Appendix D

MASSEY UNIVERSITY
Application for Approval of Request to Embargo a Thesis
(Pursuant to AC98/168 (Revised 2), Approved by Academic Board 17/02/99)

Name of Candidate: REBECCA BLOOMER ID Number: 03146057

Degree: PHD Dept/Institute/School: IFS

Thesis title: THE MOLECULAR GENETIC BASIS OF NATURAL VARIATION IN TRICHOME DENSITY IN ARABIDOPSIS THALIANA

Name of Chief Supervisor: VAUGHAN SYMONDS Telephone Ext: 2503

As author of the above named thesis, I request that my thesis be embargoed from public access until (date) FEBRUARY 2015 for the following reasons:

☐ Thesis contains commercially sensitive information.

☐ Thesis contains information which is personal or private and/or which was given on the basis that it not be disclosed.

☑ Immediate disclosure of thesis contents would not allow the author a reasonable opportunity to publish all or part of the thesis.

☐ Other (specify):

Please explain here why you think this request is justified:

Release of this thesis could allow other labs working in this competitive field to publish similar findings. A two year embargo allows time to collect additional data, which is expected to result in a stronger, higher-impact publication.

Signed (Candidate): RBL Date: 7/2/12

Endorsed (Chief Supervisor): Date: 07/02/13

Approved/Not Approved (Representative of VC): Date: 7/02/2013

Note: Copies of this form, once approved by the representative of the Vice-Chancellor, must be bound into every copy of the thesis.

The molecular genetic basis of natural variation in trichome density in *Arabidopsis thaliana*

A thesis presented in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

in

Plant Biology

Institute of Fundamental Sciences

Massey University

Palmerston North

New Zealand

Rebecca Bloomer

2013

ABSTRACT

Understanding the genetic basis of natural variation in phenotypes is a central, yet often elusive, goal in evolutionary biology. Trichome density, an herbivory defence trait in *Arabidopsis thaliana*, is a powerful model for investigating natural phenotypic variation, combining a genetically well characterised developmental pathway with a quantitatively and qualitatively variable phenotype of selective importance. Here, Quantitative Trait Locus (QTL) mapping and candidate gene analyses were undertaken to explore the genetic basis of variation in trichome density in natural accessions of *A. thaliana*, under an overarching hypothesis that allelic variation in the epidermal development pathway plays a significant role.

QTL mapping for constitutive and wounding-induced trichome density and for plasticity of density was undertaken in two newly developed mapping populations, broadening the range of allelic variation sampled in trichome density studies. This study is the first to address the genetic architecture of induced density and plasticity, finding a genetic basis to plasticity and a surprising negative response to wounding among some members of the populations used. Some QTL mapped are unique, while others appear common to both constitutive and induced density phenotypes or to overlap across mapping populations, suggesting particular loci may be key to generating variation for trichome phenotypes. Epistatic interactions and candidate genes for QTL within, up- and downstream of the epidermal development pathway are identified.

Candidate gene analyses focussed on genes within the epidermal development pathway: the trichome-specific MYB *GL1* and the pleiotropic WD-repeat *TTG1*. In both *GL1* and *TTG1*, a pattern of high frequency polymorphism correlates with variation in trichome density. In *GL1*, variation has both qualitative and quantitative effects, with both protein-coding and regulatory changes proposed as underlying bases. The *TTG1* coding region is subject to strong purifying selection, and the observed quantitative effect on density appears to be based on variation in regulatory sequence. Both QTL mapping and candidate gene approaches support the hypothesis of a key role for the epidermal development pathway in generating variation in trichome density in *A. thaliana*, and more generally a role for variation in regulatory genes contributing to natural phenotypic variation.

ACKNOWLEDGEMENTS

In researching and writing this thesis I have had the support of a fantastic group of people, without whom it may not have been started, finished, or half the fun. Many thanks first to my advisor, Dr Vaughan Symonds. I could probably count on one hand the number of times it really was “just a quick question”, and I truly appreciate your time and expertise. Thanks to both you and Dr Jen Tate for fostering a supportive and challenging lab to grow up in - it is a pleasure to be a part of a lab with such great people and diverse research interests.

A number of people provided specific assistance along the way. Thanks to Professor Alan Lloyd for kindly donating *ttg1* and *gl1* mutant lines and transgenics vectors, and Jay Jayaraman and Dr Xiao Song for donating *Agrobacterium* stocks and helping me to get transgenics up and running in our lab. Thanks also to the many others who have lent reagents and lab equipment, given helpful advice and asked thought-provoking questions. My studies were supported by a Tertiary Education Commission Top Achievers’ Doctoral Scholarship, and I am grateful also for further travel support from both the New Zealand Society of Plant Biologists and the Institute of Fundamental Sciences.

The PhD experience would have been a lonely one without my LoST Lab (& Associates) colleagues, past and present: Todd, Matt, Rowan, Tina, Cindy, Fronny, Nick, Jill, Megan, Kay, Jessie, Amir and Prashant. Thank you all for your moral support, science and life advice, and ability to find the funny side in almost anything. Thanks also to my friends in the wider world, particularly Ruth, Rechelle and Emma, for your support, and for reminding me to get out once in a while.

I am also deeply grateful to my families, the Page/Bloomers, Ridens and Pursers, who have both modelled and inspired a love of learning and been unfailingly supportive of my long career as a student. I am especially thankful to Dan for planting the idea very early on and to Helen and Vicky for their encouragement when I most needed it. Lastly and mostly, I am thankful for the love, encouragement and patience of my husband, Dean - it’s been an adventure, and I am glad I have you to share it with.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

µM	Micromolar (micromoles per litre)
µmol	Micromole
µL	Microlitre
ABRC	*Arabidopsis* Biological Resource Centre
AI-RIL	Advanced Intercross-Recombinant Inbred Line
ANOVA	Analysis of variance
bHLH	Basic Helix-Loop-Helix
bp	Base pair
CASS	Cheaply amplified size standard
cDNA	Complementary deoxyribonucleic acid
cm	Centimetre
cM	CentiMorgan
CTAB	Hexadecyltrimethylammonium bromide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
GWAS	Genome-wide association study
H^2	Broad-sense heritability
HO	Hi-0 x Ob-0
HOC	Hi-0 x Ob-0 Constitutive
HOD	Hi-0 x Ob-0 Damaged
INRA	Institut National de la Recherche Agronomique/French National Institute for Agricultural Research
kb	Kilobase pairs
LB	Luria-Bertani (agar or broth)
LD	Linkage disequilibrium
LOD	Logarithm of odds
MAGIC	Multi-parent Advanced Generation InterCross
mg	Milligram
mM	Millimolar (millimoles per litre)
mm	Millimetre
MYB	Class of transcription factor
NA	Not applicable
ng	Nanogram
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
QTL	Quantitative trait locus/loci
RIL	Recombinant inbred line
RNA	Ribonucleic acid
SD	Segregation distortion
SNP	Single nucleotide polymorphism
SOE-PCR	Splicing by overlap extension-polymerase chain reaction

SS	St-0 x Sf-2
SSC	St-0 x Sf-2 Constitutive
SSD	St-0 x Sf-2 Damaged
SSR	St-0 x Sf-2 Response
TAD	Transcriptional activation domain
UTR	Untranslated region
UV-B	Ultraviolet B
WDR	WD repeat

ABBREVIATIONS OF GENES

All genes are in *Arabidopsis thaliana* unless otherwise stated.

AtMYC1	Epidermal development pathway bHLH
C1	MYB gene regulating anthocyanin production (*Zea mays*)
CPC	*CAPRICE*; Epidermal development pathway R3 MYB repressor
CRY2	*CRYPTOCHROME 2*; light-sensing protein
EGL3	*ENHANCER OF GLABRA3*; Epidermal development pathway bHLH
ETC1	*ENHANCER OF TRYPTICHON AND CAPRICE 1*; Epidermal development pathway R3 MYB repressor
ETC2	*ENHANCER OF TRYPTICHON AND CAPRICE 2*; Epidermal development pathway R3 MYB repressor
ETC3	*ENHANCER OF TRYPTICHON AND CAPRICE 3*; Epidermal development pathway R3 MYB repressor
FLC	*FLOWERING LOCUS C*; MADS-box transcription factor, regulates floral transition
FRI	*FRIGIDA*; flowering time determinant, confers vernalisation requirement
FT	*FLOWERING LOCUS T*; promotes flowering
GA1	*GIBBERELLIC ACID REQUIRING*; gibberellin biosynthetic gene
GI	*GIGANTEA*; phytochrome B signalling pathway gene
GL1	*GLABRA 1*; Epidermal development pathway R2R3 MYB
GL2	*GLABRA 2*; Epidermal development pathway downstream target gene
GL3	*GLABRA 3*; Epidermal development pathway bHLH
JAZ2	*JASMONATE-ZIM-DOMAIN PROTEIN 2*; jasmonic acid response regulator
JAZ9	*JASMONATE-ZIM-DOMAIN PROTEIN 9*; jasmonic acid response regulator
MYB113	Epidermal development pathway R2R3 MYB
MYB114	Epidermal development pathway R2R3 MYB
MYB23	Epidermal development pathway R2R3 MYB
MYB5	Epidermal development pathway R2R3 MYB
MYBL2	*MYB-LIKE 2*; Epidermal development pathway MYB-like repressor
PAP1	*PRODUCTION OF ANTHOCYANIN 1*; Epidermal development pathway R2R3 MYB

PAP2	*PRODUCTION OF ANTHOCYANIN 2*; Epidermal development pathway R2R3 MYB
RGL1	*REPRESSOR OF GA-LIKE 1*; DELLA-protein encoding, negative regulator of gibberellin response
RPM1	*RESISTANCE TO P. SYRINGAE PV MACULICOLA 1*; Defence response
SRK	*S-Receptor Kinase*; Self-incompatibility gene (*Brassica*)
TCL1	*TRICHOMELESS 1*; Epidermal development pathway R3 MYB repressor
TCL2	*TRICHOMELESS 2*; Epidermal development pathway R3 MYB repressor
TRY	*TRYPTICHON*; Epidermal development pathway R3 MYB repressor
TT2	*TRANSPARENT TESTA 2*; Epidermal development pathway R2R3 MYB
TT8	*TRANSPARENT TESTA 8*; Epidermal development pathway bHLH
TTG1	*TRANSPARENT TESTA GLABRA 1*; Epidermal development pathway WDR
TTG2	*TRANSPARENT TESTA GLABRA 2*; Epidermal development pathway downstream target gene
URM9/SAD2	*UNARMED9/SENSITIVE TO ABA AND DROUGHT2*; importin β-like protein linking jasmonic acid signalling to trichome initiation
WDR1	WDR gene regulating floral anthocyanin production (*Ipomoea* spp)
WER	*WEREWOLF*; Epidermal development pathway R2R3 MYB

1 INTRODUCTION

1.1 Introduction

Historically, our understanding of the genetic basis of natural phenotypic variation in plants has been based on the theoretical models of population genetics and empirical evidence from mapping studies, which have provided insight into the number and effect of loci underlying traits. The increasing accessibility to molecular data has allowed the genetic basis of natural variation to be characterised at higher resolution, pinpointing the specific genes and even nucleotides underlying trait variation and identifying selection at the single gene level. However, it is still difficult to make generalisations about the genetic architecture and molecular genetic bases of such variation. In cases where the source of phenotypic variation has been characterised at the molecular level, relatively few examples are from natural plant species. The model flowering plant *Arabidopsis thaliana* (Brassicaceae) has proved to be an excellent system for studies of natural variation in plants, and is the primary focus of this introduction.

The genetic basis of trichome density in *A. thaliana* provides an ideal model system for studying the molecular genetic basis of phenotypic variation: the species is a well developed model system with unprecedented resources available for its study; trichomes and other epidermal traits have demonstrated adaptive functions; and the developmental regulation of these traits is genetically well characterised. The nature of the regulatory pathway provides opportunities to address a range of topics from the genetic architecture of phenotypes, the importance of regulatory versus structural genetic changes to the roles of epistasis and pleiotropy in generating or constraining natural phenotypic variation.

1.2 *Arabidopsis thaliana* as a model for the genetics of natural variation

The model plant species *Arabidopsis thaliana* is an ideal organism for studying the molecular genetic basis of natural variation. It occupies a diverse range of habitats, with its current global distribution thought to have originated from refugia in Central Asia and the Iberian Peninsula during the Pleistocene glaciation before spreading across Europe, and subsequently into Eastern Asia and America by human activity (Beck*, et al.*, 2008). Stocks of lineages collected from populations across its global distribution, referred to in the *Arabidopsis* community as "accessions", are available, as is a significant number of mapping lines. *Arabidopsis thaliana* has a small genome of only five chromosomes, a generation time of only a few weeks, and a wide range of molecular tools have been developed for studying the plant. Being predominantly selfing, *A. thaliana* can generally be assumed to be homozygous. The *A. thaliana* genome is fully sequenced and annotated (*Arabidopsis* Genome Initiative, 2000) in the reference accession Col-0, and genome-wide patterns of polymorphism have been identified and compared for many accessions (e.g., Nordborg, *et al.*, 2005), allowing demographic effects to be taken into account in testing for selection or association at a given locus. With resequenced genomes of >500 natural accessions from both within and among globally distributed populations now online, an unprecedented amount of genomic data is available to the *Arabidopsis* research community.

1.3 Genetic diversity and natural phenotypic variation

1.3.1 The importance of genetic diversity

Genetic diversity provides the basis for evolution; genetic differences between individuals and populations underlie the phenotypic variation on which adaptation is based. Selection can act on either standing (existing) genetic variation or new mutations, and will lead to changes in allele frequencies over time in the adapting population (Barrett & Schluter, 2008). Evolution from selection on standing variation is likely to be faster than from new mutations, due to the immediate availability and higher starting frequency of alleles and the likelihood that they have been pre-tested by previous environments (Barrett & Schluter, 2008). Populations with high genetic diversity are positioned to quickly adapt to changes in their environment, and low levels of diversity are typically associated with inbreeding and reduced individual fitness.

1.3.2 The genetic architecture of natural variation

While the importance of genetic diversity is widely accepted, traits are often genetically complex. Relatively little is known about the number and effect of genetic changes which typically underlie phenotypic variation within and among plant populations, or about the role of interactions between these loci (epistasis) in generating variation. Similarly, it is not clear whether the underlying changes typically involve regulatory or structural genes, or changes of a regulatory or structural nature within a gene.

Quantitative trait locus (QTL) mapping studies can be used to estimate the minimum number and magnitude of effect of loci involved in variation in complex traits. In addition, QTL mapping has the potential to identify epistatic interactions between loci controlling a trait of interest (Malmberg*, et al.*, 2005). One type of mapping population results from crossing two parents and repeatedly selfing for multiple successive generations, until almost completely homozygous Recombinant Inbred Lines (RILs) are produced. RILs differ from one another in the parental contribution to the composition of their genomes due to recombination. Using molecular markers, trait variation among individuals can be attributed to a number of regions of the genome (QTL), giving an estimate of the number and relative effect of loci contributing to variation. Techniques such as fine-scale mapping using near-isogenic lines can then be used to improve resolution and assist with identification of candidate genes associated with trait variation within a given QTL; one such example is the fine mapping of the promoter region of the flowering time gene *FT* as a major QTL for flowering time variation in *A. thaliana* (Schwartz*, et al.*, 2009).

QTL mapping in traditional two-parent RIL populations has several limitations. Depending on marker density and recombination rates, QTL can potentially span hundreds of genes; therefore, care needs to be taken when interpreting QTL mapping results as close linkage of loci may result in multiple loci being identified as single QTL of larger relative effect. The use of advanced inter-cross (AI) RIL populations may help overcome this problem as these crosses provide more recombination, although a higher density of markers is also required (Balasubramanian*, et al.*, 2009). QTL are also specific to the environment in which the phenotypes are measured, with environmental effects determining both the QTL identified and the magnitude of effect of QTL observed (Ungerer*, et al.*, 2003). Significantly, QTLs

identified in two-parent RIL populations represent only a limited amount of genetic variation, which may be only a fraction of the variation within a species.

Two alternative approaches to RIL mapping have been developed in *A. thaliana* which resolve some of the limitations of mapping in two-parent populations. Genome-wide association studies (GWAS) (Atwell*, et al.*, 2010) and recently developed multi-parent RIL mapping populations (Kover*, et al.*, 2009; Huang*, et al.*, 2011) incorporate a broader sample of genetic variation than two-parent populations, allowing estimation of a more global genetic architecture of traits within the species. GWAS allow a survey of a large amount of genetic variation within a species; individuals within a population or species are genotyped and phenotyped to identify non-random associations between specific alleles and the trait of interest (Shindo*, et al.*, 2007). The broad sampling of genetic variation helps to improve the resolution of QTLs as, with suitable marker density, ancestral recombination events facilitate mapping at a finer scale (Bergelson & Roux, 2010). However, a significant problem pointed out in early studies such as that of Hagenblad et al. (2004) is that population structure can provide false positive results. More recent studies, such as the 107-trait GWAS undertaken by Atwell *et al.* (2010), use statistical models that account for population structure to reduce the false-positive rate. However, these models are fairly conservative and increase the likelihood of false negatives (Brachi*, et al.*, 2010). A further limitation of mapping using GWAS is that the low frequency of rare alleles reduces the chance of identifying significant associations with phenotype, and their effects may be missed (Kover & Mott, 2012).

Multi-parent RIL lines such as those generated by Kover *et al.* (2009) and Huang *et al.* (2011) incorporate a comparatively smaller portion of the genetic variation within *A. thaliana* and may not achieve as high a resolution of QTLs as GWAS due to fewer recombination events. However, they resolve the issue of confounding population structure as intercrossing shuffles the parents' genomes; furthermore, multiparent mapping populations control for allele frequency, increasing the likelihood of detection of rare allele effects (Kover & Mott, 2012). Unfortunately, currently no QTL mapping software supports the analysis of epistasis in multi-parent mapping populations. Recent reviews (Bergelson & Roux, 2010; Kover & Mott, 2012) have proposed the use of GWAS and QTL mapping together, taking advantage of high levels of genetic diversity and high resolution mapping in GWAS but confirming true positives with a QTL mapping approach.

Mapping studies support a range of theories of the genetic architecture of adaptation, from single-locus changes of large effect (Larkin*, et al.*, 1996) through to multi-locus changes with varying magnitudes of phenotypic effects (Martin*, et al.*, 2008). In *A. thaliana*, in some cases the nucleotide changes underlying QTLs have been identified. Examples include a QTL for flowering time localised to a single nonsynonymous SNP in the cryptochrome-encoding gene *CRY2* (El-Assal*, et al.*, 2001); a nonsynonymous SNP in the epidermal development bHLH *AtMYC1* underlying a QTL for trichome density (Symonds*, et al.*, 2011); and similarly a nonsynonymous SNP in the trichome initiation repressor *ETC2* as responsible for a separate QTL for trichome density (Hilscher*, et al.*, 2009). With high resolution mapping techniques the identification of strong candidate genes, and even candidate polymorphisms, for further analysis should become increasingly straightforward.

1.3.3 Candidate genes: selection, diversity and functional variation

Selection on a given gene can act in a number of different ways. Alleles which confer a fitness benefit will be positively selected for and may ultimately sweep to fixation, replacing all other alleles in a population, leaving no genetic diversity at the locus. An allele may carry a mutation with a negative consequence, in which case it will be selected against (purifying selection) and ultimately lost from the population. In cases where heterozygous individuals have the highest fitness or in a changing environment, balancing selection will favour coexistence of the two alleles. Populations of a species may experience diversifying selection, where interpopulation differences in selective pressures favour different alleles of the gene.

Patterns of interspecies or intraspecies nucleotide diversity can be used to detect selection within genes, using statistical analyses which usually assume selective neutrality at a given locus as the null hypothesis. Tests which consider the pattern of single-nucleotide polymorphism (SNPs) across alleles at a locus, such as Tajima's D (Tajima, 1989) can be used to identify selection. A recent selective sweep at a locus results in an excess of low frequency SNPs compared to loci evolving neutrally, while balancing selection acting on a locus will result in a higher number of intermediate-frequency SNPs than expected under neutrality. Comparing polymorphism within a gene to the level of divergence at a gene between one species and a closely related outgroup can also indicate selection (Hudson*, et al.*, 1987). Under neutrality, levels of within-group polymorphism and between-group divergence within

a gene should be proportional, while deviation from this is indicative of selection acting on the gene. A similar indicator is based on the ratio of synonymous to non-synonymous substitutions within a protein-coding region (the d_n:d_s or *K*a/*K*s ratio) where a ratio of one indicates neutrality and a ratio which deviates from one indicates selection. *K*a/*K*s ratios can be calculated from within or between species data.

Nucleotide diversity, a measure of the degree of polymorphism in a gene or genome, has been investigated in a number of plant species both at loci of known phenotypic effect and at multiple unidentified loci across genomes. Two screens of nucleotide polymorphism across randomly distributed genome-wide loci in *A. thaliana* have found that the pattern of polymorphism observed does not fit standard neutral models, with high variation between genomic regions and a high level of rare alleles observed across the genome, at least partly influenced by selection (Nordborg*, et al.*, 2005; Schmid*, et al.*, 2005). Both studies found evidence of global population structure with demographic history playing a significant role in the patterns of polymorphism identified, but disagree over the role of selfing in the level of polymorphism observed. These results highlight the importance of taking demographic history into account when considering selection on a given locus. Multi-locus comparisons of patterns of polymorphism have also been used in other species, for example to estimate the number of genes involved in domestication of maize (Wright*, et al.*, 2005); to examine the contribution of demographic history to nucleotide diversity in European populations of *Pinus sylvestris* (Pyhajarvi*, et al.*, 2007); and to assess the roles of population structure and range expansion in evolution of wild tomato species (Arunyawat*, et al.*, 2007).

There are fewer cases in plants where natural allelic variation and nucleotide diversity have been measured at specific loci with known phenotypic effects. In *A. thaliana*, natural accessions have been sequenced at a number of loci with known roles. For example, a comparison of 27 plant defence response *R* genes, incorporating genome-wide SNP data across 96 accessions, found evidence of balancing selection in several genes (Bakker*, et al.*, 2006). One *R* gene, *RPM1,* is found in natural populations as a presence/absence polymorphism and has been correlated with resistance to the pathogen *Pseudomonas syringae* in transgenic lines which differ only in the presence or absence of *RPM1* (Tian*, et al.*, 2003). Comparison of allelic variation in the female self-incompatibility gene *SRK* has lead to the identification of individual codons under balancing selection in *Arabidopsis halleri*,

Arabidopsis lyrata and *Brassica*, outcrossing relatives of *A. thaliana* (Castric & Vekemans, 2007). The trichome regulatory gene *GLABRA1* (*GL1*) has been surveyed across a number of *A. thaliana* accessions, finding two haplotype clusters and only weak evidence of selection (Hauser*, et al.*, 2001). Grouping of alleles into two major haplotype clusters has been found in other analyses of nucleotide sequence divergence in *A. thaliana*, including the epidermal development pathway genes *AtMYC1* (Symonds*, et al.*, 2011) and *ETC2* (Hilscher*, et al.*, 2009), phenylpropanoid pathway (Aguade, 2001), chitinase (Kawabe*, et al.*, 1997) and R genes (Tian*, et al.*, 2002; Mauricio*, et al.*, 2003; Rose*, et al.*, 2004), and in chloroplast sequences (Yin*, et al.*, 2010).

The phenotypic consequences of allelic variation can be demonstrated using approaches such as association mapping or transgenics. A 2009 review (Alonso-Blanco*, et al.*, 2009) identified ~30 cases where genes and functional polymorphisms of known phenotypic effect have characterised effects on natural variation in *A. thaliana*. While many more examples exist, with almost 100 such genes identified in the 2009 review, these are primarily from crop species. The genetic basis of natural variation in flowering time in *A. thaliana* has been particularly well studied; the *FRIGIDA* (*FRI*) locus in *A. thaliana* is a major contributor to variation in flowering time, with multiple loss-of-function alleles segregating in natural populations and functional *FRI* alleles contributing to an allele-dependent latitudinal cline in flowering time (Stinchcombe*, et al.*, 2004). *FLC* is similarly involved in regulating flowering in response to vernalisation, with both protein truncations and independent transposon insertions in regulatory regions of *FLC* contributing to variation for this trait (Gazzani*, et al.*, 2003; Michaels*, et al.*, 2003). In the closely related species *A. lyrata*, two *FRI* alleles confer a 15 day separation of flowering time in association studies, an effect which was confirmed by transformation of the alleles into *A. thaliana* (Kuittinen*, et al.*, 2008). Other examples of characterised allelic variation include variation in the phytochrome B signalling pathway gene *GIGANTEA* (*GI*), which appears to contribute to reproductive fitness with a significant association between *GI* haplotype and fruit set (Brock*, et al.*, 2007); variation in the gibberellin biosynthesis gene *GIBERELLIC ACID REQUIRING1* (*GA1*), which underlies changes in floral organ size (Brock*, et al.*, 2012); and *CRY2*, *AtMYC1* and *ETC2* as described above.

1.3.4 Regulatory versus structural change in natural phenotypic variation

It is not clear whether regulatory or protein coding mutations play a more significant role in trait variation, or whether changes in regulatory genes such as transcription factors are more likely than changes in enzymatic or structural genes. Broadly, mutations with pleiotropic effects are considered less likely to underlie variation as they are more likely to have deleterious outcomes for one or more traits. On this basis, it has been argued that variation is likely to be based on changes in the regulatory regions of genes as their modular nature makes mutations less likely to have deleterious pleiotropic effects (Carroll, 2005), and that transcription factors are more likely to be modified than structural or enzymatic protein coding genes (Doebley & Lukens, 1998). However, recent reviews have argued for a role for protein-coding changes, particularly in transcription factors, in adaptive evolution (Hoekstra & Coyne, 2007; Lynch & Wagner, 2008).

Variation in transcription factors has been proposed to play a key role in generating phenotypic variation (Lynch & Wagner, 2008). Supporting this, in a recent review of studies of natural variation in plants (Alonso-Blanco*, et al.*, 2009), a higher than expected proportion of the genes identified as underlying variation are transcription factors (Martin*, et al.*, 2010). However, not all transcription factors are equivalent; the position of a transcription factor within a regulatory hierarchy, and hence how pleiotropic it is, may determine how likely it is to play a role in generating phenotypic variation. Transcription factors acting at an early point in a biochemical pathway or developmental process may be subject to stronger evolutionary constraint than transcription factors acting further down a pathway or on a more specialised developmental regulatory process with fewer direct downstream transcriptional targets (Martin*, et al.*, 2010). Comparison of *K*a/*K*s ratios in transcription factors interacting in the *Ipomoea* WDR:bHLH:MYB complex support this hypothesis, with the WD-repeat (WDR) protein, which is widely expressed and mediates multiple protein:protein interactions, exhibiting much stronger constraint than its interacting bHLH and MYB transcription factor partners (Streisfeld*, et al.*, 2011).

Evidence has been found for both regulatory and protein coding region changes contributing to phenotypic variation, and for both regulatory and structural genes to play a role. For example, low freezing tolerance in four *A. thaliana* accessions has been associated with

reduced expression of cold acclimation response regulatory genes, with a promoter region deletion found in two accessions, a promoter region insertion found in one accession and an insertion found in the 3' UTR of another accession (Zhen & Ungerer, 2008). Similarly, *FLC* alleles with reduced expression have arisen independently in *A. thaliana* multiple times and are correlated with changes in flowering time (Gazzani*, et al.*, 2003; Michaels*, et al.*, 2003). Natural variation in trichome density in *A. thaliana* accessions has been attributed to amino acid replacements in the transcription repressor *ETC2* (Hilscher*, et al.*, 2009) and activator *AtMYC1* (Symonds*, et al.*, 2011) as described above. While many mutations reported are regulatory, there are a number of cases of amino acid replacements in non-regulatory genes; in *A. thaliana*, the amino acid replacement described above in the light-sensing flavoprotein *CRYPTOCHROME 2* (*CRY2*) is one such example (El-Assal*, et al.*, 2001).

1.4 Trichome density in *Arabidopsis thaliana*

1.4.1 The epidermis: the interface between plant and environment

The first point of contact between the plant and its environment is the epidermis, a tissue with roles of adaptive importance. The epidermis is made up of several cell types: relatively unspecialised pavement cells which act as a boundary layer and provide mechanical strength to the epidermis; highly specialised stomatal guard cells and trichomes on stems and leaves; conical-papillate cells on petals; and root hairs on the root epidermis. Stomata allow gas exchange and regulate both temperature and water balance in the plant. Conical-papillate cells are found on petals and some other functionally specialised floral organs, suggesting a role in pollinator attraction; supporting this, conical-papillate cells improve pollination success in *Antirrhinum majus* compared to a flat-celled mutant (Glover & Martin, 1998). Root hairs greatly increase the surface area of the root, provide anchorage, facilitate water and nutrient uptake and mediate interactions between the plant and nitrogen-fixing microorganisms (Gilroy & Jones, 2000).

Secondary metabolite production in epidermal cells also has important functions. Anthocyanins play a range of proposed roles in different tissues, many of which may be of adaptive importance. These include pollinator and seed disperser attraction in flowers and fruits, photoprotection, cold and drought tolerance and protection from herbivory in shoots, leaves and roots (Chalker-Scott, 1999). Seed coat metabolites such as tannins and mucilage have proposed roles in regulating germination and in protecting the seed from predation. *Arabidopsis thaliana* mutants defective in flavonoid production show reduced seed dormancy in fresh seeds and lower seed viability in four year old seeds (Debeaujon*, et al.*, 2000), while faster germination in tomato mutants with reduced seed coat tannin has also been found (Atanassova*, et al.*, 2004).

1.4.2 The protective roles of trichomes

Trichomes are specialised epidermal cell structures which protrude from the epidermal surface. Trichomes are found in nearly all land plants with a diverse range of structures in different plant species; they can be single- or multi-celled, of simple or more complex shape and glandular or non-glandular. They are found predominantly on leaves and stems where

they have a range of plant protective functions in different species. Trichomes are also found on floral organs such as the stigma and anthers, where their roles include pollen capture, protection of pollen from predatory insects and structural roles in floral organ morphology (Martin & Glover, 2007). Trichome initiation and density are developmentally regulated, but are also induced in *A. thaliana* in response to artificial wounding and the plant hormones jasmonic acid and gibberellins (Perazza*, et al.*, 1998; Traw & Bergelson, 2003).

Adaptive roles of leaf trichomes have been studied in a number of plant species, with important functions in plant protection from both biotic and abiotic factors. Trichomes have been shown to improve drought tolerance and water use efficiency (Hogan*, et al.*, 1994; Choinski & Wise, 1999; Ennajeh*, et al.*, 2006) and to protect leaves from high light (Liakopoulos*, et al.*, 2006) and UV-B radiation (Karabourniotis*, et al.*, 1995). Trichomes also act as protection from biological threats; in some plant species, such as *Datura stromonium* (Valverde*, et al.*, 2001), trichome density has a role in protection from insect herbivory, while an additional function is in resistance to fungal infection (Lai*, et al.*, 2000).

The role of trichomes in protection from herbivory has been studied extensively in the genus *Arabidopsis*, including in *A. thaliana* (Mauricio, 1998; Handley*, et al.*, 2005), *A. lyrata* (Kivimaki*, et al.*, 2007), *A. halleri spp gemmifera* (Kawagoe*, et al.*, 2011) and *A. kamchatica* (Steets*, et al.*, 2010). Insect predation has been shown to exert positive selection on increased trichome density in *A. thaliana* (Mauricio & Rausher, 1997); however, increased trichome density also imposes fitness costs in terms of viable seed production in the absence of pests as demonstrated in *A. thaliana* (Mauricio, 1998) and in terms of standardised growth rates (Zust*, et al.*, 2011), and has also been associated with increased susceptibility to the fungal necrotroph *Botrytis cinerea* (Calo*, et al.*, 2006). A cost to trichome production under some environmental conditions is further implied by evidence of divergent selection at the locus coding for glabrousness in *A. lyrata* (Karkkainen*, et al.*, 2004). With both costs and benefits to the plant demonstrated for trichome production, trichome density is likely to be a trait of selective importance. Interestingly, trichome initiation is up-regulated in *A. thaliana* in response to artificial wounding (Traw & Bergelson, 2003), suggesting phenotypic plasticity as a mechanism by which plants manage this cost:benefit trade off. Induction by wounding has also been shown in *Mimulus guttatus* (Holeski*, et al.*, 2010), and induction as a result of herbivory in the *A. thaliana* relative *Brassica nigra* (Traw, 2002).

1.4.3 Trichome initiation shares a regulatory mechanism with other epidermal traits

Trichome initiation, non-hair root epidermal cell specification, anthocyanin biosynthesis, seed coat pigmentation and seed coat mucilage production all require interaction between a WD-repeat protein, bHLH transcription factor and an R2R3-MYB transcription factor (Figure 1.1). Together, these proteins form an activation complex that binds the promoters of target genes (Morohashi*, et al.*, 2007; Zhao*, et al.*, 2008). In addition, a number of single repeat MYBs have been identified as negative regulators of epidermal development (Wang*, et al.*, 2008). Some components of this regulatory mechanism, for example *TTG1* and *GL2*, may also influence stomatal cell patterning of the hypocotyl (Berger*, et al.*, 1998).

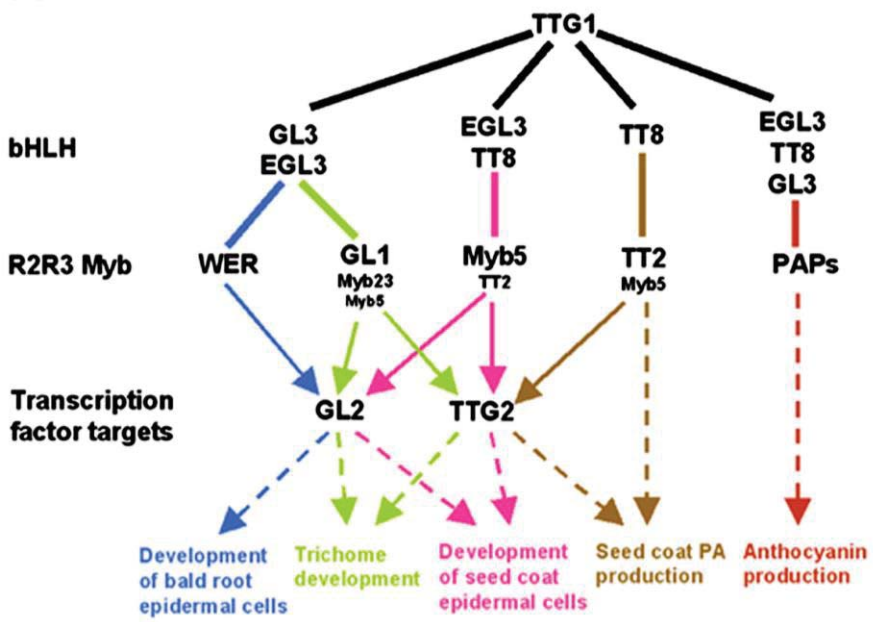

Figure 1.1: A model for interaction between regulators of *TTG1*-dependent traits in *A. thaliana*.

TTG1, in conjunction with a bHLH and a MYB, are necessary for expression of all traits. bHLHs exhibit partial functional redundancy and R2R3 MYBs are generally specific for a given trait. Figure copied from Gonzalez*, et al.* (2009).

The WD-repeat protein TTG1 is common to all epidermal pathways, including some aspects of stomatal patterning. The *ttg1* knockout mutation is pleiotropic, with plants being completely glabrous, lacking anthocyanins, seed coat pigments and mucilage, and producing more root hairs than the wild-type (Koornneef, 1981). Peculiarly, while strong mutant alleles of *ttg1* are glabrous, weak alleles exhibit a clustered trichome phenotype which seems to imply a role for TTG1 as both an activator and repressor of trichome initiation (Larkin*, et al.*, 1999). TTG1 has

no known DNA binding activity but has been shown to interact with bHLHs (Payne*, et al.*, 2000). *TTG1* is expressed at a low level in most plant tissues and once translated can move between cells in a GL3-dependent manner (Bouyer*, et al.*, 2008). In developing leaves *TTG1* is expressed throughout the epidermal layer (Bouyer*, et al.*, 2008; Zhao*, et al.*, 2008) and has been hypothesized to localise to initiating trichomes by moving into the cell and becoming trapped in a complex with GL3 (Bouyer*, et al.*, 2008; Balkunde*, et al.*, 2011), with depletion of TTG1 in the surrounding cells contributing to trichome spacing. This "depletion trapping" mechanism has been modelled mathematically and may explain the clustering phenotype of weak alleles of *TTG1* (Bouyer*, et al.*, 2008).

While all traits share TTG1 as a regulator, the trichome and non-root hair cell fate, flavonoid biosynthesis and seed coat development also require expression of one of four bHLH transcription factors so far identified. The bHLH proteins GL3, EGL3 and TT8 are closely related (Toledo-Ortiz*, et al.*, 2003) and show some trait specificity but are partially functionally redundant. Recently a fourth bHLH gene, *AtMYC1*, was shown to have roles in trichome and root hair development (Symonds*, et al.*, 2011; Zhao*, et al.*, 2012). bHLH proteins bind DNA as dimers and the bHLH proteins in the pathway could potentially act as heterodimers (Toledo-Ortiz*, et al.*, 2003). The functional redundancy of these genes may allow specific bHLHs to be involved in regulation of traits under different environmental conditions. For example *GL3*, but not *EGL3*, has been shown to regulate anthocyanin production under conditions of nitrogen depletion (Feyissa*, et al.*, 2009).

A *gl3* single mutant reduces trichome number and alters trichome branching, and overexpression increases trichome number in the wildtype and partially restores trichomes, anthocyanins and seed coat mucilage in the *ttg1* mutant background (Payne*, et al.*, 2000). However evidence suggests that *GL3* does not normally play a role in mucilage production (Payne*, et al.*, 2000). An *egl3* mutant has a reduced trichome number and a *gl3-egl3* double mutant is completely glabrous, defective in root hair spacing, with lower anthocyanin accumulation and defective seed coat mucilage (Zhang*, et al.*, 2003), indicating some redundancy between the two proteins. Mutations in *tt8* affect proanthocyanidin biosynthesis in the seed coat, with mutants lacking seed coat tannins and exhibiting lower expression of late flavonoid biosynthesis genes (Nesi*, et al.*, 2000) but normal anthocyanin biosynthesis (Zhang*, et al.*, 2003); a role for *TT8* in leaf marginal trichome formation has also been shown

(Maes*, et al.*, 2008). A *gl3-egl3-tt8* triple mutant results in a *ttg1*-like phenotype with absolutely no anthocyanin biosynthesis (Zhang*, et al.*, 2003), demonstrating that TT8 is also functionally redundant with GL3 and EGL3 for anthocyanin biosynthesis. While all other epidermal traits discussed show some overlap of bHLH protein function, only *TT8* has a demonstrated role in development of seed coat mucilage (Baudry*, et al.*, 2006). A fourth closely related bHLH, *AtMYC1*, plays a role in trichome and root hair development, (Zhao*, et al.*, 2012). *AtMYC1* appears to act as an enhancer of *GL3* and *EGL3* with natural allelic variation shown to influence trichome density; a single nonsynonymous substitution in a MYB interaction domain was found to be responsible for this phenotype, abolishing binding with TTG1 and GL1 in yeast-two-hybrid experiments (Symonds*, et al.*, 2011).

Each of the traits regulated by TTG1 is also controlled by a trait-specific R2R3 MYB transcription factor. R2R3 MYBs are transcriptional activator proteins, with an N-terminal DNA binding and protein-protein interaction domain comprising two imperfect MYB repeats, and a C-terminal domain which may be capable of transcriptional activation (Wang*, et al.*, 2008). The leaf trichome cell fate is specified by *GL1* (Marks & Feldmann, 1989) and a *gl1* mutant has completely glabrous leaf blades although retains trichomes on petioles. *MYB23* has a role in trichome branching and in initiation of trichomes on the leaf margin (Kirik*, et al.*, 2005), and *MYB5* has recently been shown to have a minor role in trichome branching (Gonzalez*, et al.*, 2009; Li*, et al.*, 2009). *WER*, a MYB with functional equivalence to but distinct expression pattern from *GL1* (Lee & Schiefelbein, 2001) specifies the non-hair cell fate in the root epidermis (Lee & Schiefelbein, 1999). In contrast to trichomes, root hairs are regulated in a positionally dependent manner (Lee & Schiefelbein, 1999). Anthocyanin biosynthesis is regulated by *PAP1*, *PAP2*, *MYB113* and *MYB114* (Gonzalez*, et al.*, 2008), while seed coat pigmentation is regulated by *TT2* (Nesi*, et al.*, 2001). Seed coat mucilage is controlled by *MYB5* (Gonzalez*, et al.*, 2009; Li*, et al.*, 2009). *MYB5* is the only *MYB* gene so far shown to affect more than one epidermal trait, with *myb5* mutants having a minor effect on seed coat pigment and leaf trichome branching phenotypes in addition to its major role in seed coat mucilage regulation (Gonzalez*, et al.*, 2009; Li*, et al.*, 2009). With the exception of *MYB5*, all *MYBs* implicated in epidermal trait regulation fall into a single monophyletic clade; *MYB5* is sister to this group (Stracke*, et al.*, 2001), which may explain its more diverse functions in contrast with the highly trait-specific MYBs.

