeared to have a shorter hypocotyl than Col-0 in the red light treatment. To verify this effect it would be important to show the short hypocotyl effect associated with the At5g45240-4 T-DNA insertion in a segregating F2 population grown in red light conditions.

We chose to measure hypocotyl length when investigating the effect of different light conditions on the morphology of T-DNA lines for NBLRR genes rather than leaf shape traits in part so that we could conduct repeated experiments quickly to determine an appropriate light intensity at which to grow the plants, and then conduct repeated experiments. Work in the literature on a NBLRR gene with a leaf shape phenotype also used hypocotyl measurements to identify a light specific phenotype (Faigón-Soverna et al., 2006), and so we reasoned this would be an appropriate start point for our work testing for light specific effects in T-DNA lines for NBLRR genes. Although in Faigón-Soverna et al., (2006), there is a correlation between hypocotyl length and leaves with a shade avoidance shape amongst the lines studied, and in our earlier work on the Bay-0 Shahdara RILs our leaf shape trait *llpPC2* was positively correlated with hypocotyl length, we cannot assume a link between hypocotyl length and leaf shape in the T-DNA lines for NBLRR genes we characterise here. As such, an interesting follow up experiment would be to test whether light conditions affecting hypocotyl length in these lines also had an effect on leaf shape. Work in the literature has identified a NBLRR gene phenotype specific to low humidity (Noutoshi et al., 2005). As such it may also be worthwhile to verify that the light specific effects on hypocotyl length we observed with seedlings grown on agar plates, and so at high humidity, persist when plants are grown in soil before measuring leaf shape for these lines in varying light conditions.

A further consideration for planning an analysis of leaf shape in different light conditions is the effect of temperature on the morphology of these T-DNA lines. Given that some of the lines had leaf shape differences specific to 24°C; such as At1g72850-1, and At5g45240-1, and some had shape effects specific to or stronger in 16°C; such as At5g45240-2 and At4g09420-1, it may be worthwhile conducting separate light condition experiments at temperatures permissive to leaf shape differences for each T-DNA line of interest.

During the project we also used complementation analysis to compare leaf shape in F1 plants created from natural accessions and either the At5g45240-1 line or the Col-0 background. Complementation analysis can be used to evaluate natural genetic variation with mutant analysis. Kobayashi et al., (2008) use complementation to investigate the *HMA5*

gene as a candidate for a root length QTL in a *Cvi* x *Ler* mapping population. These authors find that in F1 plants hybrid between a T-DNA insertion line for gene *HMA5* with a recessive short root trait, and two RIL lines with variant alleles at a root length QTL, the RIL with a *Ler* QTL allele is able to rescue the short root of the *HMA5* T-DNA line, whereas the RIL with a *Cvi* allele cannot, implicating *HMA5* as the gene responsible for this QTL (Kobayashi et al., 2008). We crossed Col-0 and the At5g45240-1 line to two groups of accessions with different alleles at the At5g45240 gene and leaf shape scores matching the estimated direction of effect of each allele in our GWAS10 work. Although results were inconclusive, and so not included in the results section of this chapter, we observed that the *lppPC2* score of F1 plants between an accession and the At5g45240-1 line was not always lower than for F1 plants created by crossing same accession to Col-0. In one instance, F1 plants created with the At5g45240-1 line and a natural accession had a higher *lppPC2* score than F1 plants created with that accession and Col-0. This suggested the effect of the At5g45240-1 insertion may vary in backgrounds of different natural accessions.

Our work characterising T-DNA lines for NBLRR genes has identified several differences in leaf shape and size and hypocotyl length specific to different light and temperature conditions. This suggests that some of the NBLRR genes associated with leaf shape traits in our GWA work may have a role in plant morphology in certain environmental conditions. Further work is required to fully characterise these lines and understand how each T-DNA insertion results in a leaf shape difference, but our work provides a solid foundation and strong starting point to understanding the roles of these NBLRR genes in plant morphology.

Chapter 6. Discussion

6.1.1 Natural leaf shape variation

There is great variation in shape amongst the leaves of flowering plants and leaves can vary dramatically between species. There are several major categories of different leaf shape and form amongst flowering plants. Species may have simple leaves such as birch trees or privet shrubs, lobed leaves such as oak leaves, compound leaves made up of multiple distinct laminae such as horse chestnut trees or tomato plants, or intricate and complex leaves, such as carrot leaves. This variety in leaf shape invites many questions as to the determination and role of natural leaf shape variation. Study of leaf shape can be approached from many aspects; such as to how leaves are formed developmentally, whether and to what conditions shape differences may be adaptive, and how the evolutionary lineage or more recent genetic variations relate to differences in leaf shape between species and accessions.

We were able to quantify leaf shape variation using a geometric morphometrics approach using LeafAnalyser (Weight et al., 2008). We could record leaf shape accurately using coordinates, and then identify the major shape variations with datasets of these co-ordinates, and score individual leaves for these shape variations, allowing us to make detailed and accurate measurements of leaf shape. In contrast to other work in the literature that uses absolute dimensions of leaves, such as width and length, as a proxy for leaf shape, using coordinate shape models allows the entire margin of a leaf to be captured and analysed for variation. This approach is well suited to analysis of quantitative natural variation, as it is able to distinguish differences in leaf shape between lines that are otherwise difficult to describe and quantify (Kieffer et al., 2011). Leaf shape has often been studied using mutant analysis in *Arabidopsis* (Bensmihen et al., 2008; Berná et al., 1999; Dkhar and Pareek, 2014; Wilson-Sánchez et al., 2014), and studies that have used natural variation to investigate leaf shape in *Arabidopsis* have used length and width measurements that do not describe the full leaf shape variation of this species (Hopkins et al., 2008; Jiménez-Gómez et al., 2010; Juenger et al., 2005). We were interested to investigate the genetic basis of natural leaf shape variation in *Arabidopsis* using a comprehensive and accurate approach to leaf shape analysis.

We chose to investigate the genetic basis of naturally occurring leaf shape differences with genetic mapping using accessions of the model plant *Arabidopsis thaliana*. There were however, a number of alternative approaches possible. Differences in leaf shape have been

genetically mapped between species (Costa et al., 2012; Ferris et al., 2015), and so one option was to choose a genus of plant capable of self fertilisation and outcrossing, and investigate the genetic basis of shape differences within the genus by creating mapping populations between species with differences in leaf shape. However, we were wary of working in a non model organism due to potential difficulties in finding suitable genetic markers. This would be a significant investment of time, and could potentially limit what we could achieve in the time available for the project. Although hybrids between species of *Arabidopsis* are possible (Nasrallah et al., 2000), we thought that mapping such between species leaf shape differences may identify loci that had morphological effects conflated with other between species differences, such as developmental timing, physiology or ecological niche, that were specific to the differences between the two species chosen. Natural accessions of *Arabidopsis* are found over a broad geographic range, including Europe, Asia and North America (Alonso-Blanco and Koornneef, 2000), and there is considerable genetic variation between these accessions (Gan et al., 2011). We aimed to identify leaf shape loci unrelated to specific differences between species by studying the within species differences in leaf shape across accessions of *Arabidopsis thaliana*, hoping to understand more as to how leaf shape varied across a broad geographic range within a single species.

Arabidopsis grows a rosette made up of a number of simple, ovate leaves before developing an inflorescence and flowering, at which point the production and growth of these leaves has ceased. We reasoned that by working in *Arabidopsis*, we would be able to achieve a great deal more than by working in another species. In part, this is due to the size and generation time of *Arabidopsis*. This allowed us to conduct multiple genetic mapping experiments, cross plants to produce double mutants, and as we were able to complete a number of mapping and growth experiments in the time available, utilise the results of completed experiments to inform our next experiments. The relatively low and determinate number of leaves grown by each plant allowed us to measure shape across every leaf the plant grew to conduct a comprehensive analysis of leaf shape in each plant. We also considered the wealth of wider work completed in the study of *Arabidopsis* to be extremely beneficial to our project, valuing the presence of a large body of literature and numerous publicly available resources and datasets, and this was a deciding factor in using *Arabidopsis* to study natural leaf shape variation. Without the publicly available RIL populations, natural accessions, and marker datasets we would have needed to create these resources during the project, and so much of the time and effort that went into the project would have been

spent achieving this. We felt that for our circumstances it was best to choose a strategy that would allow as much as possible to be achieved by a small group, and so choose to work within a species that had considerable resources already available, allowing us to work efficiently and making the most of our available time and effort.

6.1.2 Strategies for mapping leaf shape variation in *Arabidopsis*

We mapped variation in leaf shape traits to regions of the *Arabidopsis* genome in two separate mapping populations. We used QTL mapping to identify loci associated with shape variation in the Bay-0 x Shahdara RIL population, and GWA mapping to identify loci associated with shape variation across a set of over 300 natural accessions. Both approaches were successful in that we found loci associated with our leaf shape traits. Each population had specific advantages that we made use of during the project.

A major advantage for our project in working with the Bay-0 x Shahdara population was the relative ease of continuing further work on the regions associated with our leaf shape traits using Heterozygous Inbred Families (HIFs). Obtaining HIFs, collections of Near Isogenic Lines (NILs) variant at specific regions of the genome, allowed us to independently confirm the effect of regions of the genome identified in QTL analysis as having an effect on leaf shape. They could also be used for further experiments to explore the effect of variation at a specific locus. We made use of this when investigating a QTL we found associated with variation in margin morphology and leaf size. We were able to confirm the effect of this QTL in two HIFs for this locus, and so were able to explore the effect of natural variation at this locus within two different genetic backgrounds. We could test for pleiotropic effects of allelic variation at this region on other traits, such as epidermal cell size and siliques produced per plant using HIFs for this region. To investigate an effect of variation at this locus on other traits in the absence of HIFs would require measuring these traits across the Bay-0 x Shahdara RIL population to see if QTLs for these traits coincided with the margin morphology QTL, and so using a HIF approach saved considerable time and facilitated more detailed study of QTL we had identified.

There is no equivalent approach we could use in our GWA work, as the allelic variation we mapped traits to was present across over 300 different accessions, rather than between two parents as in the Bay-0 x Shahdara population. To independently confirm the association of genotypic variation at a locus to trait variation identified in a GWAS, an F2 population can be created from parent accessions varying in genotype at the locus. This has been used successfully to identify candidate genes from GWASs in the literature (Chao et al., 2012). For

loci we identified associated with leaf shape traits in our GWASs, we created F2 populations, and in two of these began genotyping individuals to compare possible association of leaf shape and variation at each region. Due to developments in other areas of the project we did not follow up work in this area further, however it is worth noting for one of the F2 populations, we found an association of leaf shape variation with genotype at a marker nearby the locus initially associated in the GWAS.

A major advantage of GWA is that the great genetic variation within a population of hundreds of natural accessions allows trait variation to be mapped to much narrower loci than is typical for QTL analysis in RIL populations such as the Bay-0 x Shahdara RILs. This allows candidate genes to be identified more easily, as the number of genes within a window of association with the trait is less. We found that several loci associated with leaf shape traits in our GWASs contained clusters of NBLRR genes. This was particularly beneficial for the project, as the reoccurrence of these NBLRR genes suggested them as possible candidate genes for each of the loci, and so targets for further investigation in our work. An alternative course of the project, without the GWA work, would have been to identify candidate genes for further investigation within leaf shape QTLs in the Bay-0 x Shahdara population. Although we applied a network analysis to identify candidate genes for the margin morphology and leaf size QTL in the Bay-0 x Shahdara RILs, T-DNA insertion lines for these candidate genes did not show any effect of leaf size or margin morphology. Fine mapping could be used to narrow regions containing QTL, and we considered this approach. However, the work required to map a QTL interval to a size containing a manageable number of candidate genes for further investigation is considerable. We reasoned that as this work would not necessarily bring us to a candidate gene for a QTL within the time available, a better approach would be to investigate the NBLRR candidate genes associated within the GWAS associated loci. Although this meant taking a step away from natural genetic variation to investigate these genes using T-DNA mutant analysis, it allowed us to ask discrete questions on the leaf shape of T-DNA lines for these genes.