Combinatorial control by bHLHs and MYBs is best characterised in flavonoid biosynthetic pathways where the MYBs and bHLH proteins involved have been identified in a number of species, including the model species *Zea mays* and *Petunia hybrida*, and recently in apple (*Malus x domestica*) (Espley*, et al.*, 2007). Work on anthocyanin regulatory genes in *Z. mays* identified the MYB domain residues in the R3 repeat required for interaction of the maize MYB C1 with the bHLH R (Grotewold*, et al.*, 2000). A conserved amino acid signature has been identified in the first two helices of the R3 MYB domain which mediates interaction between the MYB and bHLH proteins. Two residues were shown by transient promoter activation assays as being critical for stability of the interaction (Zimmermann*, et al.*, 2004). Direct interaction has been demonstrated in *A. thaliana* between MYBs and bHLHs and between bHLHs and TTG1. GL3 interacts with both GL1 and TTG1 in yeast two-hybrid assays, but TTG1 and GL1 do not interact directly (Payne*, et al.*, 2000); AtMYC1 has also been shown to interact with both TTG1 and GL1 (Zhao*, et al.*, 2012). Similarly, TTG1, TT8 and TT2 interact in yeast two- and three-hybrid analyses to form a complex that directly activates proanthocyanidin biosynthetic genes (Baudry*, et al.*, 2004). The TTG1:bHLH:MYB complex has been shown to act *in vivo*, directly targeting promoters of downstream genes such as the transcription factors *TTG2* and *GL2* and the negative pathway regulators *CPC* and *ETC1* (Zhao*, et al.*, 2008). Furthermore, transcription of the downstream gene *GL2* requires concurrent binding of both GL3 and GL1 to its promoter (Wang & Chen, 2008).

A fourth set of proteins, the single repeat R3 MYBs, act as negative regulators in the epidermal development pathway. In contrast to the R2R3 MYBs, single repeat MYBs have only an R3 MYB domain and no transcriptional activation domain (Schellmann*, et al.*, 2002), and the recognition sequence for R2R3 MYB:bHLH binding is completely conserved (Wang*, et al.*, 2008). This provides the putative basis for their function as negative regulators of epidermal development; R3 MYBs compete with R2R3 MYBs for bHLH binding, resulting in the formation of stable TTG1:bHLH:R3 MYB complexes which cannot activate transcription of downstream genes. This model fits both experimental data and mechanistic modelling of trichome spacing (Digiuni*, et al.*, 2008).

A total of eight single repeat *MYB* genes have so far been identified as negative regulators in the epidermal development pathway. Seven of these, *TRY*, *CPC*, *ETC1*, *ETC2*, *ETC3*, *TCL1* and *TCL2* (Wang*, et al.*, 2008; Gan*, et al.*, 2011; Tominaga-Wada & Nukumizu, 2012), have been

shown to have distinct but overlapping roles in suppressing trichome formation on normally glabrous organs, spacing trichomes on leaves and stems and establishing root hair patterning. An eighth R3 *MYB*, *MYBL2*, regulates anthocyanin and seed coat proanthocyanidin production but has no effect on trichome or root hair patterning or mucilage production (Dubos*, et al.*, 2008; Matsui*, et al.*, 2008). *MYBL2* is considerably longer than the other R3 MYBs, with a partial R2 domain and a repression domain (Matsui*, et al.*, 2008). In addition to *MYBL2*, *CPC* has recently been shown to have a role in regulating anthocyanin accumulation, with a possible role in regulating accumulation under stress (Zhu*, et al.*, 2009). *CPC* appears to act by competing with *PAP1* and *PAP2* for bHLH binding (Zhu*, et al.*, 2009). As yet no single repeat *MYB*s have a demonstrated role in regulation of seed coat traits.

The seven closely related R3 MYBs controlling epidermal cell fate have been shown to interact with GL3 (Wang*, et al.*, 2008), suggesting that they can all act through binding to GL3 in competition with the activating R2R3 MYBs. Both CPC and TRY can move between cells (Kurata*, et al.*, 2005; Digiuni*, et al.*, 2008; Zhao*, et al.*, 2008) and the motif identified as necessary for CPC movement (Kurata*, et al.*, 2005) is conserved in all seven, suggesting that all can move between cells. TCL1 has been shown to bind the *GL1* promoter and directly inhibit *GL1* expression (Wang*, et al.*, 2007), and TCL2 similarly suppresses *GL1* expression although interaction with the *GL1* promoter has not been demonstrated (Gan*, et al.*, 2011).

The regulation of trichome spacing has been the most extensively studied of all epidermal traits as a model for *de novo* pattern formation. Trichomes occur as evenly spaced outgrowths across the leaf epidermis, with no evidence that this spacing is determined by cell lineage (Larkin*, et al.*, 1996); all cells initially have an equal chance of becoming trichomes. Trichome spacing in *A. thaliana* can be explained by two models which are not necessarily mutually exclusive (Pesch & Hulskamp, 2009). In the activator-inhibitor model the TTG1:bHLH:GL1 complex is an activator of the trichome cell fate, promoting expression of downstream specifiers of trichome cell fate such as *GL2*. In addition, the complex can activate expression of a subset of the single repeat MYBs (Wang*, et al.*, 2008; Zhao*, et al.*, 2008), which are translated and then move to neighbouring cells. The single-repeat MYBs compete with GL1 for TTG1:GL3 binding in the neighbouring cells, and once bound are unable to activate transcription of *GL2*. As a result, the trichome cell fate is repressed in cells adjacent to the newly initiating trichome. This is largely in agreement with a theoretical model for

pattern formation first proposed by Meinhardt and Gierer (1974). However, key to this model is that the activator must also activate itself, which has not yet been experimentally demonstrated. In the activator-depletion model, TTG1 (the activator) is expressed in all cells but moves freely between them and is 'trapped' in initiating trichomes through complex formation with GL3 (Bouyer*, et al.*, 2008). This results in depletion of the activator in surrounding cells, generating the observed spacing pattern. Further layers of regulation seem likely, for example through single repeat MYBs directly binding promoters of activation complex targets to prevent transcription as shown for TCL1 (Wang*, et al.*, 2007), or through differential binding affinities and mobility of interacting proteins as shown in mobility and binding of single repeat MYBs to GL3 (Wester*, et al.*, 2009).

1.5 Trichome density in *Arabidopsis thaliana* as a model for the genetics of natural variation

Natural variation in trichome density in *A. thaliana* provides an excellent system for investigating the molecular genetic basis of natural phenotypic variation. *Arabidopsis thaliana* is an ideal model species, with accessions from globally distributed populations publically available and a wide range of molecular tools and techniques developed for its study. Trichome density is highly variable within the species and is an adaptive trait, protecting *A. thaliana* from herbivory in environments where insects are present but decreasing fitness in pest-free environments. Furthermore, the regulation of trichome density is genetically well characterised.

Evidence from previous studies suggests that naturally occurring genetic variation in the pathway contributes to variation in trichome density in *A. thaliana* and congeneric species. QTL mapping studies (Larkin*, et al.*, 1996; Mauricio, 2005; Symonds*, et al.*, 2005; Pfalz*, et al.*, 2007) and GWAS (Atwell*, et al.*, 2010) have identified many genomic regions contributing to variation in trichome density, and candidate genes from within the epidermal development pathway have been proposed for a number of these loci (Symonds*, et al.*, 2005; Hilscher*, et al.*, 2009; Atwell*, et al.*, 2010). Direct analysis of candidate genes within the pathway supports a role for pathway genes in generating natural variation: a single nucleotide change in the R3 MYB *ETC2* appears to contribute to variation in natural populations of *A. thaliana* (Hilscher*, et al.*, 2009), while more recently, a natural allele of the bHLH *AtMYC1* was shown to reduce

trichome density in some accessions (Symonds*, et al.*, 2011). In the sister species *A. lyrata* (Kivimaki*, et al.*, 2007) and *A. halleri subsp gemmifera* (Kawagoe*, et al.*, 2011), variation at the *GL1* locus is associated with an absence of trichomes in field populations; in *A. lyrata* these changes are also associated with increased susceptibility to herbivory, and analysis of sequence data from these and other field populations indicate that trichome production is under divergent selection in *A. lyrata* (Karkkainen*, et al.*, 2004).

The complex nature of the epidermal development pathway provides opportunities to address a range of topics in natural variation within a genetically well understood framework. Questions of genetic architecture include determining the number and effects of loci (as a minimum hypothesis) underlying variation in trichome density and plasticity phenotypes; whether loci underlying density are typically unique to an environment or population or have more global effects; and what role epistasis plays in variation for trichome density. With candidate gene studies, questions of interest include whether there is evidence of selection acting on the genes underlying variation in density; whether the molecular changes underlying variation in phenotypes are protein-coding or regulatory; and whether the nature of these changes differs for quantitative and qualitative changes in a trait or for pleiotropic compared with trait-specific genes. In addition, allelic variation within the pathway provides an opportunity to consider how well knowledge gleaned from functional genetics predicts the roles of genes in natural accessions. Despite a recent report that almost 10% of protein-encoding genes in *A. thaliana* are missing or knocked out in natural accessions (Clark*, et al.*, 2007), most functional genetics in *A. thaliana* is performed in a handful of common genetic backgrounds, and it is thus unclear how well findings predict gene effects in nature. Further, exploiting natural mutations may lead to an improved understanding of gene functions within the pathway.

Under an overarching hypothesis that variation in genes in the TTG1 epidermal development pathway contributes to natural variation in trichome density in *A. thaliana*, this thesis aims to address such questions using two main approaches. Here, quantitative trait locus (QTL) mapping of trichome density variation in two newly developed two-parent RIL mapping populations and two environments is used to determine the minimum number and magnitude of effect of loci affecting variation in trichome phenotypes and to identify potential genes of interest for further work. Secondly, the candidate genes *TTG1* and *GL1* are

characterised initially through sequencing in multiple natural accessions to identify genetic variation and signatures of molecular evolution. The effect of allelic variation on phenotype is assessed by association mapping, and the molecular mechanisms responsible for variation are investigated.

2 THE GENETIC ARCHITECTURE OF CONSTITUTIVE AND INDUCED TRICHOME DENSITY IN TWO NEW RECOMBINANT INBRED LINE POPULATIONS OF *ARABIDOPSIS THALIANA*: PHENOTYPIC PLASTICITY, EPISTASIS, AND LEAF DAMAGE RESPONSE

2.1 Abstract

Herbivory imposes an important selective pressure on plants. In *Arabidopsis thaliana* leaf trichomes provide a key defence against insect herbivory; however, trichome production incurs a fitness cost in the absence of herbivory. *Arabidopsis thaliana* is known to increase trichome density in response to leaf damage, suggesting a mechanism by which the cost associated with constitutively high trichome density might be mitigated. Here, we describe the mapping of quantitative trait loci (QTL) for constitutive and damage induced trichome density in two new recombinant inbred line populations. Mapping for constitutive and induced trichome density also allowed for the investigation of damage response loci. Interestingly, two of the four parental accessions and multiple RILs in each population demonstrate negative responses to leaf damage; i.e., they had lower trichome density following leaf damage. QTL were identified spanning regions unique to one population or trait, as well as loci which appear to overlap across populations or traits. Our results also identify a QTL for a trichome density response to leaf damage. For several of these mapped regions obvious candidate genes were identified drawing from the extensive literature describing trichome initiation in *A. thaliana*. These data provide further evidence that genetic variation in the TTG1 epidermal cell fate pathway likely underlies much of the variation in trichome density found in natural accessions of *A. thaliana.* Consistent with this, epistatic interactions appear to make a significant contribution to the genetic architecture of trichome density as well. Together, these results provide further insights into the genetic architecture of trichome density in *A. thaliana* specifically and to our understanding of the genetic underpinnings of natural variation generally.

2.2 Introduction

Insect herbivory is a significant selective pressure in plant populations, with herbivores consuming some 10-15% of all plant biomass produced annually (Cyr & Pace, 1993). In response, plants produce an array of deterrents, ranging from physical structures such as thorns or trichomes to a variety of unpalatable or toxic chemical defences. The model plant species *Arabidopsis thaliana* employs both physical and chemical defence strategies: most natural accessions produce both leaf trichomes and glucosinolates, a group of defensive secondary metabolites produced by members of the Brassicaceae. In natural populations of *Arabidopsis thaliana* (Mauricio, 1998) and in the closely related *Arabidopsis lyrata* (Kivimaki*, et al.*, 2007), leaf trichomes provide protection against insect herbivory. Trichome density is negatively correlated with damage resulting from herbivory (Mauricio, 1998), with predation in the field shown to exert positive selection on increased trichome density (Mauricio & Rausher, 1997). However, trichome production also has fitness costs in *A. thaliana*, both in terms of fruit production (Mauricio, 1998) and standardized growth rate (Zust*, et al.*, 2011). Similarly, a fitness cost for trichomes has been shown in the wild relatives *Arabidopsis kamchatica* (Steets*, et al.*, 2010) and *Arabidopsis halleri* ssp *gemmifera* (Kawagoe*, et al.*, 2011), with evidence of divergent selection for trichomes identified in both *A. kamchatica* and *A. lyrata* (Karkkainen*, et al.*, 2004; Steets*, et al.*, 2010). Furthermore, trichomes may increase susceptibility to pathogens (Calo*, et al.*, 2006). Reflecting these conflicting selection pressures, constitutive trichome density is highly variable among natural accessions of *A. thaliana* with a strong genetic basis to the observed variation under controlled conditions (Mauricio, 2005; Symonds*, et al.*, 2005; Bloomer*, et al.*, 2012).

Constitutive defence mechanisms are typically assumed to be costly, diverting resources away from growth and reproduction; in contrast, induced defence responses allow plants to avoid making high-level defensive investments unless required. Although induction of trichome initiation has not been demonstrated in the field in *A. thaliana* (Mauricio, 1998), trichome production is induced by artificial wounding of early leaves in *A. thaliana* (Traw & Bergelson, 2003) and by herbivory in the related species *Brassica nigra* (Brassicaceae) (Traw, 2002). Such phenotypic plasticity implies a mechanism by which *A. thaliana* may offset some of the cost of producing trichomes, investing in higher density only when required. Previous QTL mapping studies have investigated the genetic architecture of natural variation in trichome density in *A. thaliana* (Larkin*, et al.*, 1996; Mauricio, 2005; Symonds*, et al.*, 2005;

Pfalz, *et al.*, 2007; Atwell, *et al.*, 2010) but there are currently no mapping data for damage-induced trichome initiation or trichome density plasticity loci, although these are perhaps more meaningful traits in nature.

The genetic basis of trichome initiation on *A. thaliana* leaves is relatively well understood. Initiation of trichomes on the leaf lamina requires interaction between the WDR protein TRANSPARENT TESTA GLABRA (TTG1) , one of the functionally overlapping bHLH proteins GLABRA3 (GL3) or ENHANCER OF GL3 (EGL3) (Payne, *et al.*, 2000), and the trait-specific R2R3 MYB GLABRA1 (GL1) (Oppenheimer, *et al.*, 1991), forming a complex that activates downstream genes involved in trichome initiation such as GL2 (Szymanski, *et al.*, 1998) and TTG2 (Ishida, *et al.*, 2007). A suite of R3 MYBs act as suppressors of initiation in surrounding cells, generating a spacing pattern across the leaf (Wang, *et al.*, 2008). Initiation at the leaf margin is similarly controlled, with GL3 or the bHLH TT8 (Maes, *et al.*, 2008) interacting with TTG1 and another trait-specific MYB, MYB23, to activate downstream genes. Phytohormones also play a role in regulating trichome density; gibberellin is required for trichome initiation and jasmonic acid promotes increased trichome density in response to wounding, while salicylic acid reduces density. Both *GL1* and *GL3* expression are induced by gibberellins (Perazza, *et al.*, 1998; Maes, *et al.*, 2008), with the DELLA family of repressors playing a role in this signalling (Gan, *et al.*, 2007). *GL3* is up-regulated by both exogenous (Traw & Bergelson, 2003; Maes, *et al.*, 2008) and endogenous jasmonic acid (Yoshida, *et al.*, 2009) via physical interaction with JAZ proteins (Qi, *et al.*, 2011), linking induction of trichome initiation following wounding to genes within the TTG1 pathway. Previous QTL and association mapping studies have suggested TTG1 pathway genes as good candidates for trichome density variation (Symonds, *et al.*, 2005; Hilscher, *et al.*, 2009; Atwell, *et al.*, 2010), and recent studies of natural allelic variation in genes within the pathway have shown roles for the R3 MYB repressor *ETC2* (Hilscher, *et al.*, 2009), the bHLH *ATMYC1* (Symonds, *et al.*, 2011), and the R2R3 MYB *GL1* (Bloomer, *et al.*, 2012) in generating quantitative variation for trichome density in natural populations.

Here, we describe QTL mapping results for trichome density from two new RIL mapping populations available to the *Arabidopsis* research community, Hi-0 x Ob-0 (HO) and St-0 x Sf-2 (SS). Hi-0, Ob-0, St-0 and Sf-2 were chosen as parents for the development of mapping populations based on screens for genetic variability and variation in constitutive trichome

density in *A. thaliana*, among other traits; to our knowledge, these are the first publically available RIL mapping populations to include these accessions as parents. The populations are here used to assess the genetic architecture of constitutive and induced trichome density on early leaves, and to assess the genetic basis of the response of plants to damage. As a contrast, the Hi-0 and Ob-0 accessions have relatively high but differing constitutive trichome densities and the St-0 and Sf-2 accessions have relatively low and nearly identical constitutive trichome densities. Although trichome density has been mapped previously (Larkin*, et al.*, 1996; Mauricio, 2005; Symonds*, et al.*, 2005; Pfalz*, et al.*, 2007; Atwell*, et al.*, 2010), mapping in these new populations and in multiple environments (including damage) affords greater comparative analyses than previously undertaken. Of particular interest is determining whether QTL identified provide support for the recently proposed hypothesis that allelic variation within the TTG1 pathway provides much of the observed phenotypic variation in trichome density within and among natural populations of *A. thaliana* (Symonds, *et al*., 2011). Our results provide new insights into the genetic architecture of constitutive and induced trichome density in *A. thaliana* specifically and into the genetic bases of plant defence generally.

2.3 Materials and Methods

2.3.1 Plant materials

Based on preliminary screens of genetic variation and variation in constitutive trichome density and other traits in *A. thaliana*, several pairs of natural accessions were selected to serve as progenitors for the development of new recombinant inbred line (RIL) mapping populations. Among those pairs were Hi-0 / Ob-0 (HO) and St-0 / Sf-2 (SS). Members of a pair were reciprocally crossed and the resulting F_1s were confirmed to be cross progeny by genotyping several microsatellite loci and comparing the results with the parental accessions. F_1s were allowed to self-pollinate and the seed was collected. Several hundred F_2 seed were started in individual pots; this was the establishment of individual RILs. From each F_2 plant, one seed was randomly selected to begin the next generation, which was also allowed to self. This single-seed descent with selfing scheme was followed for eight generations, ending in the bulk collection of seed from F_9 plants. The bulked seed from 307 RILs of Hi-0 x Ob-0 and 261 RILs of St-0 x Sf-2 were deposited with the Arabidopsis Biological Resource Center (ABRC), where they were further bulked. Prior to genotyping, seed for each RIL population were acquired from the ABRC to assure that the genotypes matched available seed lines. The progenitor accessions and 188 randomly chosen RILs from each population were used for genotyping and phenotyping.

2.3.2 Marker screening

The parent accessions of both populations, Hi-0, Ob-0, St-0 and Sf-2, were grown under growth room conditions of 24^{o}C, 16:8 hours light:dark. Genomic DNAs were extracted from fresh tissue obtained from young rosette leaves using a modified CTAB extraction protocol (Doyle & Doyle, 1987). These DNAs were used to screen for microsatellite markers polymorphic between parental accessions (details below) initially using primers from the French National Institute for Agricultural Research (INRA) Microsatellite Database (http://www.inra.fr/internet/Produits/vast/msat.php). Additional primers were designed in Primer3 based on flanking sequences from microsatellite repeats identified from the Eukaryotic Microsatellite Database (http://www.veenuash.info/) or by our lab using the Col-0 reference genome. In excess of 150 microsatellite markers were screened to identify markers for linkage map development in the HO and SS populations; of these, many were monomorphic or amplified alleles in only one or neither parent.

An M13 primer-tailing scheme polymerase chain reaction (PCR) (Schuelke, 2000) was used for genotyping. For the initial screen for polymorphic loci, markers were amplified in 10µL PCR containing 1X NEB Thermopol buffer (New England Biolabs), 2 µmol dNTPs, 0.2 µmol M13-tailed marker-specific forward primer, 4.5 µmol reverse primer, 4.5 µmol FAM-labelled M13 primer, 0.5 U NEB Taq polymerase, and ~50 ng genomic DNA under the following cycling conditions: 95^{o}C for 3 minutes, followed by 30 cycles of 95^{o}C for 30 seconds, 52^{o}C for 40 seconds, and 72^{o}C for 40 seconds, and a final extension of 20 minutes at 72^{o}C. One microlitre of PCR product was then combined with 9 µL of a HiDi (Applied Biosystems) and CASS size standard (Symonds & Lloyd, 2004) mix. Allele size was determined by capillary separation of fluorescently labelled PCR products and CASS size standard on an ABI3730 Genetic Analyzer (Applied Biosystems) at the Massey Genome Service; parental allele sizes were called in GENEMAPPER v3.7 (Applied Biosystems). Consistently amplifiable markers polymorphic between the parents of a population were chosen for RIL population genotyping.

2.3.3 RIL genotyping and linkage map construction

A set of 188 RILs plus parents was screened in each mapping population. RIL genotyping PCRs were tailed with M13 primers labelled with one of three fluorescent dyes, FAM, VIC or NED, and PCRs carried out as described above. Three markers (each with a different fluorescent label) from the same individual were pooled for capillary separation, with markers chosen for a pooling group based on a requirement of non-overlapping allele sizes. After individual amplification reactions, markers were pooled together in a ratio dependent on the strength of amplification of each marker in parental screens. One microlitre of the pooled markers was combined with 9 µL HiDi/CASS prior to capillary separation. Results were assessed and allele calls made using GENEMAPPER v3.7 (Applied Biosystems). When homozygosity was not absolutely clear individuals were conservatively scored as heterozygotes, thereby eliminating those genotypes from further analyses.

Ultimately, the HO population was genotyped at 55 loci and the SS population at 67 loci, with 32 markers common to both populations. Markers used for each population are listed in Table 2.4 (Supplementary Material). Of the 188 RILs screened in each population, seven individuals were removed from HO and seven individuals from SS before linkage map construction and QTL mapping due to ambiguity of allele calls across a high proportion of

markers. Linkage maps for both populations were constructed in JOINMAP4 (van Ooijen, 2001) using maximum likelihood.

2.3.4 Trichome density phenotyping of the Hi-0 x Ob-0 and St-0 x Sf-2 mapping populations

The HO and SS populations were each phenotyped in separate experiments. To score trichome density phenotypes, six replicates of each RIL and of the two parents were planted in seed raising mix in 72 cell flats in a fully randomised design; approximately 5-10 seeds were sown in each cell. Seeds were vernalized for 11 days to synchronise germination times and then moved to a growth chamber at 24^{o}C under 16 hours light for 25 days post-germination. Plants were thinned to three per cell at four days post germination, and to a single plant 7 days post germination. To map "induced trichome density", three of the six replicates of each RIL were randomly selected to have leaves damaged. Plants in the leaf damage experiment were subjected to pinching of one cotyledon with serrated forceps at four days post-germination, and first and third leaves shortly after emergence at seven and ten days post-germination. As the first damage treatment was inflicted prior to complete thinning within cells, all seedlings in a "damage" cell were pinched. The remaining three (non-pinched) replicates of each RIL were used to measure "constitutive trichome density".

Trichome density was scored at 25 days post-germination. At this stage of development the fifth true rosette leaf was fully expanded and not yet senescent. Trichomes were counted on the fifth leaf in a 17 mm^2 area midway along the length of the leaf blade between the midrib and leaf margin using a dissecting microscope at 25x magnification. Broad sense heritability (H^2) was calculated independently for damaged and undamaged treatments of each population in EXCEL using mean squares from an ANOVA with RILs as groups.

2.3.5 QTL mapping

For each population, mean trichome density was calculated for each RIL in both treatments described above; henceforth these phenotypes are referred to as "constitutive trichome density" and "induced trichome density". In addition, the plasticity of trichome density in response to damage was calculated for each RIL as the difference between mean constitutive

and mean induced trichome densities; this is the “response” phenotype. The calculated RIL means for the three traits were used as phenotype values in QTL mapping.

QTL mapping was carried out on RIL means using the R/qtl package (Broman, *et al*., 2003). Initially, variation for each trait was mapped using the “em”, “haley-knott”, and “multiple imputation” (256 replicates) methods in a 1D scan (scanone) and followed by a 2D scan (scantwo) for interacting QTLs using the same methods. The results for each model were compared and found to be very similar for any one phenotype for any one population. The number of QTL for each trait was estimated based on these initial analyses and used to assign a liberal maximum number of QTL to include in “stepwiseqtl” analyses (max.qtl). 2D permutation tests of 1000 reps were run using haley-knott regression; these results were used to derive the penalties used in stepwiseqtl analyses. The stepwiseqtl analysis steps through main effect QTL models, while refining positions, and searching for interacting QTL at every step; the scan.pairs option was selected. The models are progressively more inclusive, building from one main effect QTL up to the maximum (as estimated from initial 1D and 2D scans) and then stepping back down to one main effect QTL. Model outputs are compared by a penalized LOD (pLOD) score, which are calculated at each step and the model with the highest pLOD score is taken as the best fit to the data. The pLOD approach allows one to directly compare the fit of models of different size (different numbers of QTL). As the penalties for adding epistatic interactions are quite heavy, it was rare that best fit models included any such interactions; however, the data often show strong evidence of interacting QTL in 2D scans. Ninety-five percent Bayes credible intervals for each QTL were calculated in R/qtl using the Bayesint function.

The physical positions of maximum LOD peaks and intervals were estimated from the physical position of flanking markers in the Col-0 genome, assuming a correlation between physical marker positions in the RILs with their physical positions on the Col-0 reference genome and a roughly linear relationship between physical and linkage positions. Candidate genes for trichome density QTL were identified from the extensive literature describing trichome initiation determination, trichome initiation signalling, and downstream regulatory targets. Map positions of the candidate genes identified as falling within a given 95% Bayes credible interval for a QTL were estimated using the map positions of markers with physical positions flanking the gene, based on assumptions outlined above.

2.4 Results

2.4.1 RIL population genotyping and linkage map construction for Hi-0 x Ob-0 and St-0 x Sf-2

Hi-0 x Ob-0 (HO) was genotyped with 55 markers (8-14 markers per chromosome), while St-0 x Sf-2 (SS) was genotyped with 67 markers (9-16 markers per chromosome; Table 2.4, Supplementary Material). Residual heterozygosity across all markers was 1.12% in the HO population and 1.36% in the SS population (Table 2.1). This is low and similar to that reported for other RIL populations (Pfalz*, et al.*, 2007; McKay*, et al.*, 2008; O'Neill*, et al.*, 2008) but slightly higher than the <1% expected beyond the F_7 generation, which may reflect the conservative approach taken to allele calling (described above) or heterozygote advantage. Some degree of segregation distortion was observed in both populations (Figure 2.5; Supplementary Material). HO RILs exhibited segregation distortion localised to regions of chromosomes I, IV and V with preferred parental alleles varying by genomic region. In the SS population, St-0 accounted for 60.7% of alleles observed across all markers primarily as a result of strong distortion favouring St-0 alleles across all of chromosome I, the majority of chromosome V and on localised regions of chromosome II. Localised distortion on chromosomes III and IV predominantly favoured Sf-2 alleles.

Table 2.1: Genetic analysis of RIL populations

Population	Number of RILs genotyped	Number of markers	Total map length (cM)	Average marker distance (cM)	Residual heterozygosity (%)
Hi-0 x Ob-0	181	55	479	9.60	1.12
St-0 x Sf-2	181	67	478	7.72	1.36

Linkage maps for both populations were built using the maximum likelihood mapping function in JoinMap 4 (van Ooijen, 2001). Marker order was consistent with physical position for most markers, with the exception of several tightly linked marker pairs in each population and three markers around the centromere of chromosome V in HO (Table 2.4, Supplementary Material). These markers were constrained to match their order on the physical map during linkage map construction in JoinMap 4 (van Ooijen, 2001); the likelihood of the constrained versus unconstrained marker orders was tested in R/qtl (van Ooijen, 2001; Broman*, et al.*, 2003), showing only nominally less strong support for the constrained marker order (not

shown). The linkage maps built spanned 479 cM for HO with an average marker spacing of 9.6 cM and maximum gap size of 24.3 cM. For SS, 478 cM was covered, with an average marker spacing of 7.72 cM and maximum gap size of 18.6 cM (Table 2.1, Figure 2.1).

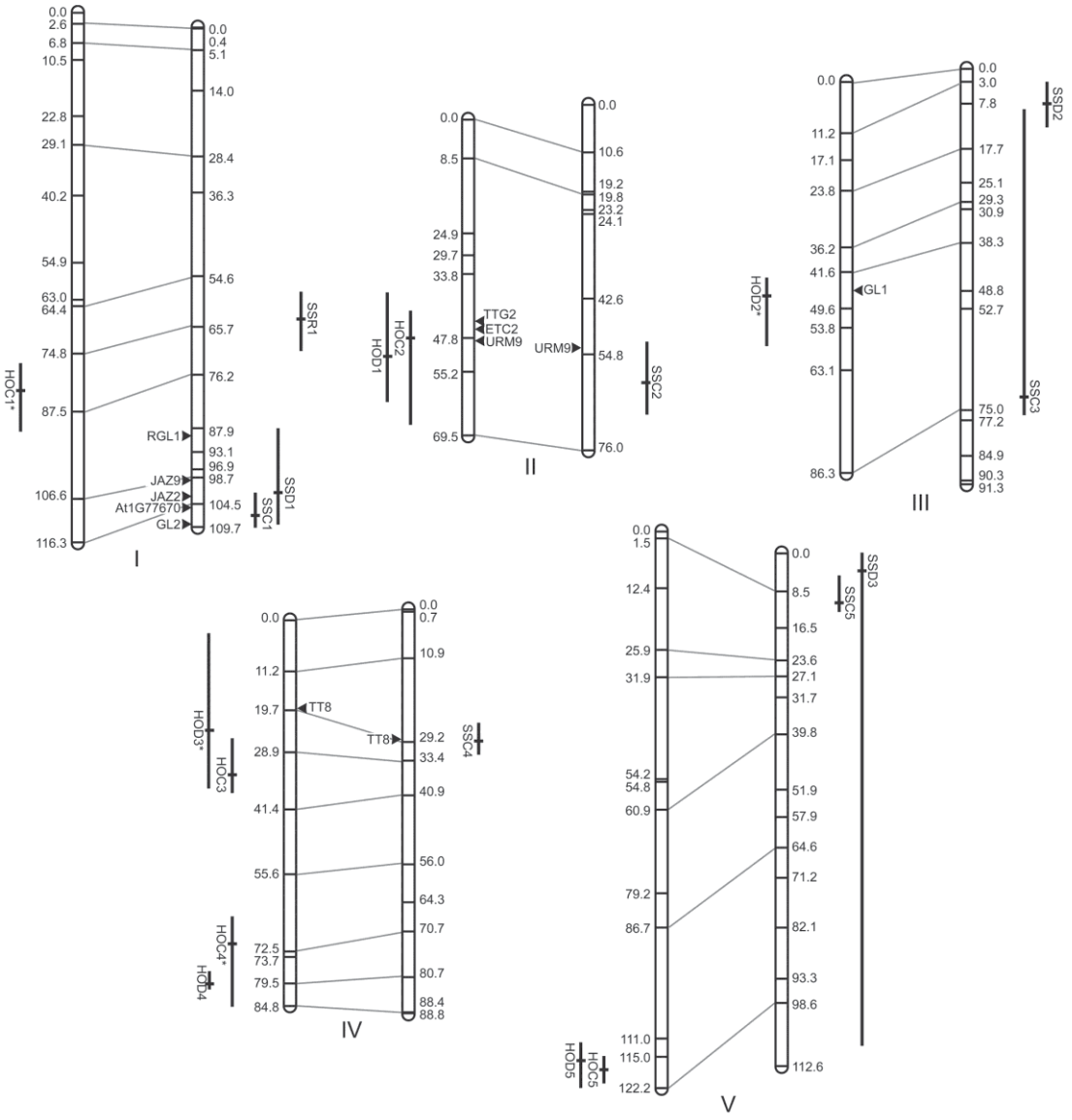

Figure 2.1: Aligned linkage maps and QTLs mapped in the Hi-0 x Ob-0 (top) and St-0 x Sf-2 (bottom) RIL mapping populations.

Marker positions are shown in cM. The peak LOD position for QTL identified for each of the three traits are indicated by short solid black horizontal bars; Bayes' credible intervals are indicated by perpendicular bars. Interacting QTL are indicated with an *. QTL are labelled by population and trait as in Table 2.3: HOC = Hi-0 x Ob-0 Constitutive; HOD = Hi-0 x Ob-0 Damage induced; SSC = St-0 x Sf-2 Constitutive; SSD = St-0 x Sf-2 Damage induced; SSR = St-0 x Sf-2 Response to leaf damage. Positions and names of candidate genes are marked with a black triangle.

2.4.2 Trichome density phenotypes of the Hi-0 x Ob-0 and St-0 x Sf-2 mapping populations

Trichome density on the fifth rosette leaf was scored in the SS and HO populations in both control (constitutive) and damaged leaf (induced) environments. The difference in trichome density scores between the two environments was calculated for each RIL as a measure of the plants' response to wounding. In the HO phenotyping experiment the parental accessions showed constitutive trichome densities of 19.67 for Hi-0 and 15.67 for Ob-0 (Table 2.2, Figure 2.2); surprisingly, both Hi-0 and Ob-0 had lower induced trichome densities than constitutive (13.5 and 12.0 respectively). The mean constitutive trichome density of the HO RILs was 12.99, increasing slightly to 13.42 when induced; this difference was weakly significant (p<0.05) as determined by two tailed paired T-test. The response to wounding of individual RILs in the HO population ranged from a decrease of 6.5 trichomes to an increase of 9.5 trichomes, with a mean increase of 0.41 trichomes. Transgressive segregation in the RILs was evident for both constitutive (4.0 to 25.5) and induced (4.0 to 29.0) trichome density phenotypes.

Table 2.2: Parental and RIL mean trichome densities, range of RIL densities, and broad sense heritability

Trait	Parental accession means	RIL mean	RIL range	H^2
Hi-0 x Ob-0				
Control	Hi-0 19.67; Ob-0 15.67	12.99*	4.0 – 25.5	0.77
Leaf damage	Hi-0 13.50; Ob-0 12.00	13.42*	4.0 – 29.0	0.79
Response	Hi-0 -6.17; Ob-0 -3.67	0.41	-6.5 – 9.5	NA
St-0 x Sf-2				
Control	St-0 6.00; Sf-2 6.00	7.01 **	2.0 – 14.0	0.68
Leaf damage	St-0 7.67; Sf-2 9.33	8.52 **	2.7 – 18.3	0.74
Response	St-0 1.67; Sf-2 3.33	1.50	-3.3 – 4.7	NA

* Means are significantly different within a population at p<0.05

**Means are significantly different within a population at p<<0.001

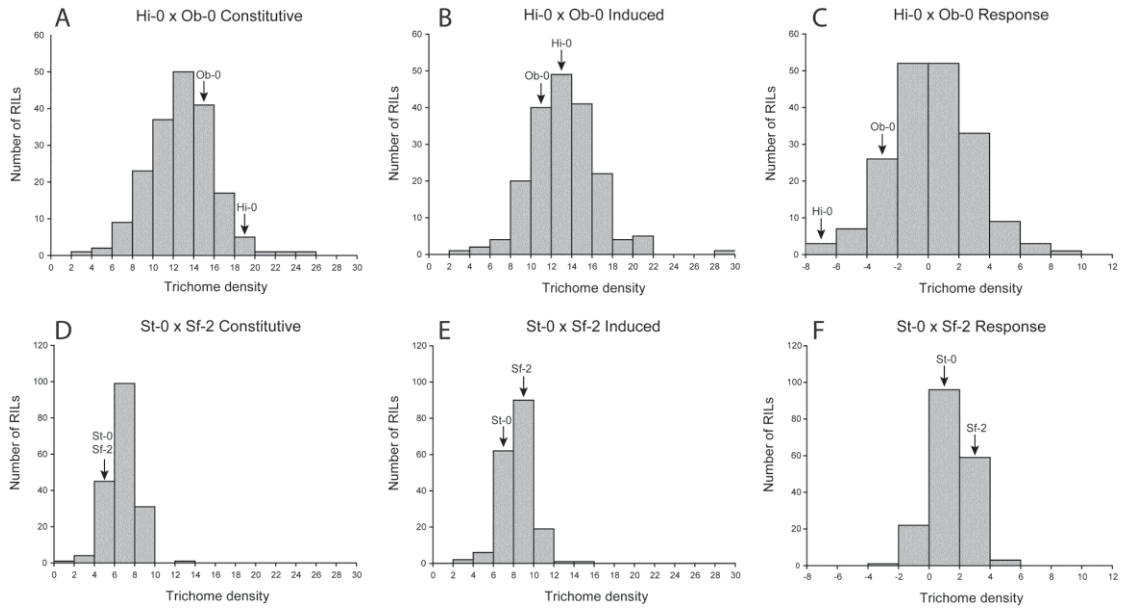

Figure 2.2: Distribution of constitutive and induced trichome densities and response to damage for the Hi-0 x Ob-0 (A-C) and St-0 x Sf-2 (D-F) RILs and population parents.

Labelled arrows indicate the parental phenotypes' positions in each distribution. Note that the Hi-0 and Ob-0 accessions and some proportion of RILs in both populations have negative responses to leaf damage.

In the SS phenotyping experiment, the parental accessions St-0 and Sf-2 had identical constitutive trichome densities of 6.0 (Table 2.2, Figure 2.2). Both St-0 and Sf-2 increased trichome density when induced to 7.67 and 9.33 respectively; this corresponds to a response to wounding of 1.67 trichomes in St-0 and 3.33 trichomes in Sf-2. Mean constitutive trichome density of the RILs was 7.02 and 8.52 when induced; this difference was highly significant (p<<0.001) as measured by two tailed paired T-test. The SS RILs displayed transgressive segregation for both phenotypes; constitutive trichome densities ranged from 2.0 to 14.0 and when induced, from 2.7 to 18.3. Most SS RILs responded to damage by increasing trichome density, with a mean wounding response of +1.5 trichomes and responses ranging from a decrease of 3.3 to an increase of 4.7 trichomes after wounding.

An ANOVA was used to assess broad sense heritability (H^2) of constitutive and induced trichome density in both populations. A strong genetic component underlies the observed variation in phenotypes. In the HO phenotyping experiment H^2 was 0.77 for constitutive and 0.79 for induced; in the SS experiment, H^2 was 0.68 for constitutive and 0.74 for induced. As the "response" score was calculated based on the mean trichome density for each individual RIL in each environment, H^2 could not be calculated.

2.4.3 QTL for trichome density

QTL mapping for constitutive and induced trichome density and response to induction was performed in R/qtl (Broman*, et al.*, 2003) for both populations. Interval mapping using the EM, Haley-Knott and Imputation methods was undertaken first for each population/trait to estimate the number of QTL and to assess the performance of each method; this information was then used to determine parameters for multiple QTL mapping using the stepwiseQTL function of R/qtl. QTL identified in each population and for each trait are provided in Table 2.3.