These two approaches to genetic mapping therefore worked complementarily for our project, GWASs led us to a set of candidate genes for further investigation, and within the Bay-0 x Shahdara RILs we were able to explore the effects of natural variation at a locus on a variety of traits beyond that for which the region was initially associated with.

6.1.3 Leaf shape and variations in wider plant morphology

We had found that some of the loci associated with variation in leaf shape traits were also associated with variation in leaf number and hypocotyl length. Two regions associated with variation in *llpPC2* score in the Bay-0 x Shahdara population were also associated with variation in published data on hypocotyl length (Loudet et al., 2008). The leaf shape trait *llpPC2* defines a change in shape between a more stunted or elongated leaf and longer hypocotyl length was found to be positively correlated with a more elongated leaf at both of these loci. Overexpression of the gene identified as responsible for the association of hypocotyl length variation at one of these loci caused an extension of all linearly elongating lateral organs of the plant (Loudet et al., 2008), suggesting a role for this gene in the wider morphology of the plant. Given that two of four of the *llpPC2* associated loci within the Bay-0 x Shahdara RILs are also associated with correlated changes in hypocotyl length, it may be relatively common that genetic loci responsible for variation in leaf shape are also associated with changes in overall plant morphology. That leaf shape variation can occur as part of overall changes in plant morphology is not unexpected. It seems reasonable to assume that factors controlling cell division and expansion or growth rate may have an effect across multiple plant organs. Members of the *ERECTA* family Leucine Rich Repeat Receptor Like Kinases show reduced and altered lateral organ growth, likely as a result of decreases in cell number (Bundy et al., 2012; Shpak et al., 2004; Torii et al., 1996). The genes *TCP14* and *TCP15* have also been found to have a function in leaf shape, internode length and pedicel length (Kieffer et al., 2011).

6.1.4 NBLRR genes associated with differences in leaf shape

Four loci associated with leaf shape variation in our GWAS contained clusters of NBLRR genes. We were intrigued by this, and investigated possible effects of these NBLRR genes on plant leaf shape through mutant analysis of T-DNA insertion lines. Through this approach we characterised several T-DNA insertion lines with light and temperature specific effects on plant morphology.

NBLRR genes are typically thought to have a role in disease response in *Arabidopsis* (McHale et al., 2006). It is possible the leaf shape differences we observed in T-DNA lines for the NBLRR candidate genes are related to disease resistance effects. Disease resistance phenotypes in the NBLRR gene *Chilling Sensitive 3 (CHS3)* are associated with decreased growth resulting in smaller more compact rosettes (Yang et al., 2010). A stunted rosette

phenotype is also associated with enhanced disease resistance in the *suppressor of npr1-1*, *constitutive 1(SNC1)* NBLRR gene (Li et al., 2001; Zhang et al., 2003). Similar effects on morphology and disease resistance have been described for with the NBLRR genes RRS1 and RPS4 (Heidrich et al., 2013), and for combinations of natural polymorphic NBLRR gene clusters within natural accessions (Alcázar et al., 2014; Bomblies et al., 2007). Analogues of salicylic acid (SA) have been known to decrease biomass in *Arabidopsis* (Canet et al., 2010), and the stunted morphology common to NBLRR mutations with altered disease resistance phenotypes may be the result of changes in SA levels in these plants (Naseem et al., 2015; Sašek et al., 2014; Vicente and Plasencia, 2011). We did not test SA levels in the T-DNA lines we studied, so it is unclear as to whether the differences in leaf morphology we observed were the result of differences in SA levels between the plants, and this may be an area for further work on from this project.

Although the majority of leaf shape differences we identified in T-DNA lines for NBLRR genes were a decrease *llpPC2* score, consistent with the stunted leaf morphology of activated disease responses phenotypes of NBLRR mutations in the literature, a T-DNA line for At1g72850, had an increased *llpPC2* score, similar to a more elongated leaf. Interestingly, an NBLRR mutation resulting in elongated rather than stunted leaves, is also more susceptible to disease (Faigón-Soverna et al., 2006). It would be interesting to test whether SA levels and disease resistance correlate with *llpPC2* score across the T-DNA lines we studied.

6.1.5 Effects of environmental variation on leaf shape in T-DNA lines for NBLRR genes

We found that several of the differences in leaf shape observed in the T-DNA lines for NBLRR genes were temperature sensitive, and varied in presence or effect size between 16°C and 24°C. Phenotypes of NBLRR genes in the literature have been found to be sensitive to temperature. However, in all cases we are aware of where the phenotypic effect of a mutation or natural variant of an NBLRR gene responds to temperature, the result is a more severe phenotype in colder temperatures (Alcázar et al., 2008; Bomblies et al., 2007; Huang et al., 2010; Zhu et al., 2010). The effect of varying temperature conditions on our leaf shape traits in the T-DNA insertion lines was not so consistent. We found that the leaf shape difference to Col-0 for two of the T-DNA lines, At5g45240-1 and At1g72850-1 was only present in 24°C conditions. The leaf shape effect of the At4g09420-1 line was only present in 16°C conditions. Interestingly, another T-DNA line for the At5g45240 gene, At5g45240-2 had

a stronger effect on leaf shape in 16°C conditions, despite the other T-DNA line for this gene, At5g45240-1 having an effect on leaf shape specific to 24°C conditions.

Although we did not test for a link between SA levels and the leaf shape of the T-DNA lines we studied, it is interesting to note that the response of *Arabidopsis* plants to temperature has been shown to vary with SA level. Differences in growth rate between Col-0 and a mutant lines with variable levels of SA suggest an increase of rosette size at lower temperatures for plants with lower SA levels (Scott et al., 2004). This may explain the temperature specific responses observed for some NBLRR mutant disease phenotypes, and it is possible a similar explanation may exist for temperature specific leaf shape differences we observed across the T-DNA insertion lines.

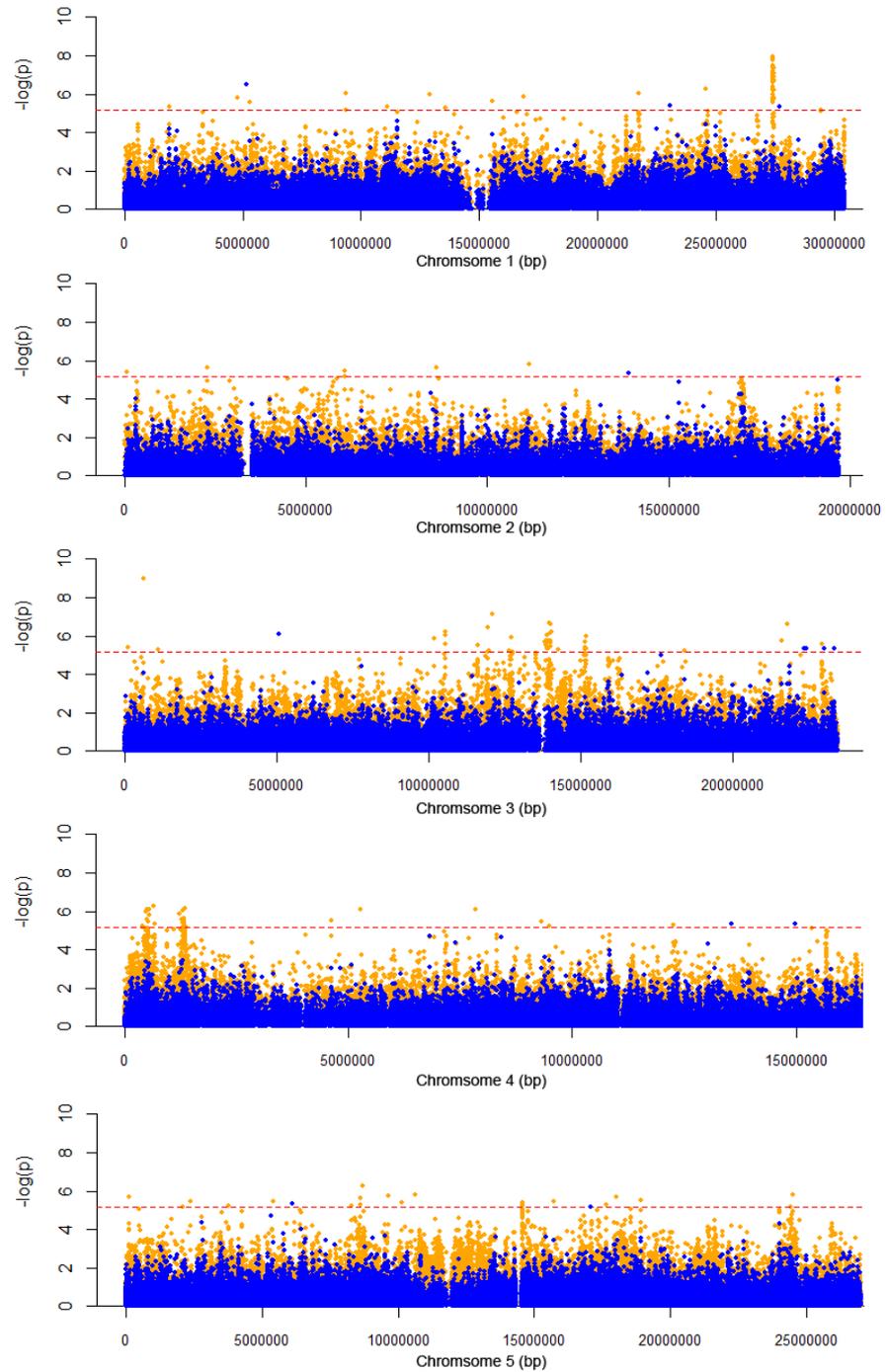
As work in the literature implicated an NBLRR gene with an effect on phytochrome B signalling (Faigón-Soverna et al., 2006), we explored possible interactions between the At5g45240-1 T-DNA insertion and a phytochrome B (PhyB) mutant, PhyB-9. Our results across hypocotyl growth experiments in different light conditions, and leaf shape analysis of plants grown in standard greenhouse conditions, did not require an interaction of PhyB-9 mutation and At5g45240-1 insertion for explanation. Instead our results suggest that within etiolating plants, due to either dark growth conditions or the presence of a PhyB-9 mutation, the decrease in llpPC2 or hypocotyl length otherwise associated with the At5g45240-1 insertion was alleviated.

Variation in red to far red ratio, or the presence of a PhyB mutation, which both result in a shade response in *Arabidopsis*, have been found to increase susceptibility of plants to pathogens (Cargnel et al., 2014; Cerrudo et al., 2012; de Wit et al., 2013). Interestingly the increase in disease susceptibility does not appear to be the result of morphological changes associated with a shade response, as mutations in other photoreceptors, causing similar shade morphologies, do not increase susceptibility (Cerrudo et al., 2012). If there is a link between the leaf shape effects we observed for the At5g45240-1 line and an as yet uncharacterised disease response for this T-DNA line, this interaction between shade response and disease resistance in the literature may explain our observations on the absence of morphology effects of the At5g45240-1 insertion in etiolating plants.

6.1.6 Natural genetic variation in NBLRR genes across *Arabidopsis* accessions

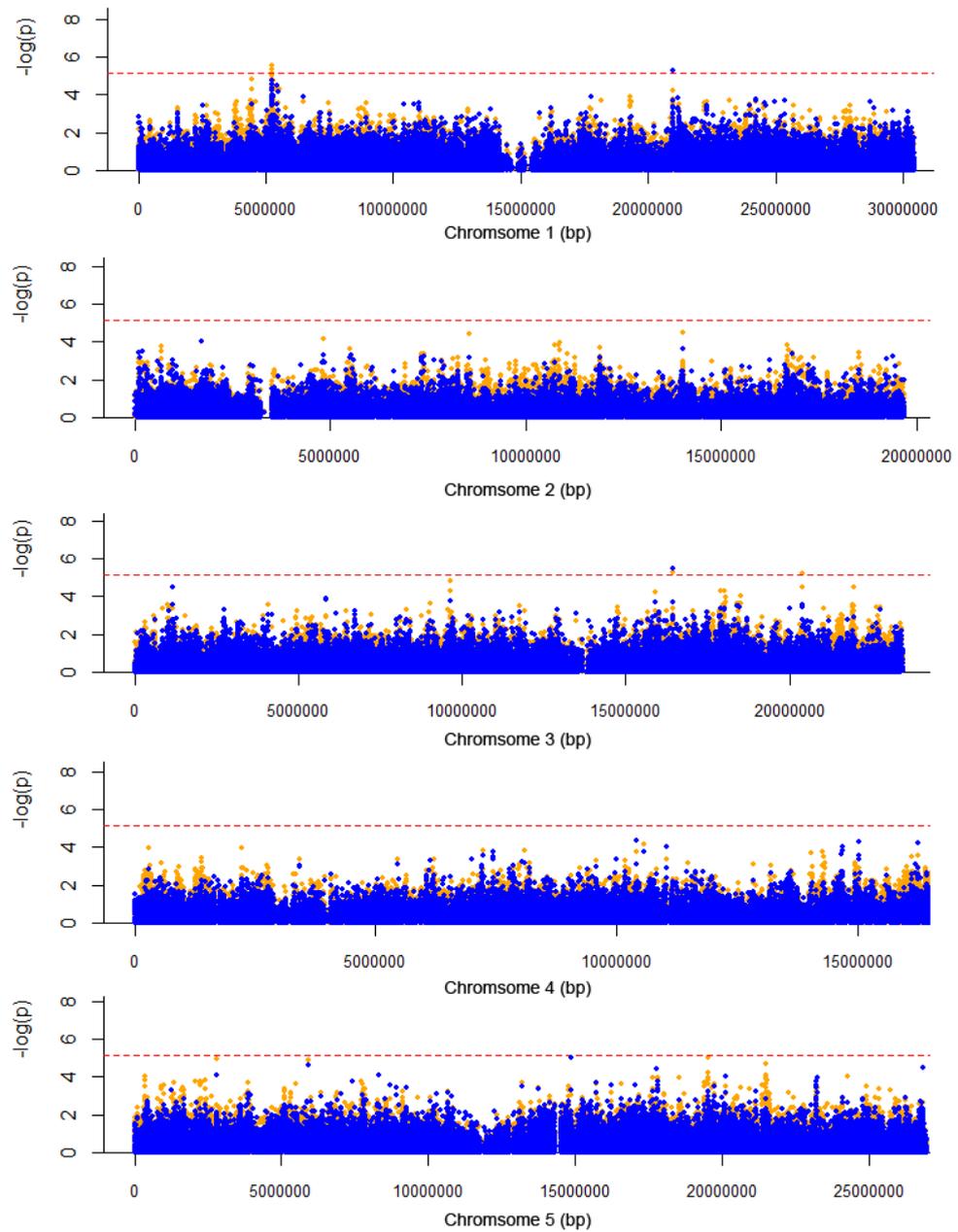
The NBLRR genes are one of the most variable gene families in *Arabidopsis*, varying greatly between different natural accessions in amino acid sequence, expression level and copy number (Gan et al., 2011). There is also great variation in presence absence polymorphisms in this gene family (Guo et al., 2011). In a set of 80 resequenced accessions, two of the NBLRR genes, At1g72850 and At5g45240, associated with leaf shape variation in our GWA work and for which T-DNA insertion lines had shown differences in leaf shape, were present in only twenty and nine of the accessions respectively (Guo et al., 2011) (<http://www.1001genomes.org>). At4g09420, a homologue of At1g72850 for which a T-DNA insertion line had a leaf shape phenotype, was present across 60 of these accessions (Guo et al., 2011) (<http://www.1001genomes.org>). GWASs typically use biallelic SNPs as genetic markers, and so more complex patterns of genetic variation are often flattened to facilitate a straightforward approach to genetic mapping. GWAPP uses imputed genotypes to fill missing alleles, and discards triallelic SNPs (Seren et al., 2012). The SNPs used in GWAPP cover the TAIR10 Col-0 reference genome, and so it is important to remember that only genetic variation for genes within this reference genome can be mapped to through GWAPP. This therefore excludes variation in genes not present in Col-0. Instances of additional NBLRR genes within clusters relative to the Col-0 reference genome have been linked to effects on disease resistance and morphology, (Alcázar et al., 2008; Staal et al., 2006), suggesting that there are NBLRR genes not present in Col-0 in which polymorphisms may be responsible for trait variation between the natural accessions. Evaluating specific genetic loci is perhaps still best done within RIL populations, where the effect of a finite number of parent alleles can be tested for associations with traits, and examined further through fine mapping experiments. Further work to test the effects of natural variation at the loci we found associated with leaf shape traits, or the genes for which we identified T-DNA lines with leaf shape effects, may use this approach to isolate and examine the role of natural genetic variation in these NBLRR genes on leaf shape.

Appendices



GWA plot for whole genome for *IlpPC2* in node 6 leaves

Association of SNPs to trait *IlpPC2* for node 6 leaves in the GWAS10 dataset. Results for two tests of association are shown, orange circles mark SNPs when tested with the KW approach, blue circles show SNPs tested with the AMM approach. The probability a SNP is associated with the trait is indicated by position on the Y axis.



GWA plot for trait llpPC2 for median node leaves

Association of SNPs to trait llpPC2 for median node leaves in the GWAS60 dataset. Results for two tests of association are shown, orange circles mark SNPs when tested with the KW approach, blue circles show SNPs tested with the AMM approach. The probability a SNP is associated with the trait is indicated by position on the Y axis.

Gene lists for loci identified in GWA

The following table includes a list of the genes within 20kb of the most highly associated SNPs for NBLRR containing loci identified during the GWA work.

Gene	Loci	Description
AT1G58370	Chr1:21.7	Encodes a protein with xylanase activity
AT1G58380	Chr1:21.7	XW6
AT1G58390	Chr1:21.7	Disease resistance protein (CC-NBS-LRR class) family
AT1G58400	Chr1:21.7	Disease resistance protein (CC-NBS-LRR class) family
AT1G58410	Chr1:21.7	Disease resistance protein (CC-NBS-LRR class) family
AT1G58420	Chr1:21.7	Uncharacterised conserved protein UCP031279
AT1G58430	Chr1:21.7	Encodes an anther-specific proline-rich protein
AT1G58440	Chr1:21.7	Encodes a putative protein that has been speculated, based on sequence similarities, to have squalene monooxygenase activity
AT1G58450	Chr1:21.7	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins (Prasad 2010, Pubmed ID: 20856808) with potential to interact with Hsp90/Hsp70 as co-chaperones
AT1G58460	Chr1:21.7	unknown protein
AT1G58470	Chr1:21.7	encodes an RNA-binding protein protein_coding RNA-BINDING PROTEIN 1 (RBP1)
AT1G58520	Chr1:21.7	RNA-BINDING PROTEIN 1 (RBP1) RXW8
AT1G72770	Chr1:27.4	mutant has ABA hypersensitive inhibition of seed germination
AT1G72780	Chr1:27.4	pre-tRNA
AT1G72790	Chr1:27.4	hydroxyproline-rich glycoprotein family protein
AT1G72800	Chr1:27.4	RNA-binding (RRM/RBD/RNP motifs) family protein
AT1G72810	Chr1:27.4	Pyridoxal-5'-phosphate-dependent enzyme family protein
AT1G72820	Chr1:27.4	Mitochondrial substrate carrier family protein
AT1G72830	Chr1:27.4	Encodes a subunit of CCAAT-binding complex, binds to CCAAT box motif present in some plant promoter sequences
AT1G72840	Chr1:27.4	Disease resistance protein (TIR-NBS-LRR class)
AT1G72850	Chr1:27.4	Disease resistance protein (TIR-NBS class)
AT1G72852	Chr1:27.4	Potential natural antisense gene, locus overlaps with AT1G72850 other_rna

AT1G72860	Chr1:27.4	Disease resistance protein (TIR-NBS-LRR class) family
AT1G72855	Chr1:27.4	Potential natural antisense gene, locus overlaps with AT1G72860 other_rna
AT1G72870	Chr1:27.4	Disease resistance protein (TIR-NBS class) Survival protein SurE-like
AT1G72880	Chr1:27.4	phosphatase/nucleotidase
AT1G72890	Chr1:27.4	Disease resistance protein (TIR-NBS class) Toll-Interleukin-Resistance (TIR) domain-containing protein
AT1G72900	Chr1:27.4	Toll-Interleukin-Resistance (TIR) domain-containing protein
AT1G72910	Chr1:27.4	Toll-Interleukin-Resistance (TIR) domain-containing protein
AT1G72920	Chr1:27.4	Toll-Interleukin-Resistance (TIR) domain family protein
AT1G72930	Chr1:27.4	Toll/interleukin-1 receptor-like protein (TIR) mRNA, protein_coding TOLL/INTERLEUKIN-1 RECEPTOR-LIKE (TIR)
		Gene Loci Description
AT5G36890	Chr5:14.57	beta glucosidase 42 (BGLU42)
AT5G36900	Chr5:14.57	unknown protein
AT5G36903	Chr5:14.57	pseudogene of protein related to self-incompatibility pseudogene
AT5G36904	Chr5:14.57	pseudogene of protein related to self-incompatibility pseudogene
AT5G36905	Chr5:14.57	transposable element gene Encodes a thionin that is expressed at a low basal level in seedlings and shows circadian variation
AT5G36910	Chr5:14.57	unknown protein
AT5G36920	Chr5:14.57	unknown protein
AT5G36925	Chr5:14.57	unknown protein
AT5G36930	Chr5:14.57	Disease resistance protein (TIR-NBS-LRR class) family
AT5G36935	Chr5:14.57	transposable element gene
AT5G36937	Chr5:14.57	transposable element gene Encodes a member of the cationic amino acid transporter (CAT) subfamily of amino acid polyamine choline transporters
AT5G36940	Chr5:14.57	acid polyamine choline transporters
AT5G36950	Chr5:14.57	Encodes a putative DegP protease
AT5G36960	Chr5:14.57	unknown protein
		Gene Loci Description
AT5G45085	Chr5:18.3	transposable element gene
AT5G45090	Chr5:18.3	phloem protein 2-A7 (PP2-A7)
AT5G45095	Chr5:18.3	unknown protein Encodes one of the BRGs (BOI-related gene) involved in resistance to Botrytis cinerea
AT5G45100	Chr5:18.3	cinerea
AT5G45105	Chr5:18.3	zinc transporter 8 precursor (ZIP8)
AT5G45110	Chr5:18.3	Encodes a paralog of NPR1 mitochondrial transcription termination factor-related / mTERF-related
AT5G45113	Chr5:18.3	factor-related / mTERF-related