In the HO population QTL were mapped for both constitutive and induced trichome density, but no QTL were identified underlying variation in the response phenotype. For constitutive trichome density, stepwiseQTL analysis produced two models with nearly identical pLOD scores. Due to their comparable pLOD values both models are presented in Table 2.3; the QTL results from Model 2, which appears to be the more comprehensive model, are presented in Figure 2.1. Model 1, with a pLOD of 7.32, identified four QTL, one each on chromosomes II and V, and two on chromosome IV, which together explained 34.26% of variation observed for this phenotype. Model 2, with a pLOD of 7.31, identified the same approximate QTL as Model 1 but included an additional QTL on chromosome I and an interaction between the chromosome I QTL and one of the two QTL identified on chromosome IV; together, the QTL and interaction identified by Model 2 explained 54.25% of the observed phenotypic variation. The highest pLOD-scoring model for induced trichome density, with a pLOD of 6.05, identified individual QTL on chromosomes II, III and V and two QTL on chromosome IV with an interaction between QTL on chromosome III and one on chromosome IV. The QTL and interactions identified by this model explain 51.56% of variation observed for the leaf damage environment.

Table 2.3: Quantitative Trait Loci and epistatic interactions determined by stepwise QTL analysis in R/qtl.

QTL	Chromosome	Position (cM)	Interval (cM)	Variation Explained (%)	Allele mean trichomes	Candidate gene(s)
Hi-0 x Ob-0 Constitutive: Model 1; pLOD=7.32						
	2	48	40-62	6.46	H **14.3**; O 12.4	
	4	33	23-39	8.88	H 12.5; O **14.5**	
	4	82	67-84.8	5.61	H **13.7**; O 13.0	
	5	118	115-121	13.31	H **14.2**; O 11.8	
	TOTAL			**34.26**		
Hi-0 x Ob-0 Constitutive: Model 2; pLOD=7.31						
HOC1	1	86*	80-95	7.07	H 13.4; **O 14.2**	-
HOC2	2	48	42-67	5.04	**H 14.3**; O 12.4	*ETC2*/*TCL1*/*TCL2*†; *TTG2*; *URM9*
HOC3	4	34	26-38	11.27	H12.5; **O 14.5**	-
HOC4	4	71*	65-84.8	11.83	**H 13.6**; O 13.1	-
HOC5	5	118	115-121	11.97	**H 14.2**; O 11.8	-
		HOC1xHOC4		7.07		
	TOTAL			**54.25**		
Hi-0 x Ob-0 Induced; pLOD = 6.05						
HOD1	2	52	38-62	5.92	**H 14.4**; O 13.1	*ETC2*/*TCL1*/*TCL2*; *TTG2*; *URM9*†
HOD2	3	47*	43-58	10.76	**H 14.4**; O 13.0	*GL1*
HOD3	4	24*	3-37	11	H 13.3; **O 14.7**	*TT8*
HOD4	4	79.5	77-81	12.72	**H 14.6**; O 13.1	-
HOD5	5	116	112-122	6	**H 14.3**; O 12.6	-
		HOD2xHOD3		5.16		
	TOTAL			**51.56**		
St-0 x Sf-2 Constitutive; pLOD = 7.32						
SSC1	1	107	102-109.7	7.84	**Sf 7.7**; St 6.7	*GL2*; *At1g77670*†; *JAZ2*
SSC2	2	61	52-68	7.15	Sf 6.6; **St 7.4**	*URM9*; *TTG2*†
SSC3	3	72	9-76	6.77	**Sf 7.4**; St 6.8	
SSC4	4	29	25-32	9.88	**Sf 7.3**; St 6.5	*TT8*
SSC5	5	11	5-13	9.84	Sf 6.6; **St 7.2**	
	TOTAL			**41.48**		
St-0 x Sf-2 Induced; pLOD = 4.0						
SSD1	1	102	87.92-109	12.34	**Sf 9.3**; St 8.3	*RGL1*; *JAZ9*†; *JAZ2*; *GL2*; *At1g77670*
SSD2	3	7.8	3-13	6.89	**Sf 8.9**; St 8.0	-
SSD3	5	4	0-108	6.99	Sf 8.2; **St 8.8**	-
	TOTAL			**26.22**		
St-0 x Sf-2 Response; pLOD = 2.51						
SSR1	1	64	58-71	11.98	**Sf 2.2**; St 1.2	-
	TOTAL			**11.98**		

† Denotes candidate gene in closest proximity to peak LOD score of the QTL

* Interacting QTL within a population/phenotype

Two epistatic interactions were detected by stepwise QTL analysis in the HO population (Figure 2.3); one in Model 2 for constitutive trichome density between main effect QTL on chromosomes I and IV, and one for induced trichome density between main effect QTL on chromosomes III and IV (Table 2.3; Figure 2.3). An effect plot of the interaction between loci on chromosomes I and IV revealed higher trichome density for RILs carrying alleles from the same parent at these loci. In contrast, an effect plot of the interaction identified in the leaf damage environment showed that the Ob-0 allele at the chromosome III locus appears to suppress the trichome density-increasing action of the Hi-0 allele (relative to the Ob-0 allele) at the chromosome IV locus.

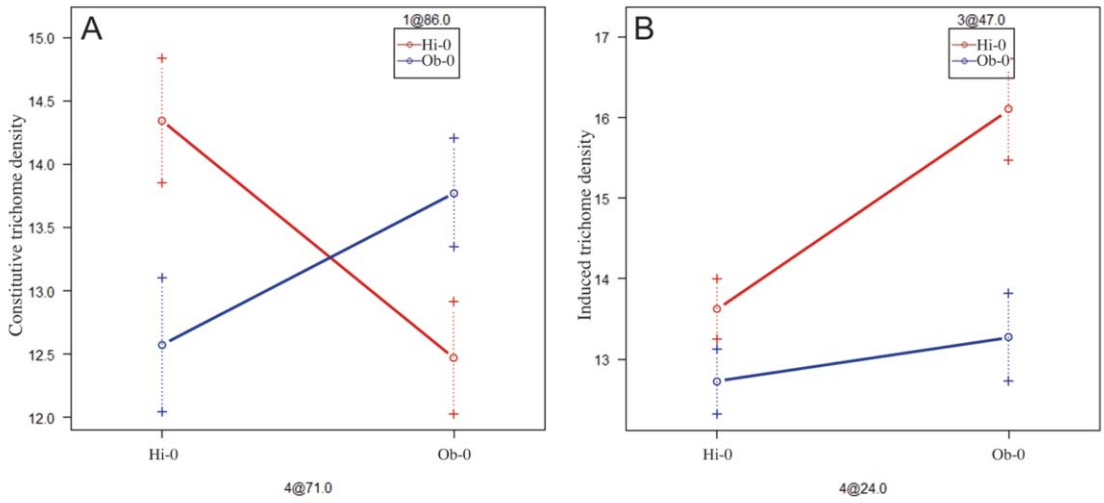

Figure 2.3: Effect plots for epistatic interactions among QTL in the Hi-0 x Ob-0 mapping population.

Panel A shows a large interaction effect for constitutive trichome density between loci at chromosome 1 at 86 cM and chromosome 4 at 71 cM; note that the highest trichome density achieved by genotypes where the alleles from the same parent co-occur. Panel B shows an interaction for induced trichome density between loci on chromosome 3 at 47 cM and chromosome 4 at 24 cM. In this case, the effect of the chromosome 4 locus appears to be masked by the Ob-0 allele on chromosome 3.

In the SS mapping population QTL were identified for all three phenotypes. StepwiseQTL mapping for constitutive trichome density revealed a highest pLOD scoring model with five QTL, one on each of the five chromosomes, and no epistatic interactions. This model had a pLOD of 7.32, with QTL identified explaining 41.48% of observed variation for this trait (Table 2.3). The highest scoring model for induced trichome density, with a pLOD of 4.0, identified only three QTL; one each on chromosomes I, III and V. These QTL together explain just 26.22% of the observed variation for this trait (Table 2.3). A single significant QTL underlying

the variation in response of plants to leaf damage in this population was identified on chromosome I, explaining 11.98% of observed variation; of interest, this QTL does not overlap with QTL for constitutive or induced trichome density.

2.4.4 Trait correlations

To explore the relationship between constitutive and induced trichome density, within each mapping population, mean values for each RIL were plotted against one another in SIGMAPLOT (Systat, Inc.). In both populations, there was a positive correlation between constitutive trichome density and induced trichome density, but the slope of the regression line was considerably less than one (Figure 2.4). Interestingly, in the population with the greater distribution of responses (HO), no QTL were mapped for the response phenotype. In the SS population, only one QTL was mapped for response. To examine the distribution of response phenotypes for the genotypes at this locus, the data were partitioned into two sets: RILs with the St-0 genotype and RILs with Sf-2 genotype at the marker nearest the QTL (msat1.42). Although both subsets revealed positive correlations, that for the Sf-2-genotype RILs was ~2/3 that of the St-0 RILS. Interestingly, the plot also reveals that only Sf-2 RILs display zero or negative responses.

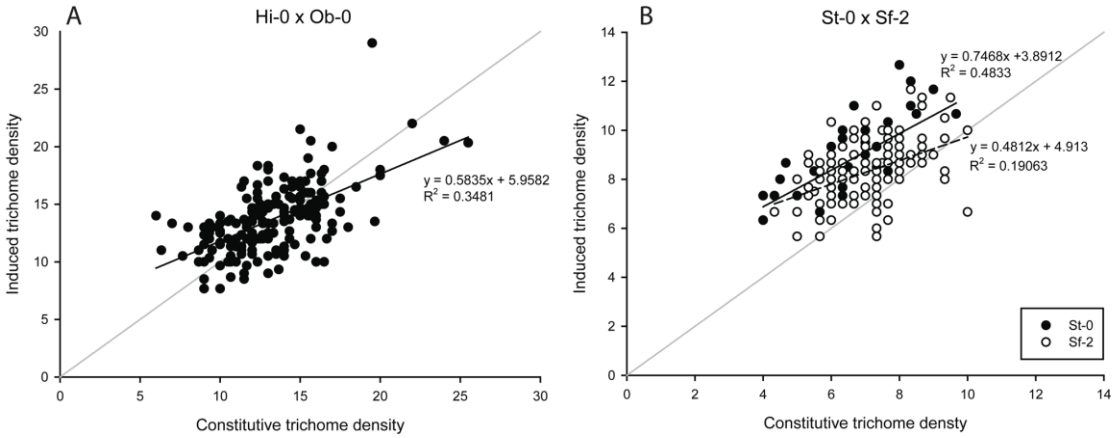

Figure 2.4: Scatterplots for constitutive versus induced trichome density in the Hi-0 x Ob-0 (A) and St-0 x Sf-2 (B) RIL populations.

Because a QTL was mapped for the response phenotype in the SS population, those data were partitioned according to the allele carried by individual RILs at the marker nearest the response QTL. The grey diagonal indicates a slope of 1 on each graph; any point above the line therefore reflects RILs with a positive response to leaf damage (i.e., they increase trichome density) and any point that lies beneath the line are negative responders, which reduce trichome density in response to leaf damage. The regression equation and R2 were calculated separately for the two genotypes in the SS population.

2.5 Discussion

Plants deploy a dynamic range of defence strategies against herbivory, including plant hairs or trichomes. Because of the cost associated with trichome production and the variability in selection pressure put upon populations by herbivory, one might expect considerable standing variation for trichome density within or among populations, individual phenotypic plasticity for trichome density, or both. Numerous mapping studies have now demonstrated considerable genetic variation for constitutive trichome density in *A. thaliana*, and although it also has been shown to be inducible, the genetic architecture of induced trichome density and response to damage have not been reported. Here, we utilized two new *A. thaliana* RIL mapping populations to investigate the genetic architecture of constitutive and inducible trichome density and the response to induction. Our results reveal intriguing variation for response to leaf damage, with parental accessions and RILs displaying both positive and negative responses. QTL were mapped for constitutive and induced trichome density and for response to damage. Overlap between QTL positions and strong candidate genes for many of the QTL identified provide support for the recently proposed hypothesis that allelic variation within the TTG1 pathway underlies considerable variation in trichome density within and among natural populations of *A. thaliana* (Symonds*, et al.*, 2011).

2.5.1 Segregation distortion

In the two new RIL populations examined here, residual heterozygosity was low and similar to that reported for other *A. thaliana* RIL populations (Pfalz*, et al.*, 2007; McKay*, et al.*, 2008; O'Neill*, et al.*, 2008). In the HO population, the distribution of parental alleles was close to the expected 1:1 ratio, with some localised segregation distortion (SD). Segregation distortion was more pronounced in the SS population than in the HO population, with strong distortion across chromosomes I and V favouring St-0 alleles. Localised SD is commonly observed in *A. thaliana* RIL populations and typically attributed to unintentional selection during RIL development, for example for traits affecting germination or flowering (McKay*, et al.*, 2008; O'Neill*, et al.*, 2008). However, SD between loci on chromosomes I and V biased toward retention of the same parental allele at both loci has been observed in a number of mapping populations involving a range of accessions (McKay*, et al.*, 2008; O'Neill*, et al.*, 2008; Simon*, et al.*, 2008; Balasubramanian*, et al.*, 2009). Previous work on mapping populations derived from Col-0 and Est-1 parental accessions suggested an incompatibility between alleles at

these loci which affects plants early in development (Balasubramanian*, et al.*, 2009). Further, an incompatibility between the bottom of chromosome I and top of chromosome V in Cvi-0 x Col-0 crosses is due to an inability to synthesise histidine when the same parental allele is not retained at both loci, resulting in embryo lethality (Bikard*, et al.*, 2009). Such incompatibilities may explain SD observed in many mapping populations, including those described here.

2.5.2 Trichome density phenotypes

For each RIL population, constitutive and induced trichome density were measured on fifth leaves in one large experiment. Heritabilities were high for both traits in both populations (Table 2.2) indicating a strong genetic component to the observed variation, and were similar to heritabilities reported for trichome density previously (Symonds*, et al.*, 2005; Atwell*, et al.*, 2010; Bloomer*, et al.*, 2012). Interestingly, heritability was slightly higher for induced trichome density than constitutive trichome density in both populations, perhaps suggesting that the damage treatment serves as a strong stimulus to the trichome initiation pathway, thereby reducing the relative effects of other environmental variables.

Trichome density distributions revealed intriguing and contrasting patterns in the two populations. The Hi-0 and Ob-0 parental accessions both had high trichome densities relative to St-0 and Sf-2 (Table 2.2; Figure 2.2). Mean constitutive trichome density of HO RILs was relatively high as well and showed a very small (<4%), though significant, mean increase in density when induced. Interestingly, both the Hi-0 and Ob-0 accessions were negative responders, displaying lower trichome densities following damage. In contrast, St-0 and Sf-2 had identical, and comparatively low, constitutive trichome densities and showed increased trichome density following damage. Likewise, the SS RILs had comparatively low mean constitutive trichome densities but showed a strong, significant increase of over 20% when damaged (induced). These observations would seem consistent with Optimal Defence Theory as outlined by Stamp (2003), which predicts a negative correlation between the level of constitutive expression of a defensive trait and its capacity for induction. Similar results have been shown in a recent mapping study of trichome production in *Mimulus guttatus* (Holeski*, et al.*, 2010), with constitutive trichome density score negatively correlated with induction. Here, this is further illustrated in plots of constitutive versus induced trichome density for RILs within each population (Figure 2.4); for both populations, the slope of the regression line

is positive but much less than one, indicating that as constitutive trichome density increases, induction capacity decreases. Plasticity for trichome density has been demonstrated in *A. thaliana* previously, but the apparent relationship between constitutive and induced trichome density observed here has not. This is further considered in the context of mapping results below.

2.5.3 QTL mapping results

The parent accessions of the two populations differ considerably for all three phenotypes. Hi-0 and Ob-0 are more different from one another for all traits than the St-0 and Sf-2 parents and yielded broader distributions in their RILs for all traits (Figure 2.2). Despite this, the total number of QTL mapped in each population was fairly similar (10 in HO and 9 in SS); however, the HO QTL explain more variation than the SS QTL for the constitutive and induced phenotypes.

Of the QTL discovered in the HO population, four were mapped to similar positions for constitutive and induced trichome density and two were unique to one environment each. Despite considerable variation for the response phenotype in this population, no QTL were identified (Table 2.3). As the response phenotype is calculated from the means of constitutive and induced trichome densities, the combined environmental effects may overwhelm any small effect genetic variation for this trait. Two strong epistatic interactions were identified in the HO population in addition to the five main effect QTL identified in each environment. Of the two interactions detected, that between HOC1 and HOC4, shows that constitutive trichome density is maximized when alleles from the same parent co-occur at these loci. The other interaction, between HOD2 and HOD3, suggests a masking effect by the Ob-0 allele at HOD2 on the effect at HOD3 for induced trichome density. Interestingly, HOC1 was not identified through one-dimensional interval mapping (data not shown) and was instead only detected when considered as an interaction with HOC4, highlighting the importance of testing for epistasis to build more comprehensive models of genetic architecture. Epistatic interactions have been shown to underlie variation in a diverse range of traits in *A. thaliana* including aluminium tolerance (Kobayashi & Koyama, 2002), fitness traits (Malmberg*, et al.*, 2005), flowering time (Juenger*, et al.*, 2005) and seed oil composition (Sanyal & Linder, 2012). These interactions appear to be a significant potential source of natural phenotypic variation,

perhaps particularly where migration and introgression introduce alleles into new genetic backgrounds.

Despite the parent accessions having nearly identical constitutive trichome densities, a total of nine QTL were identified in the SS population. Three QTL co-localised for constitutive and induced trichome density phenotypes (accounting for six of the nine QTL) and three were unique to a specific phenotype. A single QTL was mapped for response to wounding, which, interestingly, does not co-localise with QTL mapped for constitutive or induced trichome density. This contrasts with findings from work on another *A. thaliana* defensive trait that compared glucosinolate accumulation in control and methyl jasmonate treated plants, where all loci controlling phenotypic plasticity co-localised with QTL mapped in one of the two environments (Kliebenstein*, et al.*, 2002); however, unique plasticity QTL have been mapped elsewhere, for example in barley (Tetard-Jones*, et al.*, 2011), rice (Shimizu*, et al.*, 2010) and *A. thaliana* (Ungerer*, et al.*, 2003). As proposed under the gene regulation model of phenotypic plasticity (Schlichting & Pigliucci, 1993), such QTL may represent regulatory loci, controlling plasticity by affecting expression of genes with a direct effect on phenotype.

Both mapping populations possessed RILs with positive and negative responses to damage. The frequency of negative responders in the HO population was much greater than that in the SS population; this isn't terribly surprising given that both parental accessions of HO are negative responders and both parents of the SS population are positive responders. Although no QTL for response was mapped in the HO population, one QTL (SSR1) was mapped in the SS population, indicating that there is genetic variation for the response phenotype (plasticity). Interestingly, when the SS RIL data are partitioned according to genotype at the marker nearest SSR1 and constitutive versus induced trichome densities are plotted, it is revealed that only RILs carrying the Sf-0 allele demonstrate zero or negative responses. It is not immediately clear why plants would have a negative trichome density response to wounding. However, this might suggest that for plants that have a high constitutive trichome density, making more trichomes isn't necessarily a good strategy, therefore, plants may instead switch between defence strategies (e.g., producing more glucosinolates). Clearly, further work that focuses on the distribution of naturally occurring positive and negative responders, identifying the genetic basis of the switch, and determining whether negative responders induce defence by other means would be of interest.

2.5.4 Candidate genes for trichome density QTL

Typically, mapped QTL span fairly large intervals containing many genes and, as such, different loci may underlie QTL mapped to similar positions in different environments; similarly, multiple contributing loci could be contained within the QTL intervals identified within an environment. To provide a framework for identifying the genes that underlie mapped QTL, we estimated physical positions of LOD intervals from the physical positions of markers flanking the interval. Based on the extensive literature around the molecular genetic pathway for trichome initiation in *A. thaliana*, we were able to identify a number of strong candidate genes for the QTL mapped here.

Three QTL, HOC2, HOD1, and SSC2, were mapped in close proximity to one another on chromosome II. QTL have been mapped in this region previously in four different mapping populations (Larkin*, et al.*, 1996; Mauricio, 2005; Symonds*, et al.*, 2005; Hilscher*, et al.*, 2009), and in a genome-wide association mapping analysis of 191 *A. thaliana* natural accessions (Atwell*, et al.*, 2010) for trichome traits including number, density and density following damage. This locus explains between 11 and 73% of trichome variation observed in previous mapping studies. The interval covered by HOC2/HOD2/SSC2 includes five candidate genes: an array of trichome initiation repressor R3 MYB genes, *ETC2*, *TCL1*, and *TCL2* (Wang*, et al.*, 2008); *TTG2*, a downstream target of the TTG1:bHLH:MYB complex (Zhao*, et al.*, 2008); and *URM9*/*SAD2*, an importin β-like protein which links jasmonic acid signalling to trichome initiation via regulation of *GL3, GL1, TTG1* and *GL2* (Gao*, et al.*, 2008; Yoshida*, et al.*, 2009). The peak LOD score of HOC2 resides on or very near the tightly linked R3 MYB trichome initiation repressors *ETC2*, *TCL1* and *TCL2*, and a lysine to glutamine mutation in *ETC2* has been suggested as the underlying quantitative trait nucleotide for this locus (Hilscher*, et al.*, 2009). While sequence data from the 1001 Genomes Project (http://www.1001genomes.org/; data not shown) reveal that this polymorphism is segregating between Hi-0 and Ob-0 and between St-0 and Sf-2, it is not segregating between the accessions Ler and No-0 and cannot be determined due to missing data between Bay-0 and Shah, parents of mapping populations where a similarly located QTL has been identified (Symonds*, et al.*, 2005). So, while ETC2 may be a good candidate for QTL mapped to this region, it is not likely to be the only variable trichome initiation gene in the region.

In contrast to HOC2, the peak LOD of the overlapping QTL under induced conditions, HOD1, is positioned closer to *URM9*/*SAD2*. Given its role in jasmonic acid signalling and trichome induction (Gao*, et al.*, 2008; Yoshida*, et al.*, 2009), *URM9*/*SAD2* is a good candidate for this QTL for the induced trichome density trait. The estimated physical position of the SSC2 locus excludes *ETC2*, *TCL1* and *TCL2* while the maximum LOD score for SSC2 is instead located in close proximity to the estimated position of *TTG2*. We consider the *ETC2*/*TCL1*/*TCL2* array, *URM9*/*SAD2* and *TTG2* to be strong (and not mutually exclusive) candidate genes for HOC2, HOD1 and SSC2, respectively. The effect of a combination of tightly linked variable loci might explain why the reported percentage of variation reported for this locus is so variable among mapping populations and so high in particular studies. Interestingly, while *ETC2* seems a likely candidate gene for constitutive trichome density in the HO population, *URM9/SAD2* appears to be a better candidate for a similarly positioned QTL for induced trichome density. Similarly, *ETC2* is a good candidate for constitutive trichome density in the SS population but no such QTL is identified for induced trichome density in that population. This may suggest that *ETC2* underlies some variation in constitutive trichome density but does not play a role in trichome initiation as an induced defence in *A. thaliana*.

HOD2, mapped on chromosome III, appears to co-localise with QTL mapped in three RIL mapping populations including SS, Col x Ler (Symonds*, et al.*, 2005) and Da(1)-12 x Ei-2 (Pfalz *et al.* 2007), a genome wide association mapping study of 191 accessions (Atwell*, et al.*, 2010), and an association mapping study of 94 accessions (Bloomer*, et al.*, 2012). The TTG1 pathway MYB gene *GL1* sits very near the estimated physical position of HOD2 and is a strong candidate gene for this locus. *GL1*, which is required for trichome initiation (Marks & Feldmann, 1989), has been suggested as a candidate in previous QTL (Symonds*, et al.*, 2005) and association mapping (Atwell*, et al.*, 2010) studies of trichome density, with natural variation at this locus shown to contribute to qualitative and quantitative variation in natural accessions of *A. thaliana* (Bloomer*, et al.*, 2012). However, the high frequency pattern of SNP variation in *GL1* which has been reported to underlie trichome density variation in natural accessions (Bloomer*, et al.*, 2012) is not polymorphic between Hi-0 and Ob-0; in fact, no coding changes separate the two alleles, suggesting that this QTL may represent a regulatory polymorphism at *GL1* or simply variation in a different gene.

HOD3, which partially overlaps with SSC4 and a region previously mapped in L*er* x No-0 RILs (Symonds*, et al.*, 2005; Pfalz*, et al.*, 2007), spans the physical position of the TTG1 pathway bHLH *TT8*. *TT8* does not yet have a demonstrated role in regulating trichome initiation on the leaf lamina but our mapping results, together with evidence for a role in trichome initiation on leaf margins and expression in the leaf lamina in response to jasmonic acid (Maes*, et al.*, 2008), suggests that such a role merits further study. Overlapping QTL on chromosome I, SSC1 and SSD1, are positioned near *GL2*, a direct downstream target of the TTG1:bHLH:MYB complex (Zhao*, et al.*, 2008) and *At1G77670*, an aminotransferase which is a direct target of GL3/GL1 in ChIP studies (Morohashi & Grotewold, 2009). SSD1 spans a larger interval than SSC1 that also includes a DELLA protein, *RGL1*, and two regulators of the jasmonic acid response, *JAZ2* and *JAZ9. B*oth *JAZ2* and *JAZ9* can interact with TTG1 pathway genes in yeast-2-hybrid assays (Qi*, et al.*, 2011) and are potential candidate genes for induced trichome density variation.

2.5.5 The TTG1 pathway and natural variation for trichome density

Natural variation in genes within the TTG1 epidermal cell fate pathway has been proposed to underlie much of observed variation in trichome density in natural accessions of *A. thaliana* (Symonds, *et al.*, 2011; Bloomer, *et al.*, 2012). Indeed, QTL and association mapping studies have proposed a number of candidate genes comprising the TTG1:bHLH:MYB complex, including the master regulator *TTG1* (Symonds*, et al.*, 2005; Atwell*, et al.*, 2010), the bHLH *GL3* (Symonds*, et al.*, 2005)*, GL1* (Symonds*, et al.*, 2005; Atwell*, et al.*, 2010)*,* and *ETC2* (Hilscher*, et al.*, 2009; Atwell*, et al.*, 2010) and its tightly linked co-repressors *TCL1* and *TCL2* (Atwell*, et al.*, 2010), as well as a downstream target of the complex, *TTG2* (Symonds*, et al.*, 2005). Confirming the hypothesis, candidate gene analyses have identified natural allelic variation in the R3 MYB repressor *ETC2* (Hilscher*, et al.*, 2009), the bHLH *AtMYC1* (Symonds*, et al.*, 2011), and the R2R3 MYB *GL1* (Bloomer*, et al.*, 2012) as contributing to natural qualitative and quantitative variation in trichome density. Here, we identify the *ETC2*/*TCL1*/*TCL2* R3 MYB array, the bHLH *TT8*, *GL1*, and the downstream targets *TTG2, GL2* and *At1G77670* as candidates for loci mapped for trichome density variation. Considering recent findings linking hormonal signalling to constitutive and induced trichome density, we also propose the gibberellin response regulator *RGL1* and the jasmonic acid response regulators *URM9*/*SAD2*, *JAZ2* and *JAZ9* as possible candidates. Clearly, many of the QTL for trichome density identified both here and elsewhere may be independent of the pathway;

however, mounting evidence suggests that genes within, upstream and downstream of the TTG1 pathway play an important role in shaping trichome density variation in nature.

2.6 Conclusions

In this work, we have mapped QTL for constitutive and induced trichome density and for trichome density response to damage in two new RIL populations of *A. thaliana*. Our results confirm QTL in positions mapped previously and identify new QTL in novel positions. Importantly, we have identified QTL x QTL interactions and QTL for the response to wounding that appear to be independent of constitutive and induced trichome density QTL. The results identify many strong candidate genes that support the hypothesis that much of standing variation for trichome density in *A. thaliana* is determined by allelic variation in within the TTG1 pathway and in integrating genetic pathways. Future efforts should focus on identifying the polymorphisms that underlie these QTL and extending this work beyond trichome density in *A. thaliana* to determine whether the broadly conserved TTG1 pathway underlies variation in other epidermal traits and across plant groups.

2.7 Acknowledgements

RHB screened markers and collected genotypes, developed linkage maps, performed trichome density experiments and preliminary QTL analyses and interpreted results. Prashant Joshi provided technical assistance with collection of the ~23000 genotypes used in this study. Vaughan Symonds developed the mapping populations, assisted with experimental design and assisted with QTL analyses in R/qtl. Funding for this work was provided by a National Science Foundation grant (IBN-0344200), and a New Zealand Royal Society Marsden Fund grant (09-MAU-114).

2.8 Supplementary Materials

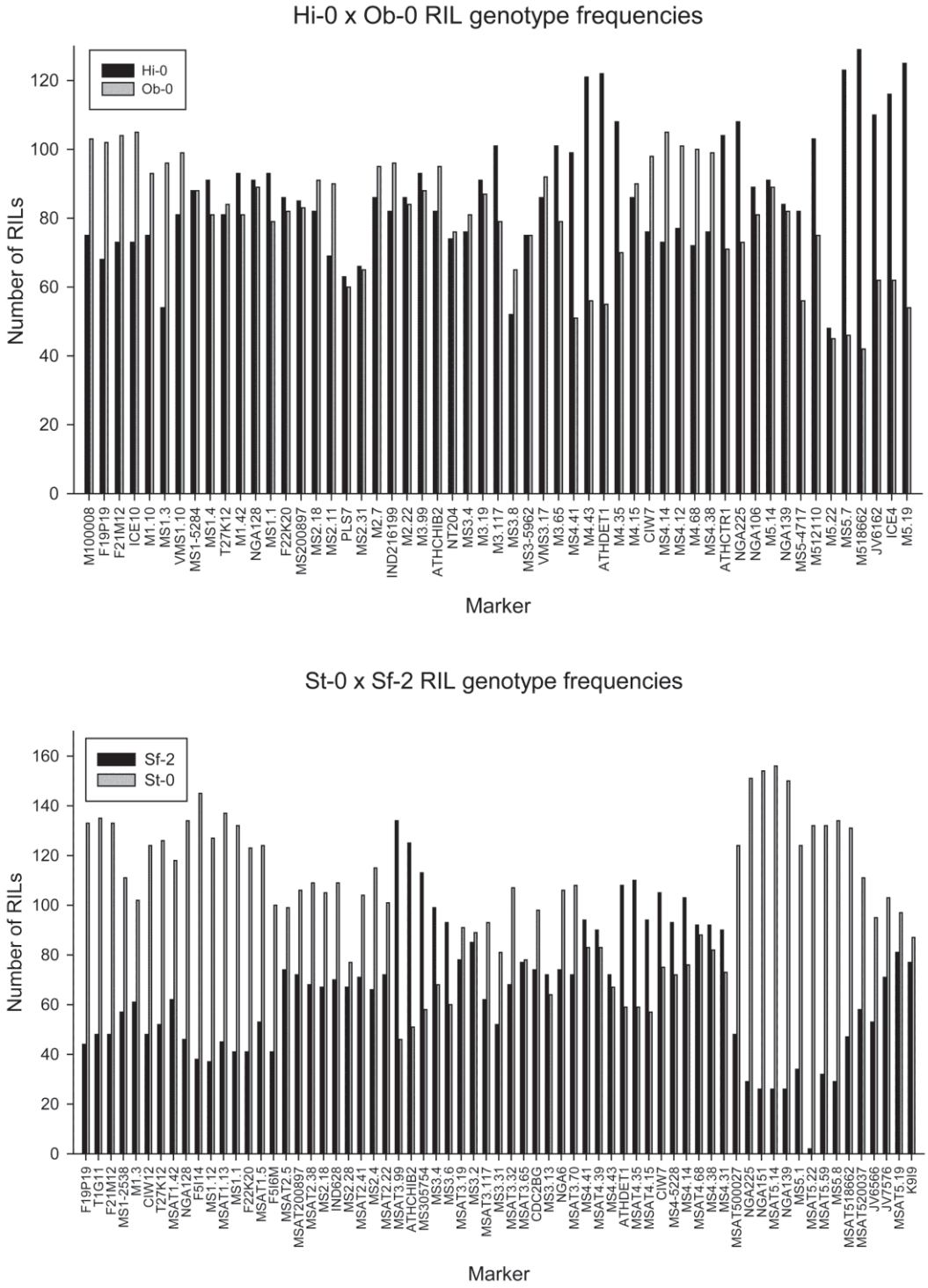

Figure 2.5: Genotype frequencies for each marker in each population; Supplementary Material.

In the Hi-0 x Ob-0 population (top), the number of RILs carrying the Hi-0 allele at each marker is indicated by the black bar, while the number of RILs carrying the Ob-0 allele is given by the grey bar. In St-0 x Sf-2, St-0 alleles at each marker are shown by the grey bar and Sf-2 alleles by the black bar. Segregation distortion is present in both populations but is considerably stronger in the St-0 x Sf-2 population; for significance of difference in allele frequencies see Table 2.4, Supplementary Material.

Table 2.4: Microsatellite markers used to generate linkage maps for Hi-0 x Ob-0 and St-0 x Sf-2 RILs; Supplementary Material.

			HO		SS				
Marker	**Linkage Group**	**Physical Position (bp)**	**Map position (cM)**	**Significant SD**	**Map position (cM)**	**Significant SD**	**Source**	**Forward primer**	**Reverse primer**
M100008	1	8639	0	**			VAST	CCACGCATCTGTGTTCACTC	GCAAAAGTGGGCCTAACAAA
F19P19	1	1157727	2.598	***	0†	*****	VAST	CCACGTAGGTCAAGAAGAAGAAG	TGTCTGCTGCGATAGAGAGAG
T1G11	1	1243352			0.424†	*****	VAST	GAAGACAAAGCTCTGCAGTAATG	AATTGCATAAGGCACTTGAAAG
F21M12	1	3212191	6.784	**	5.105	*****	VAST	GGCTTTCTCGAAATCTGTCC	TTACTTTTTGCCTCTTGTCATTG
ICE10	1	4015381	10.47	**			VAST	AACATCCACAAGTTTCTAAAACAATC	GACTCTTATGGGTAAGCTCCTTG
MS1-2538	1	6165662			13.956	*****	VAST	TGGGTTTTTAATGGCGTTTG	CCGGCGACTTATGCTTACC
MSAT1.10	1	7296649	22.834	-			VAST	ATGGTGAGATACTGAGATTAT	CGAGAAGGTCTAAAGGTA
MSAT1.3	1	8322175	29.082	*****	28.392	****	VAST	GGAACTGTTGTCTGGGTAAG	CGATTGCACTAAAAGCTCTC
CIW12	1	9621344			36.254	*****	VAST	AGGTTTTATTGCTTTTCACA	CTTTCAAAAGCACATCACA
VMS1.10	1	10	40.157	-			This paper	ACCAACTCGGATTCGCCTCG	TTTGCTGGCGATATGAGTCACTTG
MS1-5284	1	12511232	54.917	-			EuMicrosat	CCTGAAAACCCCGTTAAAAG	CGGGACCTTTGTGTGATATTC
MSAT1.4	1	14160180	63.02†	-			VAST	CTAAACTAGAACCAGGGGTAA	ACAAAAATCGTGGTGATAATA
T27K12	1	15926702	64.389†	-	54.593	*****	VAST	GGAGGCTATACGAATCTTGACA	GGACAACGTCTCAAACGGTT
MSAT1.42	1	18153083	74.759	-	65.695	*****	VAST	TAACTAAAATTGCACCC	ACAAATAAAATTGAATATCAG

NGA128	1	20633251	87.511	-	76.226	*****	VAST	GGTCTGTTGATGTCGTAAGTCG	ATCTTGAAACCTTTAGGGAGGG
F5I14	1	24374008			87.92	*****	VAST	CTGCCTGAAATTGTCGAAAC	GGCATCACAGTTCTGATTCC
MSAT1.12	1	25686587			93.066	*****	VAST	TTAGAGATTCGCCAACCTC	CGTGTGCCCAACCA
MSAT1.13	1	25827433			96.928	*****	VAST	CAACCACCAGGCTC	GTCAAACCAGTTCAATCA
MSAT1.1	1	26398711	106.553	-	98.67	*****	VAST	ATACGATAAGATTTATTAGCA	CCCATGCTCTTTTTGTGAAA
F22K20	1	28895040	116.332	-	104.549	*****	VAST	TTTTTGGTGAGATTTTAAGCCC	ATATCTCCATCGCTGCAACC
MSAT1.5	1	29016886			104.557	*****	VAST	GCATCGCTCTTAAACAACCAT	CGTTGCAAAACCGTATCAGAA
F5I6M	1	30192367			109.661	*****	VAST	GACCCTAGAGGCAAACCAAATG	AGTGAGTAAATGGTGGGGAATCG
MSAT2.5	2	208178			0	*	VAST	TGAGAGGGACAGATAGGAA	ATCAAAAGGGATACTGACAA
MSAT200897	2	897197	0	-	10.632	**	VAST	AAAAATTTTAACCCATGTTGTAGCA	AGAAGTCGGTGGATGAGCAC
MSAT2.38	2	2457014			19.183†	****	VAST	TGTAACGCTAATTTAATTGG	CGCTCTTTCGCTCTG
MSAT2.18	2	2799644	8.528	-	19.819†	****	VAST	TAGTCTCTTTTGGTGCGCATA	AGCCTCTCCAAGCTTAGGTCT
IND628	2	6280231			23.184†	****	VAST	AATTTAAGAAAGAACACAATTTTGAGA	CCATTTCAGCTCTTTACACTTCTACA
MSAT2.28	2	6409845			24.096†	-	VAST	AATAGAAATGGAGTTCGACG	TGAACTTGTTGTGAGCTTTG
MSAT2.11	2	8227826	24.945	*			VAST	GATTTAAAAGTCCGACCTA	CCAAAGAGTTGTGCAA
PLS7	2	9813750	29.735	-			VAST	GATGAATCTTCTCGTCCAAAAT	GACAAACTAAACAACATCCTTCTT
MSAT2.31	2	10778033	33.798	-			VAST	GCTCCTCTTTGCCGCTAG	GCGATTTCATCTTGTGCATC

MSAT2.41	2	11095452			42.648	**	VAST	GACTGTTTCATCGGATCCAT	ACAAACCATTGTTGGTCGTG
MSAT2.7	2	13192607	47.843	-			VAST	CTCAAATCAAGAACGCTGAC	CCCGATATAGACAACGACAA
MSAT2.4	2	13831870			54.808	*****	VAST	TGGGTTTTTGTGGGTC	GTATTATTGTGCTGCCTTTT
IND216199	2	16199330	55.207	-			VAST	TGGCTCATTTCACTCACAGG	GGCTATCTATGGCCGTTATGA
MSAT2.22	2	19632943	69.455	-	75.988	**	VAST	CGATCCAATCGGTCTCTCT	TGGTAACATCCCGAACTTC
MSAT3.99	3	1514825	0	-	0	*****	VAST	CCAACAAGAACAAGACATT	AGTGAAATCTGAAGCGATA
ATHCHIB2	3	3963391	11.193	-	3.039	*****	VAST	GGATCCAAGTGCTCATATATAC	CTTTCGTTTCTAAATATGAGAAGC
NT204	3	5570082	17.117	-			VAST	TGGAAGCTCTAGAAACGATCG	ACCACCTAAACCGAGAATTGG
MSAT305754	3	5754608			7.814	*****	VAST	AGAAGAAGCGCACATGTTTTT	AAAATTGTACGTTTGATTGTTGAA
MSAT3.4	3	7672410	23.793	-	17.718	**	VAST	TGTAACATTCGCATTTTATG	CCTCTTCGGTCTATATGACT
MSAT3.6	3	8467190			25.118	***	VAST	CTACTTTACGCCTTTTATTG	TGAATTGTCCCGAGAC
MSAT3.19	3	8808167	36.21	-	29.328	-	VAST	TAATTCGATCCAATTGACAT	TGGCTTGGCACAAAC
MSAT3.2	3	9055511			30.85	-	VAST	AAGGTACGGCGGTGGATATTG	CGGGGATTTCTTCTTCCTGTG
MSAT3.117	3	9860185	41.648	-	38.296	**	VAST	CCCGTTTAGAAATATGTGT	CGTCATACCATTTTATCAG
MSAT3.8	3	11002949	49.586	-			VAST	ATGTTAAAAACCCGTGTTGG	TTTAACCTTATCCGGGAAAG
MSAT3.31	3	11037360			48.772	**	VAST	CTCTTTTTCTTATGCCAA	AGAGTGGGTTGATACCAT
MSAT3.32	3	11208231			52.728	****	VAST	GCACTTGCAGCTTAACTT	CGTGACTGTCAAACCG
MS3-5962	3	15392985	53.817	-			EuMicrosat	CACGTGCGAACGACTACAAC	GCTCCAAAACGGAACCCTAC
VMS3.17	3	17	63.063	-			This paper	GCCGGGCTACAAGCCCAGAAAAG	ACCGGTGAGGAATGGCCAGC