AT5G45116	Chr5:18.3	transposable element gene
AT5G45120	Chr5:18.3	Eukaryotic aspartyl protease family protein
AT5G45130	Chr5:18.3	small GTP binding protein protein_coding RAB HOMOLOG 1 (RHA1) ROOT HANDEDNESS 1 (RHA1) Encodes a subunit of RNA polymerase III (aka RNA polymerase C)
AT5G45140	Chr5:18.3	RNAse THREE-like protein 3 (RTL3)
AT5G45150	Chr5:18.3	Root hair defective 3 GTP-binding protein (RHD3)
AT5G45160	Chr5:18.3	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein
AT5G45170	Chr5:18.3	Flavin-binding monooxygenase family protein
AT5G45180	Chr5:18.3	Encodes a cyclin T partner CYCT1
AT5G45190	Chr5:18.3	Disease resistance protein (TIR-NBS-LRR class) family
AT5G45200	Chr5:18.3	Disease resistance protein (TIR-NBS-LRR class) family
AT5G45210	Chr5:18.3	Disease resistance protein (TIR-NBS-LRR class) family
AT5G45220	Chr5:18.3	Disease resistance protein (TIR-NBS-LRR class) family
AT5G45230	Chr5:18.3	Disease resistance protein (TIR-NBS-LRR class) family
AT5G45240	Chr5:18.3	Disease resistance protein (TIR-NBS-LRR class) RPS4 belongs to the Toll/interleukin-1 receptor (TIR)-nucleotide binding site (NBS)-Leu-rich repeat (LRR) class of disease resistance (R) genes
AT5G45250	Chr5:18.3	Confers resistance to Ralstonia solanacearum
AT5G45260	Chr5:18.3	Major facilitator superfamily protein
AT5G45275	Chr5:18.3	

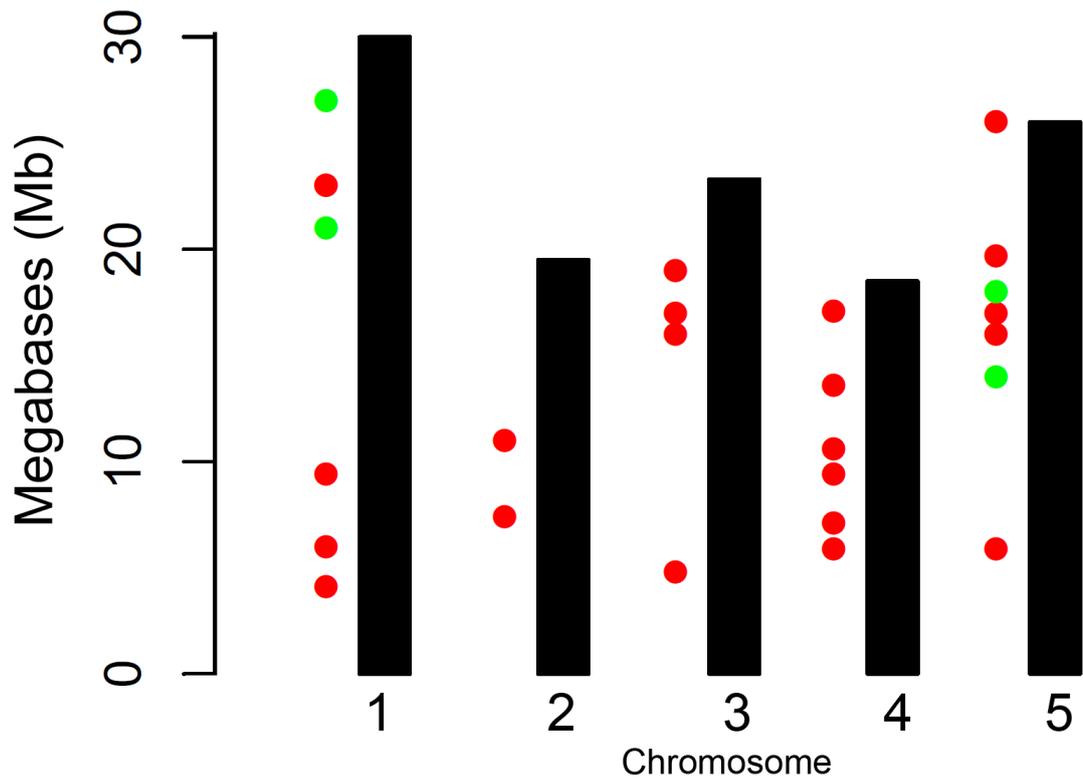
The following table includes a list of the genes within 20kb of the most highly associated SNPs for non NBLRR containing loci identified during the GWA work.

Gene	Loci	Description
AT1G14520	Chr1:4	Encodes MIOX1
AT1G14530	Chr1:4	TOM THREE HOMOLOG 1 (THH1)
AT1G14540	Chr1:4	Peroxidase superfamily protein
AT1G14549	Chr1:4	unknown protein
AT1G14550	Chr1:4	Peroxidase superfamily protein
AT1G14560	Chr1:4	Mitochondrial substrate carrier family protein
AT1G14570	Chr1:4	UBX domain-containing protein
AT1G14580	Chr1:4	C2H2-like zinc finger protein
AT1G14590	Chr1:4	Nucleotide-diphospho-sugar transferase family protein
AT1G14600	Chr1:4	Homeodomain-like superfamily protein
AT1G14610	Chr1:4	Required for proper proliferation of basal cells
AT1G14620	Chr1:4	DECOY (DECOY)
AT1G14630	Chr1:4	unknown protein
AT1G14640	Chr1:4	SWAP (Suppressor-of-White-APricot)/surp domain-containing protein
AT1G14642	Chr1:4	unknown protein
AT1G14650	Chr1:4	SWAP (Suppressor-of-White-APricot)/surp domain-containing protein / ubiquitin family protein
Gene	Loci	Description
AT1G33060	Chr1:12	NAC 014 (NAC014)
AT1G33070	Chr1:12	MADS-box family protein
AT1G33080	Chr1:12	MATE efflux family protein
AT1G33090	Chr1:12	MATE efflux family protein
AT1G33100	Chr1:12	MATE efflux family protein
AT1G33102	Chr1:12	unknown protein
AT1G33110	Chr1:12	MATE efflux family protein
AT1G33120	Chr1:12	Ribosomal protein L6 family
AT1G33130	Chr1:12	transposable element gene
AT1G33135	Chr1:12	transposable element gene
AT1G33140	Chr1:12	Encodes ribosomal protein L9
AT1G33160	Chr1:12	pseudogene, similar to actin, blastp match of 74% identity and 8 S-adenosyl-L-methionine-dependent methyltransferases superfamily
AT1G33170	Chr1:12	protein
Gene	Loci	Description
AT1G71050	Chr1:26	Heavy metal transport/detoxification superfamily protein

AT1G71060	Chr1:26	Tetratricopeptide repeat (TPR)-like superfamily protein
AT1G71070	Chr1:26	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
AT1G71080	Chr1:26	RNA polymerase II transcription elongation factor
AT1G71090	Chr1:26	Auxin efflux carrier family protein Encodes a ribose 5-phosphate isomerase involved in the formation of uridine used for the synthesis of
AT1G71100	Chr1:26	UDP-sugars
AT1G71110	Chr1:26	unknown protein
AT1G71120	Chr1:26	Contains lipase signature motif and GDSL domain encodes a member of the ERF (ethylene response factor) subfamily B-5 of ERF/AP2 transcription factor
AT1G71130	Chr1:26	family
AT1G71140	Chr1:26	MATE efflux family protein
AT1G71150	Chr1:26	unknown protein
Gene	Loci	Description
AT2G10330	Chr2:4	transposable element gene
AT2G10340	Chr2:4	unknown protein
AT2G10350	Chr2:4	transposable element gene
AT2G10360	Chr2:4	transposable element gene
AT2G10370	Chr2:4	transposable element gene
AT2G10380	Chr2:4	transposable element gene
AT2G10390	Chr2:4	transposable element gene
AT2G10400	Chr2:4	transposable element gene
AT2G10405	Chr2:4	transposable element gene Member of Sadhu non-coding retrotransposon family transposable_element_gene SADHU NON-CODING RETROTRANSPOSON 1-1 (SADHU1-1) SADHU NON-CODING
AT2G10410	Chr2:4	RETROTRANSPOSON 1-1 (SADHU1-1)
AT2G10420	Chr2:4	transposable element gene pseudogene, similar to putative reverse transcriptase, blastp match of
AT2G10430	Chr2:4	34% identity and 1
AT2G10440	Chr2:4	unknown protein
AT2G10450	Chr2:4	14-3-3 family protein
AT2G10460	Chr2:4	transposable element gene
Gene	Loci	Description
AT4G05616	Chr4:3	transposable element gene Galactose oxidase/kelch repeat
AT4G05620	Chr4:3	superfamily protein
AT4G05630	Chr4:3	unknown protein

AT4G05631	Chr4:3	unknown protein
AT4G05635	Chr4:3	transposable element gene
AT4G05632	Chr4:3	unknown protein
AT4G05633	Chr4:3	transposable element gene
AT4G05634	Chr4:3	transposable element gene
AT4G05636	Chr4:3	transposable element gene
AT4G05638	Chr4:3	transposable element gene
AT4G05640	Chr4:3	transposable element gene
Gene	Loci	Description
AT4G08657	Chr4:5	transposable element gene
AT4G08660	Chr4:5	transposable element gene
		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G08670	Chr4:5	transposable element gene
AT4G08680	Chr4:5	Encodes a protein, expressed in leaves, with similarity to pollen allergens
AT4G08685	Chr4:5	Sec14p-like phosphatidylinositol transfer family protein
AT4G08690	Chr4:5	unknown protein
AT4G08691	Chr4:5	transposable element gene
AT4G08692	Chr4:5	pseudogene, similar to ribosomal protein L2, blastp match of 66% identity and 2
AT4G08695	Chr4:5	Member of a family of proteins related to PUP1, a purine transporter
AT4G08700	Chr4:5	transposable element gene
AT4G08710	Chr4:5	transposable element gene
AT4G08720	Chr4:5	transposable element gene
AT4G08730	Chr4:5	RNA-binding protein
AT4G08740	Chr4:5	unknown protein
Gene	Loci	Description
AT5G15330	Chr5:5	SPX domain gene 4 (SPX4)
		Pentatricopeptide repeat (PPR) superfamily protein
AT5G15340	Chr5:5	early nodulin-like protein 17 (ENODL17)
AT5G15350	Chr5:5	unknown protein
AT5G15360	Chr5:5	Encodes methyltransferase involved in the de novo DNA methylation and maintenance of asymmetric methylation of DNA sequences
AT5G15380	Chr5:5	tRNA/rRNA methyltransferase (SpoU) family protein
AT5G15390	Chr5:5	U-box domain-containing protein
AT5G15400	Chr5:5	'defense, no death' gene (DND1) encodes a mutated cyclic nucleotide-gated cation channel
AT5G15410	Chr5:5	unknown protein
AT5G15420	Chr5:5	unknown protein

AT5G15430	Chr5:5	Plant calmodulin-binding protein-related
AT5G15440	Chr5:5	EID1-like 1 (EDL1)
AT5G15450	Chr5:5	Encodes a chloroplast-targeted Hsp101 homologue
AT5G15460	Chr5:5	membrane-anchored ubiquitin-fold protein 2 (MUB2)
AT5G15470	Chr5:5	Encodes a protein with putative galacturonosyltransferase activity
AT5G15480	Chr5:5	C2H2-type zinc finger family protein



NBLRR clusters across the *Arabidopsis thaliana* genome

Figure shows the five chromosomes of *Arabidopsis* represented as black bars. The y axis indicates the position in megabases on each chromosome. The green and red circles each show a cluster of NBLRR genes, using data from Guo et al., (2011). Green circles show clusters that were within towers of associated SNPs in our GWASs for leaf shape.