MSAT3.65	3	19835420	86.335	-	74.988	-	VAST	CCAACTCTTCTTTCAAACA	TCTCTTAGCAAACGAATGTC
CDC2BG	3	20070710			77.153	*	VAST	GGGAAAAACGAAGTGACGTG	ATTGAACTGTGTTGGTTTCTGG
MSAT3.13	3	21377089			84.888	-	VAST	TTGTGTGTTTGCGATC	CATATCCGTTTTTATGTTTT
NGA6	3	23042025			90.284	**	VAST	TGGATTTCTTCCTCTCTTCAC	ATGGAGAAGCTTACACTGATC
MSAT3.70	3	23449179			91.296	***	VAST	CACCCATAGACACACAACC	TCAAAACCTCCAACCATAC
MSAT4.41	4	42208	0	*****	0†	-	VAST	CCTACACTCATCAGATTCAG	TAATTAGGGTTTGATCATTC
MSAT4.39	4	89498			0.739†	-	VAST	GTTATCACATTAAAATCACC	CCAATTGTAATATATGAACA
MSAT4.43	4	2575724	11.23	*****	10.854	-	VAST	GCATTGGACCATAAGGG	TCAAGACGGTTTCAAGTTG
ATHDET1	4	6346349	19.741	*****	29.233	*****	VAST	TTCAAACACCAATATCAGGCC	GGTGAAAATGGAGGAGACGA
MSAT4.35	4	7549254	28.94	****	33.361	*****	VAST	CCCATGTCTCCGATGA	GGCGTTTAATTTGCATTCT
MSAT4.15	4	9362588	41.419	-	40.877	****	VAST	TTTCTTGTCTTTCCCCTGAA	GACGAAGAAGGAGACGAAAA
CIW7	4	11524362	55.634	*	56.022	**	VAST	AATTTGGAGATTAGCTGGAAT	CCATGTTGATGATAAGCACAA
MS4-5228	4	13421331			64.347	-	EuMicrosat	GCTCCTCGAGAAAAAGAACG	TCTCACATCACGGGACAATC
MSAT4.14	4	15212515	72.547	**	70.668	**	VAST	GACCGTTTCTAGTGCTCACA	ACGGAATAAGCGGAGGA
MSAT4.12	4	15960078	73.708	*			VAST	AAAGGAAGAAGAAGACTGTT	AGAAGAAGAAGCGAGATT
MSAT4.68	4	17261718	79.464	**	80.69	-	VAST	GGTCTCTTTTTCCCTTAAC	CATTGTCCATCATCAGAAC
MSAT4.38	4	18412248	84.821	*	88.353	-	VAST	GCCTTATAGTACACCCAAA	CCACTCCACTCTCGAA
MSAT4.31	4	18573159			88.848	-	VAST	AGGGATATGGATTGAGA	GCCGTATAACTATTGGTT
MSAT500027	5	27394			0	*****	VAST	AAAAGGAAGACAAAGCTTGACG	AACCCCACATGGCCTTATTT
ATHCTR1A	5	979763	0	**			VAST	TATCAACAGAAACGCACCGAG	CCACTTGTTTCTCTCTCTAG
NGA225	5	1507104	1.475	***	8.504	*****	VAST	GAAATCCAAATCCCAGAGAGG	TCTCCCCACTAGTTTTGTGTCC
NGA151	5	4669932			16.457	*****	VAST	GTTTTGGGAAGTTTTGCTGG	CAGTCTAAAAGCGAGAGTATG

									ATG
NGA106	5	5397351	12.377	-			VAST	GTTATGGAGTTTCTAGGGCACG	TGCCCCATTTTGTTCTTCTC
MSAT5.14	5	7498509	25.871	-	23.619	*****	VAST	AACAACCCTATCTTCTTCTG	TGTGACCCCTTACTCAATA
NGA139	5	8428136	31.946	-	27.093	*****	VAST	GGTTTCGTTTCACTATCCAGG	AGAGCTACCAGATCCGATGG
MS5-4717	5	10960119	54.19†	**			EuMicrosat	GCCAATCCGAGTAAATCGTG	GGTGGACCGTTCCATTTTG
MSAT512110	5	12109959	54.759†	**			VAST	GATCTGATATACCTCGATTTTC	TTTAATAAGTCTTTTCTCTACGC
MSAT5.1	5	13924203			31.696	*****	VAST	TCTAGCTTTATCCTTCTTTGA	GCCTATTCCGAGATTCA
MSAT5.22	5	13961058	60.913†	-	39.808	*****	VAST	AGAACAAGTTAGGTGGCT	GGGACAAGAATGGAGT
MSAT5.59	5	15461361			51.898	*****	VAST	CATAAGTCAGTTGATTCCA	GCTAAAACTAGAGCGAGAG
MSAT5.7	5	17071433	79.241	*****			VAST	GCTTGTACATTACCGGACA	CCAAGCGTACTCGATCA
MSAT5.8	5	17163228			57.879	*****	VAST	GAAACATTTCATGGGCTTT	GAGAGCAGAGGCAAGTCA
MSAT518662	5	18662765	86.687	*****	64.629	*****	VAST	GCTCCAAAACCCAACAGGTA	TGGCCTCAAAGCTACCCTTA
MSAT520037	5	20037472			71.238	*****	VAST	CTTTCGGTGGATAAATCATGTTA	CGCATACAAATTTTCTGCTATGT
JV6566	5	22027830			82.103	*****	VAST	TTTCTTTCACGGTTTCTAACTTTT	TCATCTTCCCTTTAATTCTGATGA
JV6162	5	22527623	110.994	*****			VAST	CGCTTTCCTTGTGTCATTCC	AAATGCAAATATTGATGTGTGAAA
ICE4	5	23837141	115.024	*****			VAST	CACGAGGAATCTGGCATGGTCG	AGCGATTGCAAGCGGCTCAAG
JV7576	5	23879358			93.282	**	VAST	CACAATCAGAGGGGGTTGAT	AAATTTTGGGGGAAATGAAA
MSAT5.19	5	25924795	122.18	*****	98.609	-	VAST	AACGCATTTGCTGTTTCCCA	ATGGTTATCTCATCTGGTCT
K9I9	5	26983961			112.624	-	VAST	GATTCAGAGTTACGGTAACAG	CTATTGACAGTAAAGAGTCG

3 NATURAL VARIATION IN *GL1* AND ITS EFFECTS ON TRICHOME DENSITY IN *ARABIDOPSIS THALIANA*

This chapter published as:

Bloomer, R. H., Juenger, T. E., & Symonds, V. V. (2012). Natural Variation in *GL1* and Its Effects on Trichome Density in *Arabidopsis thaliana*. *Molecular Ecology, 21*(14), 3501-3515.

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: REBECCA BLOOMER

Name/Title of Principal Supervisor: VAUGHAN SYMONDS

Name of Published Research Output and full reference: Bloomer, RH, Juenger, TE & Symonds, VV (2012). Natural variation in GL1 and its effects on trichome density in Arabidopsis thaliana. Molecular Ecology 21(14), 3501-3515

In which Chapter is the Published Work: 3

Please indicate either:

- **The percentage of the Published Work that was contributed by the candidate:** 95%
  and / or
- **Describe the contribution that the candidate has made to the Published Work:**

RB
Candidate's Signature

7 FEB 2012
Date

Principal Supervisor's signature

07/02/13
Date

GRS Version 3– 16 September 2011

3.1 Abstract

The ultimate understanding of how biological diversity arises, is maintained, and lost depends on identifying the genes responsible. Although a good deal has been discovered about gene function over the past few decades, far less is understood about gene effects; i.e., how natural variation in a gene contributes to natural variation in phenotypes. Trichome density in *Arabidopsis thaliana* is an ideal trait for studies of natural molecular and phenotypic variation, as trichome initiation is genetically well-characterised and trichome density is highly variable in and among natural populations. Here, we show that variation at *GLABRA1* (*GL1*), an R2R3 *MYB* transcription factor gene, which has a role in trichome initiation, has qualitative and likely quantitative effects on trichome density in natural accessions of *A. thaliana.* Specifically, we characterize four independent loss-of-function alleles for *GL1*, each of which yields a glabrous phenotype. Further, we find that a pattern of common polymorphisms confined to the *GL1* locus is associated with quantitative variation for trichome density. While mutations resulting in a glabrous phenotype are primarily coding changes, the pattern resulting in quantitative variation spans both coding and regulatory regions. These data show that *GL1* is an important source of trichome density variation within *A. thaliana* and, along with recent reports, suggest that the TTG1 epidermal cell fate pathway generally may be the key molecular genetic source of natural trichome density variation and an important model for the study of molecular evolution.

3.2 Introduction

The identification and characterization of the genes that underlie natural variation is central to our understanding of the origins and maintenance of biological diversity, and remains a significant challenge for evolutionary biology. To date, our understanding of gene function primarily is derived from studies utilizing gene knockout or overexpression. While these approaches are powerful, it is unclear how well an expectedly extreme knockout or overexpression phenotype reflects the potential range of effects that natural allelic variation in the same gene may confer (Tonsor*, et al.*, 2005). Indeed while we now have a "function" ascribed to several thousand genes across many taxa, in only a small fraction of these has a phenotypic effect due to natural mutations been described. Despite any potential limitations of functional genetics, those data remain our best entry point for identifying the molecular genetic bases of biological variation.

Here we examine natural variation in the R2R3 MYB transcription factor gene *GLABRA1* (*GL1*), which was originally identified through a forward genetic screen (Oppenheimer*, et al.*, 1991). GL1 is a member of the *Arabidopsis thaliana* TTG1 epidermal cell fate pathway with a specific role in trichome initiation. Formation of trichomes in *A. thaliana* requires interactions between the master regulator WD-repeat protein TTG1, one of the functionally overlapping bHLH encoding transcription factors, GL3 or EGL3, and GL1 to form a transcriptional activation complex that induces expression of downstream genes such as *GL2* (Szymanski*, et al.*, 1998) and *TTG2* (Ishida*, et al.*, 2007). TTG1 and GL1 are known to physically interact with the bHLHs, but not with one another. The initiation process is impeded in cells adjacent to trichomes by a suite of single repeat MYBs, which outcompete GL1 for binding with the bHLHs (Wester*, et al.*, 2009). Plants homozygous for a loss-of-function allele of *GL1* (e.g., *gl1-1*) initiate no trichomes on the leaf blade and a weak mutant allele, *gl1*-2, missing the last 27 amino acid residues, has a reduced trichome density phenotype (Esch*, et al.*, 1994). Interestingly, the original *GL1* mutant, *gl1-1* (Col-5), is purportedly a natural "glabrous" allele of *GL1* introgressed into the Col-0 background but has never been fully characterized at the molecular level. Natural accessions and populations of *A. thaliana* show both qualitative and quantitative variation for leaf trichome density with leaf blades varying from completely glabrous to densely hairy (Symonds*, et al.*, 2005); however, the extent to which *GL1* may underlie this variation has not been clear.

Trichomes are single- or multi-cellular projections from the plant aerial epidermis, found in most land plant lineages, although not all trichomes are homologous (Werker, 2000). Trichomes have a diverse range of adaptive functions among species, from protection against insect herbivory (Mauricio, 1998), high light (Liakopoulos*, et al.*, 2006) and UV radiation (Karabourniotis*, et al.*, 1995) to improvement of water use efficiency and drought tolerance (Hogan*, et al.*, 1994; Ennajeh*, et al.*, 2006). A particularly good example of trichome diversification can be found within the Brassicaceae, where many different trichome types have evolved apparently multiple times (Beilstein*, et al.*, 2008). In *A. thaliana*, leaf trichomes provide protection from herbivory, with insect predation shown to exert positive selection on increased trichome density (Mauricio & Rausher, 1997). However, increased density has a fitness cost in the absence of such predators (Mauricio, 1998; Karkkainen*, et al.*, 2004; Steets*, et al.*, 2010) and further may increase susceptibility to fungal infection (Calo*, et al.*, 2006); the "cost" of trichome production was recently characterized by Zust *et al.* (2011), showing that glabrous plants had higher size standardized growth rates relative to trichome producing plants in the same genetic background. In wild populations of the closely related species *A. lyrata*, trichome production is also under divergent selection (Karkkainen*, et al.*, 2004), and similar results have been shown for the allopolyploid *A. kamchatica*, with fitness consequences of trichome production differing between populations in different environments (Steets*, et al.*, 2010). The cost of trichome production has also been measured in *A. halleri*; however, in this case, no corresponding fitness advantage was found under insect predation (Kawagoe*, et al.*, 2011). In the latter case, selection was intense, potentially overwhelming any trichome advantage under herbivory. In *A. thaliana*, the demonstrated alternative forms of selection should be expected to maintain genetic variation, phenotypic plasticity, or both for trichome density within or among populations. Although mapping results have demonstrated a strong genetic basis for trichome density variation (Larkin*, et al.*, 1996; Mauricio, 2005; Symonds*, et al.*, 2005; Atwell*, et al.*, 2010), relatively little is known about the molecular bases of this variation in nature.

Thus far, natural variation in just two genes has been shown to be associated with quantitative trichome density variation and both genes are in the TTG1 genetic pathway. High frequency variation in the coding region of the single repeat MYB, *ETC2*, has been linked to trichome density variation in *A. thaliana*, purportedly explaining a large proportion of natural trichome density variation (Hilscher*, et al.*, 2009); however, *ETC2* is tightly linked with two

other known trichome initiation genes and it is unclear whether the reported effect is due to one or several of these loci. Variation in one of the pathway's four bHLH transcription factors, *AtMYC1*, has also been shown to underlie significant trichome density variation in *A. thaliana* (Symonds*, et al.*, 2011). Mapping results have located QTL near several other TTG1 pathway genes (Symonds*, et al.*, 2005; Atwell*, et al.*, 2010), including *GL1*, but none of these have been confirmed.

Analysis of *A. thaliana* laboratory mutants has shown that the qualitative loss of trichomes on leaf blades can be caused by the loss of function of either *TTG1* or *GL1*. The *ttg1* mutant is strongly pleiotropic with knockouts affecting a range of adaptively important epidermal traits (Koornneef, 1981), while *GL1* knockouts are trichome-specific (Marks & Feldmann, 1989) or only weakly pleiotropic, with a suggested role in cuticle development (Xia*, et al.*, 2010), and so is considerably more likely to persist in natural populations; indeed, no natural *ttg1* loss-of-function mutant has been described, while one such mutation in *GL1* has been characterised previously (Hauser*, et al.*, 2001). Although no evidence for a quantitative effect on trichome density was associated with *GL1* variation by Hauser *et al.* (2001), this may be due to a small sample size. The frequency of glabrousness in wild *A. thaliana* is unknown. Several other natural accessions described as glabrous have been collected but the molecular basis of the phenotype thus far has been unidentified. While mutations in *GL1* would seem likely to be responsible for the glabrous phenotype in these accessions, as was found in *A. lyrata* (Kivimaki*, et al.*, 2007), this has not been closely examined. Further, if *GL1* is typically responsible for glabrousness in *A. thaliana*, it is unclear whether different accessions might carry independent mutations or a single allele has arisen and spread to other populations.

Direct examination of molecular variation in *GL1* from a broad sample of *A. thaliana* allows us to address: (1) To what extent does variation in *GL1* underlie glabrousness in *A. thaliana*? (2) Is there a single *GL1* mutation that has spread to multiple populations or are there multiple independent mutations that underlie the glabrous phenotype? (3) Does allelic variation in *GL1* have a quantitative effect on trichome density? (4) What type(s) of mutations (coding or regulatory) underlie trichome density variation? (5) Is there a difference between the types of mutations that underlie qualitative versus quantitative phenotypic variation? Our results better define the natural role of *GL1* in generating trichome density variation specifically and expand on our understanding of the molecular nature of biological diversity generally.

3.3 Materials and methods

3.3.1 Molecular variation at *GL1* in trichome producing accessions

GL1 sequences from 115 trichome producing *A. thaliana* accessions were analysed for sequence variation; these sequences comprise Panel 1 (Table 3.4, Supplementary Material). This panel is made up of sequences from the common, globally distributed 96 accession set (CS22660) and other natural accessions; a "natural accession" in the *A. thaliana* research community consists of seed collected from one or more individuals from a natural population. As *A. thaliana* is predominantly self-pollinating, seed for a particular accession typically represent one genotype. Ninety-two of the *GL1* sequences are new (see below) and 23 were acquired from Genbank (Table 3.4, Supplementary Material).

3.3.2 *GL1* sequencing

Genomic DNAs were isolated from fresh rosette leaf tissue using a modified CTAB protocol (Doyle & Doyle, 1987) from 92 accessions (Table 3.4, Supplementary Material). The *GL1* locus was amplified by PCR in three overlapping fragments, including approximately 420 bp upstream of the start codon, the coding region, and 470 bp downstream of the stop codon. Sequences for primers used for PCR are supplied in Supplementary Material (Table 3.5, Supplementary Material). PCR reactions were carried out in a 25 μL reaction volume containing 0.2 μL of Accuprime Taq DNA polymerase as supplied, 1X Accuprime Taq Buffer II, 0.4 μM of each primer and ~50 ng genomic DNA template. Amplification was carried out using a cycling scheme of 3 minutes at 94^{o}C; 32 cycles of denaturation at 94^{o}C for 30 seconds, primer annealing at 52^{o}C for 30 seconds and extension at 68^{o}C for 3 minutes; this was followed by ten minutes at 68^{o}C and a 4^{o}C hold.

Several accessions contained an uninterrupted poly-A repeat in intron two which resulted in poor quality sequence; in these accessions, the amplified fragment was cloned into *E. coli* using a TOPO-TA cloning kit (Invitrogen) and colonies were screened for positive transformants by PCR with kit primers. Plasmids were extracted from positive colonies using a Qiagen plasmid miniprep kit (Qiagen, Inc.), with extracted plasmids serving as template for sequencing reactions.

The *GL1* locus was sequenced from PCR or plasmid templates using BigDye Terminator v3.1 chemistry and protocols (Applied Biosystems). Contiguous sequences were assembled for each accession using SEQUENCHER v4.7 (Gene Codes) and double-checked and corrected manually.

While analysing our sequence data, several sets of "resequenced genomes" from *A. thaliana* accessions became available via the 1001 Genomes Project (1001genomes.org). We selected the largest of these data sets and acquired *GL1* sequences from those 261 lines at: http://signal.salk.edu/atg1001/3.0/gebrowser.php; the 261 lines are listed in Supplementary Material (Table 3.6, Supplementary Material). Fifty-five accessions were in common between our data set and the resequence data set; these make up Panel 2 (Table 3.4, Supplementary Material). To determine whether our data and the resequence data should be analysed together, pairwise sequence comparisons were made between the Sanger-sequencing-derived data and the resequence-derived *GL1* sequences for Panel 2. Because of the high frequency of discrepancies in these comparisons (described in the results section), we focused on analysing our panel of 115 accessions (Panel 1) and present a subset of analyses on the data from the resequence data as supplementary material (see Results).

3.3.3 Molecular data analyses

An alignment of *GL1* genomic sequences from 115 trichome producing accessions (Panel 1), including approximately 65 bp of 5' flanking sequence and 83 bp of 3' flanking sequence, was generated using CLUSTALL W in BIOEDIT v7.0.9.0 (Hall, 1999). To assess nucleotide diversity in the genomic *GL1* alignment a sliding window analysis of π was generated in DNASP v5.10 (Rozas*, et al.*, 2003), using a window of 50 bp and a step size of 10 bp. These analyses only included sequences from trichome producing accessions; glabrous accessions likely carry pseudogenes of *GL1* which could artificially inflate diversity statistics.

Using intron/exon boundaries from Columbia *GL1* cDNA sequence, an inferred cDNA alignment of *GL1* from 120 accessions was produced which included 115 trichome-producing accessions (Panel 1) and five glabrous accessions (described in a later section and listed in Table 3.4, Supplementary Material as Panel 3); *GL1* sequences from glabrous accessions were

included here to evaluate their placement relative to all functional haplotypes. A haplotype file for the 120 accessions was generated using DNASP v.4.5 (Rozas*, et al.*, 2003) from the cDNA nucleotide alignment. The haplotype file was generated with gaps considered and exported to NETWORK v.4.5 (Fluxus Engineering) to create a haplotype network using median-joining (Bandelt*, et al.*, 1999).

Departure of *GL1* sequences from the neutral model of evolution was tested by calculating Tajima's D (Tajima, 1989), and Fu and Li's D* and F* (Fu & Li, 1993). All calculations were made on genomic and cDNA alignments of the 115 accession panel (Panel 1), the resequence data from 261 accessions, and the 55-accession panel (Panel 2). Based on a pattern revealed by the haplotype network, departure from neutrality was also tested by comparing the alleles of two predominant haplogroups. To this end, a sliding window analysis of *K*a/*K*s ratios across a cDNA alignment of Panel 1 sequences, with a window of 30 bp and step size of 9 bp using DNASP v5.10 (Rozas*, et al.*, 2003). The two major haplogroups were used as "interspecific" and "intraspecific" datasets. Very low polymorphism in parts of the gene, yielding *K*s values of zero, meant that a ratio could not be calculated using sliding window values of *K*s across much of the gene. To overcome this, sliding window values for *K*a only were used and the ratio calculated using a gene-wide value for *K*s as in Shiu *et al.* (2004).

3.3.4 *GL1* and quantitative trichome density variation

3.3.4.1 Trichome density phenotyping

To screen trichome density in the set of 96 accessions (CS22660; Panel 4 in Table 3.4, Supplementary Material), seeds were planted in 4 cm cells of seed raising mix (Oderings Nurseries, NZ) in 72-cell flats (Hummert International, Inc.), with approximately five seeds planted per cell and thinned to one plant per cell 7 days post-germination. The core set of 96 accessions is particularly useful for association mapping as they have been well studied with regard to population structure (Nordbord *et al.* 2005); if not accounted for, population structure is a confounding source of false positives in association mapping (Lander & Schork, 1994). Five replicates of each accession were planted in a fully randomized layout across seven flats. Potted seeds were vernalised for 8 days at 4^{o}C then transferred to a plant growth room at approximately 23^{o}C under an 18:6 hour light:dark cycle. All flats were placed on a

single shelf in a plant growth room after vernalisation and were rotated in position and orientation daily to control for potential variation in light and temperature across the shelf. Trichome density was scored at 27 days post-germination; at this point, trichomes are still living cells. At this stage of development the fifth true rosette leaf was fully expanded and not yet senescent. Trichomes were counted on the fifth true rosette leaf in a 17 mm^2 area midway along the length of the leaf blade between the midrib and leaf margin using a dissecting microscope at 25x magnification according to Symonds *et al.* (2005). Mean trichome density scores were calculated for each accession and broad-sense heritability (H^2) was calculated from ANOVA results.

3.3.4.2 Association mapping

Association mapping was conducted using our trichome density and *GL1* sequence data. Three GL1 polymorphisms were chosen for association mapping. The two highest frequency polymorphisms were examined: (1) a serine deletion at amino acid position 148 (S148-), and (2) a valine/isoleucine polymorphism (V224I), representing a pattern of high frequency variation that spans several polymorphisms and defines two predominant *GL1* haplogroups (later referred to as haplogroups A and B). The third polymorphism examined is a proline/arginine replacement (P176R). The P176R polymorphism was of interest, as it differentiates the Col and L*er GL1* alleles; a QTL has been mapped near *GL1* in the Col x L*er* mapping population previously (Mauricio, 2005; Symonds*, et al.*, 2005; Atwell*, et al.*, 2010). The 94 accessions (of the starting 96, one was glabrous and one didn't germinate in the experiment) were sorted according to haplotype at each of the three polymorphisms considered, and each polymorphism was tested independently for an association between haplotype and trichome density. The data were analysed using a unified mixed model association mapping approach that included various forms of statistical control for population structure (Yu*, et al.*, 2006). In the simplest model, we completed a "naive" analysis fitting independent linear models for each fixed effect SNP using a standard error model in PROC MIXED in SAS (Littell, 2006). In subsequent models, we included a random effect representing the kinship (K or K*) matrix relationships among accessions or a fixed population structure (Q) matrix, or both (QK) as in (Zhao*, et al.*, 2007) with Proc Mixed as implemented in JMP Genomics 4 (SAS institute, Cary NC). Here, the QK terms reduce confounding due to population structure and thus improve our test for association between polymorphisms in *GL1* and trichome density. Our K and Q matrices were derived from the genome-wide

polymorphism screen of the 96 accessions as published in Nordborg *et al.* (2005) and as utilized in QK model association mapping in Zhao *et al.* (2007). To explore the structure of polymorphism surrounding the *GL1* locus we made use of the 250K SNP chip genotypes project dataset (v3.06) (http://www.gramene.org/db/diversity/). We extracted all SNP polymorphisms 20 kb to either side of GL1 for the 1,307 available accessions. The pattern of pairwise linkage disequilibrium was explored by calculating r^2 between pairs of SNPs using a Pearson correlation in the statistical package R (http://www.r-project.org/).

3.3.5 *GL1* and qualitative trichome density variation

3.3.5.1 Plant materials

All 34 accessions listed by the ABRC as "glabrous" (Table 3.4, Supplementary Material) were obtained and grown to confirm the phenotype. Of these 34, only seven represented unique accessions with completely glabrous rosette leaf blades (Table 3.4, Supplementary Material); the remainder either produced trichomes or represented an already sampled lineage (e.g., Wil-2 and Wil-3). One representative of each lineage purported to be glabrous but found not to be (four accessions) were included in the *GL1* sequencing described in the previous section. Glabrous lines were grown in seed raising mix (Oderings Nurseries, NZ) in a growth room at 24^{o}C under 16 hour days. Tissue was harvested at the early rosette stage and DNAs extracted as described in the previous section.

3.3.5.2 Molecular methods

PCR amplification from the seven unique glabrous accessions (Panel 3) was carried out as described above for trichome-producing accessions. Contiguous sequences for each product were assembled using SEQUENCHER v4.7 (Gene Codes), and the resulting sequences for all accessions were aligned with the Col-0 *GL1* sequence using CLUSTALL (BIOEDIT v.7.0.90, Hall 1999). Publically available Col-0 cDNA sequence was used as the reference accession to infer intron-exon boundaries and to generate a coding region alignment.

In the accessions Wil-2 and Est, PCRs with four different primer pairs within the *GL1* locus failed to amplify a product with gene loss inferred as a result. The borders of the putatively

deleted region were initially investigated by PCR of short intragenic fragments in genes 25 kb, 10 kb and 5 kb up- and downstream of *GL1*. As all of these successfully amplified from Wil-2 and Est, new primers were designed in the intergenic regions between *GL1* and the proximal up- and downstream genes (At3g27910 and At3g27925); these were used to narrow down the region and ultimately to amplify a 1520 bp product spanning the presumed deletion site. The amplified fragment was cloned into *E. coli* using a TOPO-TA cloning kit (Invitrogen) and colonies screened for positive transformants by PCR with kit primers. Plasmids were extracted from liquid cultures of two positive colonies from each of the two accessions using a Plasmid Miniprep kit (Qiagen). Those plasmids served as template for subsequent sequencing reactions. Sequences from Wil-2 and Est were aligned to Col-0 to define the deleted region. As the original *gl1-1* allele in Col-5 was putatively introgressed into Col-0 from Est (although we have not found confirmation of this in the literature), we also amplified this region from the Col-5 accession.

3.3.5.3 Genetic complementation tests and GL1 expression

Genetic complementation tests were performed to determine whether the glabrous *A. thaliana* accessions carry mutations in the same gene, and to assess the potential dominance of these mutations. To assess dominance, four accessions, representing each of the four presumptive loss-of-function alleles (9354, Fran-3, Est and Br-0), were crossed to the trichome producing Col-0 accession. The accessions 9354, Fran-3 and Wil-2 (carrying the same allele as Est) also were crossed to Br-0, representing the fourth (previously identified) loss-of-function allele to confirm that each accession carried a mutation in the same gene. The resulting F_1 plants from all crosses were grown to check for glabrous or trichome-producing mature leaves. To confirm that the cross progeny were the result of crossing and not selfing, each F_1 was genotyped with a minimum of four microsatellite markers polymorphic between the parents. The accession 9354 could not be distinguished from Col-0 based on five microsatellite loci, so *GL1* was sequenced from cross progeny and known polymorphisms in the gene used to confirm cross status.

To determine whether *GL1* is still expressed in the glabrous accessions, replicates of each of the seven glabrous accessions as well as the trichome-producing accessions Col-0 and Bor-4 (as positive controls), were grown. At 12 days post-germination tissue was pooled from two

to three plants/accession for RNA extractions with an RNeasy Plant kit (Qiagen), with on column DNAse treatment. Because *GL1* is expressed at low levels (Larkin*, et al.*, 1994; Hauser*, et al.*, 2001), expression of *GL1* was tested using a nested PCR approach, with *GL1* primers in exons one and three for the initial PCR and internal primers in exons two and three for the second round of PCR using products of the first PCR as template. The products obtained were sequenced to confirm that they were *GL1* cDNA. Actin expression was also tested as a positive control using primers described by Gonzalez *et al*. (2009).

3.4 Results

3.4.1 Nucleotide diversity and patterns of polymorphism at the *GL1* locus

Gene-wide, the nucleotide diversity (π) for *GL1* from the 115 trichome-producing accessions surveyed was 0.0049. A sliding window analysis of π was performed across the *GL1* locus from approximately 65 bp upstream of the start codon to approximately 83 bp downstream of the stop codon (Figure 3.1). The results identified regions of low variation in the first, second, and beginning of the third exons, corresponding with the R2R3 MYB domain, as well as in the first intron. Higher levels of diversity were observed in the upstream sequence, second intron and 3' end of the third exon, before a large peak of diversity in the final 35 bases of the third exon and in the 3' flank. The peak observed in the 3' end of *GL1* was due to a pattern of high frequency polymorphism (36 accessions vs. 79 accessions) identified in the second intron (one SNP) and third exon (five SNPs). The pattern identified was even more pronounced downstream of *GL1*, where two SNPs and two deletions of eight base pairs each were revealed in the 83 bp of 3' flank examined in this alignment. The six polymorphisms within *GL1* and four polymorphisms in the immediate flank are completely linked in our data set.

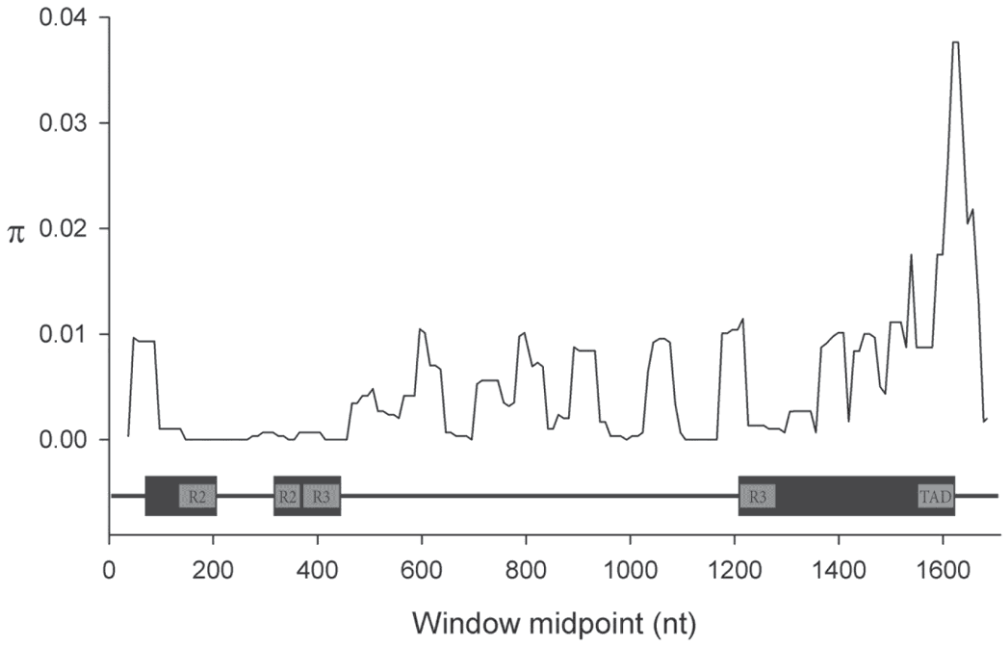

Figure 3.1: Sliding window analysis of nucleotide diversity (π) for an alignment of *GL1* genomic sequences from 115 trichome-producing accessions of *Arabidopsis thaliana*.

Pi is plotted against window midpoint. The alignment includes ~65 bp of 5' flanking sequence, *GL1*, and 83 bp of 3' flanking sequence. The underlying schematic indicates positions of the three *GL1* exons

(black boxes), introns and flanking sequences (black lines) with positions of the R2R3 MYB and transcriptional activation (TAD) domains indicated by the underlying grey boxes.

3.4.2 Allelic variation – haplotypes

A total of 120 sequences from trichome-producing (115) and glabrous (five) natural accessions of *A. thaliana* were used to create a coding region alignment. Based on this alignment, the 120 sequences revealed 22 haplotypes, each of which was found in one to 44 accessions (Figure 3.2). The 22 haplotypes are split into two major haplogroups based on five high frequency SNPs in the third exon, three of which are nonsynonymous. These haplogroups are referred to as haplogroup A (eight haplotypes from 37 accessions, one of which is glabrous) and haplogroup B (14 haplotypes from 83 accessions, four of which are glabrous). Using only sequence from trichome producing accessions, nucleotide diversity was calculated across the whole alignment separately for each of the two haplogroups and was found to be somewhat lower in the A haplogroup (π = 0.0020) compared with the B haplogroup (π = 0.0033).

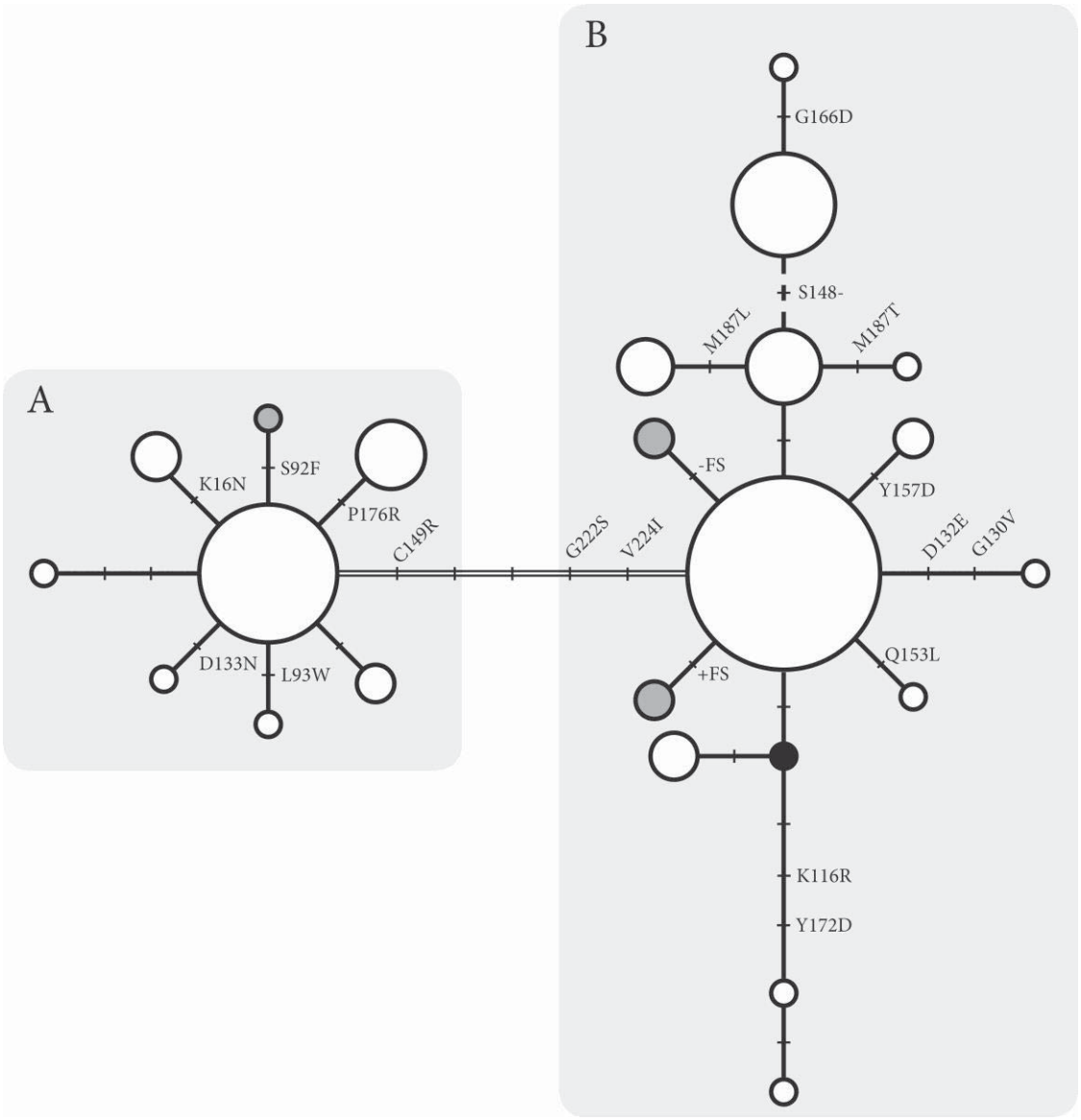

Figure 3.2: Median-joining haplotype network of *GL1* coding region alleles.

Twenty-two haplotypes were identified based on inferred cDNA nucleotide sequence from 120 *Arabidopsis thaliana* natural accessions; haplotypes are represented by circles, and their frequencies indicated by relative circle size. Trichome-producing alleles are shown as open circles and glabrous accessions' alleles are indicated by grey-filled circles. The black-filled circle represents a theoretical, unsampled haplotype required to complete the network. Branch lengths and tick marks reflect the number of nucleotide changes separating haplotypes, with the exception of the dashed line; this represents a 3 bp indel that is interpreted as a single mutation event. Alleles belonging to haplogroups A and B are circumscribed by shaded boxes labelled A and B; the haplogroups are distinguished by five SNPs, three of which are nonsynonymous.

3.4.3 Coding variation in *GL1*

The GL1 protein is 228 amino acids long, encoding an R2R3 MYB domain (AA14-AA117) and, based on sequence similarity with the functionally equivalent MYB, WEREWOLF (WER; Lee & Schiefelbein 2001), a potential C-terminal transcriptional activation domain (amino acids 203-228); these regions are shown in Figure 3.3.

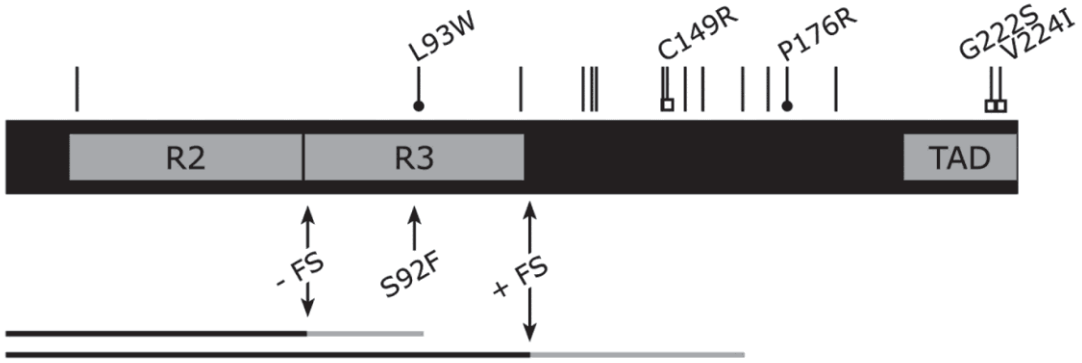

Figure 3.3: Schematic representation of the GL1 protein showing positions of amino acid replacements and potentially functionally significant polymorphisms.

The large horizontal black bar represents the full length protein, with the positions of the MYB domains (R2 and R3) and putative C-terminal transcriptional activation domain (TAD) shown as grey boxes. Each small vertical bar above the schematic indicates the position of an amino acid replacement in the 115 trichome-producing accessions (Table 3.1). Amino acid replacements with likely functional consequences are indicated with filled black circles at the base and the three linked replacements that define haplogroups A and B and are associated with trichome density variation are shown with open squares at the base. Below the schematic are shown the three mutations identified in the glabrous accessions: '-FS' is the frameshift caused by a single bp deletion, '+FS' is the frameshift caused by a single bp insertion, and 'S92F' is the serine to phenylalanine replacement. The in-frame (black) and out-of-frame (grey) portions of the putatively truncated proteins produced by the frameshift alleles are shown as horizontal lines below each mutation.

3.4.3.1 MYB domain variation

In our sample the MYB domain is highly conserved with only three replacements found in six accessions (Table 3.1; Figure 3.3). The replacements K16N and K116R are found in three and two accessions, respectively (Table 3.1); these replacements occur near the N- and C-terminal ends of the R2R3 MYB domain, respectively, and are not obvious candidates for a functional effect on trichome density, although such an effect remains plausible. In contrast, the L93W replacement, identified only in accession C24, falls within a 20 amino acid region of the R3 MYB domain which has been identified as a conserved signature among MYB proteins known

to interact with bHLHs (Zimmermann*, et al.*, 2004). This 20 amino acid region is completely conserved in GL1 in all 114 other accessions used to generate the coding region alignment. Furthermore, all bHLH-interacting MYB genes (from Col-0) surveyed by Zimmerman *et al.* (2004) have a leucine residue at L93. Notably, accession C24 has one of the lowest trichome densities of the trichome-producing accessions surveyed.

Table 3.1: GL1 amino acid replacements and indels identified from 120 *Arabidopsis thaliana* accessions.

Protein region	Polymorphism*	Minority ecotypes
MYB domain	K16N	Got-22, Kas-1, CS1264
	N68 (FS)†	Br-0, Mir-0
	S92F†	9354 (CS22468)
	L93W	C24
	K116R	Nok-3, Kin-0
"Undefined" domain	L118 (FS) †	Fran-3, PHW-2
	G130V	Can
	D132E	Can
	D133N	Wa-1
	S148-	Ms-0, Eden-2, Fab-2, Fab-4, Spr1-6, Zdr-6, Bor-1, Bor-4, Pu2-7, Lp2-2, CS22491, Ga-0, Ct-1, CS1288
	R149C‡	Haplogroup A
	Q153L	Cvi-0
	Y157D	Ei-2, St-0
	G166D	CS1288
	Y172D	Nok-3, Kin-0
	P176R	Di-0, Ler, Omo-2, Zdr-1, Ren-11, CS1314
	M187L	Lov-5, Bil-5, Bil-7, CS22574
	M187T	Edi-0
C-terminal domain	S222G‡	Haplogroup A
	I224V‡	Haplogroup A

*The minority allele is shown to the right of the amino acid position.
†Mutations in the glabrous accessions; for the frameshift-causing indels (FS), the first out-of-frame position is reported.
‡Polymorphisms that distinguish haplogroups A and B.

3.4.3.2 "Undefined region" variation

In our data set, 16 SNPs occur between the MYB domain and the putative C-terminal activation domain, 11 of which are nonsynonymous (Table 3.1; Figure 3.3). This "undefined region" spans amino acid positions 118-202. In addition, 14 of the 115 accessions share a three base pair in-frame deletion (S148-). One of the 11 amino acid replacements, C149R, is

in complete LD with two replacements in the C-terminal domain (see below) and at relatively high frequency (36 C and 79 R); together, these represent three of the five polymorphisms (Figure 3.3) that separate accessions into the two major haplogroups (Figure 3.2). The other 10 replacements in this region are found in only one or a few accessions. Overall, this internal region is not highly conserved between GL1, MYB23 and WER in Col-0 (Lee & Schiefelbein, 2001). The P176R replacement is of interest as it is the only coding difference between the Col-0 and L*er* alleles, which share the same haplogroup (haplogroup A); two independent studies have mapped a QTL near *GL1* in Col x L*er* RIL populations (Mauricio, 2005; Symonds*, et al.*, 2005), implying that some difference in Col and L*er GL1* alleles may have a quantitative effect on trichome density. L*er* is one of six accessions in our panel that carry this mutation.

3.4.3.3 C-terminal domain

The last 26 amino acids of GL1 align with the 24 C-terminal amino acids identified as the transcriptional activation domain of WER (Lee & Schiefelbein, 2001), suggesting a likely role for this region of GL1 in transcriptional activation. Supporting this, the truncated *gl1-2* allele (missing the last 27 amino acids) confers a reduced trichome density phenotype (Esch*, et al.*, 1994). The 26 C-terminal amino acids of GL1 include a 17 amino acid signature which is conserved across the epidermal cell fate regulators GL1, WER and MYB23 in Col-0 (Kranz*, et al.*, 1998). In our sample, this 17 amino acid signature is completely conserved across all 115 trichome producing accessions. Only two amino acid replacements, G222S and V224I, occur in the last 26 amino acids of GL1 in the accessions surveyed; these are two of the three replacements that define the haplogroups, but fall outside the 17 amino acid signature region. G222 is conserved between GL1, MYB23 and WER in the reference accession Col-0 but V224 is not (Lee & Schiefelbein, 2001). G222 is one of six amino acids deleted in trichome-producing *A. lyrata* individuals (Hauser*, et al.*, 2001; Kivimaki*, et al.*, 2007), so is probably not essential for trichome initiation but could have a quantitative effect.

3.4.4 Pairwise *GL1* sequence comparisons between two panels of 55 accessions

Pairwise comparisons between sequences from Sanger data and resequence data across a 1686 bp alignment for the same 55 accessions revealed 31 discrepancies in *GL1* sequences (including SNPs, small indels, and a whole locus deletion) (Table 3.7, Supplementary Material). As the Sanger sequences were produced by sequencing both strands for each accession,

rechecked by hand, and possessed consistent polymorphism patterns (in multiple independent accessions/sequences) not present in the set of 261 sequences, the inconsistencies between the two sets would seem most likely to be a result of the resequencing/assembly (sequence mapping) process or contamination of seed stocks. The former is consistent with a recent report that directly compares Sanger- and resequence-derived sequence from L*er* (Lai*, et al.*, 2011). While much is consistent between the two datasets for *GL1* and analyses of the two sets yielded qualitatively similar results (Table 3.8, Supplementary Material), some were sufficiently different that it would be inappropriate to analyse all data together. Therefore, data analyses for our data are presented here and analyses from the resequence set and comparisons between the sets are presented as supplementary material. While the resequenced genomes are an invaluable resource, direct comparisons suggest that their use should include validation and one should be wary of analyses that are sensitive to the frequencies of rare versus common polymorphisms.

3.4.5 Molecular evolution of *GL1*

To evaluate potential departure from neutral expectation for *GL1*, Tajima's D, and Fu and Li's D* and F* were calculated on the panel of 115 accessions and on the panel of 261 (results for the panel of 261 are presented in Table 3.8, Supplementary Material). For the panel of 115 accessions, Tajima's D was not significantly different from 0 but Fu and Li's D* and F* were. A sliding window analysis of *K*a/*K*s across *GL1* comparing haplogroups A and B (Figure 3.4, Supplementary Material) revealed evidence for purifying selection (*K*a/*K*s <<1) across the majority of the coding region; however, two peaks of *K*a/*K*s>1 were observed. The first peak, spanning three overlapping windows covering bases 412-476 of the alignment, centres on the high frequency nonsynonymous polymorphism (which yields the C149R replacement) identified in the variable region of *GL1*. The second, higher peak is centred over the last 55 bp of the coding region; in this region two nonsynonymous high frequency polymorphisms are found (resulting in the G222S and V224I replacements) while all other nucleotide positions are conserved. Interestingly, the three polymorphisms that underlie the two *K*a/*K*s peaks define the two predominant haplogroups.

3.4.6 Quantitative effects of *GL1* in natural populations

3.4.6.1 Natural variation in trichome density

Trichome density was surveyed in the common (CS22660) set of 96 natural accessions (Nordborg*, et al.*, 2005). Data for two of the 96 accessions were not included: replicates of Got-7 did not germinate and Br-0 is glabrous (and potentially confounding), so was excluded. Trichomes were counted on a total of 376 plants, averaging four replicates for each of 94 accessions. The accessions exhibited a ~five-fold range of trichome densities (Figure 3.5, Supplementary Material). Trichome-producing accessions' mean density scores ranged from 5.2 trichomes in accession C24, to 25.0 trichomes in Fab-4. Variation was continuous and followed a roughly normal distribution with a positive skew. Broad-sense heritability (H^2) was 0.71 in this experiment, indicating that a strong genetic component underlies trichome density variation between accessions under our experimental conditions; this falls within the range of broad sense heritabilities for trichome density and trichome number traits reported in previous QTL (0.58-0.90; Mauricio, 2005; Symonds, *et al.,* 2005), and association mapping studies (0.88; Atwell, *et al,.* 2010).

3.4.6.2 Association between GL1 and trichome density

Both sequence data and trichome density phenotypes were ultimately obtained for 94 of the 96 accessions in the CS22660 set. Three polymorphisms of interest were identified for association mapping with trichome density: deletion of a serine residue at position 148 (S148-), shared by 13 of the 94 accessions; substitution of a proline residue for arginine at position 176 (P176R) in four accessions; and the high frequency pattern of polymorphism identified in the third exon and 3' flank, with 28 accessions possessing haplogroup A and the remaining 66 accessions carrying haplogroup B (Table 3.4, Supplementary Material). No significant association was found between trichome density and S148- or P176R for any of the six models used in mapping (Table 3.2). In contrast, there was a significant association between mean trichome density and haplogroup, with a maximum $p<0.025$ across all six models tested (Table 3.2). Mean trichome densities were 10.82 and 13.13 for haplogroups A and B, respectively, yielding a 17.6% difference between groups and accounting for ~11% of the trichome density variation observed in the set of 94 lines. The observed polymorphism pattern that distinguishes haplogroups A and B does not extend beyond the *GL1* locus (Figure

3.6, Supplementary Materials); i.e., linkage disequilibrium decays prior to the most proximal genes up- and downstream of *GL1*.

Table 3.2: Association test results (*p*-values) for *GL1* polymorphisms and trichome density

Polymorphism	Model*					
	Naïve	QK	Q	K	QK*	K*
S:148:-	0.89790	0.59310	0.71400	0.77730	0.82820	0.97560
P:176:R	0.51380	0.71810	0.81950	0.57650	0.67670	0.67780
Haplogroups A & B	0.01910	0.02010	0.02240	0.02270	0.02010	0.01870

*Association models are described in the text.

3.4.7 Qualitative effects of *GL1* in natural populations

3.4.7.1 GL1 molecular variation in glabrous accessions

Of the 34 accessions described by the ABRC as glabrous, only seven represented independent accessions (i.e., not synonyms of the same lineage) with completely glabrous rosette leaf blades. These seven accessions originate from sites distributed across Eurasia (Table 3.3) and revealed four independent mutations in *GL1* that explain the glabrous phenotype (Figure 3.3).

Table 3.3: Summary of data for glabrous accessions and *GL1* loss-of-function alleles.

Stock centre no.	Accession	Location	Mutation	*GL1* expression	Complementation results (x Br-0)*	Complementation results (x Col-0)*
CS1378	Mir-0	Italy	1bp deletion (cDNA:203)	+	NA	NA
CS22628	Br-0	Czech Republic	1bp deletion (cDNA:203)	+	NA	+
CS75673	Fran-3	Netherlands	1bp insertion (cDNA:352)	+	-	+
CS6002	PHW-2	Italy	1bp insertion (cDNA:352)	+	NA	NA
CS22468	9354	Kazakhstan	S92F	+	-	+
CS1596	Wil-2	Lithuania	Locus deletion	-	-	NA
CS911	Est	Estonia	Locus deletion	-	NA	+

*Complementation results indicate whether trichomes were (+) or were not (-) produced in F_1 plants.

Sequencing of *GL1* from Mir-0 and Br-0 confirmed the presence of a previously identified (Hauser*, et al.*, 2001) single base deletion at cDNA position 203 (AA position 68) in the second exon. The deletion results in a reading frame shift, leading to 22 missense amino acids partway through the R3 MYB region and premature termination of the protein (Figure 3.3). The resultant protein has an incomplete R3 MYB domain, no signature for interaction with bHLHs, and no putative C-terminal activation domain.

Accessions Fran-3 and PHW-2 shared a second, distinct mutation in *GL1*, a single base insertion at cDNA position 352 (amino acid position 118) in exon three resulting in a reading frame shift at the beginning of the third exon (Figure 3.3). This frame shift occurs immediately after the R3 MYB domain, resulting in a prematurely terminated protein with 47 missense amino acids and no C-terminal activation domain. The MYB amino acids identified as necessary for bHLH interaction (Zimmermann*, et al.*, 2004) are retained in the Fran-3/PHW-2 allele but the missense amino acids immediately beyond the R3 MYB repeat and the large portion of missing protein likely have deleterious effects on domain folding and interaction stability.

A third, likely causal, mutation identified was a SNP at cDNA position 275 in exon three in accession 9354, which results in a replacement at amino acid position 92 (Figure 3.3). Across the entire sequenced region including up and downstream sequences and both introns and exons, 9354 was otherwise identical to the functional Col-0 allele. This nonsynonymous change, which replaces a serine residue for phenylalanine, occurs in the R2R3 MYB region (Figure 3.3) within a 20 residue motif required for bHLH interaction (Zimmermann*, et al.*, 2004). The mutated site is otherwise conserved both in all 115 functional copies of *GL1* analysed here and in all R2R3 *MYB*s with known roles in epidermal cell fate (Zimmermann*, et al.*, 2004). Furthermore, serine residues are targets for phosphorylation while phenylalanine is not; phosphorylation has been identified as a mechanism of post-translational modification in other R2R3 MYB proteins in plants (Moyano*, et al.*, 1996; Morse*, et al.*, 2009). Although there is no published evidence that phosphorylation of S92 is required for bHLH interaction it suggests a second possible mechanism through which this mutation could act.

In the Wil-2 and Est accessions, the *GL1* locus was discovered to be entirely missing, here characterised as a deletion spanning a 4758 bp region beginning 948 bp upstream of the Col-0 *GL1* coding region and finishing 2248 bp downstream. The deletion therefore spans not only the coding region but also the likely promoter region and a previously identified 3' enhancer region (Larkin*, et al.*, 1993). The original *GL1* mutant, *gl1-1* (Col-5), was confirmed to carry the same Est deletion allele (data not shown).

3.4.7.2 Complementation tests and GL1 expression

Crosses between the trichome producing Col-0 accession and glabrous accessions representing each of the four putative loss-of-function alleles identified here, 9354, Fran-3, Est, and Br-0 each resulted in trichome-producing F_1 offspring (Table 3.3), indicating that the mutations causing glabrousness are all (at least partially) recessive to the Col-0 *GL1* allele. Genetic complementation test crosses between Br-0 and Fran-3, 9354, and Wil-2 (carrying the same allele as Est) all resulted in glabrous F_1 offspring (Table 3.3). Together, these results indicate that the glabrous accessions all possess loss-of-function mutations in the same gene, *GL1*.

A nested PCR was used to assay for expression in twelve day old seedlings to determine whether regulatory changes may have eliminated *GL1* expression in the glabrous accessions still carrying a copy of *GL1*. *GL1* expression was detected in Mir-0, Br-0, PHW-2, Fran-3 and 9354, which still possess the *GL1* locus, as well as in the trichome-producing (positive control) accessions Col-0 and Bor-4. Sequencing confirmed that the products were indeed *GL1* transcripts. Expression was not detected in Wil-2 and Est, accessions with a whole-locus deletion of *GL1* (Table 3.3). Actin expression was detected in all samples (data not shown).

3.5 Discussion

Our understanding of how evolution proceeds depends upon the identification and characterization of specific genes underlying biological variation. While observations from functional genetic studies provide an initial indication of gene function, it remains unclear how allelic variation in those same genes shape phenotypic variation in nature. Here, we investigated the role(s) of the trichome initiation gene, *GL1*, in generating quantitative and qualitative variation for trichome density in *A. thaliana*.

3.5.1 Allelic variation at *GL1* in trichome-producing accessions

3.5.1.1 Nucleotide diversity and patterns of polymorphism in GL1

The gene-wide estimate of nucleotide diversity from our sample (π =0.0049) is roughly half that reported by Hauser *et al.* (2001) for *GL1* (π =0.0109); however, Hauser *et al.*'s analysis included a greater proportion of noncoding sequence and approximately one-quarter the number of accessions compared with the current analysis so direct comparisons are difficult to make. Comparing more broadly, our results are highly consistent with genome-wide average estimates of π for *A. thaliana* (π =0.0047; Bakker, *et al.,* 2006). Across *GL1* nucleotide diversity is not uniform, however, with known functional domains showing the highest sequence conservation. Exons one, two and the beginning of exon three, which encode for the R2R3 MYB domain of *GL1* (Figure 3.1), are highly conserved, reflecting the importance of domain structure and interaction stability for protein:protein and protein:DNA interactions in this region. Similarly, π is low in the last 60bp of the third exon, a region conserved in the MYB family subgroup that includes *GL1* and two other epidermis-specific, functionally equivalent but differentially expressed MYBs, *WERWOLF* (*WER*) and *MYB23* (Stracke*, et al.*, 2001). The 24 terminal amino acids of WER form a transcriptional activation domain (Lee & Schiefelbein, 2001), a function likely shared by GL1 and MYB23, which have high sequence similarity with WER in this region (data not shown). Pi is also fairly low in the first intron of *GL1*, which is likely conserved due to its regulatory role in directing trichome-specific enhancement and non-trichome specific repression of *GL1* expression (Wang*, et al.*, 2004).

Gene-wide, Tajima's D for *GL1* was -1.15 but not significantly different from the neutral expectation of zero (Table 3.8, Supplementary Material); this value is also similar to the *A.*

thaliana genome-wide average of -0.80 (Nordborg*, et al.*, 2005). Conversely, the less conservative Fu and Li's D* and F* did reject selective neutrality for *GL1*, instead suggesting a role for purifying selection or population expansion, both of which are plausible (more below). These common inferences drawn from Tajima's D and Fu and Li's D* and F* test results may seem incongruent with the high frequency haplogroup pattern described here for *GL1*, but the haplogroups are distinguished by relatively few polymorphisms compared to a preponderance of rare segregating sites.

3.5.2 *GL1* haplotype variation

The 115 trichome producing and five glabrous accessions analysed here revealed 22 *GL1* coding region alleles (haplotypes): 19 intact alleles and three loss-of-function alleles. The 22 haplotypes fall into two major haplogroups (A and B; Figure 3.3) based on five common, completely linked, polymorphisms in the third exon, three of which are nonsynonymous. This sequence dimorphism was previously reported for *GL1* by Hauser *et al.* (2001) and a similar pattern has been demonstrated at many other loci in *A. thaliana*, including chitinase genes (Kawabe*, et al.*, 1997), the *Adh* locus (Innan*, et al.*, 1996), *R* genes (Tian*, et al.*, 2002; Mauricio*, et al.*, 2003; Rose*, et al.*, 2004), and genes in the phenylpropanoid pathway (Aguade, 2001). *Arabidopsis thaliana* may have been isolated during the Pleistocene in two or more glacial refugia, and recolonized central Europe from these refugia during subsequent glacial contraction (Sharbel *et al.* 2000; Beck *et al.* 2008), giving rise to the dimorphism commonly observed. Consistent with the results from Fu and Li's tests, a *K*a/*K*s sliding window comparison between haplogroups A and B (from trichome producing accessions) also revealed evidence for purifying selection across most of *GL1* (Figure 3.4, Supplementary Material); however spikes in the ratio that coincide with the two coding regions that differentiate the haplogroups may reflect positive selection on certain regions of *GL1*. Given fluctuating selection on trichome density in *A. thaliana* and the putative separation of multiple metapopulations during glaciation, it is plausible that positive selection has acted on either trichome density directly or on differentially evolving TTG1 genetic pathways in isolated populations; consistent with the latter, a biallelic pattern similar to that observed in *GL1* has been identified recently in other members of the TTG1 pathway (Hilscher*, et al.*, 2009; Symonds*, et al.*, 2011). While preliminary analyses have shown no evidence of LD among these loci (VVS, unpub. data), one might suspect that the cumulative effect of

shuffling many loci with similar functionally significant biallelic patterns could explain much of quantitative epidermal trait variation in *A. thaliana*.

3.5.3 Evidence for a quantitative role for *GL1*

Of the three polymorphisms tested for association with trichome density, only the haplogroup variation showed a significant association (Table 3.2). The pattern that differentiates the haplogroups includes three amino acid replacements (described above) and extends beyond the coding region and through a known 3' enhancer region, ~1 kb downstream of *GL1*. This enhancer of *GL1* expression has a role in directing expression of *GL1* to leaf primordia; in its absence, the 5' promoter region directs expression only in the stipules (Larkin*, et al.*, 1993). No sequence obtained in this study extended into the 3' enhancer; however, publically available sequence that spans this region is available for accessions of both major haplogroups and reveals several substitutions and indels that further differentiate the two haplogroups (not shown). Specifically, an examination of LD within and around *GL1* showed that the haplogroup pattern decays rapidly outside of the *GL1* locus, but includes downstream regions (Figure 3.6, Supplementary Material). Interestingly, Atwell *et al.* (2010) recently reported an association between trichome number and a polymorphism ~1 kb down-stream of *GL1*, which would fall within or very near the known but undefined 3' enhancer region. Further work is required to determine whether amino acid polymorphisms, regulatory differences, or both are responsible for the observed trichome density variation between haplogroups but the data show unequivocally that a pattern of variation, localised to the *GL1* locus, is correlated with naturally occurring quantitative trichome density variation in *A. thaliana*.

3.5.4 A qualitative role for *GL1*

Functional analysis has long indicated a qualitative role for *GL1* in trichome initiation, with laboratory knockouts shown to have a glabrous phenotype (Oppenheimer*, et al.*, 1991; Hauser*, et al.*, 2001); however, the extent to which this qualitative role is played in natural populations has been unknown. Of the 34 accessions listed as glabrous in the ABRC stock centre, we found only seven to be independent collections and actually glabrous. From these seven lineages, our data identify four independent loss-of-function *GL1* alleles, each with a different type of mutation (Figure 3.3). Subsequent complementation experiments confirmed

that each glabrous lineage is due to a nonfunctional *GL1* allele. Interestingly, our cDNA sequence data show that each naturally glabrous accession (except those with the whole locus deletion) continues to express *GL1,* indicating that the coding changes identified here are likely the causal mutations for glabrousness in these lines. Although changes in expression timing or placement of *GL1* remain potential explanations, the coding region mutations are sufficient to explain the glabrous phenotypes observed.

Our results show that glabrousness is not monophyletic within *A. thaliana*, having arisen at least four times independently. This degree of homoplasy within a single species should serve as a strong warning against the use of trichome presence/absence as a character in phylogenetic reconstructions, a practice that remains common. Indeed, there are many glabrous species within the Brassicaceae (Beilstein, 2008), clearly the result of multiple transitions. Interestingly, none of the loss-of-function alleles identified in *A. thaliana* show evidence of degradation beyond the likely causal mutations and so may represent relatively recent events; each only differs by one polymorphism from the most common haplotype in each haplogroup (Figure 3.2). Kawagoe *et al.* (2011) report similar findings for a single *GL1* allele in glabrous *A. halleri* from one population. However, in *A. thaliana*, three of the four alleles have subsequently spread to other, geographically distinct populations (Table 3.3); the fourth was discovered in an area not well-collected so may have expanded in range as well. Microsatellite data indicate that the whole locus deletion allele resides in accessions with similar genetic backgrounds; the same is true for the single base deletion allele (Symonds & Lloyd, 2003). In contrast, those same markers (data not shown) indicate that the single base insertion allele is present in two very distinct backgrounds, Fran-3 and PHW-2, suggesting that this allele has introgressed from its original background into another.

While variation in selection pressures (Mauricio & Rausher, 1997; Mauricio, 1998; Karkkainen*, et al.*, 2004; Calo*, et al.*, 2006; Steets*, et al.*, 2010; Zust*, et al.*, 2011) may be expected to allow glabrous plant lineages to persist, it is unclear for what period of time. It seems likely that alleles that confer glabrousness in *A. thaliana* are conditionally deleterious and contribute to short-term standing variation but may not persist over longer evolutionary timescales. Indeed, current evidence suggests that null (loss-of-function) mutations generally have a greater contribution to short-term evolution than to long-term evolution (Stern & Orgogozo, 2009). Consistent with this, the diversification of each *GL1* haplogroup (A and B) from trichome

producing accessions suggests that the pattern of sequence variation that, in part, underlies quantitative variation for trichome density must be considerably older, therefore more likely to contribute to longer term evolution.

3.6 Conclusions

Here, we present evidence that the trichome initiation gene *GL1* has qualitative and likely quantitative effects on trichome density in natural populations of *A. thaliana*. While the qualitative role seems to have arisen several times independently via three protein coding changes and one whole-locus deletion, the molecular nature of the quantitative role remains equivocal. The observed high-frequency pattern of polymorphism associated with quantitative variation includes completely linked amino acid replacements and mutations in a known but undefined enhancer region, but does not extend beyond the *GL1* locus. Although a single loss-of-function allele of *GL1* had been identified previously, our data reveal considerable natural variation at the *GL1* locus with effects on trichome initiation. As such, *GL1* joins two other loci in the TTG1 epidermal cell fate pathway, *ETC2* (Hilscher*, et al.*, 2009) and *ATMYC1* (Symonds*, et al.*, 2011), both of which have been implicated in trichome density variation in natural accessions of *A. thaliana*. These results suggest that much of the standing variation for trichome density in *A. thaliana* may be due to allelic variation in the TTG1 epidermal cell fate pathway and provide an important insight into the molecular nature of morphological variation and the duration of large and small effect polymorphisms.

3.7 Acknowledgements

RHB collected all data and performed most analyses and interpretation. Trichome density phenotyping and preliminary GL1 sequencing data for the set of 96 accessions used here were collected by RHB in 2008 as part of a BSc (Hons) project. The mixed-model association mapping analysis was performed by Thomas Juenger (University of Texas, Austin). Funding for this work was provided by a New Zealand Royal Society Marsden Fund grant (09-MAU-114).

3.8 Supplementary Materials

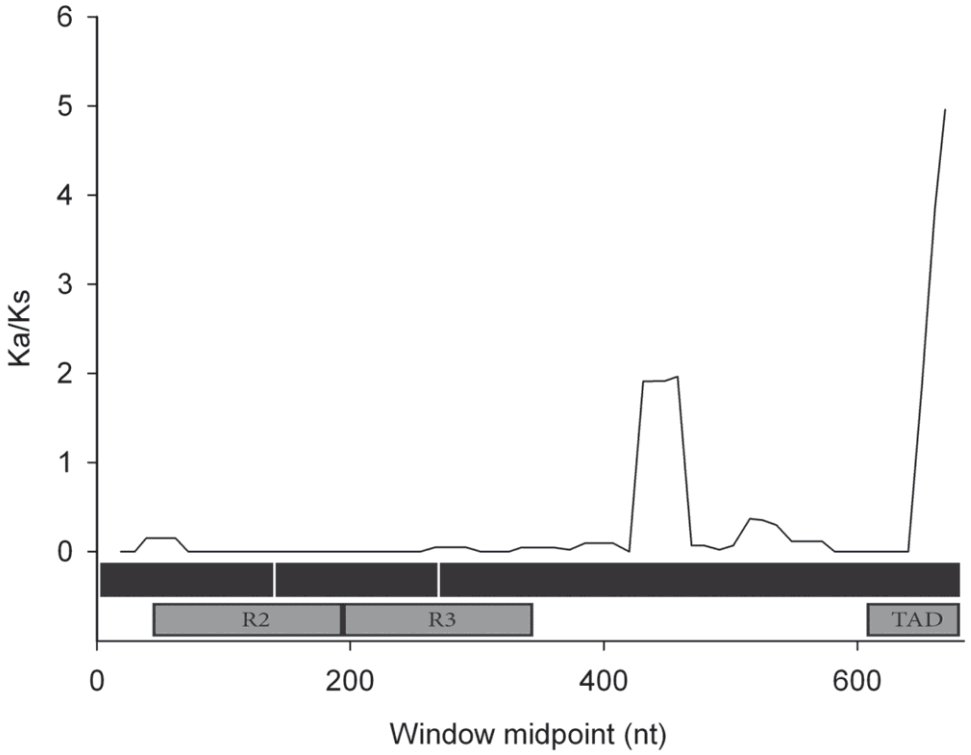

Figure 3.4: Sliding window analysis of *K*a/*K*s ratio comparing haplogroups A and B for the 684 bp *GL1* coding region; Supplementary Material.

Ratio is plotted against window midpoint (in bp). The ratio was calculated using a sliding window value of *K*a and a locus-wide summary value of *K*s. The underlying schematic depicts exons as black bars, with R2R3 MYB and transcriptional activation (TAD) domains indicated by the underlying grey boxes.

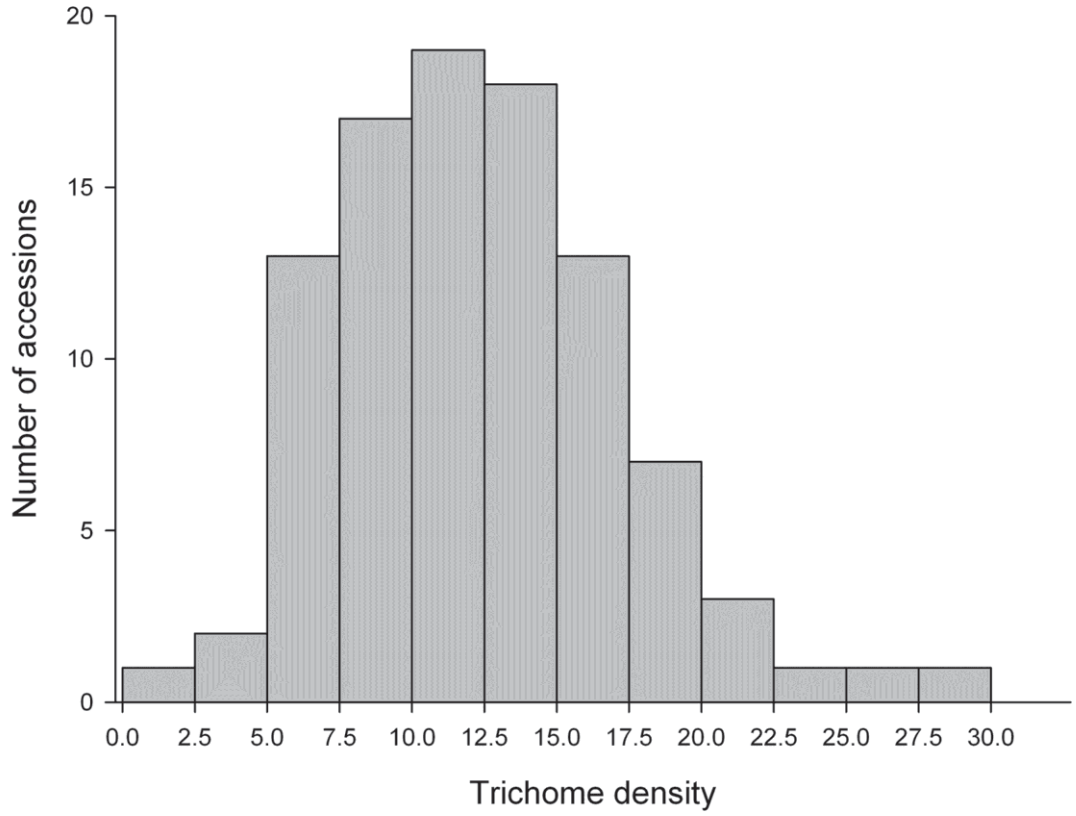

Figure 3.5: The range and distribution of natural variation for trichome density in the screened set of 94 accessions; Supplementary Material.

Trichomes were counted in a 17 mm^2 area centred between the midrib and leaf margin at the widest part of the fifth true leaf at 27 days post-germination. An average of four replicates per accession was counted and averaged for each accession in the set.

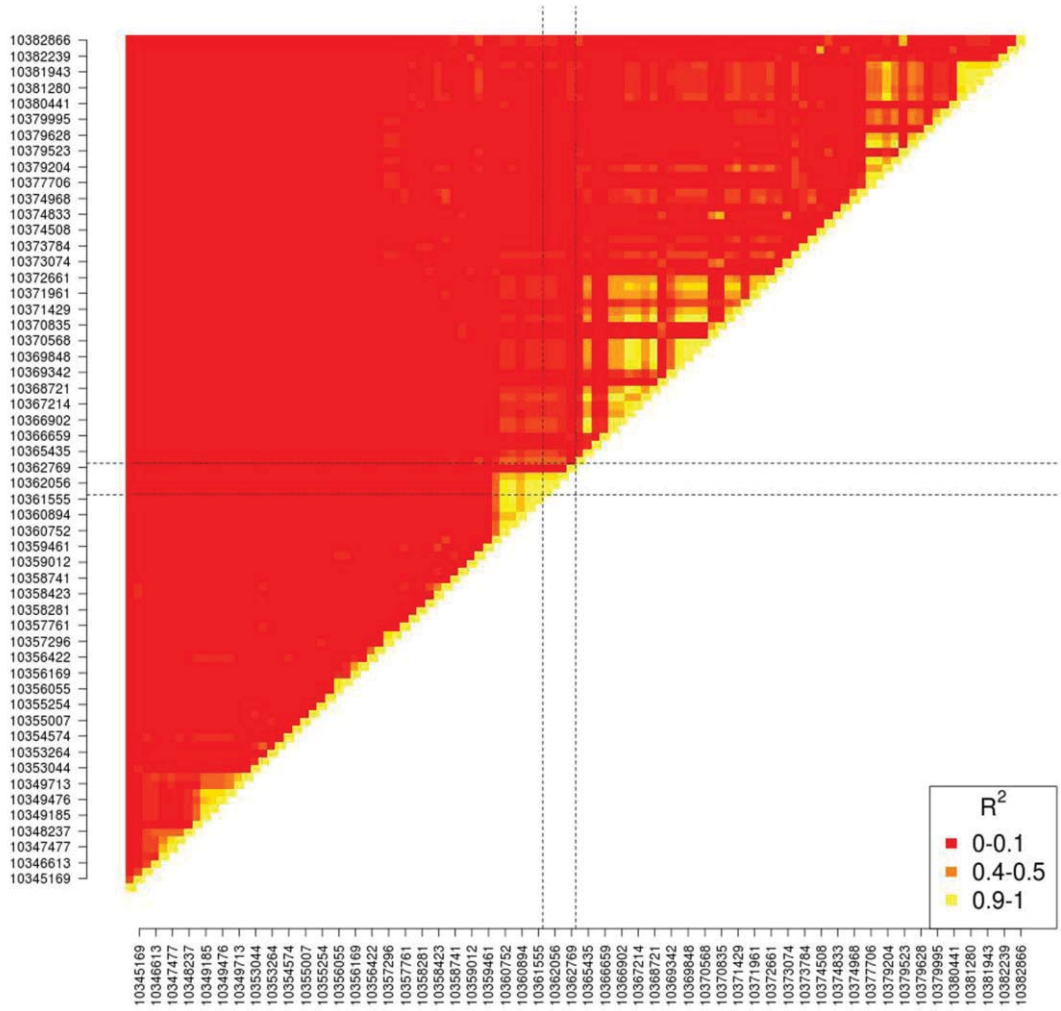

Figure 3.6: Heat map of linkage disequilibrium (LD) analysis (r2) of 40 kb centred on the *GL1* locus; Supplementary Material.

Based on SNP polymorphisms from the 250k dataset (v3.06) and 1,307 accessions. The pair of broken lines indicate the location of the *GL1* reading frame on each axis. The results show a block of LD confined to within and just downstream of *GL1*. Other blocks of LD are shown within the 40 kb window but do not involve *GL1*.

Table 3.4: Information on all *A. thaliana* accessions and panels of accessions used in this study; Supplementary Material.