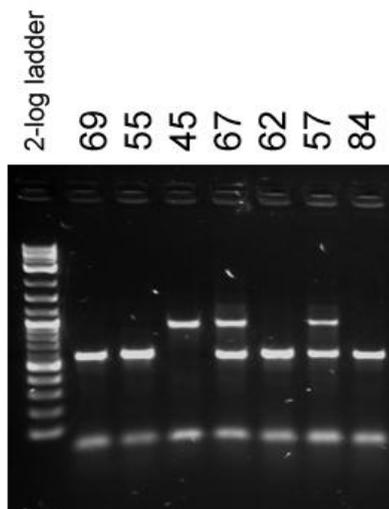


Part one of *Arabidopsis* NBLRR phylogenetic tree

Part one of *Arabidopsis* NBLRR phylogenetic tree, containing predominantly Coiled Coiled domain NBLRRs. The tree was produced using PhyML (<http://atgc.lirmm.fr/phyml/>) with *Arabidopsis* genes identified as containing a nucleotide binding domain. The sequence used to gather NB domain *Arabidopsis* genes was taken from Guo et al., (2011). The tree was drawn using the ape package in R (Paradis et al., 2004).

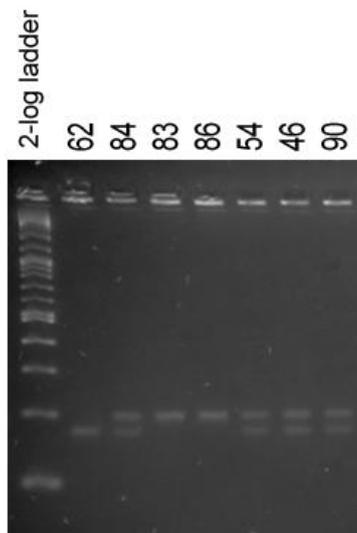
Creation of At5g45240-1 PhyB-9 double mutant

To produce the At5g45240-1 PhyB-9 double mutant a F2 population was produced using the two individual mutant lines as parents. This F2 was then genotyped for the At5g45240-1 insertion using the same primers and methods used to confirm the annotated At5g45240-1 insertion in the parent T-DNA line. Lines with a band size matching the insertion at 600bp, but no wild type band at 1000bp were considered homozygous for the At5g45240-1 insertion and selected for further genotyping. To test for the PhyB-9 insertion, the method described in Ward et al., (2005) was used. First primers were used to amplify the region containing the PhyB-9 mutation, and the product was cut with the MnlI restriction enzyme. DNA with the PhyB-9 mutation contains a restriction site for this enzyme and is cut, producing two products, one roughly 25bp smaller than the wild type product.



F2 lines genotyped for At5g45240-1 insertion

A band around 1000bp indicates wild type sequence, a band around 600bp indicates the presence of the At5g45240-1 T-DNA insertion. Lines 69, 55, 62 and 84 are homozygous for the At5g45240-1 insertion.

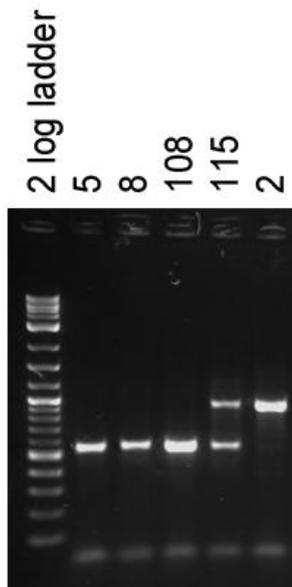


F2 lines genotyped for the PhyB-9 mutation

A band around 200bp indicates wild type sequence, and a band around 180bp indicates the presence of the PhyB-9 mutation, as the mutant sequence will be cut by the enzyme MnlI. Line 62 is homozygous for the PhyB-9 mutation.

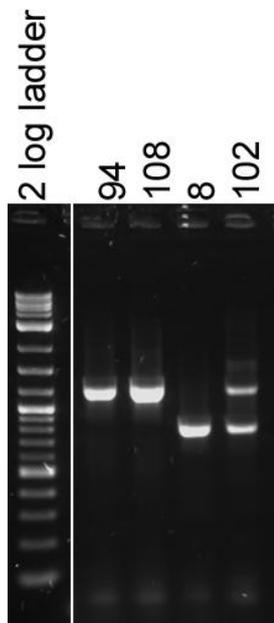
Creation of the At5g45240-1 At1g72850-1 double mutant

An At5g45240-1 At1g72850-1 double mutant was produced by creating a F2 population using the two individual mutant lines as parents. The F2 was then genotyped for the At5g45240-1 insertion. Lines with a band size matching the insertion at 600bp were selected for further genotyping for the At1g72850-1 insertion, using the same primers used to confirm the annotated insertion in the original parent line.



F2 lines genotyped for the At5g45240-1 insertion

Primers used are those used to confirm presence of annotated insertion for the original T-DNA line previously in the thesis. A band around 1000bp indicates wild type sequence, a band around 600bp indicates the presence of the At5g45240-1 insertion. Lines 5, 8 and 108 are homozygous for the At5g45240-1 insertion.



F2 lines genotype for the At1g72850-1 insertion

A band around 1000bp indicates wild type sequence, a band around 600bp indicates the presence of the At1g72850-1 T-DNA insertion. Line 8 is homozygous for the At1g72850-1 insertion.

Abbreviations

Abbreviation	Definition
ADR1	Advanced Disease Resistance 1 (gene)
AMM	Accelerated Mixed Model
AMPRIL	Arabidopsis multiparent RIL
ATTEDJP	http://atted.jp , collection of microarray datasets
avg	average
BLAST	Basic Local Alignment Search Tool
bp	base pairs
CC	Coiled-Coiled
cDNA	complementary DNA
Chr	Chromosome
CHS3	Chilling Sensitive 3
Col-0	Columbia-0, commonly used Arabidopsis accession
CSA1	Constitutive Shade Avoidance 1 (gene)
Cvi	Cape Verde Islands, Arabidopsis accession
DNA	Deoxyribonucleic acid
DOG1	Delay Of Germination 1 (gene)
ELF3	Early Flowering 3
EMMA	Efficient Mixed-Model Association
eQTL	expression QTL
G09_pPC	Procrustes fitted PC created using GWAS09 leaf dataset
G10_pPC	Procrustes fitted PC created using GWAS10 leaf dataset
G60_pPC	Procrustes fitted PC created using GWAS60 leaf dataset
GO	Gene Ontology
GWAPP	A Web Application for Genome-Wide Association Mapping in Arabidopsis
GWAS	Genome Wide Association Study
hetPC	A heteroblasty PC
HIF	Heterozygous Inbred Family
HMA5	Heavy Metal ATPase 5 (gene)
INRA	French National Institute for Agricultural Research
kb	kilobase
KNOX	Knotted1-like Homeobox (gene family)
KW	Wilcoxon rank sum test
L2_PC	PC created using PCA on first two leaves of a dataset
LBI1.3	Primer for left border of SALK T-DNA insertion
Ler	Landsberg erecta, commonly used Arabidopsis accession
llpPC	A procrustes fitted PC created from a PCA on the leaf library leaves
LOD	Logarithm Of Odds
LP	Left Primer
MAGIC	Multiparent Advanced Generation Inter-Cross
MAX1	More Axillary Branches 1 (gene)
mn	median node
MQM	Multiple QTL Mapping
N6PCA_PC	A PC created from node 6 leaves
NASC	Nottingham Arabidopsis Stock Centre

NBLRR	Nucleotide Binding Leucine Rich Repeat
NIL	Near Isogenic Line
PAR	Photosynthetically Active Radiation
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PhyB	Phytochrome B (gene)
QTL	Quantitative Trait Loci
R/qtl	QTL package for R
RIL	Recombinant Inbred Line
RNA	Ribonucleic acid
RP	Right Primer
RPP4	Recognition of Peronospora Parasitica 4
RPS4	Resistant to Psuedonomas Syringae 4
RRS1	Resistant to Ralstonia Solanacearum1 (gene)
RTPCR	Reverse Transcriptase PCR
SA	Salicylic acid
SALK	Salk Institute for biological studies
SD	Standard Deviation
SIL3	Simple Leaf 3 (gene)
SNC1	Suppressor of NPR1-1, Constitutive 1
SNP	Single Nucleotide Polymorphism
TAIR	The Arabidopsis Information Resource
TCP	Teosinte Branched, Cycloidea and PCF (gene family)
T-DNA	Transfer DNA
TIR	domain resembling Drosophila Toll and mammalian IL-1 receptors
TZP	At5g43630

References

- Adams, D.C., Rohlf, F.J., and Slice, D.E. (2004). Geometric morphometrics: ten years of progress following the “revolution.” *Ital. J. Zool.* *71*, 5–16.
- Agren, J., and Schemske, D.W. (2012). Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. *New Phytol.* *194*, 1112–1122.
- Alcázar, R., García, A.V., Parker, J.E., and Reymond, M. (2008). Incremental steps toward incompatibility revealed by *Arabidopsis* epistatic interactions modulating salicylic acid pathway activation. *Proc. Natl. Acad. Sci. pnas.0811734106*.
- Alcázar, R., von Reth, M., Bautor, J., Chae, E., Weigel, D., Koornneef, M., and Parker, J.E. (2014). Analysis of a Plant Complex Resistance Gene Locus Underlying Immune-Related Hybrid Incompatibility and Its Occurrence in Nature. *PLoS Genet* *10*, e1004848.
- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-Wide Insertional Mutagenesis of *Arabidopsis thaliana*. *Science* *301*, 653–657.
- Alonso-Blanco, C., and Koornneef, M. (2000). Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. *Trends Plant Sci.* *5*, 22–29.
- Alonso-Blanco, C., Peeters, A.J., Koornneef, M., Lister, C., Dean, C., van den Bosch, N., Pot, J., and Kuiper, M.T. (1998). Development of an AFLP based linkage map of Ler, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population. *Plant J. Cell Mol. Biol.* *14*, 259–271.
- Alonso-Blanco, C., Bentsink, L., Hanhart, C.J., Blankestijn-de Vries, H., and Koornneef, M. (2003). Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. *Genetics* *164*, 711–729.
- Alonso-Blanco, C., Aarts, M.G.M., Bentsink, L., Keurentjes, J.J.B., Reymond, M., Vreugdenhil, D., and Koornneef, M. (2009). What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? *Plant Cell Online* *21*, 1877–1896.
- Arends, D., Prins, P., Jansen, R.C., and Broman, K.W. (2010). R/qtl: high-throughput multiple QTL mapping. *Bioinforma. Oxf. Engl.* *26*, 2990–2992.
- Atwell, S., Huang, Y.S., Vilhjalmsón, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* *465*, 627–631.
- Baker, R.L., Leong, W.F., Brock, M.T., Markelz, R.J.C., Covington, M.F., Devisetty, U.K., Edwards, C.E., Maloof, J., Welch, S., and Weigand, C. (2015). Modeling development and quantitative trait mapping reveal independent genetic modules for leaf size and shape. *New Phytol.* n/a – n/a.
- Bakker, E.G., Stahl, E.A., Toomajian, C., Nordborg, M., Kreitman, M., and Bergelson, J. (2006). Distribution of genetic variation within and among local populations of *Arabidopsis thaliana* over its species range. *Mol. Ecol.* *15*, 1405–1418.