Stock #*	Accession ID	Accession Origin	Sequence origin	Panel 1 115 sequence Alignment	Panel 2 55 sequence sets	Panel 3 Glabrous accessions	Panel 4 Association Mapping	*GL1* haplogroup
22564	RRS-7	CS22660	Bloomer et al.	X	X		X	A
22565	RRS-10	CS22660	Bloomer et al.	X	X		X	B
22566	Knox-10	CS22660	Bloomer et al.	X			X	B
22567	Knox-18	CS22660	Bloomer et al.	X	X		X	A
22568	Rmx-A02	CS22660	Bloomer et al.	X	X		X	B
22569	Rmx-A180	CS22660	Bloomer et al.	X			X	B
22570	Pna-17	CS22660	Bloomer et al.	X	X		X	B
22571	Pna-10	CS22660	Bloomer et al.	X			X	B
22572	Eden-1	CS22660	Bloomer et al.	X			X	B
22573	Eden-2	CS22660	Bloomer et al.	X			X	B
22574	Lov-1	CS22660	Bloomer et al.	X			X	B
22575	Lov-5	CS22660	Bloomer et al.	X			X	B
22576	Fab-2	CS22660	Bloomer et al.	X			X	B
22577	Fab-4	CS22660	Bloomer et al.	X			X	B
22578	Bil-5	CS22660	Bloomer et al.	X			X	B
22579	Bil-7	CS22660	Bloomer et al.	X			X	B
22580	Var2-1	CS22660	Bloomer et al.	X			X	B
22581	Var2-6	CS22660	Bloomer et al.	X			X	A
22582	Spr1-2	CS22660	Bloomer et al.	X			X	B
22583	Spr1-6	CS22660	Bloomer et al.	X			x	B
22584	Omo2-1	CS22660	Bloomer et al.	X			x	A
22585	Omo2-3	CS22660	Bloomer et al.	X			x	B

22586	Ull2-5	CS22660	Bloomer et al.	X		x	A
22587	Ull2-3	CS22660	Bloomer et al.	X		x	A
22588	Zdr-1	CS22660	Bloomer et al.	X	X	x	A
22589	Zdr-6	CS22660	Bloomer et al.	X		x	B
22590	Bor-1	CS22660	Bloomer et al.	X	X	x	B
22591	Bor-4	CS22660	Bloomer et al.	X	X	x	B
22592	Pu2-7	CS22660	Bloomer et al.	X	X	x	B
22593	Pu2-23	CS22660	Bloomer et al.	X	X	x	B
22594	Lp2-2	CS22660	Bloomer et al.	X	X	x	B
22595	Lp2-6	CS22660	Bloomer et al.	X		x	B
22596	HR-5	CS22660	Bloomer et al.	X		x	B
22597	HR-10	CS22660	Bloomer et al.	X		x	B
22598	NFA-8	CS22660	Bloomer et al.	X		x	B
22599	NFA-10	CS22660	Bloomer et al.	X		x	B
22600	Sq-1	CS22660	Bloomer et al.	X		x	B
22601	Sq-8	CS22660	Bloomer et al.	X	X	x	B
22602	CIBC-5	CS22660	Bloomer et al.	X		x	B
22603	CIBC-17	CS22660	Bloomer et al.	X		x	B
22604	Tamm-2	CS22660	Bloomer et al.	X	X	x	B
22605	Tamm-27	CS22660	Bloomer et al.	X		x	B
22606	Kz-1	CS22660	Bloomer et al.	X		x	B
22607	Kz-9	CS22660	Bloomer et al.	X	X	x	B
22608	Got-7	CS22660					
22609	Got-22	CS22660	Bloomer et al.	X		x	A
22610	Ren-1	CS22660	Bloomer et al.	X		x	B
22611	Ren-11	CS22660	Bloomer et al.	X		x	A
22612	Uod-1	CS22660	Bloomer et al.	X	X	x	B

22613	Uod-7	CS22660	Bloomer et al.	X			x	B
22614	Cvi-0	CS22660	Bloomer et al.	X	x		x	B
22615	Lz-0	CS22660	Bloomer et al.	X			x	A
22616	Ei-2	CS22660	Bloomer et al.	X	x		x	B
22617	Gu-0	CS22660	Bloomer et al.	X	x		x	A
22618	Ler-1	CS22660	Genbank	X	x		x	A
22619	Nd-1	CS22660	Bloomer et al.	X			x	B
22620	C24	CS22660/"glabrous"†	Bloomer et al.	X	x		x	A
22621	CS22491/N13	CS22660	Bloomer et al.	X			x	B
22622	Wei-0	CS22660	Bloomer et al.	X	x		x	B
22623	Ws-0	CS22660	Bloomer et al.	X			x	B
22624	Yo-0	CS22660	Genbank	X	x		x	B
22625	Col-0	CS22660	Genbank	X	x		x	A
22626	An-1	CS22660	Genbank	X	x		x	B
22627	Van-0	CS22660	Bloomer et al.	X	x		x	B
22628	Br-0	CS22660/" glabrous" (=CS994)	Bloomer et al.		x	X		B
22629	Est-1	CS22660	Bloomer et al.	X	x		x	B
22630	Ag-0	CS22660	Bloomer et al.	X	x		x	A
22631	Gy-0	CS22660	Bloomer et al.	X	x		x	B
22632	Ra-0	CS22660	Bloomer et al.	X	x		x	A
22633	Bay-0	CS22660	Bloomer et al.	X			x	B
22634	Ga-0	CS22660	Bloomer et al.	X	x		x	B
22635	Mrk-0	CS22660	Bloomer et al.	X			x	A
22636	Mz-0	CS22660	Bloomer et al.	X	x		x	B
22637	Wt-5	CS22660	Bloomer et al.	X	x		x	A
22638	Kas-1	CS22660	Bloomer et al.	X	x		x	A

22639	Ct-1	CS22660	Bloomer et al.	X			x	B
22640	Mr-0	CS22660	Bloomer et al.	X			x	A
22641	Tsu-1	CS22660	Genbank	X			x	B
22642	Mt-0	CS22660	Genbank	X			x	B
22643	Nok-3	CS22660	Bloomer et al.	X	X		x	B
22644	Wa-1	CS22660	Bloomer et al.	X	X		x	A
22645	Fei-0	CS22660	Bloomer et al.	X	X		x	A
22646	Se-0	CS22660	Bloomer et al.	X	X		x	A
22647	Ts-1	CS22660	Bloomer et al.	X	X		x	A
22648	Ts-5	CS22660	Bloomer et al.	X			x	B
22649	Pro-0	CS22660	Bloomer et al.	X	X		x	B
22650	LL-0	CS22660	Bloomer et al.	X			x	A
22651	Kondara	CS22660	Bloomer et al.	X	X		x	A
22652	Sakhdara	CS22660	Bloomer et al.	X			x	A
22653	Sorbo	CS22660	Bloomer et al.	X	X		x	A
22654	Kin-0	CS22660	Bloomer et al.	X	X		x	B
22655	Ms-0	CS22660	Genbank	X	X		x	B
22656	Bur-0	CS22660	Genbank	X	X		x	B
22657	Edi-0	CS22660	Bloomer et al.	X			x	B
22658	Oy-0	CS22660	Genbank	X			x	B
22659	Ws-2	CS22660	Bloomer et al.	X			x	B
CS22468	9354	"glabrous"	Bloomer et al.			X		A
CS994	Br-0	"glabrous"						
CS6626	Br-0 parent line	"glabrous"	Bloomer et al.			X		B
CS995	Br-0 progeny of 994	"glabrous"						
CS906	C24 22620 parent line	"glabrous"						
CS911	Est	"glabrous"	Deletion		X	x		

			characterised					
CS6173	Est progeny of 911	"glabrous"						
CS1148	Est-0	"glabrous"						
CS1149	Est-0 progeny of 1148	"glabrous"						
CS6700	Est-0 progeny of 1148	"glabrous"						
CS75673	Fran-3	"glabrous"	Bloomer et al.				X	B
CS75878	Fran-3 progeny line	"glabrous"						
CS1264	Kas-2	"glabrous"	Bloomer et al.	X				A
CS6751	Kas-2	"glabrous"						
CS1265	Kas-2 progeny of 1264	"glabrous"						
CS1288	Ko-0	"glabrous"	Bloomer et al.	X				B
CS1289	Ko-0 progeny of 1288	"glabrous"						
CS6763	Ko-2	"glabrous"						
CS3886	Krot-0	"glabrous"	Bloomer et al.	X	x			A
CS3887	Krot-1	"glabrous"						
CS3888	Krot-2	"glabrous"						
CS1314	Li2:1	"glabrous"	Bloomer et al.	X	x			A
CS6772	LI2:1 progeny of 1314	"glabrous"						
CS1378	Mir-0	"glabrous"	Bloomer et al.				X	B
CS1379	Mir-0 progeny of 1378	"glabrous"						
CS6798	Mir-0 progeny of 1378	"glabrous"						
CS6002	PHW-2	"glabrous"	Bloomer et al.				X	B
CS6044	PHW-2 progeny of 6002	"glabrous"						
CS1596	Wil-2	"glabrous"	Deletion characterised				X	
CS6889	Wil-2 progeny of 1596	"glabrous"						

CS1598	Wil-3	"glabrous"					
CS6890	Wil-3 progeny of 1598	"glabrous"					
CS38906	St-0	Additional natural lines	Bloomer et al.	X			B
CS6816	Ob-0	Additional natural lines	Bloomer et al.	X	X		B
CS6857	Sf-2	Additional natural lines	Bloomer et al.	X			A
	RLD		Genbank	X			B
	Can		Genbank	X			B
	Ita-0		Genbank	X			B
	Hi-0		Genbank	X			B
	No-0		Genbank	X			B
	Gr-1		Genbank	X	X		B
	Aa-0		Genbank	X	X		B
	Es-0		Genbank	X	X		B
	Ba-1		Genbank	X	X		A
	Bla-1		Genbank	X	X		A
	Te-0		Genbank	X			A
	Di-0		Genbank	X			A
	Nd-0		Genbank	X			B
	Ws		Genbank	X	X		B

* Stock # refers to the ABRC/TAIR identification number (www.arabidopsis.org)

† ”glabrous” refers to accessions described by the ABRC as glabrous.

Table 3.5: Primers used for PCR and sequencing *GL1*; Supplementary Material

Region amplified	Primer name	Sequence	Primer location
Genomic DNA PCR 5' flank	GL1-f41-F	Agaatgtaagtgtatacagtgc	5' flank
	GL1-i18-R	Acgatgcggttccattggcc	Exon 1
Genomic DNA PCR coding region	GL1-start	Atgagaataaggagaagaga	Exon 1
	GL1-stop	Ctaaaggcagtactcaacatc	Exon 3
Genomic DNA PCR 3' flank	GL1-i15-F	Acagataaccaagtcaagaac	Exon 3
	GL1-1669-R	Ttatccaactactagtcgcg	3' flank
cDNA amplification	GL1_cons_i1F	Gaatcaagaatacaagaaagg	Exon 1
	GL1-i13-R	Ttttgctgagatgagtgttcca	Exon 2
	GL1-i44-F	Cctcattattcgtctccac	Exon 3
	GL1_cons_i1R	Cyyggatcattagtagttgc	Exon 3

Table 3.6: Resequence set of 261 accessions analysed for *GL1*; Supplementary Material

Accession ID
Aa_0, Abd_0, Agu-1, Ag_0, Ak_1, Alst_1, Altai_5, Amel_1, Ang_0, Anholt_1, Ann_1, Anz_0, An_1, Appt_1, Baa_1, Bak-2, Bak-7, Ba_1, Bch_1, Bd_0, Benk_1, Ber, Bik_1, Bla_1, Bl_1, Boot_1, Bor_1, Bor_4, Br_0, Bsch_0, Bs_1, Bur-0, Bu_0, C24, Cal_0, Ca_0, Cdm-0, Cerv_1, Chat_1, Chi_0, Cnt_1, Col-0, Com_1, Co_1, Cvi-0, Da1_12, Db_1, Del-10, Di_G, Dja_1, Dog-4, Don-0, Do_0, Dra_0, Dr_0, Durh_1, Eil-0, Ei_2, El_0, Ema_1, En_2, En_D, Er_0, Est-1, Est, Es_0, Etna_2, Et_0, Ey15-2, Fei-0, Fi_0, Fr_2, Ga_0, Gel_1, Gie_0, Gifu_2, Got_7, Gre_0, Gr_1, Gu_0, Gy_0, Hau_0, Ha_0, Hey_1, Hh_0, HKT2.4, Hn_0, HR_5, Hs_0, ICE1, ICE102, ICE104, ICE106, ICE107, ICE111, ICE112, ICE119, ICE120, ICE127, ICE130, ICE134, ICE138, ICE150, ICE152, ICE153, ICE163, ICE169, ICE173, ICE181, ICE21, ICE212, ICE213, ICE216, ICE226, ICE228, ICE29, ICE33, ICE36, ICE49, ICE50, ICE60, ICE61, ICE63, ICE7, ICE70, ICE71, ICE72, ICE73, ICE75, ICE79, ICE91, ICE92, ICE93, ICE97, ICE98, In_0, Istisu-1, Is_0, Je_0, Jl_3, Jm_0, Kar_1, Kastel-1, Kas_1, Kelsterbach_4, Kil_0, Kin_0, KI_5, Knox_18, Koch-1, Kondara, Ko_2, Krot_0, Kro_0, Kyoto, Kz_9, Lag2.2, Lan_0, La_0, Lc-0, Leo-1, Ler-1, Lerik1-3, Le_0, Lip_0, Litva, Li_2:1, Lm_2, Lp2_2, Mc_0, Mer-6, Mh_0, Mnz_0, Ms_0, Mz_0, Nc_1, Nemrut-1, Neo_6, Nie1-2, Nok_3, Np_0, Nw_0, Ob_0, Old_1, Ove_0, Ped-0, Per_1, Pi_0, Pla_0, Pna_17, Pog_0, Pra-6, Pro_0, Pt_0, Pu2_23, Pu2_7, Qar_8a, Qui-0, Ragl_1, Ra_0, Rd_0, Rennes_1, Rhen_1, Rld_1, Rmx_A02, Rome_1, Rou_0, RRs_10, RRS_7, Rubeznhoe_1, Rue3-1-31, Sav-0, Seattle_0, Sei_0, Se_0, Sg_1, Sha, Si_0, Sorbo, Sp_0, Sq_8, Star-8, Stw_0, Sus_1, Su_0, Tamm_2, Ta_0, Tha_1, Ting_1, Tol_0, Tscha_1, Tsu-1, Ts_1, TueSB30-3, Tuescha9, TueV13, TueWa1-2, Tul_0, Tu_0, Ty_0, Uk_1, Uod_1, Utrecht, Van_0, Vash-1, Ven_1, Vie-0, Vind_1, WalhaesB4, Wa_1, Wc_1, Wei_0, Westkar_4, Wl_0, Ws_2, Wt_5, Xan-1, Yeg-1, Yo_0, Zal_1, Zdr_1

Table 3.7: Discrepancies between *GL1* sequences from Sanger- and resequence-derived data; Supplementary Material

Alignment position	Region	Col-0 sequence	"1001 genomes" resequence	Bloomer *et al.* sequence	Accessions with discrepancy
198	Intron 1	C	C	_	Wa-1
199-202	Intron 1	GCGC	TCTC	_ _ _ _	Bur-0, Lp2-2, Kz-9, Kin-0
			GCGC	_ _ _ _	Yo-0, RmxA02, Cvi-0, Gy-0, Pro-0, Br-0
			GCTC	_ _ _ _	Ws-2, Es-0, Ga-0, Ler-1
			_ _ _ G	_ _ _ _	Bor-1
			TCGC	_ _ _ _	Uod-1
			_ _ GC	_ _ _ _	Wei-0
508	Intron 2	G	A	G	Ms-0
784	Intron 2	T	A	T	Es-0
964	Intron 2	A	A	_	C24
980	Intron 2	A	A	_	C24
1020	Intron 2	T	A	_	C24
1026	Intron 2	T	T	_	C24
1032	Intron 2	A	_	T	Zdr-1, Li2:1
1174	Intron 2	C	C	T	Es-0
1300	Exon 3	C	C	T	Wei-0
1358-1360	Exon 3	TCT	TCT	_ _ _	Ms-0, Lp2-2
			TC_	_ _ _	Bor-1
			_C_	_ _ _	Bor-4
1361	Exon 3	T	T	C	Ms-0, RmxA02, Bor-1, Bor-4, Lp2-2, Uod-1, Kin-0
1537	Exon 3	C	C	T	Aa-0
1580	Exon 3	G	G	A	An-1, Gr-1, Yo-0, Ws-2, Aa-0, Ms-0, Es-0, RRS10, RmxA02, Pna17, Pu2-23, Lp2-2, Sq-8, Tamm-2, Kz-9, Uod-1, Wei-0, Van-0, Gy-0, Ga-0, Mz-0, Nok-3, Kin-0,

					Ob-0, Pro-0, Br-0
1586	Exon 3	G	G	A	Gr-1, Ws-2, Aa-0, Ms-0, Es-0, RRS10, RmxA02, Pna-17, Pu2-23, Lp2-2, Sq-8, Tamm-2, Kz-9, Uod-1, Wei-0, Ga-0, Nok-3, Kin-0, Ob-0, Pro-0, Br-0
1606	3' flank	G	G	A	An-1, Bur-0, Gr-1, Yo-0, Ws-2, Aa-0, Ms-0, Es-0, RRS10, RmxA02, Pna-17, Bor-1, Bor-4, Pu2-7, Pu2-23, Lp2-2, Sq-8, Tamm-2, Kz-9, Uod-1, Cvi-0, Ei-2, Wei-0, Van-0, Gy-0, Ga-0, Mz-0, Nok-3, Kin-0, Ob-0, Pro-0, Br-0
1608-1615	3' flank	_ _ _ _ _ _ _ _	_ _ _ _ _ _ _ _	ATTGTTCA	An-1, Bur-0, Gr-1, Ws-2, Aa-0, Ms-0, Es-0, RRS10, RmxA02, Pna-17, Bor-1, Bor-4, Pu2-7, Pu2-23, Lp2-2, Tamm-2, Kz-9, Uod-1, Cvi-0, Ei-2, Wei-0, Van-0, Est-1, Gy-0, Ga-0, Nok-3, Kin-0, Ob-0, Pro-0, Br-0
	3' flank	_ _ _ _ _ _ _ _	_ _ _ _ _ _ _ _	ATAGTTCA	Yo-0, Sq-8, Mz-0
1621	3' flank	T	T	A	Pu2-23, Uod-1, Van-0, Pro-0, Br-0
1625	3' flank	T	A	T	An-1, Gr-1, Yo-0, Ws-2, Aa-0, Ms-0, Es-0, RRS10, RmxA02, Pna-17, Bor-1, Bor-4, Pu2-7, Pu2-23, Lp2-2, Sq-8, Tamm-2, Kz-9, Uod-1, Ei-2, Wei-0, Van-0, Gy-0, Ga-0, Mz-0, Nok-3, Kin-0, Ob-0, Pro-0, Br-0
1676-1682	3' flank	_ _ _ _ _ _ _	_ _ _ _ _ _ _	AGTAATA	Bur-0, Gr-0, Yo-0, Ws-2, Aa-0, Ms-0, Es-0, Bor-1, Bor-4, Pu2-7, Pu2-23, Lp2-2, Sq-8, Tamm-2, Kz-9, Uod-1, Cvi-0, Ei-2, Wei-0, Van-0, Est-1, Gy-0, Ga-0, Mz-0, Nok-3, Kin-0, Ob-0, Pro-0, Br-0
		_ _ _ _ _ _ _	_ _ _ _ _ _ _	AGTAAGA	An-1, RRS10, RmxA02, Pna-17
1-1686	whole locus		Identical to Col-0	Whole locus deletion	Est

Table 3.8: Comparison between *GL1* datasets for tests of molecular evolution; Supplementary Material

	Tajima's D	Fu and Li's D*	Fu and Li's F*
55 Sanger gDNA	-1.15108 $p>0.1$	-0.73658 $p>0.1$	-1.06179 $p>0.1$
55 Reseqeuence gDNA	-1.11693 $p>0.1$	-1.30130 $p>0.1$	-1.47179 $p>0.1$
55 Sanger cDNA	-0.52662 $p>0.1$	-1.08396 $p>0.1$	-1.05606 $p>0.1$
55 Resequence cDNA	-0.81839 $p>0.1$	-1.08396 $p>0.1$	-1.17666 $p>0.1$
Sanger complete (115)	-1.54605 $p>0.1$	-2.99051 $p<0.05$	-2.85753 $p<0.05$
Resequence complete (261)	-1.84507 $p<0.05$	-4.08234 $p<0.02$	-3.61552 $p<0.02$

4 VARIATION IN THE PLEIOTROPIC REGULATORY GENE *TTG1* IMPACTS TRICHOME DENSITY

4.1 Abstract

The characterisation of variation in genes and pathways with known phenotypic effects can provide insight into the genetic changes that underlie natural variation in phenotypes. The epidermal development pathway in *Arabidopsis thaliana* is a powerful model for such studies, providing opportunities to address a diverse range of questions in evolutionary biology. *TTG1* plays a key role in epidermal development as a structural component of the TTG1:bHLH:MYB transcriptional activation complex and is highly pleiotropic; *ttg1* mutants display aberrant phenotypes for trichome and root hair initiation, anthocyanin, seed coat pigment and mucilage production. Here, *TTG1* allelic variation is surveyed in a set of 96 natural accessions of *A. thaliana*. A strong pattern of high frequency polymorphism separates sequences into two haplogroups with evidence of an underlying geographic distribution. Despite considerable variation in nucleotide sequence, the TTG1 protein is highly conserved; only four amino acid replacements found across all accessions, none reflecting the pattern of polymorphism. The observed pattern associates with trichome density in both association and F_2 mapping, suggesting that regulatory variation in *TTG1* contributes to natural variation in density. These results support the proposal that variation in regulatory genes may make a significant contribution to phenotypic variation, and that variation in regulatory sequences could allow pleiotropic genes to impact phenotype with minimal deleterious consequences.

4.2 Introduction

Elucidating the genetic basis of natural variation in phenotypes is a long-standing goal of evolutionary biology, and relies on the characterisation of variation in genes and pathways with known phenotypic effects. Regulatory mutations, both in cis-regulatory regions and in trans-acting factors, may be central to the generation of natural phenotypic variation (Doebley & Lukens, 1998; Carroll, 2005; Hoekstra & Coyne, 2007; Lynch & Wagner, 2008), with recent reviews emphasizing a role for transcriptional regulators (Lynch & Wagner, 2008; Martin*, et al.*, 2010) as key players underlying variation in traits. This emphasis on regulatory change follows from the assumption that such mutations are less likely to have deleterious pleiotropic consequences than structural variation, particularly in non-regulatory genes. As such, within regulatory networks those transcription factors with many target genes or which impact multiple traits are expected to be more constrained in their evolution than transcription factors with a narrower range of effects (Martin*, et al.*, 2010).

The epidermis mediates interactions between the plant and its environment, with epidermal traits regulated by the TTG1:bHLH:MYB complex playing a number of vital roles in nature. Root hairs are involved in nutrient and water acquisition (Gilroy & Jones, 2000); anthocyanins play roles in attraction of pollinators and seed dispersers as well as being implicated in a range of stress responses (Chalker-Scott, 1999); and seed coat tannins and mucilage are important for seed protection, dormancy and germination (Debeaujon*, et al.*, 2000). Trichomes are found across most land plant lineages, fulfilling a range of functions in different plant species from improved drought tolerance (Ennajeh*, et al.*, 2006) and water use efficiency (Hogan*, et al.*, 1994) to UV-B (Karabourniotis*, et al.*, 1995) and herbivory protection (Mauricio, 1998). In *A. thaliana* and its relatives trichomes act as a defence against insect herbivory. Trichomes have both fitness costs and benefits in *A. thaliana*; insect predation exerts positive selection for increased trichome density (Mauricio & Rausher, 1997) although at a cost in terms of standardised growth rate (Zust*, et al.*, 2011) and susceptibility to fungal pathogens (Calo*, et al.*, 2006). In the sister species *A. lyrata* trichome production is under divergent selection (Karkkainen*, et al.*, 2004), while in *A. kamchatica* the fitness consequences of trichome production are environmentally dependent (Steets*, et al.*, 2010). Variation in trichome density has a strong genetic basis in *A. thaliana* (Symonds*, et al.*, 2005; Bloomer*, et al.*, 2012), and the selective pressures observed in this species have been

suggested to maintain genetic variation and/or phenotypic plasticity for this trait within or among populations (Bloomer*, et al.*, 2012).

The regulation of trichome initiation and density in *Arabidopsis thaliana* has long been a model for the genetic basis of cellular differentiation and pattern formation, and shares a common regulatory pathway with a number of other epidermal cell fates. A relatively well genetically characterised and ecologically important trait, leaf trichome density in *A. thaliana* is emerging as a model system for evolutionary biologists addressing the genetic architecture (Symonds*, et al.*, 2005) and molecular genetic basis (Hilscher*, et al.*, 2009; Symonds*, et al.*, 2011; Bloomer*, et al.*, 2012) of natural variation in plants. Trichome initiation is controlled via a regulatory pathway common to a range of epidermal traits including root hair development, anthocyanin, seed coat tannin and mucilage production. In this pathway, the WD40 repeat protein TTG1 (Walker*, et al.*, 1999) interacts with bHLH and R2R3 MYB transcription factors (Zhang*, et al.*, 2003) to form a transcriptional activation complex, targeting downstream genes and leading to cell fate determination. While TTG1 is not a transcription factor, having no known DNA binding activity, it appears to mediate interaction between bHLH and MYB transcription factors and is required for the transcriptional activation activity of the TTG1:bHLH:MYB complex. TTG1 is required for all cell fates, four functionally overlapping bHLHs each act as activators of a subset of the traits, and the R2R3 MYBs are highly trait-specific (Gonzalez*, et al.*, 2009). For trichome production, TTG1 interacts with the bHLHs GL3, EGL3 (Zhang*, et al.*, 2003) or AtMYC1 (Symonds*, et al.*, 2011) and the trait-specific MYB GL1 (Oppenheimer*, et al.*, 1991).

Trichome spacing is achieved by lateral inhibition, through the action of a suite of R3 MYBs which have bHLH binding capability but lack a transcriptional activation domain (Wang*, et al.*, 2008). R3 MYB expression is activated by the TTG1:bHLH:GL1 complex in the initiating trichome. Following translation, the R3 MYBs move into neighbouring cells where they compete with GL1 for binding, preventing trichome initiation in trichome-adjacent cells. In addition to the activator-inhibitor model, a depletion trapping mechanism has been proposed to generate spacing (Bouyer*, et al.*, 2008); TTG1 is initially expressed in all epidermal cells but can move between them, and is trapped in developing trichomes through direct interaction with GL3 (Balkunde*, et al.*, 2011). While the role of *TTG1* in determining trichome initiation is clear, its role in trichome spacing is somewhat more cryptic. Strong loss-of-function *ttg1*

alleles result in glabrousness, while weak alleles have an observed trichome clustering phenotype (Larkin*, et al.*, 1994; Larkin*, et al.*, 1999). Mathematical modelling of depletion trapping by GL3 indicates that this mechanism may play a role in the trichome clustering phenotype of weak *TTG1* alleles (Bouyer*, et al.*, 2008).

While loss of function mutations in either *TTG1* or the R2R3 MYB *GL1* result in glabrous plants, *ttg1* mutants also display aberrant root hair patterning and lack anthocyanins, seed coat tannins and mucilage (Koornneef, 1981). All natural glabrous accessions of *A. thaliana* thus far characterised have loss-of-function mutations in *GL1* (Bloomer*, et al.*, 2012). Given its broad expression pattern (Bouyer*, et al.*, 2008) and involvement in a range of epidermal traits, all of which play important roles in the mediation of plant:environment interactions, it is not surprising that no loss-of-function mutations in *TTG1* have been described in natural accessions. Pleiotropic genes are commonly thought to be more evolutionarily constrained than those affecting few or single traits, suggesting that *TTG1* should be highly conserved. Despite this, quantitative trait loci (QTL) have been mapped near *TTG1* in *A. thaliana* (Symonds*, et al.*, 2005; Atwell*, et al.*, 2010), hinting at a role for *TTG1* in contributing to trichome density variation in nature. Here, we describe allelic variation in *TTG1*, the highly pleiotropic master regulator of epidermal phenotypes (Koornneef, 1981), in a set of 96 natural accessions of *A. thaliana* and report a surprising role for variation at *TTG1* in determining trichome density. The results support a role for *TTG1*, together with other genes in the epidermal development pathway, in contributing to natural variation in trichome density in *A. thaliana*, and more generally for molecular variation in regulatory genes in shaping natural phenotypic variation.

4.3 Materials and methods

4.3.1 Natural molecular variation at *TTG1*

A set of 96 globally distributed natural accessions of *A. thaliana* (CS22660; Table 4.4, Supplementary Material) was obtained from the *Arabidopsis* Biological Resource Centre (ABRC). DNA was extracted from early rosette leaves of these plants using a modified CTAB extraction protocol (Doyle & Doyle, 1987).

Two PCR reactions were used to amplify an approximately 2.3 kb region of the *TTG1* locus, spanning the 5' UTR, coding region, intron and 3'UTR. The upstream, coding and intron regions were amplified in one PCR in a 20 µL reaction volume containing 1X NEB buffer, 0.25 mM dNTPs, 0.4 µM each of primers TTG1-5pf7-F and TTG1-3pf19-R (primer sequences given in Table 4.5, Supplementary Material), one unit of NEB Taq polymerase (New England Biolabs) and ~50 ng genomic DNA template using the following cycling conditions: 95^{o}C for 3 minutes, followed by 30 cycles of 95^{o}C for 40 seconds, 52^{o}C for 40 seconds and 72^{o}C for 3 minutes, a 10 minute extension at 72^{o}C and a final holding temperature of 4^{o}C. A second PCR, using the same reaction conditions but primers TTG1-i12-F and TTG1-3'-3R, was used to obtain product spanning the remaining intron and 3'UTR. Accuprime DNA Polymerase (Invitrogen) was used for several difficult-to-amplify templates in a 25 µL reaction volume containing 1X Accuprime Buffer II, 0.4 µM each of the primers TTG1-5pf7-F and TTG1-3pf19-R, 0.35 µL Accuprime DNA Polymerase as supplied by the manufacturer and ~50 ng genomic DNA template using the following cycling conditions: 94^{o}C for 3 minutes, followed by 30 cycles of 95^{o}C for 30 seconds, 52^{o}C for 30 seconds and 68^{o}C for 2 minutes and 30 seconds, a 10 minute extension at 68^{o}C and a final holding temperature of 4^{o}C. PCR products were used as template for sequencing *TTG1* using BigDye Terminator v3.1 chemistry and manufacturer's protocols (Applied Biosystems); primers used for sequencing are given in Table 4.5 (Supplementary Material). Sequencing products were aligned using SEQUENCHER v. 4.7 (Gene Codes) and manually checked and corrected for quality. SEQUENCHER was used for assembly of contiguous sequences, which were then exported to BIOEDIT v7.0.9.0 (Hall, 1999).

4.3.2 Analyses of molecular variation

A full length alignment of *TTG1* genomic sequences from the 96 accessions, spanning the 5' UTR, exon, intron and 3' UTR, was generated using ClustallW in BIOEDIT v7.0.9.0 (Hall, 1999). This alignment was edited to give an alignment of the *TTG1* coding region as inferred from the Col-0 reference genome. Nucleotide diversity (π) was examined in the full length genomic alignment using a sliding window analysis of π, calculated in DNASP v5.10 (Rozas*, et al.*, 2003) using a window of 50 bp and step size of 10 bp with gaps considered; in addition, a gene-wide value of π was calculated for both the full length genomic and coding region alignments. The coding region alignment was used to create a haplotype file in DNASP v5.10 (Rozas*, et al.*, 2003); this was subsequently exported to NETWORK v4.5 (Fluxus Engineering) and a haplotype network was generated using median-joining (Bandelt*, et al.*, 1999). A pattern of high frequency polymorphism separated the sequences into two major haplogroups, henceforth referred to as "A" (Col-0 haplogroup) and "B" haplogroups following Bloomer *et al.* (2012). These were defined as distinct sequence group sets and used as "interspecific" or intraspecific" datasets to perform a sliding window analysis of *K*a/*K*s in DNASP, using a window size of 45 bp and step size of 9 bp, as a test of departure from neutral evolution across the *TTG1* coding region. Due to extremely low levels of polymorphism in the coding region, a sliding window value of *K*a and a gene-wide value of *K*s were used to determine the *K*a/*K*s ratio for these analyses following Shiu *et al.* (2004). As further tests for departure from neutrality at *TTG1*, Tajima's D and Fu and Li's D* and F* were calculated across both the full length genomic and coding region alignments in DNASP for the full alignment of 96 sequences, and across the coding region alignment for the A and B haplogroups separately.

4.3.3 Trichome density phenotyping in natural accessions

The trichome density phenotyping method used for the 96 natural accessions has been described previously (Chapter 3; (Bloomer*, et al.*, 2012). Briefly, three replicates of each accession were potted in soil in 72-cell flats in a fully randomised layout, vernalized to synchronise germination and then grown in a growth room until 27 days post-germination. At this point the fifth true leaf was fully expanded but not yet senescent, and was used to score trichome density in a 17 mm^2 area midway along the length of the leaf blade between the midrib and leaf margin using a dissecting microscope at 25x magnification according to Symonds, *et al.* (2005).

4.3.4 Association mapping in natural accessions

The pattern of high frequency polymorphism differentiating *TTG1* haplogroups A and B was tested for an association with trichome density using analyses described previously (Chapter 3, Section 3.3.4.2). The association was tested using a unified mixed model approach (Yu*, et al.*, 2006) as used by Zhao *et al.* (2007); the analyses include a naïve model and models incorporating either kinship or population structure matrices or both. This approach was taken to reduce confounding due to population structure in the set of accessions used, minimising the chance of identifying a false positive association.

4.3.5 F_2 association mapping

Association mapping models such as those used here can lead to false negative results. To complement the association mapping analysis described above, an F_2 mapping population was developed from a cross between the *TTG1* haplogroup A accession Bor-4 and the haplogroup B accession Sorbo, with Sorbo used as the maternal parent. The F_1 seed resulting from the initial cross were genotyped with four microsatellite markers to confirm the cross, and a single plant was selfed to obtain F_2 seed. Trichome density phenotypes were scored for 191 F_2 plants following the method described above (Section 4.3.3). DNA was extracted from the plants after phenotyping and used to genotype plants for *TTG1* haplogroup by PCR detection of a 7 bp size difference between alleles. An M13 primer-tailing scheme was used to fluorescently label PCR products in a 10 µL PCR containing 1X Buffer BD (Solis BioDyne), 2.5 mM $MgCl_2$, 2 µmol dNTPs, 0.2 µmol of the M13-tailed marker-specific forward primer TTG1_3pf31_M13F, 4.5 µmol TTG1_3pf30_R (Table 4.5, Supplementary Material), 4.5 µmol FAM-labelled M13 primer (M13-FAM), 0.25 units of FIREPol DNA polymerase (Solis BioDyne), and ~50 ng genomic DNA under the following cycling conditions: 95^{o}C for 3 minutes, followed by 30 cycles of 95^{o}C for 30 seconds, 52^{o}C for 40 seconds, and 72^{o}C for 40 seconds, and a final extension of 20 minutes at 72^{o}C. One microlitre of PCR product was then combined with 9 µL of a HiDi (Applied Biosystems) and CASS size standard (Symonds & Lloyd, 2004) mix. Allele size was determined by capillary separation of the PCR products on an ABI3730 Genetic Analyzer (Applied Biosystems) at the Massey Genome Service, and allele sizes were scored in GENEMAPPER v3.7 (Applied Biosystems). Ultimately, both phenotypes and genotypes were obtained for 182 F_2s.

4.3.6 Transgenic analysis of functional variation in *TTG1*

Our association mapping results suggested an association between trichome density and a pattern of high frequency polymorphism that differentiates the alleles into two haplotype groups. Of particular interest was a highly variable region spanning the *TTG1* intron. Constructs were developed to test the effect of this intron using transgenics. PCR was used to amplify a 5.2 kb region of the *TTG1* locus spanning a promoter region shown previously to be sufficient for complementation of *ttg1-1* mutants with *TTG1* (Walker*, et al.*, 1999), the 5' UTR, exon, intron, the 3' UTR and downstream flanking sequence that extends to the beginning of the next gene. This region was amplified from four natural accessions (Ler-1 and Col-0 (haplogroup A) and Sorbo and RRS-7 (haplogroup B); Table 4.4, Supplementary Material), in a 50 µL reaction volume containing 1X HF buffer, 0.2 mM dNTPs, 0.48 µM Gateway (Invitrogen) attB-tailed primers TTG1_5'_3F and TTG1_3'_3R (Table 4.4, Supplementary Material) and one unit of Phusion DNA Polymerase (Finnizymes) using the following cycling conditions: 98^{o}C for 3 minutes, followed by 35 cycles of 98^{o}C for 10 seconds, 55^{o}C for 20 seconds and 72^{o}C for 2 minutes and 30 seconds, then a 10 minute extension at 72^{o}C and a 4^{o}C hold.

In addition, two constructs in which the haplogroup A and haplogroup B introns were spliced into an opposite haplogroup allele were constructed using Splicing by Overlap Extension (SOE) PCR (Horton*, et al.*, 1989). The *TTG1* locus defined above was amplified in three PCRs from L*er*-1 (haplogroup A) and RRS-7 (haplogroup B) to amplify 1) the promoter region, 5' UTR and exon; 2) the intron; and 3) the 3' UTR and 3' downstream sequence, using 30-40 bp primers designed to overlap by 15-20 bp at the splicing points (Table 4.5, Supplementary Material). Each PCR was carried out in a 25 µL reaction volume containing Phusion 1X HF buffer, 0.2 mM dNTPs, 0.48 µM of each primer, and 0.5 units of Phusion DNA Polymerase (Finnizymes) with the following cycling conditions: 98^{o}C for 2 minutes, followed by 35 cycles of 98^{o}C for 10 seconds, 58^{o}C for 20 seconds and 72^{o}C for either 90 seconds (PCR 1) or 20 seconds (PCRs 2 and 3). The resulting PCR products were purified by precipitation in 10% PEG/$MgCl_2$ to remove primer dimers, then pooled together in a secondary PCR reaction: haplogroup A PCRS 1 and 3 were pooled together with haplogroup B PCR 2, and vice versa. Initially, a 50 µL reaction containing 1X HF buffer, 0.2 mM dNTPs, 1 unit of Phusion DNA Polymerase (Finnizymes) and pooled PCR products was heated to 98^{o}C for 2 minutes, then amplified for four cycles of 98^{o}C for 10 seconds, 52^{o}C for 20 seconds and 72^{o}C for 2 minutes, with the overlapping ends of the first-round PCRs annealing and acting as primers to extend a full

length product. The primers TTG1_5'_3F and TTG1_3'_3R were added to the reaction which was then heated for 3 minutes at 72°C, followed by a further 34 cycles of 98°C for 10 seconds, 58°C for 20 seconds and 72°C for 2 minutes and 30 seconds, followed by a 10 minute extension at 72°C, to amplify the full length intron-spliced construct.

Native allele and intron-spliced PCR constructs were purified by precipitation in 10% PEG/$MgCl_2$ to remove primer dimers and then recombined into the Gateway donor plasmid pDONR-zeo (Invitrogen) using half Gateway BP reactions according to the manufacturer's guidelines. The plasmids were cloned into *E. coli* TOP10 cells (Invitrogen) and selected on low salt LB/Agar containing 50 µg/mL zeocin (Invitrogen). Positive transformants were identified by colony PCR using primers TTG1-5pf7-F and TTG1-3pf19-R and reaction conditions described above, and used to start 10 mL low salt LB/zeocin overnight cultures. Plasmids were isolated from liquid cultures using a QiaSpin plasmid miniprep kit (Qiagen), and the *TTG1* inserts sequenced. Half Gateway LR recombination reactions (Invitrogen) were used to transfer the constructs into the binary plasmid pBARN-RFB; the plasmids were then cloned into *E. coli* TOP10 cells and selected on LB/Agar containing 50 µg/mL kanamycin. Positive clones were identified by PCR and used to start 10mL LB/ 50 µg/mL kanamycin overnight cultures, with the plasmids isolated using a QiaSpin plasmid miniprep kit (Qiagen). The recovered plasmids, and an empty pBARN-RFB vector, were transformed into *Agrobacterium tumifaciens* strain GV3101 by electroporation and selected on LB/Agar containing 50 µg/mL kanamycin.

The six *TTG1* constructs, together with the empty pBARN-RFB vector as a negative control (constructs LLC-001 – LLC-007, Table 4.3), were transformed into the *A. thaliana* wildtype accession L*er*-0 and the *ttg1-1* mutant (a *TTG1* loss-of-function allele in the L*er* backgroiund), using a modified floral dip method (Clough & Bent, 1998). Transformed *A. tumifaciens* strain GV3101 was grown in a LB/50 µg/mL kanamycin culture for 48 hours, then pelleted and resuspended in 5% w/v sucrose solution containing 0.05% v/v Pulse (>800 g/L modified polydimethylsiloxane; Nufarm) as a wetting agent. Three pots (10 plants/pot) each of flowering L*er*-0 and *ttg1-1* were dipped in *Agrobacteria*/sucrose/Pulse solution for 15 minutes with occasional agitation. Plants were laid prostrate down in a covered flat in a darkened growth room overnight to dry, and then sat upright, bagged to prevent crossing and grown as normal until seed set.

4.3.7 Selection and screening of *TTG1* transgenics for trichome density

Primary transformants (T_1s) were screened using a glufosinate-ammonium resistance marker. For each of the seven constructs, approximately 6400 T_1 seeds in the *ttg1-1* background and 3200 seeds in the L*er*-0 background were potted at ~100 seeds per pot. Seedlings were sprayed with 125 mg/L glufosinate-ammonium at approximately 3 days post-germination, with additional spray treatments at 6 and 9 days post-germination. Surviving plants' trichome densities were scored on both fifth and seventh leaves in a 17 mm^2 area midway along the length of the leaf blade as outlined above. In addition, trichome clustering (defined as two or more trichomes emerging from adjacent cells in the area scored for density) and branching (the average number of branches per trichome in the area scored for density) were scored on both leaves. As plants germinated at different times (within a range of approximately one week), in order to standardise developmental stage trichome density was scored on fifth leaves on the day the plant bolted and on seventh leaves 2 days later.

4.4 Results

4.4.1 Nucleotide diversity at the *TTG1* locus

The extent and distribution of variation at the *TTG1* locus was assessed using both gene-wide and sliding window measures of nucleotide diversity (π) across the full length genomic alignment of 96 sequences spanning the 5' UTR, exon, intron and 3' UTR. Gene-wide π across the 96 sequences was 0.0135 (Table 4.1), while a sliding window analysis revealed that this variation was distributed unevenly across the locus (Figure 4.1). The 5' UTR, exon and 3' UTR were overall relatively conserved although localised peaks in π of between 0.03 and 0.04 were identified in both the exon and 3' UTR. In contrast, regions of very high π, between 0.07 and 0.08, were observed in the intron of *TTG1*. Across the coding region, gene-wide π was 0.0058.

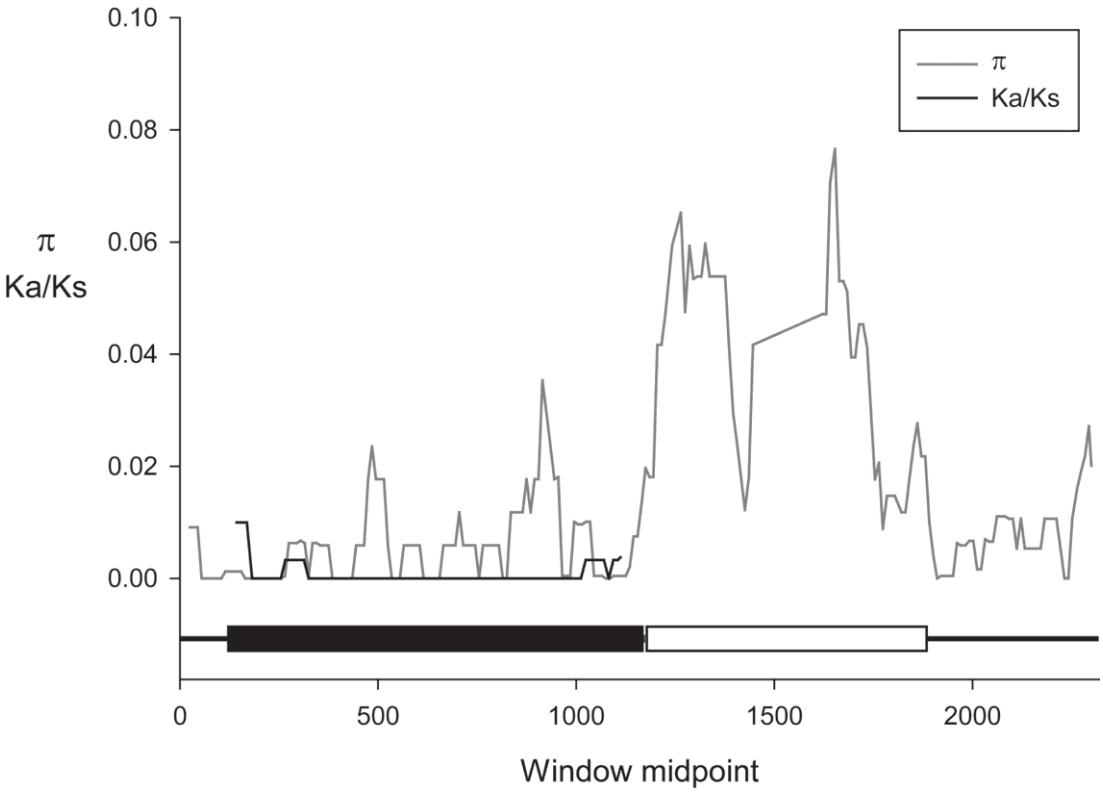

Figure 4.1: Sliding window analysis of nucleotide diversity (π) and *K*a/*K*s across the *TTG1* locus.