- Baxter, I., Brazelton, J.N., Yu, D., Huang, Y.S., Lahner, B., Yakubova, E., Li, Y., Bergelson, J., Borevitz, J.O., Nordborg, M., et al. (2010). A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter *AtHKT1;1*. *PLoS Genet.* 6, e1001193.
- Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188.
- Bensmihen, S., Hanna, A.I., Langlade, N.B., Micol, J.L., Bangham, A., and Coen, E.S. (2008). Mutational spaces for leaf shape and size. *HFSP J.* 2, 110–120.
- Bentsink, L., Hanson, J., Hanhart, C.J., Vries, H.B., Coltrane, C., Keizer, P., El-Lithy, M., Alonso-Blanco, C., Andrés, M.T. de, Reymond, M., et al. (2010). Natural variation for seed dormancy in *Arabidopsis* is regulated by additive genetic and molecular pathways. *Proc. Natl. Acad. Sci.* 107, 4264–4269.
- Berardini, T.Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L.A., Yoon, J., Doyle, A., Lander, G., et al. (2004). Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiol.* 135, 745–755.
- Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E. (2015). The *Arabidopsis* information resource: Making and mining the “gold standard” annotated reference plant genome. *Genesis* 53, 474–485.
- Bergelson, J., Stahl, E., Dudek, S., and Kreitman, M. (1998). Genetic Variation Within and Among Populations of *Arabidopsis thaliana*. *Genetics* 148, 1311–1323.
- Berná, G., Robles, P., and Micol, J.L. (1999). A Mutational Analysis of Leaf Morphogenesis in *Arabidopsis thaliana*. *Genetics* 152, 729–742.
- Bomblies, K., Lempe, J., Epple, P., Warthmann, N., Lanz, C., Dangl, J.L., and Weigel, D. (2007). Autoimmune Response as a Mechanism for a Dobzhansky-Muller-Type Incompatibility Syndrome in Plants. *PLoS Biol* 5, e236.
- Botto, J.F., and Coluccio, M.P. (2007). Seasonal and plant-density dependency for quantitative trait loci affecting flowering time in multiple populations of *Arabidopsis thaliana*. *Plant Cell Environ.* 30, 1465–1479.
- Bundy, M.G.R., Thompson, O.A., Sieger, M.T., and Shpak, E.D. (2012). Patterns of Cell Division, Cell Differentiation and Cell Elongation in Epidermis and Cortex of *Arabidopsis* pedicels in the Wild Type and in *erecta*. *PLoS ONE* 7, e46262.
- Campitelli, B.E., and Stinchcombe, J.R. (2013). Natural selection maintains a single-locus leaf shape cline in Ivyleaf morning glory, *Ipomoea hederacea*. *Mol. Ecol.* 22, 552–564.
- Canet, J.V., Dobón, A., Ibáñez, F., Perales, L., and Tornero, P. (2010). Resistance and biomass in *Arabidopsis*: a new model for salicylic acid perception. *Plant Biotechnol. J.* 8, 126–141.
- Cao, J., Schneeberger, K., Ossowski, S., Gunther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., et al. (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* 43, 956–963.
- Cargnel, M.D., Demkura, P.V., and Ballaré, C.L. (2014). Linking phytochrome to plant immunity: low red : far-red ratios increase *Arabidopsis* susceptibility to *Botrytis cinerea* by

- reducing the biosynthesis of indolic glucosinolates and camalexin. *New Phytol.* *204*, 342–354.
- Cartolano, M., Pieper, B., Lempe, J., Tattersall, A., Huijser, P., Tresch, A., Darrah, P.R., Hay, A., and Tsiantis, M. (2015). Heterochrony underpins natural variation in *Cardamine hirsuta* leaf form. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 10539–10544.
- Cerrudo, I., Keller, M.M., Cargnel, M.D., Demkura, P.V., de Wit, M., Patitucci, M.S., Pierik, R., Pieterse, C.M.J., and Ballaré, C.L. (2012). Low red/far-red ratios reduce *Arabidopsis* resistance to *Botrytis cinerea* and jasmonate responses via a COI1-JAZ10-dependent, salicylic acid-independent mechanism. *Plant Physiol.* *158*, 2042–2052.
- Challis, R., Hepworth, J., Mouchel, C., Waites, R., and Leyser, O. (2013). A role for MAX1 in evolutionary diversity in strigolactone signalling upstream of MAX2. *Plant Physiol.*
- Chao, D.-Y., Silva, A., Baxter, I., Huang, Y.S., Nordborg, M., Danku, J., Lahner, B., Yakubova, E., and Salt, D.E. (2012). Genome-Wide Association Studies Identify Heavy Metal ATPase3 as the Primary Determinant of Natural Variation in Leaf Cadmium in *Arabidopsis thaliana*. *PLoS Genet* *8*, e1002923.
- Chardon, F., Jasinski, S., Durandet, M., Lécureuil, A., Soulay, F., Bedu, M., Guerche, P., and Masclaux-Daubresse, C. (2014). QTL meta-analysis in *Arabidopsis* reveals an interaction between leaf senescence and resource allocation to seeds. *J. Exp. Bot.* *65*, 3949–3962.
- Chiang, G.C.K., Bartsch, M., Barua, D., Nakabayashi, K., Debieu, M., Kronholm, I., Koornneef, M., Soppe, W.J.J., Donohue, K., and De MEAUX, J. (2011). DOG1 expression is predicted by the seed-maturation environment and contributes to geographical variation in germination in *Arabidopsis thaliana*. *Mol. Ecol.* *20*, 3336–3349.
- Chitwood, D.H., Ranjan, A., Kumar, R., Ichihashi, Y., Zumstein, K., Headland, L.R., Ostria-Gallardo, E., Aguilar-Martínez, J.A., Bush, S., Carriedo, L., et al. (2014). Resolving Distinct Genetic Regulators of Tomato Leaf Shape within a Heteroblastic and Ontogenetic Context. *Plant Cell Online tpc.114.130112*.
- Clarke, J.H., and Dean, C. (1994). Mapping FRI, a locus controlling flowering time and vernalization response in *Arabidopsis thaliana*. *Mol. Gen. Genet.* *MGG 242*, 81–89.
- Clarke, J.H., Mithen, R., Brown, J.K., and Dean, C. (1995). QTL analysis of flowering time in *Arabidopsis thaliana*. *Mol. Gen. Genet.* *MGG 248*, 278–286.
- Coluccio, M.P., Sanchez, S.E., Kasulin, L., Yanovsky, M.J., and Botto, J.F. (2010). Genetic mapping of natural variation in a shade avoidance response: ELF3 is the candidate gene for a QTL in hypocotyl growth regulation. *J. Exp. Bot.*
- Copenhaver, G.P., Nickel, K., Kuromori, T., Benito, M.-I., Kaul, S., Lin, X., Bevan, M., Murphy, G., Harris, B., Parnell, L.D., et al. (1999). Genetic Definition and Sequence Analysis of *Arabidopsis* Centromeres. *Science* *286*, 2468–2474.
- Costa, M.M.R., Yang, S., Critchley, J., Feng, X., Wilson, Y., Langlade, N., Copsey, L., and Hudson, A. (2012). The genetic basis for natural variation in heteroblasty in *Antirrhinum*. *New Phytol.* *196*, 1251–1259.
- Danisman, S., van der Wal, F., Dhondt, S., Waites, R., de Folter, S., Bimbo, A., van Dijk, A.D.J., Muino, J.M., Cutri, L., Dornelas, M.C., et al. (2012). *Arabidopsis* class I and class II TCP

transcription factors regulate jasmonic acid metabolism and leaf development antagonistically. *Plant Physiol.* *159*, 1511–1523.

Dkhar, J., and Pareek, A. (2014). What determines a leaf's shape? *EvoDevo* *5*, 47.

Drost, D.R., Puranik, S., Novaes, E., Novaes, C.R.D.B., Dervinis, C., Gailing, O., and Kirst, M. (2015). Genetical genomics of *Populus* leaf shape variation. *Bmc Plant Biol.* *15*, 166.

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 Res.* *19*, 1349.

Elmore, C.D. (1986). Mode of Reproduction and Inheritance of Leaf Shape in *Ipomoea hederacea*. *Weed Sci.* *34*, 391–395.

Faigón-Soverna, A., Harmon, F.G., Storani, L., Karayekov, E., Staneloni, R.J., Gassmann, W., Más, P., Casal, J.J., Kay, S.A., and Yanovsky, M.J. (2006). A Constitutive Shade-Avoidance Mutant Implicates TIR-NBS-LRR Proteins in *Arabidopsis* Photomorphogenic Development. *Plant Cell Online* *18*, 2919–2928.

Feng, X., Wilson, Y., Bowers, J., Kennaway, R., Bangham, A., Hannah, A., Coen, E., and Hudson, A. (2009). Evolution of Allometry in *Antirrhinum* RID C-2115-2011. *Plant Cell* *21*, 2999–3007.

Ferris, K.G., Rushton, T., Greenlee, A.B., Toll, K., Blackman, B.K., and Willis, J.H. (2015). Leaf shape evolution has a similar genetic architecture in three edaphic specialists within the *Mimulus guttatus* species complex. *Ann. Bot.* *116*, 213–223.

Filiault, D.L., and Maloof, J.N. (2012). A Genome-Wide Association Study Identifies Variants Underlying the *Arabidopsis thaliana* Shade Avoidance Response. *PLoS Genet.* *8*.

Fonseca, C.R., Overton, J.M., Collins, B., and Westoby, M. (2000). Shifts in trait-combinations along rainfall and phosphorus gradients. *J. Ecol.* *88*, 964–977.

Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T., and Jones, G.H. (2000). Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell* *100*, 367–376.

Gailing, O. (2008). QTL analysis of leaf morphological characters in a *Quercus robur* full-sib family (*Q. robur* × *Q. robur* ssp. *slavonica*). *Plant Biol.* *10*, 624–634.

Gan, X., 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.

Gao, Y., and Zhao, Y. (2013). Epigenetic suppression of T-DNA insertion mutants in *Arabidopsis*. *Mol. Plant* *6*, 539–545.

Gassmann, W., Hinsch, M.E., and Staskawicz, B.J. (1999). The *Arabidopsis* RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *Plant J. Cell Mol. Biol.* *20*, 265–277.

- Glennon, K.L., and Cron, G.V. (2015). Climate and leaf shape relationships in four *Helichrysum* species from the Eastern Mountain Region of South Africa. *Evol. Ecol.* *29*, 657–678.
- Gonzalez, N., and Inzé, D. (2015). Molecular systems governing leaf growth: from genes to networks. *J. Exp. Bot.* eru541.
- Guo, Y.-L., Fitz, J., Schneeberger, K., Ossowski, S., Cao, J., and Weigel, D. (2011). Genome-Wide Comparison of Nucleotide-Binding Site-Leucine-Rich Repeat-Encoding Genes in *Arabidopsis*. *Plant Physiol.* *157*, 757–769.
- Hay, A., and Tsiantis, M. (2006). The genetic basis for differences in leaf form between *Arabidopsis thaliana* and its wild relative *Cardamine hirsuta*. *Nat. Genet.* *38*, 942–947.
- Heidrich, K., Tsuda, K., Blanvillain-Baufumé, S., Wirthmueller, L., Bautor, J., and Parker, J.E. (2013). *Arabidopsis* TNL-WRKY domain receptor RRS1 contributes to temperature-conditioned RPS4 auto-immunity. *Front. Plant Sci.* *4*.
- Hepworth, J., and Lenhard, M. (2014). Regulation of plant lateral-organ growth by modulating cell number and size. *Curr. Opin. Plant Biol.* *17*, 36–42.
- Hinsch, M., and Staskawicz, B. (1996). Identification of a new *Arabidopsis* disease resistance locus, RPs4, and cloning of the corresponding avirulence gene, avrRps4, from *Pseudomonas syringae* pv. pisi. *Mol. Plant-Microbe Interact.* *MPMI 9*, 55–61.
- Hoekenga, O.A., Vision, T.J., Shaff, J.E., Monforte, A.J., Lee, G.P., Howell, S.H., and Kochian, L.V. (2003). Identification and Characterization of Aluminum Tolerance Loci in *Arabidopsis* (*Landsberg erecta* × *Columbia*) by Quantitative Trait Locus Mapping. A Physiologically Simple But Genetically Complex Trait. *Plant Physiol.* *132*, 936–948.
- Hoekenga, O.A., Maron, L.G., Piñeros, M.A., Cançado, G.M.A., Shaff, J., Kobayashi, Y., Ryan, P.R., Dong, B., Delhaize, E., Sasaki, T., et al. (2006). AtALMT1, which encodes a malate transporter, is identified as one of several genes critical for aluminum tolerance in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 9738–9743.
- Hopkins, R., Schmitt, J., and Stinchcombe, J.R. (2008). A latitudinal cline and response to vernalization in leaf angle and morphology in *Arabidopsis thaliana* (Brassicaceae). *New Phytol.* *179*, 155–164.
- Huala, E., Dickerman, A.W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M., Huang, W., et al. (2001). The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* *29*, 102–105.
- Huang, X., Li, J., Bao, F., Zhang, X., and Yang, S. (2010). A Gain-of-Function Mutation in the *Arabidopsis* Disease Resistance Gene RPP4 Confers Sensitivity to Low Temperature. *Plant Physiol.* *154*, 796–809.
- Huang, X., 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. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 4488–4493.
- Jiménez-Gómez, J.M., Wallace, A.D., and Maloof, J.N. (2010). Network Analysis Identifies ELF3 as a QTL for the Shade Avoidance Response in *Arabidopsis*. *PLoS Genet* *6*, e1001100.

- Jimenez-Gomez, J.M., Corwin, J.A., Joseph, B., Maloof, J.N., and Kliebenstein, D.J. (2011). Genomic Analysis of QTLs and Genes Altering Natural Variation in Stochastic Noise. *PLoS Genet* 7, e1002295.
- 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.
- Juenger, T., Pérez-Pérez, J.M., Bernal, S., and Micol, J.L. (2005). Quantitative trait loci mapping of floral and leaf morphology traits in Arabidopsis thaliana: evidence for modular genetic architecture. *Evol. Dev.* 7, 259–271.
- 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.
- Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42, 348–354.
- Kieffer, M., Master, V., Waites, R., and Davies, B. (2011). TCP14 and TCP15 affect internode length and leaf shape in Arabidopsis. *Plant J. Cell Mol. Biol.* 68, 147–158.
- Kim, T.-H., Kunz, H.-H., Bhattacharjee, S., Hauser, F., Park, J., Engineer, C., Liu, A., Ha, T., Parker, J.E., Gassmann, W., et al. (2012). Natural variation in small molecule-induced TIR-NB-LRR signaling induces root growth arrest via EDS1- and PAD4-complexed R protein VICTR in Arabidopsis. *Plant Cell* 24, 5177–5192.
- Klingenberg, C.P. (2011). MorphoJ: an integrated software package for geometric morphometrics. *Mol. Ecol. Resour.* 11, 353–357.
- Klingenberg, C., and McIntyre, G. (1998). Geometric morphometrics of developmental instability: Analyzing patterns of fluctuating asymmetry with procrustes methods RID B-3470-2010. *Evolution* 52, 1363–1375.
- Klingenberg, C., Leamy, L., Routman, E., and Cheverud, J. (2001). Genetic architecture of mandible shape in mice: Effects of quantitative trait loci analyzed by geometric morphometrics RID B-3470-2010. *Genetics* 157, 785–802.
- Kobayashi, Y., Kuroda, K., Kimura, K., Southron-Francis, J.L., Furuzawa, A., Kimura, K., Iuchi, S., Kobayashi, M., Taylor, G.J., and Koyama, H. (2008). Amino Acid Polymorphisms in Strictly Conserved Domains of a P-Type ATPase HMA5 Are Involved in the Mechanism of Copper Tolerance Variation in Arabidopsis. *Plant Physiol.* 148, 969–980.
- Koike, T., Kitao, M., Maruyama, Y., Mori, S., and Lei, T.T. (2001). Leaf morphology and photosynthetic adjustments among deciduous broad-leaved trees within the vertical canopy profile. *Tree Physiol.* 21, 951–958.
- Koornneef, M., Hanhart, C.J., and van der Veen, J.H. (1991). A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana. *Mol. Gen. Genet. MGG* 229, 57–66.
- Koornneef, M., Alonso-Blanco, C., and Vreugdenhil, D. (2004). NATURALLY OCCURRING GENETIC VARIATION IN ARABIDOPSIS THALIANA. *Annu. Rev. Plant Biol.* 55, 141–172.

- Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9, 29.
- Kougioumoutzi, E., Cartolano, M., Canales, C., Dupré, M., Bramsiede, J., Vlad, D., Rast, M., Ioio, R.D., Tattersall, A., Schnittger, A., et al. (2012). SIMPLE LEAF3 encodes a ribosome-associated protein required for leaflet development in *Cardamine hirsuta*. *Plant J.* n/a – n/a.
- 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 Genet.* 5.
- Kroymann, J., and Mitchell-Olds, T. (2005). Epistasis and balanced polymorphism influencing complex trait variation. *Nature* 435, 95–98.
- Ku, L.X., Zhang, J., Guo, S.L., Liu, H.Y., Zhao, R.F., and Chen, Y.H. (2012). Integrated multiple population analysis of leaf architecture traits in maize (*Zea mays* L.). *J. Exp. Bot.* 63, 261–274.
- Li, X., Clarke, J.D., Zhang, Y., and Dong, X. (2001). Activation of an EDS1-Mediated R-Gene Pathway in the *snc1* Mutant Leads to Constitutive, NPR1-Independent Pathogen Resistance. *Mol. Plant. Microbe Interact.* 14, 1131–1139.
- Lockwood, C., Kimbel, W., and Lynch, J. (2004). Morphometrics and hominoid phylogeny: Support for a chimpanzee-human clade and differentiation among great ape subspecies. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4356–4360.
- Long, Q., Rabanal, F.A., Meng, D., Huber, C.D., Farlow, A., Platzer, A., Zhang, Q., Vilhjálmsson, B.J., Korte, A., Nizhynska, V., et al. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet. advance online publication.*
- Loudet, O., Chaillou, S., Camilleri, C., Bouchez, D., and Daniel-Vedele, F. (2002). Bay-0 × Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. *Theor. Appl. Genet.* 104, 1173–1184.
- Loudet, O., Michael, T.P., Burger, B.T., Mett , C.L., Mockler, T.C., Weigel, D., and Chory, J. (2008). A zinc knuckle protein that negatively controls morning-specific growth in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* 105, 17193–17198.
- Lovell, J.T., Juenger, T.E., Michaels, S.D., Lasky, J.R., Platt, A., Richards, J.H., Yu, X., Easlon, H.M., Sen, S., and McKay, J.K. (2013). Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation. *Proc. R. Soc. B Biol. Sci.* 280.
- Luo, Y., Widmer, A., and Karrenberg, S. (2015). The roles of genetic drift and natural selection in quantitative trait divergence along an altitudinal gradient in *Arabidopsis thaliana*. *Heredity* 114, 220–228.
- Magliano, T.M.A., Botto, J.F., Godoy, A.V., Symonds, V.V., Lloyd, A.M., and Casal, J.J. (2005). New *Arabidopsis* recombinant inbred lines (*Landsberg erecta* x *Nossen*) reveal natural variation in phytochrome-mediated responses. *Plant Physiol.* 138, 1126–1135.
- Matser V.A. (2014). The role of class I *TCP* genes in determining leaf shape and size. PhD thesis.

- McDonald, P.G., Fonseca, C.R., Overton, J.M., and Westoby, M. (2003). Leaf-size divergence along rainfall and soil-nutrient gradients: is the method of size reduction common among clades? *Funct. Ecol.* *17*, 50–57.
- McHale, L., Tan, X., Koehl, P., and Michelmore, R.W. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* *7*, 212.
- Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-Wide Analysis of NBS-LRR-Encoding Genes in *Arabidopsis*. *Plant Cell* *15*, 809–834.
- Michaels, S.D., and Amasino, R.M. (1999). FLOWERING LOCUS C Encodes a Novel MADS Domain Protein That Acts as a Repressor of Flowering. *Plant Cell* *11*, 949–956.
- Morgan, D.C., and Smith, H. (1976). Linear relationship between phytochrome photoequilibrium and growth in plants under simulated natural radiation. *Nature* *262*, 210–212.
- Naseem, M., Kaldorf, M., and Dandekar, T. (2015). The nexus between growth and defence signalling: auxin and cytokinin modulate plant immune response pathways. *J. Exp. Bot.* *66*, 4885–4896.
- Nasrallah, M.E., Yogeewaran, K., Snyder, S., and Nasrallah, J.B. (2000). *Arabidopsis* Species Hybrids in the Study of Species Differences and Evolution of Amphiploidy in Plants. *Plant Physiol.* *124*, 1605–1614.
- Nicotra, A.B., Cosgrove, M.J., Cowling, A., Schlichting, C.D., and Jones, C.S. (2008). Leaf shape linked to photosynthetic rates and temperature optima in South African *Pelargonium* species. *Oecologia* *154*, 625–635.
- Nicotra, A.B., Leigh, A., Boyce, C.K., Jones, C.S., Niklas, K.J., Royer, D.L., and Tsukaya, H. (2011). The evolution and functional significance of leaf shape in the angiosperms. *Funct Plant Biol* *38*, 535–552.
- Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., et al. (2005). The Pattern of Polymorphism in *Arabidopsis thaliana*. *PLoS Biol* *3*, e196.
- Noutoshi, Y., Ito, T., Seki, M., Nakashita, H., Yoshida, S., Marco, Y., Shirasu, K., and Shinozaki, K. (2005). A single amino acid insertion in the WRKY domain of the *Arabidopsis* TIR-NBS-LRR-WRKY-type disease resistance protein SLH1 (sensitive to low humidity 1) causes activation of defense responses and hypersensitive cell death. *Plant J. Cell Mol. Biol.* *43*, 873–888.
- Obayashi, T., Hayashi, S., Saeki, M., Ohta, H., and Kinoshita, K. (2009). ATTED-II provides coexpressed gene networks for *Arabidopsis*. *Nucleic Acids Res.* *37*, D987–D991.
- Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N., and Weigel, D. (2008). Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* *18*, 2024–2033.
- Poethig, R.S. (2010). The Past, Present, and Future of Vegetative Phase Change. *Plant Physiol.* *154*, 541–544.