Pi (π; grey line) was calculated across the full genomic region alignment, while *K*a/*K*s (black line) was calculated for the coding region. The underlying schematic representation of *TTG1* shows the 5' and 3' UTR as black lines, the exon as a black bar and the intron as a white bar.

Table 4.1: Nucleotide haplotypes, diversity and tests for molecular evolution in *TTG1* coding region and full length alignments

	No. nt haplotypes	π	Tajima's D	Fu and Li's D*	Fu and Li's F*
All accessions, full length	23	0.0135	1.01024 $p>0.10$	1.31442 $p>0.10$	1.42351 $p>0.10$
All accessions, coding	8	0.0058	0.52779 $p>0.10$	0.04210 $p>0.10$	0.27298 $p>0.10$
Haplogroup A, coding	7	0.0004	-1.70562 $p<0.10$	-3.00732 $p<0.05$	-3.03847 $p<0.02$
Haplogroup B, coding	1	0	NA*	NA*	NA*

*NA: Could not be calculated as sequences were completely conserved

4.4.2 Coding haplotypes and functional variation in *TTG1*

An alignment of *TTG1* coding region sequences from the 96 accessions was used as the basis for a haplotype network reflecting nucleotide and amino acid diversity across the coding region (Figure 4.2). Eight haplotypes were identified based on nucleotide differences between accessions, with between one and 62 accessions in each group. A pattern of high frequency polymorphism based on 19 high frequency SNPs found across the coding region separated the haplotypes into two major groups, referred to as haplogroups A and B (Table 4.4, Supplementary Material), with 79 and 17 sequences respectively. An additional seven SNPs were found at lower frequency, occurring in between one and ten accessions. Seven haplotypes were found in haplogroup A while just one haplotype, representing all 17 haplogroup B accessions, was found in haplogroup B (Figure 4.2. The comparative conservation of haplogroup B alleles may reflect sampling, with only 17 haplogroup B accessions identified compared with 79 haplogroup A accessions. The pattern of polymorphism separating the sequences into haplogroups A and B across the coding region continued into the intron and was responsible for much of the nucleotide diversity observed in the sliding window analysis; in the 758 aligned base pairs of the intron, 61 SNPs and 10 indels with lengths varying from 1 to 166 bp distinguished the two haplogroups. In addition to this pattern of high frequency polymorphism, a further 13 SNPs and one indel, shared by between one and 31 accessions respectively, were found in the *TTG1* intron. Interestingly, the *TTG1* intron was highly conserved amongst the 17 haplogroup B accessions with only a single SNP found, in accession Kz-1.

If only amino acid changes were considered just four haplotypes are identified, with between one and 91 accessions in each group. The pattern of polymorphism distinguishing haplogroups A and B was based entirely on synonymous SNPs so collapsed when only amino acid replacements were considered. The amino acid sequence of TTG1 was highly conserved with only four amino acid replacements found across the 341 amino acids encoded by *TTG1,* shared among just five of the 96 accessions sequenced (Figure 4.2; Table 4.4, Supplementary Material). One replacement, L9I, was common to three accessions, Var2-6, Spr1-2 and Spr1-6. Another, A311T, was found only in accession Lp2-2. Finally, accession Tsu-1 had two replacements, D59N and A333S.

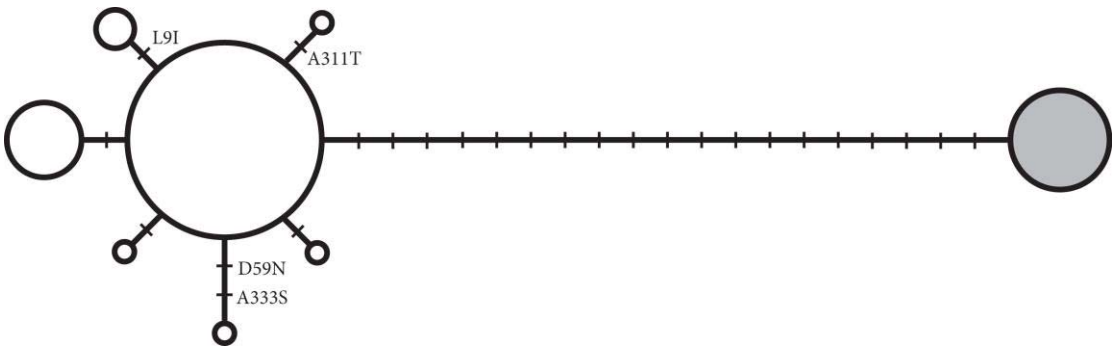

Figure 4.2: Median-joining haplotype network of *TTG1* coding region alleles.

Each circle represents a distinct nucleotide haplotype, with relative circle size indicating the number of accessions sharing a given haplotype. Nucleotide changes are indicated by line length, with the number of changes shown by perpendicular lines; nonsynonymous nucleotide changes are indicated with amino acid replacements given. Haplogroup A haplotypes are shown as unfilled circles, while the single haplogroup B haplotype is represented by the filled grey circle.

4.4.3 Molecular evolution at *TTG1*

To investigate molecular evolution at *TTG1*, Tajima's D and Fu and Li's D* and F* were calculated for both coding and full length *TTG1* alignments of the 96 accessions' sequences, and for haplogroups A and B separately across the coding alignment. When calculated from the full length and coding alignments of all 96 accessions' sequences, positive values were obtained for all statistics, with higher values found for the full length alignment (Table 4.1).

No test result was significantly different from zero. In contrast, D, D* and F* were negative across the haplogroup A coding region alignment, with D* and F* significantly different from zero; D, D* and F* could not be calculated for the haplogroup B coding alignment as the sequences were completely conserved (Table 4.1). In addition, a sliding window analysis of *K*a/*K*s was performed across the coding region alignment of all 96 accessions using a site specific *K*a and gene-wide *K*s; this test suggested strong purifying selection acting on the *TTG1* coding region with *K*a/*K*s<<1 (Figure 4.1).

4.4.4 Association mapping in natural accessions

Trichome density phenotypes for 94 of the 96 accessions sequenced at the *TTG1* locus were obtained previously and are presented elsewhere (Figure 3.5, Chapter 3 Supplementary Material; (Bloomer*, et al.*, 2012)); of the two accessions not included, one is glabrous and left out of the association mapping analyses presented here, and a second did not germinate in the phenotyping experiment. A range of association mapping models (described in Section 3.3.4.2; (Bloomer*, et al.*, 2012)) were used to test for an association between trichome density and the A and B *TTG1* haplogroups (Table 4.2). Mean trichome density was 11.47 for haplogroup A accessions and 16.86 for haplogroup B accessions (Figure 4.3 A), a difference which was highly significant (p<0.001) according to unpaired T-test. In contrast, the significance of this difference varied by association mapping model used in the analysis. The naïve model and models which accounted for population structure using either population structure as a fixed effect (Q) or kinship as a random effect (K or K*) determined a highly significant association between haplogroups and trichome density, as did model QK. One model, QK*, yielded a marginally insignificant association, with p=0.061 (Table 4.2).

Table 4.2: Association test results (*p*-values) for *TTG1* haplogroups and trichome density

***TTG1* haplogroup**	**Trichome density**	**Association model**					
		Naïve	**QK**	**Q**	**K**	**QK***	**K***
A/B	11.47 / 16.86	<0.0001	0.0024	0.0010	0.0007	0.0610	0.0077

4.4.5 F_2 association mapping in the Sorbo x Bor-4 population

As a further test of an association between the A and B haplogroups and trichome density, density phenotypes were scored in an F_2 population derived from haplogroup A (Bor-4) and haplogroup B (Sorbo) parental accessions; this approach should eliminate confounding due to population structure. Phenotypes and genotypes were obtained for 182 F_2 plants. Segregation distortion was observed in the F_2s, with fewer haplogroup A alleles observed than expected; 35 F_2 plants were homozygous for haplogroup A *TTG1*, compared with 57 homozygous for haplogroup B and 90 F_2s heterozygous. Both *TTG1* and a number of QTL for seed dormancy are found on Chromosome V in *A. thaliana*, which may explain the difference in frequency of haplogroup A and B alleles here (Alonso-Blanco*, et al.*, 2003). Plants homozygous for haplogroup A *TTG1* had an average trichome density of 7.69, while haplogroup B homozygotes averaged 8.79 (Figure 4.3 B); the difference in densities between the two haplogroups was significant (p<0.01) as determined by unpaired T-test. Heterozygotes averaged 8.90, significantly different from plants homozygous for haplogroup A (p<0.01), but not for haplogroup B (p=0.78), suggesting that the haplogroup B allele is dominant.

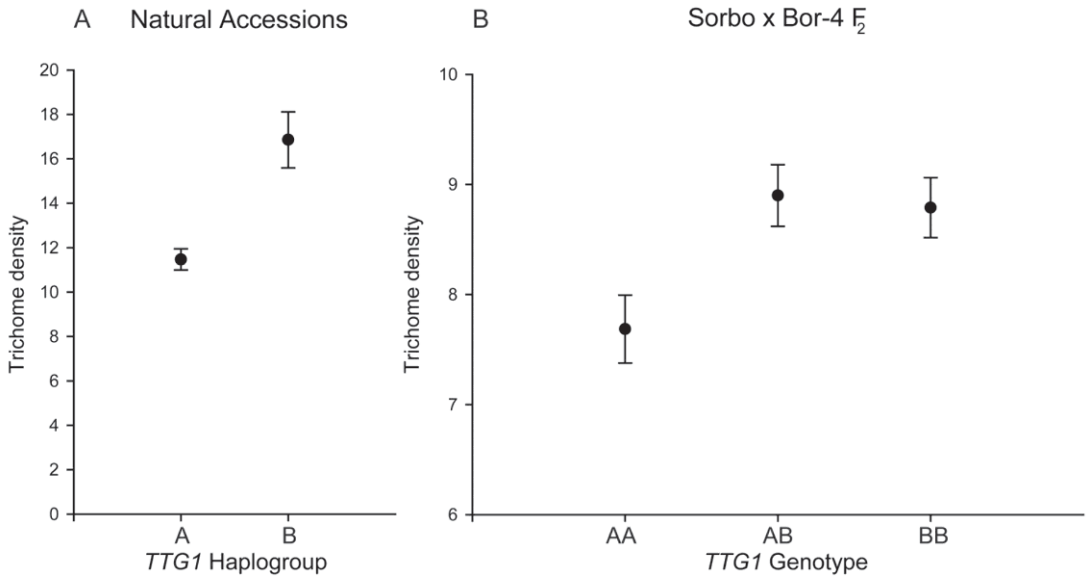

Figure 4.3: Mean trichome densities of A and B haplogroup plants in natural accessions of the CS22660 set and in the F_2 mapping population Sorbo x Bor-4.

Error bars indicate standard error of the mean. The difference in densities between A and B haplogroups is significant at p<0.001 in the natural accessions and p<0.01 in the Sorbo x Bor-4 F_2 mapping population.

4.4.6 Transgenic analysis of *TTG1*

To further assess the effect of the variation distinguishing the A and B haplogroups, *ttg1-1*, a *TTG1* loss-of-function mutation in the L*er* background, and L*er* wildtype plants were transformed with native *TTG1* alleles of both A and B haplogroups. In addition, the role of the intron in generating the difference in trichome densities was tested using spliced gene constructs, with an haplogroup A intron spliced into a B-haplogroup native allele including promoter, 5'UTR, exon and 3'UTR, and vice versa. Transformation efficiency was fairly poor overall, with between zero and eight primary transformants (T_1s) recovered for each construct in each background (Table 4.3). Although such low numbers make comparisons between constructs difficult, trichome density, clustering and branching were scored on both fifth and seventh leaves in the T_1 generation to assess evidence of trends in the data and to determine whether to follow this work further. Preliminary results show that the constructs can complement the *ttg1-1* mutation, with all *ttg1-1* plants transformed with functional *TTG1* alleles producing trichomes on the leaf blade. Similar densities were observed for both *ttg1-1* and wildtype L*er* plants transformed with *TTG1*, suggesting that wildtype-level trichome density phenotypes were restored in *ttg1-1* plants (Table 4.3). Control *ttg1-1* plants, transformed with pBARN-RFB with no insert, did not produce trichomes (Table 4.3). While previous studies have reported clustering of up to 30% of trichomes with some mutant alleles (Larkin*, et al.*, 1994; Larkin*, et al.*, 1999), only one example of trichome clustering was observed here, in a L*er* wildtype plant transformed with LLC-004 (Sorbo *TTG1*; haplogroup B); two of three trichomes in the area of the fifth leaf scored for density formed a single cluster. Trichome branching phenotypes were similar for all constructs.

Trichome density scores were very low for all transformants on both fifth and seventh leaves (Table 4.3), as has been observed previously in the L*er* wildtype accession. Due to the very small number of T_1 plants obtained for most constructs (Table 4.3), a meaningful comparison of phenotypes between individual constructs cannot be made. Trichome densities of haplogroup A (LLC-001 and LLC-002) and haplogroup B (LLC-003 and LLC-004) native allele transformants in the *ttg1-1* background were pooled to compare phenotypes. Mean trichome density was slightly higher in haplogroup A than haplogroup B transformants for fifth leaves, while the reverse was found on seventh leaves (Table 4.3); neither difference was significant (data not shown). Eight primary transformants were obtained for two constructs in the *ttg1-1* background: LLC-001, the L*er* native allele (haplogroup A promoter, coding region,

intron and downstream sequence; AAA); and LLC-005, a spliced intron construct comprising the L*er* promoter, coding region and downstream sequence with RRS-7 (haplogroup B) intron (ABA). Comparison of these two groups of transformants aims to test the effect of variation between haplogroup A and B alleles in the intron. Trichome density was higher, although not significantly so, in the LLC-005 transformants than in LLC-001 transformants for fifth leaves and identical for seventh leaves (Table 4.3).

Table 4.3: Trichome phenotypes of *ttg1-1* and L*er* plants transformed with *TTG1* constructs

						Fifth leaf average			Seventh leaf average	
Background	**Construct**	**Insert**	***TTG1* haplotype***	**Number of T_1 plants**	**Trichome density**	**Trichome clustering**	**Trichome branch number**	**Trichome density**	**Trichome clustering**	**Trichome branch number**
ttg1-1	LLC-001	Ler native allele	AAA	8	2.38	0	3.27	2.75	0	3.07
	LLC-002	Col-0 native allele	AAA	1	3	0	3	1	3	3
	LLC-003	RRS-7 native allele	BBB	1	1	0	3	4	0	3.25
	LLC-004	Sorbo native allele	BBB	3	2	0	2.56	2.67	0	3.22
	LLC-005	Ler with RRS-7 intron	ABA	8	2.5	0	3.21	2.75	0	3.15
	LLC-006	RRS-7 with Ler intron	BAB	0	-	-	-	-	-	-
	LLC-007	pBARN-RFB; no insert	NA	6	0	NA	NA	0	NA	NA
	Haplogroup A	Pooled LLC-001/LLC-002	AAA	9	2.44	0	3.24	2.67	0	3.12
	Haplogroup B	Pooled LLC-003/LLC-004	BBB	4	1.75	0	2.67	3	0	3.23
Ler	LLC-001	Ler native allele	AAA	1	3	0	3.33	3	0	3
	LLC-002	Col-0 native allele	AAA	4	2.75	0	3.33	4.25	0	3.27
	LLC-003	RRS-7 native allele	BBB	2	1.5	0	3.5	2.5	0	3.17
	LLC-004	Sorbo native allele	BBB	3	3	0.22	3	3.33	0	3.17
	LLC-005	Ler with RRS-7 intron	ABA	1	1	0	3	3	0	3
	LLC-006	RRS-7 with Ler intron	BAB	0	-	-	-	-	-	-
	LLC-007	pBARN-RFB; no insert	NA	0	-	-	-	-	-	-

* Three-letter code indicates the pattern of high-frequency polymorphism across the cloned region of the *TTG1* locus as defines the A and B haplogroups: the first letter indicates the haplotype of the promoter, 5'UTR and exon; the second letter the haplotype of the intron; and the third the haplotype of the 3'UTR and downstream sequence. For example, AAA represents a haplogroup A native allele, while ABA represents a haplogroup A allele with a haplogroup B intron spliced in place of the native A-haplogroup intron.

4.5 Discussion

Investigating the consequences of allelic variation at loci with known phenotypic effects is key to understanding the genetic basis of phenotypic variation in nature. The epidermal development pathway in *Arabidopsis thaliana* is an excellent model system for such studies, combining a well-understood genetic pathway with environmentally important phenotypes. The pathway provides opportunities to investigate the importance of regulatory genes in phenotypic variation; the molecular nature of the changes underlying variation in phenotypes; and the way pleiotropy, functional overlap and trait specificity of genes constrain such changes. Here, we investigate the role of natural allelic variation in the highly pleiotropic regulator of epidermal phenotypes, *TTG1*, and its consequences for trichome density.

4.5.1 Molecular variation and evolution in *TTG1*

Nucleotide diversity across the full-length alignment of *TTG1* was relatively high (π=0.01348) in comparison with the genome-wide estimate of nucleotide diversity for *A. thaliana* (π=0.0047 (Bakker*, et al.*, 2006)). While also high in comparison with the gene-wide value for the trichome-specific R2R3 MYB transcription factor *GL1* (π=0.0049; (Bloomer*, et al.*, 2012)), a similarly high value of 0.011 was previously reported for *AtMYC1* (Symonds*, et al.*, 2011), a bHLH transcription factor acting in concert with *TTG1* and *GL1* in trichome initiation. Nucleotide diversity is variable across the *TTG1* locus, with peaks in π as high as 0.07657 found within the intron (Figure 4.1). However, π is considerably lower in the 5' UTR, exon and 3' UTR; the gene-wide value of π across the coding region (exon) is similar to the genome-wide estimate at 0.0058. Peaks of nucleotide diversity in the sliding window analysis reflect a pattern of high frequency polymorphism which is particularly pronounced in the intron, separating *TTG1* alleles into two major haplogroups referred to here as A and B. Such allelic dimorphism is commonly observed in *A. thaliana*, including in other epidermal development genes such as the R3 repressor *ETC2* (Hilscher*, et al.*, 2009), *AtMYC1* (Symonds*, et al.*, 2011),and *GL1* (Bloomer*, et al.*, 2012), and in a number of others including the phenylpropanoid pathway (Aguade, 2001), chitinase (Kawabe*, et al.*, 1997) and R genes (Tian*, et al.*, 2002; Mauricio*, et al.*, 2003; Rose*, et al.*, 2004), and chloroplast sequences (Yin*, et al.*, 2010). *Arabidopsis thaliana* is thought to have been isolated in two glacial refugia before subsequently spreading back into central Europe (Sharbel *et al.* 2000; Beck *et al.* 2008), and divergence, secondary contact and recombination between diverged genomes has been

suggested as an explanation for the observed patterns of allelic dimorphism (Aguade, 2001; Symonds*, et al.*, 2011).

With the exception of the high frequency pattern described above, the *TTG1* coding region was highly conserved both in terms of nucleotide and amino acid sequence. None of the 19 coding region SNPs reflecting the pattern of high frequency polymorphism on which the A and B haplogroups are based are nonsynonymous. Of the four nonsynonymous SNPs elsewhere in the coding region, all are at positions conserved across *A. thaliana*, *A. lyrata* and three *Brassica* species based on cDNA sequences available from Genbank (data not shown). Notably, the replacement A311T is at a position completely conserved across both AtTTG1-type and PhAN11-type WD40 proteins from a range of both dicot and monocot taxa (Lu*, et al.*, 2009; Pang*, et al.*, 2009). Truncation of the C-terminal 25 amino acids of TTG1 is sufficient to abolish bHLH interaction (Payne*, et al.*, 2000) and is the mutation underlying the strongest *ttg1* mutant allele; two of the replacements identified here, A333S and A311T, fall near or within this region. Interestingly, the accession Lp2-2, which carries the A311T mutation, has one of the lowest trichome densities amongst the accessions phenotyped (data presented in Supplementary Material to Chapter 3, Figure 3.5 (Bloomer*, et al.*, 2012)). This mutation is a candidate for a role in reducing trichome density, perhaps reducing TTG1 binding affinity with its bHLH partners.

In agreement with its highly pleiotropic role and with the level of conservation of amino acid sequence observed, a sliding window of *K*a/*K*s across the single exon of *TTG1* suggests that evolution of the protein is highly constrained. *K*a/*K*s is much less than one across the entire coding region, indicating purifying selection acting on *TTG1*. Conversely, when calculated using all 96 accessions' sequences, significant departure from the neutral model of evolution was not detected at *TTG1* using either Tajima's D or Fu and Li's D* and F* statistics. *TTG1* instead shows non-significant positive values for D, D* and F*, which could be interpreted as indicative of balancing selection. However, D, D* and F* are influenced by both selection and by demographic processes. An alternative interpretation, particularly given the hypothesised history of the species (Sharbel *et al.* 2000; Beck *et al.* 2008), is that the positive values obtained here reflect the admixture of two diverged populations, masking any signature of selection. Supporting this interpretation, when calculated separately for haplogroups A and B the values of D, D* and F* indicate purifying selection at *TTG1.* D* and F* are significantly

negative for haplogroup A across the coding region and cannot be calculated for haplogroup B due to absolute conservation of sequence among the accessions sampled. Consistent with the *K*a/*K*s analysis, these results suggest purifying selection acting on *TTG1*.

WDR proteins often act as protein scaffolds, interacting with and mediating interactions of a range of other proteins. Homologues of *A. thaliana TTG1*, which have generally conserved functions in regulating anthocyanin production (and a broader set of epidermal traits in some species) via interaction with bHLH and MYB transcription factors, show strong conservation of amino acid sequence across a diverse range of monocot and dicot plant species (Lu*, et al.*, 2009; Pang*, et al.*, 2009). Cross-species comparison of amino acid sequences has shown that that homologous WDR proteins are more similar to one another than homologous bHLHs (Carey*, et al.*, 2004); further, the functional homologue of *TTG1* in *Ipomoea* species, *WDR1*, exhibits a higher level of evolutionary constraint than its less pleiotropic bHLH and MYB partners (Streisfeld*, et al.*, 2011). Similarly, the *K*a/*K*s analysis undertaken here indicates strong purifying selection at *TTG1* while less pleiotropic genes within the pathway, the bHLH *AtMYC1* and the MYB *GL1*, show evidence of positive or balancing selection with spikes of *K*a/*K*s>>1 (Symonds*, et al.*, 2011; Bloomer*, et al.*, 2012). This supports the proposal that, within regulatory pathways, those genes higher in regulatory hierarchies will be subject to stronger constraint (Martin*, et al.*, 2010).

4.5.2 Variation at *TTG1* impacts trichome density

Based on both QTL and genome-wide association mapping and candidate gene studies, it has recently been proposed that variation in epidermal development pathway genes might underlie much of the observed variation in epidermal phenotypes in *A. thaliana*. Allelic dimorphism has now been observed at all levels within the epidermal development pathway, including at *TTG1*, the bHLH *AtMYC1* (Symonds*, et al.*, 2011), the R2R3 MYB *GL1* (Bloomer*, et al.*, 2012) and the R3 MYB repressor *ETC2* (Hilscher*, et al.*, 2009). In *ETC2* and *GL1*, polymorphism associated with this pattern has been associated with variation for trichome density in natural accessions. Given the striking difference in sequences between the *TTG1* A and B haplogroups, in particular in the intron, an association between haplogroups and trichome density was investigated using association mapping in natural accessions and an F_2 mapping population, and transgenic approaches.

Association mapping was undertaken in a set of 94 natural accessions for which trichome phenotypes had been obtained previously, following the unified mixed models method of Yu *et al.* (2006) as implemented by Zhao *et al.* (2007) in this set of accessions in *A. thaliana*. A large difference in mean trichome densities between the A and B haplogroups was shown by our association mapping results to be significant, although different methods of accounting for population structure produced somewhat different results. Four of the five models determined a highly significant association (p<0.01) between trichome density and A and B haplogroup variation (Table 4.2). In contrast, the model QK*, which accounts for population structure using ancestral populations determined by STRUCTURE as a fixed effect and kinship as determined by haplotype sharing as a random effect, found a marginally insignificant association. Models which account for population structure in association mapping are fairly conservative and can result in false negative results when underlying population structure is strong (Hagenblad*, et al.*, 2004). Population structure is likely to be a factor here as the haplogroup B sample size is small (17 out of 94 accessions) and the accessions carrying haplogroup B alleles are geographically clustered towards the western end of *A. thaliana*'s native range, collected predominantly from Central Asian and Northern European locations (Table 4.4, Supplementary Material). *Arabidopsis thaliana* shows evidence of an East-West pattern of isolation by distance and, interestingly, many of the haplogroup B accessions share an ancestral group as determined by STRUCTURE when K=2-5 (Nordborg*, et al.*, 2005), which may explain why the QK* model found a marginally insignificant association. To confirm the association mapping results, the association between haplogroups was tested in an F_2 mapping population developed from the haplogroup A and B accessions Bor-4 and Sorbo. Mapping in an F_2 population should resolve the issue of population structure, with recombination shuffling alleles at other loci which may otherwise contribute to the difference in density observed. The F_2 mapping results identified a significant association between the haplogroups and trichome density, suggesting that the effect of *TTG1* is not an artefact of population structure. Causal variation at a tightly linked locus is not ruled out by this work, but could be addressed through development of markers in nearby loci. The difference in mean trichome densities between A and B alleles in the F_2 population is smaller than that observed in the 94 natural accessions. This could be a result of the population structure described above, with haplogroup B accessions potentially sharing high trichome density alleles at other loci.

Transgenics is considered the "gold standard" approach for verification of the effects of natural variants on phenotype, but its utility in assessing small quantitative changes in phenotype is limited. A significant limitation to testing *TTG1* specifically is that the only available mutant is in the L*er* background, which has extremely low transformation efficiency (0.21%; (Desfeux*, et al.*, 2000)). In addition, with a loss-of-function allele at *AtMYC1* (Symonds*, et al.*, 2011) and a putatively disrupted *GL1* allele (Bloomer*, et al.*, 2012), L*er* has a low wildtype trichome density, magnifying the difficulty in detecting small quantitative differences in density in this background. Here, native *TTG1* alleles of both haplogroups and spliced intron constructs were transformed into both wildtype L*er* and *ttg1-1* mutant lines to test the feasibility of a transgenic approach to confirming and further characterising the genetic basis of the observed variation in density.

In this preliminary screen, no significant support was found for the difference in trichome densities between A and B *TTG1* haplogroups. No transformants were obtained for construct LLC-006 (RRS-7 allele with L*er* intron spliced) in either background, or for the empty vector (LLC-007) in the L*er* background; for most other constructs between one and four T_1s were obtained, making meaningful comparison of these transformants' phenotypes impossible. Grouping all native allele transformants by *TTG1* haplogroup in the *ttg1-1* background, the highest-density haplogroup was not consistent between fifth and seventh leaves and there was no significant difference in trichome density between haplogroups A and B. Similarly, a comparison of phenotypes for constructs LLC-001 (L*er* (haplogroup A) native allele), and LLC-005 (L*er* allele with RRS-7 (haplogroup B) intron spliced), with eight T_1s each, revealed no significant difference.

A number of factors may play a role in the insignificant differences found here. The very low trichome density of L*er*, the only genetic background for which a *ttg1* mutant is available, is a major factor; with such low densities, a small difference between haplogroups will be difficult to detect. While TDNA insertions near *TTG1* are available in Col-0, a much higher-density accession, none of these has a *ttg1* mutant phenotype. The low number of biological replicates for each construct is another key issue; even aggregated as haplogroups A and B, sample sizes are very small. This compounds the problems of variation inherent in transgenic analyses due to insert number (dosage), the possibility of transgene silencing, and positional effects (Schubert*, et al.*, 2004), with too few replicates to "average out" these variables

across constructs. Floral dips need be repeated to increase the number of individual transformation events for each construct and transgenic plants screened in later generations to allow more biological replicates of each transformant before a final decision is made on the feasibility of this approach.

While evidence of the pattern underlying the variation between haplogroups A and B is found in SNPs both upstream of the translation start site and within the coding sequence, none of the SNPs distinguishing the haplogroups in our sequences result in amino acid replacements or disrupt predicted cis-regulatory motifs in the AGRIS AtcisDB database (Palaniswamy*, et al.*, 2006) (data not shown). Thus, variation in either the intron or an as-yet-unidentified downstream enhancer region of *TTG1* seems likely to underlie the variation in trichome density observed. Although not a significant result, L*er* native allele transformants had slightly lower trichome densities than transformants with a L*er* allele with a haplogroup B intron spliced, suggesting that this line of evidence is worth following further. Previous studies have used *TTG1* cDNA to complement *ttg1* mutants (Bouyer*, et al.*, 2008), suggesting that the intron is not essential for *TTG1* expression; however, both introns and downstream sequences can contain cis-regulatory elements such as transcription factor binding sites.

4.5.3 *TTG1*, trichome density and pleiotropy in nature

Given the pleiotropic role of *TTG1* shown by functional genetic studies and the potentially deleterious consequences of changes in the epidermal phenotypes which it regulates, an association between *TTG1* variation and trichome density was unexpected. The A and B haplogroups are not distinguished by any coding changes, suggesting that variation in expression of *TTG1* likely underlies the observed difference in trichome densities; a possible expression difference should be investigated in future work. Association of the two haplogroups with variation in other epidermal traits is beyond the scope of this study, but investigation of other epidermal phenotypes in later-generation transgenics could be undertaken to address this possibility directly. For such allelic variation to persist in nature, three alternative explanations are proposed. One possibility is that, given the relatively small effect of *TTG1* variation on trichome density, even if other traits are similarly affected this variation may be effectively neutral. Alternatively, non-trichome phenotypes may be less susceptible than trichome density to variation in *TTG1* expression levels. The *ttg1-10* allele,

which has a single SNP in the 5' UTR and shows significantly reduced expression relative to wildtype *TTG1*, has near-wildtype trichome number with trichome clustering, increased root hair number, reduced anthocyanin production and no seed coat pigmentation, although reportedly normal mucilage production (Larkin*, et al.*, 1999), suggesting not all phenotypes are equally affected by this mutation. Finally, variation distinguishing the haplogroups may be limited to a trichome-specific regulatory element of *TTG1*, circumventing the potentially deleterious consequences of alterations in *TTG1* expression on other phenotypes. Timing and localisation of expression may differ for each trait, with root hair and seed phenotypes requiring expression of *TTG1* in distinct organs and foliar anthocyanins being produced later in development than trichomes.

4.6 Conclusions

The TTG1 protein sequence is highly conserved, with evidence of purifying selection across the coding region. In contrast, variation in *TTG1* nucleotide sequence is extensive, is particularly pronounced in the intron, and contributes to a pattern of allelic dimorphism as has been observed in other genes within the epidermal development pathway. This dimorphism is not reflected in amino acid replacements but is significantly associated with trichome density in association mapping studies, both in natural accessions and an F_2 mapping population, suggesting a role for expression changes. Allelic variation at all levels within the TTG1 epidermal development pathway has now been shown to affect trichome density, supporting the hypothesis that variation within the pathway plays a key role in generating variation in trichome density in nature. Further work should be undertaken to determine whether the molecular variation observed contributes to other epidermal phenotypes regulated by *TTG1* or is limited to an effect on trichome density.

4.7 Acknowledgements

RHB collected all data and performed most analyses and interpretation. Trichome density phenotyping data for the set of 96 accessions used here were collected by RHB in 2008 as part of a BSc (Hons) project. The mixed-model association mapping analysis was performed by Thomas Juenger (University of Texas, Austin). The *ttg1-1* mutant, pDONR-zeo and pBARN-RFB vectors used here were gifted by Alan Lloyd (University of Texas, Austin), and the *Agrobacterium* strain GV3101 was from Xiao Song (Massey University). Funding for this work was provided by a New Zealand Royal Society Marsden Fund grant (09-MAU-114).

4.8 Supplementary Materials

Table 4.4: Accessions used in this study (CS22660 set)

Stock centre #	Name	Geographic origin	Trichome density	*TTG1* haplogroup	AA replacements	Analyses
22564	RRS-7	USA North Liberty, Indiana	13.0	B		Sequence analysis; Association mapping; Transgenics
22565	RRS-10	USA North Liberty, Indiana	12.0	A		Sequence analysis; Association mapping
22566	Knox-10	USA Knox, IN	16.3	A		Sequence analysis; Association mapping
22567	Knox-18	USA Knox, IN	5.7	A		Sequence analysis; Association mapping
22568	Rmx-A02	USA St. Joseph, Michigan	11.7	A		Sequence analysis; Association mapping
22569	Rmx-A180	USA St. Joseph, Michigan	16.7	A		Sequence analysis; Association mapping
22570	Pna-17	USA Benton Harbor, Michigan	21.8	B		Sequence analysis; Association mapping
22571	Pna-10	USA Benton Harbor, Michigan	16.4	A		Sequence analysis; Association mapping
22572	Eden-1	Sweden Eden, N Sweden	13.0	B		Sequence analysis; Association mapping
22573	Eden-2	Sweden Eden, N Sweden	20.3	B		Sequence analysis; Association mapping
22574	Lov-1	Sweden Lovvik, N Sweden	9.3	A		Sequence analysis; Association mapping
22575	Lov-5	Sweden Lovvik, N Sweden	9.7	A		Sequence analysis; Association mapping
22576	Fab-2	Sweden Faberget, N Sweden	14.8	B		Sequence analysis; Association mapping
22577	Fab-4	Sweden Faberget, N Sweden	26.7	B		Sequence analysis; Association mapping
22578	Bil-5	Sweden Billaberget, N Sweden	12.8	B		Sequence analysis; Association mapping
22579	Bil-7	Sweden Billaberget, N Sweden	28.0	B		Sequence analysis; Association mapping
22580	Var2-1	Sweden Varhallarna, S Sweden	16.3	A		Sequence analysis; Association mapping
22581	Var2-6	Sweden Varhallarna, S Sweden	17.0	A	L9I	Sequence analysis; Association mapping
22582	Spr1-2	Sweden Spratteboda, S Sweden	9.3	A	L9I	Sequence analysis; Association mapping
22583	Spr1-6	Sweden Spratteboda, S Sweden	13.5	A	L9I	Sequence analysis; Association mapping

22584	Omo2-1	Sweden Ostra Mocklo, S Sweden	13.8	A		Sequence analysis; Association mapping
22585	Omo2-3	Sweden Ostra Mocklo, S Sweden	9.3	A		Sequence analysis; Association mapping
22586	Ull2-5	Sweden Ullstorp, S Sweden	6.0	A		Sequence analysis; Association mapping
22587	Ull2-3	Sweden Ullstorp, S Sweden	9.0	A		Sequence analysis; Association mapping
22588	Zdr-1	Zdarec, Czech Republic	5.6	A		Sequence analysis; Association mapping
22589	Zdr-6	Zdarec, Czech Republic	4.8	A		Sequence analysis; Association mapping
22590	Bor-1	Borky, Czech Republic	8.3	A		Sequence analysis; Association mapping
22591	Bor-4	Borky, Czech Republic	12.3	A		Sequence analysis; Association mapping; F_2 mapping
22592	Pu2-7	Prudka, Czech Republic	18.0	A		Sequence analysis; Association mapping
22593	Pu2-23	Prudka, Czech Republic	16.8	A		Sequence analysis; Association mapping
22594	Lp2-2	Lipovec, Czech Republic	5.0	A	A311T	Sequence analysis; Association mapping
22595	Lp2-6	Lipovec, Czech Republic	17.5	A		Sequence analysis; Association mapping
22596	HR-5	United Kingdom Ascot, Berks	14.5	A		Sequence analysis; Association mapping
22597	HR-10	United Kingdom Ascot	10.0	A		Sequence analysis; Association mapping
22598	NFA-8	United Kingdom Ascot	9.5	A		Sequence analysis; Association mapping
22599	NFA-10	United Kingdom Ascot	10.3	A		Sequence analysis; Association mapping
22600	Sq-1	United Kingdom Ascot	9.5	A		Sequence analysis; Association mapping
22601	Sq-8	United Kingdom Ascot	14.8	A		Sequence analysis; Association mapping
22602	CIBC-5	United Kingdom Ascot, Berks	11.2	A		Sequence analysis; Association mapping
22603	CIBC-17	United Kingdom Ascot, Berks	14.2	A		Sequence analysis; Association mapping
22604	Tamm-2	Finland Tammisari	12.6	B		Sequence analysis; Association mapping
22605	Tamm-27	Finland Tammisari	17.3	B		Sequence analysis; Association mapping
22606	Kz-1	Kazakhstan Karagandy	12.7	B		Sequence analysis; Association mapping
22607	Kz-9	Kazakhstan Karagandy	16.6	B		Sequence analysis; Association mapping
22608	Got-7	Germany Goettingen		A		Sequence analysis; Association mapping

22609	Got-22	Germany Goettingen	5.5	A	Sequence analysis; Association mapping
22610	Ren-1	France Rennes	19.0	A	Sequence analysis; Association mapping
22611	Ren-11	France Rennes	15.8	A	Sequence analysis; Association mapping
22612	Uod-1	Austria Ottenhof	7.0	A	Sequence analysis; Association mapping
22613	Uod-7	Austria Ottenhof	7.8	A	Sequence analysis; Association mapping
22614	Cvi-0	Cape Verdi	22.8	A	Sequence analysis; Association mapping
22615	Lz-0	France Lezoux	6.0	A	Sequence analysis; Association mapping
22616	Ei-2	Germany Eifel	8.4	A	Sequence analysis; Association mapping
22617	Gu-0	Germany Glueckingen	14.5	A	Sequence analysis; Association mapping
22618	Ler-1	Germany	8.3	A	Sequence analysis; Association mapping; Transgenics
22619	Nd-1	Germany Niederzenz	14.2	A	Sequence analysis; Association mapping
22620	C24		5.2	A	Sequence analysis; Association mapping
22621	CS22491/N13	Russia Konchezero	15.5	A	Sequence analysis; Association mapping
22622	Wei-0	Switzerland Weiningen	14.8	A	Sequence analysis; Association mapping
22623	Ws-0	Russia Wassilewskija	16.2	A	Sequence analysis; Association mapping
22624	Yo-0	USA Yosemite Nat. Park	11.0	A	Sequence analysis; Association mapping
22625	Col-0	USA Columbia	14.5	A	Sequence analysis; Association mapping; Transgenics
22626	An-1	Belgium Antwerpen	6.0	A	Sequence analysis; Association mapping
22627	Van-0	Canada University of British Columbia	10.8	A	Sequence analysis; Association mapping
22628	Br-0	Czech Republic Brunn	0.0	A	Sequence analysis
22629	Est-1	Russia Estland	10.8	A	Sequence analysis; Association mapping
22630	Ag-0	France Argentat	6.0	A	Sequence analysis; Association mapping
22631	Gy-0	France La Miniere	16.4	A	Sequence analysis; Association mapping
22632	Ra-0	France Randan, Puy-de-Dome	10.0	A	Sequence analysis; Association mapping

22633	Bay-0	Germany Bayreuth	10.4	A		Sequence analysis; Association mapping
22634	Ga-0	Germany Gabelstein	9.3	A		Sequence analysis; Association mapping
22635	Mrk-0	Germany Markt/Baden	7.6	A		Sequence analysis; Association mapping
22636	Mz-0	Germany Merzhausen/Ts.	14.0	A		Sequence analysis; Association mapping
22637	Wt-5	Germany Wietze	9.3	A		Sequence analysis; Association mapping
22638	Kas-1	India Kashmir	10.5	B		Sequence analysis; Association mapping
22639	Ct-1	Italy Catania	7.0	A		Sequence analysis; Association mapping
22640	Mr-0	Italy Monte/Tosso	13.0	A		Sequence analysis; Association mapping
22641	Tsu-1	Japan Tsushima	8.5	A	D59N; A333S	Sequence analysis; Association mapping
22642	Mt-0	Libya Martuba/Cyrenaika	12.3	A		Sequence analysis; Association mapping
22643	Nok-3	Netherlands Noordwijk	5.0	A		Sequence analysis; Association mapping
22644	Wa-1	Poland Warsaw	5.6	A		Sequence analysis; Association mapping
22645	Fei-0	Portugal St. Maria d. Feiria	18.8	A		Sequence analysis; Association mapping
22646	Se-0	Spain San Eleno	10.4	A		Sequence analysis; Association mapping
22647	Ts-1	Spain Tossa del Mar	9.8	A		Sequence analysis; Association mapping
22648	Ts-5	Spain Tossa del Mar	8.2	A		Sequence analysis; Association mapping
22649	Pro-0	Spain Proaza, Asturios	17.5	A		Sequence analysis; Association mapping
22650	LL-0	Spain Llagostera	10.3	A		Sequence analysis; Association mapping
22651	Kondara	Tadjikistan Khurmatov	11.8	B		Sequence analysis; Association mapping
22652	Sakhdara	Tadjikistan Pamiro-Alay	20.7	B		Sequence analysis; Association mapping
22653	Sorbo	Tadjikistan	19.5	B		Sequence analysis; Association mapping; F_2 mapping; Transgenics
22654	Kin-0	USA Kindalville, MI	10.0	A		Sequence analysis; Association mapping
22655	Ms-0	Russia	14.8	B		Sequence analysis; Association mapping
22656	Bur-0	Ireland Burren	12.3	A		Sequence analysis; Association mapping
22657	Edi-0	United Kingdom Edinburgh	16.0	A		Sequence analysis; Association mapping

22658	Oy-0	Norway Oystese	10.2	A	Sequence analysis; Association mapping
22659	Ws-2	Russia Wassilewskija	17.7	A	Sequence analysis; Association mapping

Table 4.5: Primers used for PCR, sequencing, transgenic construct development and genotyping; Supplementary Material

Primer	Sequence	Application
TTG1_5pf7_F	CTAGTGATCCAATAATTAGGCC	PCR; Sequencing
TTG1_3pf19_R	ACCAAATTCCCATTGGAG	PCR; Sequencing
TTG1_i12_F	AGTCCTCCTTCTCTCCGTC	PCR; Sequencing
TTG1_5'_3F	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGTATTCAATTGCTACGTGG	PCR; Sequencing; Transgenic construct development (attB tailed)
TTG1_3'_3R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTTTAACCTCTTTGTACTC	PCR; Sequencing; Transgenic construct development (attB tailed)
TTG1_3pf14_R	GGCATACAGCACAGTCAGCAG	Sequencing
TTG1_3pf15_R	ACCTGTCGCTAATAAGAGCC	Sequencing
TTG1_3pf21_R	TTCACATCTGCACCTGAC	Sequencing
TTG1_i22_F	TTTACGAGAGTCCTCAGCC	Sequencing
TTG1_5p5_F	AAACCAGCCCATAGAC	Sequencing
TTG1_5p6_F	TTCGTTATGTCCCTTTCC	Sequencing
TTG1_3'_2R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTAAGTTTTGAGGACGTAGG	Sequencing
TTG1_intron5p_F	GCAGCTCCTTAGAGTTTGAGGTGAGAG	SOE PCR; Transgenic construct development
TTG1_intron5p_R	CTCTCACCTCAAACTCTAAGGAGCTGC	SOE PCR; Transgenic construct development
TTG1_intron3p_F	CAGGTGCAGATGTGAAGTGATCAATAAGG	SOE PCR; Transgenic construct development
TTG1_intron3p_R	CCTTATTGATCACTTCACATCTGCACCTG	SOE PCR; Transgenic construct development
TTG1_Isplice_F	GCTCCTTAGAGTTTGAGGTGCAGATGTGAAGTG	SOE PCR; Transgenic construct development
TTG1_Isplice_R	CACTTCACATCTGCACCTCAAACTCTAAGGAGC	SOE PCR; Transgenic construct development
M13-FAM	GTAAAACGACGGCCAG	F_2 population genotyping (FAM dye labelled)
TTG1_3pf31_M13F	CACGACGTTGTAAAACGACTCTTTGTATGTGTGCTTCAGG	F_2 population genotyping (M13 tailed)
TTG1_3pf30_R	ACCTACTTACGCACATGACG	F_2 population genotyping

5 General Discussion

5.1 General discussion

The epidermis in *Arabidopsis thaliana* mediates a range of plant:environment interactions, from nutrient acquisition and seed germination to biotic and abiotic stress protection, with trichomes playing a specific role in the prevention of damage due to herbivory. A range of epidermal cell fates, including trichome density, share a common and genetically well-characterised regulatory pathway in *A. thaliana*. The WDR protein TTG1 is absolutely required for all cell fates, acting together with several functionally overlapping bHLHs and trait-specific R2R3 MYBs to form a transcriptional activation complex targeting downstream genes. Trichome density and the epidermal development pathway are used here to address a range of questions regarding the genetic architecture and molecular genetic basis of natural phenotypic variation, utilising both QTL mapping and candidate gene approaches.