- Pouteau, S., Ferret, V., Gaudin, V., Lefebvre, D., Sabar, M., Zhao, G., and Prunus, F. (2004). Extensive Phenotypic Variation in Early Flowering Mutants of *Arabidopsis*. *Plant Physiol.* *135*, 201–211.
- Reed, J.W., Elumalai, R.P., and Chory, J. (1998). Suppressors of an *Arabidopsis thaliana* phyB mutation identify genes that control light signaling and hypocotyl elongation. *Genetics* *148*, 1295–1310.
- Rodríguez, M.C., Wawrzyńska, A., and Sirko, A. (2014). Intronic T-DNA insertion in *Arabidopsis* NBR1 conditionally affects wild-type transcript level. *Plant Signal. Behav.* *9*, e975659.
- Rohlf, F.J. (1986). Relationships among eigenshape analysis, Fourier analysis, and analysis of coordinates. *Math. Geol.* *18*, 845–854.
- Royer, D.L. (2012). Leaf Shape Responds to Temperature but Not CO₂ in *Acer rubrum*. *PLoS ONE* *7*, e49559.
- Royer, D.L., McElwain, J.C., Adams, J.M., and Wilf, P. (2008). Sensitivity of leaf size and shape to climate within *Acer rubrum* and *Quercus kelloggii*. *New Phytol.* *179*, 808–817.
- Royer, D.L., Meyerson, L.A., Robertson, K.M., and Adams, J.M. (2009). Phenotypic Plasticity of Leaf Shape along a Temperature Gradient in *Acer rubrum*. *PLoS ONE* *4*, e7653.
- Sack, L., Melcher, P.J., Liu, W.H., Middleton, E., and Pardee, T. (2006). How strong is intracopy leaf plasticity in temperate deciduous trees? *Am. J. Bot.* *93*, 829–839.
- Sašek, V., Janda, M., Delage, E., Puyaubert, J., Guivarc'h, A., López Maseda, E., Dobrev, P.I., Caius, J., Bóka, K., Valentová, O., et al. (2014). Constitutive salicylic acid accumulation in pi4kIIIβ1β2 *Arabidopsis* plants stunts rosette but not root growth. *New Phytol.* *203*, 805–816.
- Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* *37*, 501–506.
- Schmitt, J. (1997). Is photomorphogenic shade avoidance adaptive? Perspectives from population biology. *Plant Cell Environ.* *20*, 826–830.
- Schmitz, R.J., Schultz, M.D., Urich, M.A., Nery, J.R., Pelizzola, M., Libiger, O., Alix, A., McCosh, R.B., Chen, H., Schork, N.J., et al. (2013). Patterns of population epigenomic diversity. *Nature* *495*, 193–198.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* *9*, 671–675.
- Scott, I.M., Clarke, S.M., Wood, J.E., and Mur, L.A.J. (2004). Salicylate Accumulation Inhibits Growth at Chilling Temperature in *Arabidopsis*. *Plant Physiol.* *135*, 1040–1049.
- Segura, V., Vilhjálmsson, B.J., Platt, A., Korte, A., Seren, Ü., Long, Q., and Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* *44*, 825–830.

- Seren, Ü., Vilhjálmsón, B.J., Horton, M.W., Meng, D., Forai, P., Huang, Y.S., Long, Q., Segura, V., and Nordborg, M. (2012). GWAPP: A Web Application for Genome-wide Association Mapping in *A. thaliana*. arXiv:1212.0661.
- Shalit, A., Rozman, A., Goldshmidt, A., Alvarez, J.P., Bowman, J.L., Eshed, Y., and Lifschitz, E. (2009). The flowering hormone florigen functions as a general systemic regulator of growth and termination. *Proc. Natl. Acad. Sci.* *106*, 8392–8397.
- Shindo, C., Aranzana, M.J., Lister, C., Baxter, C., Nicholls, C., Nordborg, M., and Dean, C. (2005). Role of FRIGIDA and FLOWERING LOCUS C in Determining Variation in Flowering Time of *Arabidopsis*. *Plant Physiol.* *138*, 1163–1173.
- Shirano, Y., Kachroo, P., Shah, J., and Klessig, D.F. (2002). A gain-of-function mutation in an *Arabidopsis* Toll Interleukin1 receptor-nucleotide binding site-leucine-rich repeat type R gene triggers defense responses and results in enhanced disease resistance. *Plant Cell* *14*, 3149–3162.
- Shpak, E.D., Berthiaume, C.T., Hill, E.J., and Torii, K.U. (2004). Synergistic interaction of three ERECTA-family receptor-like kinases controls *Arabidopsis* organ growth and flower development by promoting cell proliferation. *Dev. Camb. Engl.* *131*, 1491–1501.
- Smith, H., and Whitelam, G.C. (1997). The shade avoidance syndrome: multiple responses mediated by multiple phytochromes. *Plant Cell Environ.* *20*, 840–844.
- Sohn, K.H., Segonzac, C., Rallapalli, G., Sarris, P.F., Woo, J.Y., Williams, S.J., Newman, T.E., Paek, K.H., Kobe, B., and Jones, J.D.G. (2014). The Nuclear Immune Receptor RPS4 Is Required for RRS1SLH1-Dependent Constitutive Defense Activation in *Arabidopsis thaliana*. *PLoS Genet.* *10*, e1004655.
- Staal, J., Kaliff, M., Bohman, S., and Dixelius, C. (2006). Transgressive segregation reveals two *Arabidopsis* TIR-NB-LRR resistance genes effective against *Leptosphaeria maculans*, causal agent of blackleg disease. *Plant J.* *46*, 218–230.
- Stinchcombe, J.R., Weinig, C., Ungerer, M., Olsen, K.M., Mays, C., Halldorsdottir, S.S., Purugganan, M.D., and Schmitt, J. (2004). A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 4712–4717.
- Tan, X., Meyers, B.C., Kozik, A., West, M.A.L., Morgante, M., St Clair, D.A., Bent, A.F., and Michelmore, R.W. (2007). Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in *Arabidopsis*. *BMC Plant Biol.* *7*, 56.
- Telfer, A., Bollman, K.M., and Poethig, R.S. (1997a). Phase change and the regulation of trichome distribution in *Arabidopsis thaliana*. *Development* *124*, 645–654.
- Telfer, A., Bollman, K.M., and Poethig, R.S. (1997b). Phase change and the regulation of trichome distribution in *Arabidopsis thaliana*. *Development* *124*, 645–654.
- Tisné, S., Barbier, F., and Granier, C. (2011). The ERECTA gene controls spatial and temporal patterns of epidermal cell number and size in successive developing leaves of *Arabidopsis thaliana*. *Ann. Bot.* *108*, 159–168.

- 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.
- Tuinstra, M.R., Ejeta, G., and Goldsbrough, P.B. (1997). Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *Theor. Appl. Genet.* 95, 1005–1011.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 40, e115.
- Valentine, M.E., Wolyniak, M.J., and Rutter, M.T. (2012). Extensive Phenotypic Variation among Allelic T-DNA Inserts in Arabidopsis thaliana. *PLoS ONE* 7, e44981.
- Van Ooijen J.W., (2004). MapQTL Version 5, Software for the mapping of Quantitative Trait Loci in Experimental Populations.
- Vicente, M.R.-S., and Plasencia, J. (2011). Salicylic acid beyond defence: its role in plant growth and development. *J. Exp. Bot.* 62, 3321–3338.
- Wang, Y.H. (2008). How effective is T-DNA insertional mutagenesis in Arabidopsis? *J. Biochem. Technol.* 1, 11–20.
- Wang, S., Basten C.J., Zeng, Z.B. (2011). Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh.
- Weigel, D. (2012). Natural Variation in Arabidopsis: From Molecular Genetics to Ecological Genomics. *Plant Physiol.* 158, 2–22.
- Weight, C., Parnham, D., and Waites, R. (2008a). LeafAnalyser: a computational method for rapid and large-scale analyses of leaf shape variation RID B-4877-2008. *Plant J.* 53, 578–586.
- Weight, C., Parnham, D. (dan@openillusionist.org.uk), and Waites, R. (rw18@york.ac.uk). (2008b). LeafAnalyser: a computational method for rapid and large-scale analyses of leaf shape variation. *Plant J.* 55, 578–586.
- West, M.A.L., van Leeuwen, H., Kozik, A., Kliebenstein, D.J., Doerge, R.W., Clair, D.A.S., and Michelmore, R.W. (2006). High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. *Genome Res.* 16, 787–795.
- West, M.A.L., Kim, K., Kliebenstein, D.J., van Leeuwen, H., Michelmore, R.W., Doerge, R.W., and St Clair, D.A. (2007). Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. *Genetics* 175, 1441–1450.
- Wilson, I.W., Schiff, C.L., Hughes, D.E., and Somerville, S.C. (2001). Quantitative Trait Loci Analysis of Powdery Mildew Disease Resistance in the Arabidopsis thaliana Accession Kashmir-1. *Genetics* 158, 1301–1309.
- Wilson-Sánchez, D., Rubio-Díaz, S., Muñoz-Viana, R., Pérez-Pérez, J.M., Jover-Gil, S., Ponce, M.R., and Micol, J.L. (2014). Leaf phenomics: a systematic reverse genetic screen for Arabidopsis leaf mutants. *Plant J.* n/a – n/a.

- Wingler, A., Purdy, S.J., Edwards, S.-A., Chardon, F., and Masclaux-Daubresse, C. (2010). QTL analysis for sugar-regulated leaf senescence supports flowering-dependent and -independent senescence pathways. *New Phytol.* *185*, 420–433.
- de Wit, M., Spoel, S.H., Sanchez-Perez, G.F., Gommers, C.M.M., Pieterse, C.M.J., Voesenek, L.A.C.J., and Pierik, R. (2013). Perception of low red:far-red ratio compromises both salicylic acid- and jasmonic acid-dependent pathogen defences in *Arabidopsis*. *Plant J.* *75*, 90–103.
- Wu, R., Bradshaw, H., and Stettler, R. (1997). Molecular genetics of growth and development in *Populus* (Salicaceae). v. mapping quantitative trait loci affecting leaf variation. *Am. J. Bot.* *84*, 143–143.
- Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., et al. (2003). Empirical Analysis of Transcriptional Activity in the *Arabidopsis* Genome. *Science* *302*, 842–846.
- Yang, H., Shi, Y., Liu, J., Guo, L., Zhang, X., and Yang, S. (2010). A mutant CHS3 protein with TIR-NB-LRR-LIM domains modulates growth, cell death and freezing tolerance in a temperature-dependent manner in *Arabidopsis*. *Plant J. Cell Mol. Biol.* *63*, 283–296.
- Yang, L., Conway, S.R., and Poethig, R.S. (2011). Vegetative phase change is mediated by a leaf-derived signal that represses the transcription of miR156. *Dev. Camb. Engl.* *138*, 245–249.
- Zhang, X.-C., and Gassmann, W. (2003). RPS4-Mediated Disease Resistance Requires the Combined Presence of RPS4 Transcripts with Full-Length and Truncated Open Reading Frames. *Plant Cell* *15*, 2333–2342.
- Zhang, X.-C., and Gassmann, W. (2007). Alternative Splicing and mRNA Levels of the Disease Resistance Gene RPS4 Are Induced during Defense Responses. *Plant Physiol.* *145*, 1577–1587.
- Zhang, Y., Goritschnig, S., Dong, X., and Li, X. (2003). A Gain-of-Function Mutation in a Plant Disease Resistance Gene Leads to Constitutive Activation of Downstream Signal Transduction Pathways in suppressor of npr1-1, constitutive 1. *Plant Cell* *15*, 2636–2646.
- Zhao, K., Aranzana, M.J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., et al. (2007). An *Arabidopsis* Example of Association Mapping in Structured Samples. *PLoS Genet.* *3*.
- Zhu, Y., Qian, W., and Hua, J. (2010a). Temperature Modulates Plant Defense Responses through NB-LRR Proteins. *PLoS Pathog* *6*, e1000844.
- Zhu, Z., Xu, F., Zhang, Y., Cheng, Y.T., Wiermer, M., Li, X., and Zhang, Y. (2010b). *Arabidopsis* resistance protein SNC1 activates immune responses through association with a transcriptional corepressor. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 13960–13965.
- Zwieniecki, M.A., Boyce, C.K., and Holbrook, N.M. (2004). Hydraulic limitations imposed by crown placement determine final size and shape of *Quercus rubra* L. leaves. *Plant Cell Environ.* *27*, 357–365.
- The *Arabidopsis* genome initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* *408*, 796–815.

Zotz G., Wilhelm K., Becker A. (2011). Heteroblasty – A review. *The Botanical Review* 77: 109-151.