5.1.1 The genetic architecture of trichome density and plasticity

Constitutive trichome density, induced density and plasticity of trichome density in response to wounding were mapped here in two newly characterised mapping populations, Hi-0 x Ob-0 and St-0 x Sf-2. Multiple loci underlie quantitative variation in trichome density with between three and five QTL mapped each for constitutive and induced trichome density in both Hi-0 x Ob-0 and St-0 x Sf-2; this is broadly in line with previous mapping results in other populations (Larkin*, et al.*, 1996; Mauricio, 2005; Symonds*, et al.*, 2005; Pfalz*, et al.*, 2007; Atwell*, et al.*, 2010). A number of loci mapped appeared common to both constitutive and induced trichome density, including loci with downstream targets of the TTG1:bHLH:GL1 complex, *GL2* and *TTG2*, as candidates. Some loci mapped also appear common to multiple populations, both here and in previous studies (Larkin*, et al.*, 1996; Mauricio, 2005; Symonds*, et al.*, 2005; Pfalz*, et al.*, 2007; Atwell*, et al.*, 2010). As with QTLs overlapping for constitutive and induced density within populations, candidate genes for several of those overlapping across populations are genes within the epidermal development pathway, including *TT8*, *GL1* and the R3 MYB repressors *ETC2*, *TCL1* and *TCL2*. Care needs to be taken in interpreting such results as QTL which appear to overlap can certainly represent distinct loci. However, this result suggests that some loci may commonly be responsible for trichome density variation among natural accessions, playing a key role in the generation of such variation.

Although trichome phenotypes have been mapped previously, this work is the first to map QTLs for trichome density following damage and for trichome plasticity. These are possibly

more meaningful traits in nature, allowing plants to avoid the costs associated with high trichome density unless herbivory is a significant pressure. Although it was not possible to calculate heritability, plasticity of trichome density in response to wounding appears to be at least partially under genetic control with one locus mapped for response in the St-0 x Sf-2 population. This locus does not overlap with any of the loci mapped for constitutive or induced density; as such, it may represent a regulatory locus which controls plasticity via regulation of expression of genes with direct effects on trichome density (Schlichting & Pigliucci, 1993). The negative response to wounding of some high density RILs is particularly interesting and does not appear to have been reported previously.

5.1.2 Qualitative and quantitative changes in trichome density

A broad range of densities are observed in natural accessions, and a comparison of quantitative changes (density) with qualitative changes (glabrousness) in phenotype identifies both similarities and differences in the underlying molecular basis of such variation. Quantitative variation in trichome density is polygenic, with between three and five loci identified underlying variation in trichome density phenotypes in the QTL mapping populations examined here. Multiple loci for trichome phenotypes have been found in all other QTL mapping populations and genome-wide association studies (GWAS) (Larkin*, et al.*, 1996; Mauricio, 2005; Symonds*, et al.*, 2005; Pfalz*, et al.*, 2007; Atwell*, et al.*, 2010). Candidate gene analysis so far has identified four genes with natural allelic variants contributing to quantitative changes in trichome density, all from within the epidermal development pathway. The specific molecular changes underlying the quantitative shift in phenotype vary: in *TTG1* a regulatory mutation is likely causal; in *GL1*, both regulatory and protein coding changes are plausible candidates; in *ETC2*, an amino acid replacement is associated with trichome density variation (Hilscher*, et al.*, 2009); while in *AtMYC1* a single amino acid replacement generates a loss-of-function allele (Symonds*, et al.*, 2011), reducing density but not resulting in glabrousness due to functional overlap with *GL3* and *EGL3*. In addition, in both *GL1* and *TTG1* other amino acid replacements have been identified as plausible candidates for an effect on trichome density in one or few accessions, although this has not been demonstrated.

Conversely, the qualitative change in trichome density in the seven glabrous accessions of *A. thaliana* surveyed here appears to be a result of four independent loss-of-function mutations

in the same single locus, *GL1*. All polymorphisms identified as causing glabrousness in *A. thaliana* disrupt protein function (Hauser*, et al.*, 2001; Bloomer*, et al.*, 2012). A whole-gene deletion, two frameshift mutations and an amino acid replacement at a key bHLH-binding residue are proposed causal mutations. With the exception of the whole locus deletion, all *GL1* alleles were shown to be expressed; however, a change in timing or localisation of *GL1* expression was not ruled out by this work. While a loss-of-function allele at *TTG1* would similarly result in glabrousness, no such natural mutations were identified in this study; the pleiotropic nature of *TTG1* increases the likelihood that loss-of-function mutations would be deleterious, presumably explaining this observation.

5.1.3 Pleiotropy and phenotypic variation

Pleiotropic genes within regulatory pathways are likely to be subject to stronger constraint than genes affecting few or single traits, potentially limiting their contribution to natural variation in phenotypes (Stern & Orgogozo, 2009; Martin*, et al.*, 2010). Variation within the epidermal development pathway suggests an important role for regulatory genes as a source of phenotypic variation, although the prediction that pleiotropic genes will contribute less to the generation of such variation appears to only partially hold. TTG1, which affects all epidermal cell fates regulated by the pathway, is more constrained in protein sequence than either GL1 or AtMYC1 (Symonds*, et al.*, 2011) as determined by *K*a/*K*s ratio, in agreement with previous findings in *Ipomoea* (Streisfeld*, et al.*, 2011). Despite this, variation in *TTG1* plays a surprising role in contributing to quantitative differences in trichome density. With no nonsynonymous polymorphisms distinguishing the two *TTG1* haplogroups which show an association with density, the contribution of *TTG1* is likely based on expression changes. The impact of variation in *TTG1* on epidermal traits besides trichome density has not been examined here, but if expression changes are responsible the effect could potentially be limited to trichome phenotypes, for example if molecular variation is limited to a trichome-specific enhancer region.

In contrast to *TTG1*, *GL1* has a highly specific role in trichome initiation. While amino acid sequence is highly conserved in TTG1, the quantitative effect of *GL1* on trichome density may be due to amino acid replacements, with three such replacements distinguishing the *GL1* haplogroups which associate with trichome density variation. In addition, a number of replacements identified in *GL1* in different accessions are suggested to have quantitative

effects on trichome density through disruption of protein:protein interactions or protein folding. All naturally glabrous accessions surveyed here appear to be a result of mutations in *GL1*, perhaps because loss-of-function mutations in *TTG1* have potentially deleterious consequences due to the number of phenotypes affected.

Mutations in the MYB component of the WDR:bHLH:MYB complex underlie natural phenotypic variation in a range of species. Glabrousness in close relatives of *A. thaliana*, *A. lyrata* (Kivimaki*, et al.*, 2007) and *A. halleri ssp. gemmifera* (Kawagoe*, et al.*, 2011), has been attributed to coding region mutations in *GL1*. In *Petunia*, at least five independent loss-of-function mutations in the MYB *AN2* underlie the white-flowered phenotype of *P. axillaris* (Quattrocchio*, et al.*, 1999; Hoballah*, et al.*, 2007). Similarly, in both *Antirrhinum* (Schwinn*, et al.*, 2006) and *Mimulus* (Lowry*, et al.*, 2012), variation in floral patterning phenotypes has been attributed to variation in MYB genes. These less pleiotropic genes within the pathway appear to be a common source of natural phenotypic variation.

5.1.4 Molecular evolution in the TTG1 pathway

Varying selection pressures influence trichome density in *A. thaliana*; trichomes provide a fitness benefit under conditions of insect herbivory (Mauricio, 1998), but impose costs including reduced growth rate (Zust*, et al.*, 2011) and increased susceptibility to fungal pathogens (Calo*, et al.*, 2006) in the absence of herbivores. Both *GL1* and *TTG1* show evidence of selection acting on protein coding sequence. *K*a/*K*s ratios of much less than one, indicating purifying selection, are found across the entire *TTG1* coding region and localised to the R2R3 MYB and C-terminal transcriptional activation domain of *GL1*. Overall strong conservation of the coding sequence of *GL1* and *TTG1* likely reflects the need to maintain interactions with other members of the TTG1:bHLH:GL1 complex and with target regulatory sequences. Two spikes of *K*a/*K*s above one in the *GL1* sliding window analysis of *K*a/*K*s indicate that the amino acid replacements reflecting the pattern of high frequency polymorphism in *GL1* may be under positive selection. Despite the high level of conservation of *TTG1* coding sequence, Tajima's D test results are not significantly different from neutral expectations for either *GL1* or *TTG1.* However, the less conservative Fu and Li's D* and F*tests are significantly negative for both *GL1* and for each of the two *TTG1* haplogroups analysed separately, suggesting purifying selection at both loci.

5.1.5 Complementary approaches: Functional genetic studies and gene effects in nature

For both *GL1* and *TTG1*, functional genetic studies indicate that a glabrous phenotype is associated with loss-of-function mutations (Oppenheimer*, et al.*, 1991; Larkin*, et al.*, 1999; Walker*, et al.*, 1999). In addition, weak alleles which reduce trichome production relative to wild-type have been described for both genes. In *GL1*, the weak allele identified is a truncation of the terminal 27 amino acids (Esch*, et al.*, 1994), while in *TTG1* weak alleles have been attributed to both reduced expression and to amino acid replacements (Larkin*, et al.*, 1999). Despite a limited range of genetic backgrounds sampled, functional genetic analyses previously undertaken for both *TTG1* and *GL1* serve as a good indicator of the role of these genes in nature.

Previous characterisations of *TTG1* and *GL1* (and other MYBs within the pathway) have identified key functional domains and conserved amino acid signatures involved in protein:protein interactions and DNA binding. In addition, in *GL1* a downstream enhancer region with a role in correct localisation of *GL1* expression has been described. These studies, described in more detail in the relevant chapters, were drawn on here to identify polymorphisms with potentially significant impacts on *GL1* and *TTG1* function or expression and thus on trichome density. Conversely, surveys of natural variation in genes and in epidermal phenotypes, both here and in previous studies, provide a rich source of mutations and phenotypes in a wide range of genetic backgrounds and have improved our understanding of how genes within the pathway function. QTL mapping and GWAS (here; (Larkin*, et al.*, 1996; Mauricio, 2005; Symonds*, et al.*, 2005; Pfalz*, et al.*, 2007; Atwell*, et al.*, 2010)) have identified loci and interactions influencing trichome density, some with no obvious candidate genes. These loci suggest targets for follow-up work, potentially leading to the identification of new genes or an enhanced understanding of genes already described. In addition, new amino acids of possible structural/functional importance and potential sources of cis-regulatory variation are highlighted for further investigation. Here, examples include the P176R and high-frequency replacement polymorphisms in GL1 or the A311T mutation in TTG1, and the intron and downstream sequences of *TTG1*. Previous studies have similarly identified key amino acids in epidermal pathway genes (Hilscher*, et al.*, 2009; Symonds*, et al.*, 2011), as well as ascribed functions to genes not previously characterised (*AtMYC1*; (Symonds*, et al.*, 2011)). Thus, functional genetics and studies of natural variation are complementary approaches to understanding gene functions, interactions, and the resulting consequences for phenotypes.

5.2 Future Directions

5.2.1 The genetic architecture of trichome density and plasticity

As QTL mapping in two-parent populations limits the range of genetic variation sampled, the work presented here is being expanded on in an 19-parent intercrossed mapping population (data not presented here), known as the MAGIC lines (Kover*, et al.*, 2009). Preliminary analysis of trichome density and plasticity phenotypes in the MAGIC lines has identified both previously mapped and novel loci. The plasticity of trichome density in the MAGIC lines is genetically controlled and polygenic with four significant QTL identified for this trait, including one which appears to overlap with the locus for plasticity mapped in St-0 x Sf-2. High levels of recombination, the large number of SNP markers used in genotyping this set, and the availability of genome sequences for all parents mean strong candidate genes for many QTL should be identified. Ultimately trichome density traits mapped in the MAGIC lines will be integrated with data on other epidermal traits mapped in this set, providing new insight into the complex and interacting nature of their genetic control.

Of interest in the results presented here was the negative relationship between constitutive and induced trichome density observed in both mapping populations, and in particular the negative response to wounding of some high-density RILs which does not appear to have been reported previously. It would be informative to look at natural variation in density and plasticity in a large set of natural accessions (as has been done here for constitutive trichome density) to determine whether this is a general trend or one specific to a limited range of genetic backgrounds.

5.2.2 Mutation in *GL1*: Quantitative and qualitative effects

Allelic dimorphism in *GL1* is associated with natural variation in trichome density. Possible causal polymorphisms reflecting this pattern include three amino acid replacements, two of which are within the functionally important C-terminal transcriptional activation domain, and SNPs and indels in an enhancer region located approximately 1kb downstream of *GL1*. Native *GL1* sequences of both haplogroups A and B from natural accessions and spliced gene constructs, combining haplogroup A coding sequence with haplogroup B downstream sequence and vice versa, have been developed and will be transformed into a *gl1* mutant background to test whether coding or enhancer region variation is responsible for the

observed difference in density. This approach will also be taken to investigate the roles of two amino acid replacements, L93W and P176R, which seem strong candidates for reducing trichome density; reciprocal site-directed mutagenesis of these residues in each allele has been completed, and these constructs, along with native *GL1* alleles, will be analysed using transgenics.

Preliminary analysis of F_1 hybrids between Col-0 (with a functional copy of *GL1*) and glabrous natural accessions carrying loss-of-function mutations in *GL1* indicates that these crosses, while trichome-producing, have reduced trichome density compared to plants homozygous for functional *GL1*. Two of the loss-of-function alleles, the whole-gene deletion and the amino acid replacement (S92F), suggest a simple dosage effect, with F_1 phenotypes of about half the trichome density of Col-0. The two loss-of-function frameshift mutations generate truncated proteins that may retain capacity to bind DNA or to interact with bHLHs; F_1 phenotypes suggest a repression effect for these two alleles, with lower trichome density observed than in crosses with gene-loss or S92F mutations. These observations will be confirmed by analysis of phenotypes in F_2 populations and transgenic plants. Yeast-two-hybrid experiments will be used to investigate the consequences of the mutations for bHLH interaction. This project should provide some insight into the role, evolution and maintenance of these alleles in *A. thaliana* specifically, and of MYB repressors more generally.

5.2.3 Variation in *TTG1* impacts trichome density

Analyses presented here show that a pattern of allelic dimorphism in *TTG1* impacts trichome density phenotypes in natural accessions. Prior to publication of this work, further experiments will be performed. These will include analysis of trichome density in two additional F_2 mapping populations derived from *TTG1* haplogroup A and B parental accessions (currently in progress), and re-scoring of trichome density phenotypes in additional transformants for some constructs and in later (~T_3) generations, where phenotypes are expected to be more stable and replicates of each transformant can be used. Given the extremely low trichome density of the L*er* background used here, ideally another *ttg1* mutant in an accession with a higher density would also be transformed, although this is not currently feasible; while TDNA insertions near *TTG1* are available in Col-0, none results in a *ttg1* mutant phenotype. In these later-generation transgenics, it would be of considerable interest to score some or all of the other epidermal phenotypes regulated by *TTG1* in order to

determine whether the effect documented here is limited to trichome density. Fine mapping could also be considered to better localise the causal mutations underlying the trichome density variation observed between haplogroups A and B.

5.3 Conclusion

5.3.1 The epidermal development pathway plays a key role in natural variation in trichome density

The overarching hypothesis underlying the research presented here is that variation in genes within the epidermal development pathway contributes to variation in trichome density in *Arabidopsis thaliana*. Quantitative trait locus (QTL) mapping and candidate gene analyses were undertaken to investigate this, with both approaches finding support for the hypothesis.

Previous QTL mapping and Genome Wide Association Studies (GWAS) have identified a number of candidate genes from within the epidermal development pathway as candidate genes for loci mapped for trichome density and number in natural accessions. Here, loci for constitutive and induced trichome density were mapped in proximity to a number of genes within the pathway, including near a tandem array of R3 repressors, *ETC2*, *TCL1* and *TCL2*; the bHLH *TT8*; and the trichome-specific R2R3 MYB *GL1*. Further candidates identified here include both the upstream hormonal response regulators, *RGL1*, *JAZ2* and *JAZ9*, and downstream targets of the TTG1:bHLH:GL1 complex, *GL2*, *TTG2* and *At1G77670*.

The candidate gene analyses described here demonstrate roles for both *TTG1* and *GL1* in determining trichome density, confirming proposed roles as candidate genes in QTL mapping and GWAS (here; Symonds, *et al.*, 2005; Atwell, *et al.*, 2010). In both genes a pattern of high frequency polymorphism, proposed to have arisen during the Pleistocene glaciation, associates with variation in trichome density in natural accessions. In *TTG1* this association is presumably based on expression variation as no nonsynonymous SNPs reflect the associated pattern of high frequency polymorphism. In *GL1* it is unclear whether this association is based on changes in expression of *GL1* or of downstream target genes; both amino acid replacements within the transcriptional activation domain of *GL1* and variation in a downstream enhancer region with a role in the localisation of *GL1* expression correspond to

the pattern observed. While the amino acid sequence of TTG1 is more conserved than in GL1, both genes' coding sequences appear to be subject to purifying selection. However, amino acid replacements with possible functional effects have been identified affecting one or few accessions in both genes. With candidate gene studies indicating a role for both the bHLH *AtMYC1* (Symonds*, et al.*, 2011) and the R3 repressor *ETC2* (Hilscher*, et al.*, 2009) in determining trichome density in nature, allelic variation affecting density has now been identified at all levels within the epidermal development pathway. This research demonstrates a key role for the TTG1:bHLH:GL1 complex and its up- and downstream genes in regulating trichome density, and more generally for allelic variation in regulatory genes in generating phenotypic variation within species.

REFERENCES

Aguade, M. (2001). Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the FAH1 and F3H genes, in *Arabidopsis thaliana*. *Molecular Biology and Evolution, 18*(1), 1-9.

Alonso-Blanco, C., Aarts, M. G. M., Bentsink, L., Keurentjes, J. J. B., Reymond, M., Vreugdenhil, D., et al. (2009). What has natural variation taught us about plant development, physiology, and adaptation? [Review]. *Plant Cell, 21*(7), 1877-1896.

Alonso-Blanco, C., Bentsink, L., Hanhart, C. J., Vries, H. B. E., & Koornneef, M. (2003). Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. *Genetics, 164*(2), 711-729.

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature, 408*(6814), 796-815.

Arunyawat, U., Stephan, W., & Stadler, T. (2007). Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. *Molecular Biology and Evolution, 24*(10), 2310-2322.

Atanassova, B., Shtereva, L., Georgieva, Y., & Balatcheva, E. (2004). Study on seed coat morphology and histochemistry in three anthocyaninless mutants in tomato (*Lycopersicon esculentum* Mill.) in relation to their enhanced germination. *Seed Science and Technology, 32*(1), 79-90.

Atwell, S., Huang, Y. S., Vilhjalmsson, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature, 465*(7298), 627-631.

Bakker, E. G., Toomajian, C., Kreitman, M., & Bergelson, J. (2006). A genome-wide survey of *R* gene polymorphisms in *Arabidopsis*. *Plant Cell, 18*(8), 1803-1818.

Balasubramanian, S., Schwartz, C., Singh, A., Warthmann, N., Kim, M. C., Maloof, J. N., et al. (2009). QTL mapping in new *Arabidopsis thaliana* Advanced Intercross-Recombinant Inbred Lines. *Plos One, 4*(2).

Balkunde, R., Bouyer, D., & Hulskamp, M. (2011). Nuclear trapping by GL3 controls intercellular transport and redistribution of TTG1 protein in *Arabidopsis*. *Development, 138*(22), 5039-5048.

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution, 16*(1), 37-48.

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution, 23*(1), 38-44.

Baudry, A., Caboche, M., & Lepiniec, L. (2006). *TT8* controls its own expression in a feedback regulation involving *TTG1* and homologous *MYB* and *bHLH* factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*. *Plant Journal, 46*(5), 768-779.

Baudry, A., Heim, M. A., Dubreucq, B., Caboche, M., Weisshaar, B., & Lepiniec, L. (2004). *TT2*, *TT8*, and *TTG1* synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant Journal, 39*(3), 366-380.

Beck, J. B., Schmuths, H., & Schaal, B. A. (2008). Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. *Molecular Ecology, 17*(3), 902-915.

Beilstein, M. A., Al-Shehbaz, I. A., Mathews, S., & Kellogg, E. A. (2008). Brassicaceae phylogeny inferred from Phytochrome A znd NDHF sequence data: Tribes and trichomes revisited. *American Journal of Botany, 95*(10), 1307-1327.

Bergelson, J., & Roux, F. (2010). Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. [Review]. *Nature Reviews Genetics, 11*(12), 867-879.

Berger, F., Linstead, P., Dolan, L., & Haseloff, J. (1998). Stomata patterning on the hypocotyl of *Arabidopsis thaliana* is controlled by genes involved in the control of root epidermis patterning. *Developmental Biology, 194*(2), 226-234.

Bikard, D., Patel, D., Le Mette, C., Giorgi, V., Camilleri, C., Bennett, M. J., et al. (2009). Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science, 323*(5914), 623-626.

Bloomer, R. H., Juenger, T. E., & Symonds, V. V. (2012). Natural variation in *GL1* and its effects on trichome density in *Arabidopsis thaliana*. *Molecular Ecology, 21*(14), 3501-3515.

Bouyer, D., Geier, F., Kragler, F., Schnittger, A., Pesch, M., Wester, K., et al. (2008). Two-dimensional patterning by a trapping/depletion mechanism: The role of *TTG1* and *GL3* in *Arabidopsis* trichome formation. *Plos Biology, 6*(6), 1166-1177.

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., et al. (2010). Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *Plos Genetics, 6*(5), 17.

Brock, M. T., Kover, P. X., & Weinig, C. (2012). Natural variation in *GA1* associates with floral morphology in *Arabidopsis thaliana*. *New Phytologist, 195*(1), 58-70.

Brock, M. T., Tiffin, P., & Weinig, C. (2007). Sequence diversity and haplotype associations with phenotypic responses to crowding: *GIGANTEA* affects fruit set in *Arabidopsis thaliana*. *Molecular Ecology, 16*(14), 3050-3062.

Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics, 19*(7), 889-890.

Calo, L., Garcia, I., Gotor, C., & Romero, L. C. (2006). Leaf hairs influence phytopathogenic fungus infection and confer an increased resistance when expressing a Trichoderma alpha-1,3-glucanase. *Journal of Experimental Botany, 57*(14), 3911-3920.

Carey, C. C., Strahle, J. T., Selinger, D. A., & Chandler, V. L. (2004). Mutations in the pale aleurone color1 regulatory gene of the *Zea mays* anthocyanin pathway have distinct phenotypes relative to the functionally similar *TRANSPARENT TESTA GLABRA1* gene in *Arabidopsis thaliana*. *Plant Cell, 16*(2), 450-464.

Carroll, S. B. (2005). Evolution at two levels: On genes and form. *PLoS Biol, 3*(7), e245.

Castric, V., & Vekemans, X. (2007). Evolution under strong balancing selection: how many codons determine specificity at the female self-incompatibility gene SRK in Brassicaceae? *Bmc Evolutionary Biology, 7*.

Chalker-Scott, L. (1999). Environmental significance of anthocyanins in plant stress responses. [Review]. *Photochemistry and Photobiology, 70*(1), 1-9.

Choinski, J. S., & Wise, R. R. (1999). Leaf growth and development in relation to gas exchange in *Quercus marilandica* Muenchh. *Journal of Plant Physiology, 154*(3), 302-309.

Clark, R. M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., et al. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science, 317*(5836), 338-342.

Clough, S. J., & Bent, A. F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal, 16*(6), 735-743.

Cyr, H., & Pace, M. L. (1993). Magnitude and patterns of herbivory in aquatic and terrestrial ecosystems. *Nature, 361*(6408), 148-150.

Debeaujon, I., Leon-Kloosterziel, K. M., & Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Physiology, 122*(2), 403-413.

Desfeux, C., Clough, S. J., & Bent, A. F. (2000). Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the *Arabidopsis* floral-dip method. *Plant Physiology, 123*(3), 895-904.

Digiuni, S., Schellmann, S., Geier, F., Greese, B., Pesch, M., Wester, K., et al. (2008). A competitive complex formation mechanism underlies trichome patterning on *Arabidopsis* leaves. *Molecular Systems Biology, 4*.

Doebley, J., & Lukens, L. (1998). Transcriptional regulators and the evolution of plant form. *The Plant Cell Online, 10*(7), 1075-1082.

Doyle, J. J., & Doyle, L. J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin, 19*, 11-15.

Dubos, C., Le Gourrierec, J., Baudry, A., Huep, G., Lanet, E., Debeaujon, I., et al. (2008). *MYBL2* is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. *Plant Journal, 55*(6), 940-953.

El-Assal, S. E. D., Alonso-Blanco, C., Peeters, A. J. M., Raz, V., & Koornneef, M. (2001). A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nature Genetics, 29*(4), 435-440.

Ennajeh, M., Vadel, A. M., Khemira, H., Ben Mimoun, M., & Hellali, R. (2006). Defense mechanisms against water deficit in two olive (*Olea europaea* L.) cultivars 'Meski' and 'Chemlali'. *Journal of Horticultural Science & Biotechnology, 81*(1), 99-104.

Esch, J. J., Oppenheimer, D. G., & Marks, M. D. (1994). Characterization of a weak allele of the *GL1* gene of *Arabidopsis-thaliana*. *Plant Molecular Biology, 24*(1), 203-207.

Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., & Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant Journal, 49*(3), 414-427.

Feyissa, D. N., Lovdal, T., Olsen, K. M., Slimestad, R., & Lillo, C. (2009). The endogenous *GL3*, but not *EGL3*, gene is necessary for anthocyanin accumulation as induced by nitrogen depletion in *Arabidopsis* rosette stage leaves. *Planta, 230*(4), 747-754.

Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics, 133*(3), 693-709.

Gan, L. J., Xia, K., Chen, J. G., & Wang, S. C. (2011). Functional characterization of *TRICHOMELESS2*, a new single-repeat R3 MYB transcription factor in the regulation of trichome patterning in *Arabidopsis*. *BMC Plant Biology, 11*.

Gan, Y., Yu, H., Peng, J., & Broun, P. (2007). Genetic and molecular regulation by DELLA proteins of trichome development in *Arabidopsis*. *Plant Physiology, 145*(3), 1031-1042.

Gao, Y., Gong, X. M., Cao, W. H., Zhao, J. F., Fu, L. Q., Wang, X. C., et al. (2008). *SAD2* in *Arabidopsis* functions in trichome initiation through mediating *GL3* function and regulating *GL1*, *TTG1* and *GL2* expression. *Journal of Integrative Plant Biology, 50*(7), 906-917.

Gazzani, S., Gendall, A. R., Lister, C., & Dean, C. (2003). Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiology, 132*(2), 1107-1114.

Gilroy, S., & Jones, D. L. (2000). Through form to function: root hair development and nutrient uptake. *Trends in Plant Science, 5*(2), 56-60.

Glover, B. J., & Martin, C. (1998). The role of petal cell shape and pigmentation in pollination success in *Antirrhinum majus*. *Heredity, 80*, 778-784.

Gonzalez, A., Mendenhall, J., Huo, Y., & Lloyd, A. (2009). *TTG1* complex MYBs, *MYB5* and *TT2*, control outer seed coat differentiation. *Developmental Biology, 325*(2), 412-421.

Gonzalez, A., Zhao, M., Leavitt, J. M., & Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant Journal, 53*(5), 814-827.

Grotewold, E., Sainz, M. B., Tagliani, L., Hernandez, J. M., Bowen, B., & Chandler, V. L. (2000). Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proceedings of the National Academy of Sciences of the United States of America, 97*(25), 13579-13584.

Hagenblad, J., Tang, C. L., Molitor, J., Werner, J., Zhao, K., Zheng, H. G., et al. (2004). Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. *Genetics, 168*(3), 1627-1638.

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series, 41*, 95-98.

Handley, R., Ekbom, B., & Agren, J. (2005). Variation in trichome density and resistance against a specialist insect herbivore in natural populations of *Arabidopsis thaliana*. *Ecological Entomology, 30*(3), 284-292.

Hauser, M. T., Harr, B., & Schlotterer, C. (2001). Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: Molecular analysis of the candidate gene *GLABROUS1*. *Molecular Biology and Evolution, 18*(9), 1754-1763.

Hilscher, J., Schlotterer, C., & Hauser, M. T. (2009). A single amino acid replacement in *ETC2* shapes trichome patterning in natural *Arabidopsis* populations. *Current Biology, 19*(20), 1747-1751.

Hoballah, M. E., Gubitz, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., et al. (2007). Single gene-mediated shift in pollinator attraction in *Petunia*. *Plant Cell, 19*(3), 779-790.

Hoekstra, H. E., & Coyne, J. A. (2007). The locus of evolution: Evo devo and the genetics of adaptation. *Evolution, 61*(5), 995-1016.

Hogan, K. P., Smith, A. P., Araus, J. L., & Saavedra, A. (1994). Ecotypic differentiation of gas exchange responses and leaf anatomy in a tropical forest understorey shrub from areas of contrasting rainfall regimes. *Tree Physiology, 14*(7-9), 819-831.

Holeski, L. M., Chase-Alone, R., & Kelly, J. K. (2010). The genetics of phenotypic plasticity in plant defense: Trichome production in *Mimulus guttatus*. *American Naturalist, 175*(4), 391-400.

Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., & Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes - gene-splicing by overlap extension. *Gene, 77*(1), 61-68.

Huang, X. Q., Paulo, M. J., Boer, M., Effgen, S., Keizer, P., Koornneef, M., et al. (2011). Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population. *Proceedings of the National Academy of Sciences of the United States of America, 108*(11), 4488-4493.

Hudson, R. R., Kreitman, M., & Aguade, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics, 116*(1), 153-159.

Innan, H., Tajima, F., Terauchi, R., & Miyashita, N. T. (1996). Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana*. *Genetics, 143*(4), 1761-1770.

Ishida, T., Hattori, S., Sano, R., Inoue, K., Shirano, Y., Hayashi, H., et al. (2007). *Arabidopsis TRANSPARENT TESTA GLABRA2* is directly regulated by R2R3 MYB transcription factors and is involved in regulation of *GLABRA2* transcription in epidermal differentiation. *Plant Cell, 19*(8), 2531-2543.

Juenger, T. E., Sen, S., Stowe, K. A., & Simms, E. L. (2005). Epistasis and genotype-environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica, 123*(1-2), 87-105.

Karabourniotis, G., Kotsabassidis, D., & Manetas, Y. (1995). Trichome density and its protective potential against ultraviolet-B radiation damage during leaf development. *Canadian Journal of Botany-Revue Canadienne De Botanique, 73*(3), 376-383.

Karkkainen, K., Loe, G., & Agren, J. (2004). Population structure in *Arabidopsis lyrata*: Evidence for divergent selection on trichome production. *Evolution, 58*(12), 2831-2836.

Kawabe, A., Innan, H., Terauchi, R., & Miyashita, N. T. (1997). Nucleotide polymorphism in the acidic chitinase focus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. *Molecular Biology and Evolution, 14*(12), 1303-1315.

Kawagoe, T., Shimizu, K. K., Kakutani, T., & Kudoh, H. (2011). Coexistence of trichome variation in a natural plant population: A combined study using ecological and candidate gene approaches. *Plos One, 6*(7).

Kirik, V., Lee, M. M., Wester, K., Herrmann, U., Zheng, Z. G., Oppenheimer, D., et al. (2005). Functional diversification of *MYB23* and *GL1* genes in trichome morphogenesis and initiation. *Development, 132*(7), 1477-1485.

Kivimaki, M., Karkkainen, K., Gaudeul, M., Loe, G., & Agren, J. (2007). Gene, phenotype and function: *GLABROUS1* and resistance to herbivory in natural populations of *Arabidopsis lyrata*. *Molecular Ecology, 16*(2), 453-462.

Kliebenstein, D. J., Figuth, A., & Mitchell-Olds, T. (2002). Genetic architecture of plastic methyl jasmonate responses in *Arabidopsis thaliana*. *Genetics, 161*(4), 1685-1696.

Kobayashi, Y., & Koyama, H. (2002). QTL analysis of Al tolerance in recombinant inbred lines of *Arabidopsis thaliana*. *Plant and Cell Physiology, 43*(12), 1526-1533.

Koornneef, M. (1981). The complex syndrome of *TTG* mutants. *Arabidopsis Information Service, 18*.

Kover, P. X., & Mott, R. (2012). Mapping the genetic basis of ecologically and evolutionarily relevant traits in *Arabidopsis thaliana*. *Current Opinion in Plant Biology, 15*(2), 212-217.

Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., et al. (2009). A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in *Arabidopsis thaliana*. *Plos Genetics, 5*(7).

Kranz, H. D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R. C., et al. (1998). Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*. *Plant Journal, 16*(2), 263-276.

Kuittinen, H., Niittyvuopio, A., Rinne, P., & Savolainen, O. (2008). Natural variation in *Arabidopsis lyrata* vernalization requirement conferred by a *FRIGIDA* indel polymorphism. *Molecular Biology and Evolution, 25*(2), 319-329.

Kurata, T., Ishida, T., Kawabata-Awai, C., Noguchi, M., Hattori, S., Sano, R., et al. (2005). Cell-to-cell movement of the CAPRICE protein in *Arabidopsis* root epidermal cell differentiation. *Development, 132*(24), 5387-5398.

Lai, A., Cianciolo, V., Chiavarini, S., & Sonnino, A. (2000). Effects of glandular trichomes on the development of *Phytophthora infestans* infection in potato (S-tuberosum). *Euphytica, 114*(3), 165-174.

Lai, A. G., Denton-Giles, M., Mueller-Roeber, B., Schippers, J. H. M., & Dijkwel, P. P. (2011). Positional information resolves structural variations and uncovers an evolutionarily divergent genetic locus in accessions of *Arabidopsis thaliana*. *Genome Biology and Evolution, 3*, 627-640.

Lander, E. S., & Schork, N. J. (1994). Genetic dissection of complex traits. *Science, 265*(5181), 2037-2048.

Larkin, J. C., Oppenheimer, D. G., Lloyd, A. M., Paparozzi, E. T., & Marks, M. D. (1994). Roles of the *GLABROUS1* and *TRANSPARENT TESTA GLABRA* genes in *Arabidopsis* trichome development. *Plant Cell, 6*(8), 1065-1076.

Larkin, J. C., Oppenheimer, D. G., Pollock, S., & Marks, M. D. (1993). *Arabidopsis GLABROUS1* gene requires downstream sequences for function. *Plant Cell, 5*(12), 1739-1748.

Larkin, J. C., Walker, J. D., Bolognesi-Winfield, A. C., Gray, J. C., & Walker, A. R. (1999). Allele-specific interactions between *ttg* and *gl1* during trichome development in *Arabidopsis thaliana*. *Genetics, 151*(4), 1591-1604.

Larkin, J. C., Young, N., Prigge, M., & Marks, M. D. (1996). The control of trichome spacing and number in *Arabidopsis*. *Development, 122*(3), 997-1005.

Lee, M. M., & Schiefelbein, J. (1999). *WEREWOLF*, a MYB-related protein in arabidopsis, is a position-dependent regulator of epidermal cell patterning. *Cell, 99*(5), 473-483.

Lee, M. M., & Schiefelbein, J. (2001). Developmentally distinct MYB genes encode functionally equivalent proteins in *Arabidopsis*. *Development, 128*(9), 1539-1546.

Li, S. F., Milliken, O. N., Pham, H., Seyit, R., Napoli, R., Preston, J., et al. (2009). The *Arabidopsis MYB5* transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *Plant Cell, 21*(1), 72-89.

Liakopoulos, G., Nikolopoulos, D., Klouvatou, A., Vekkos, K. A., Manetas, Y., & Karabourniotis, G. (2006). The photoprotective role of epidermal anthocyanins and surface pubescence in young leaves of grapevine (*Vitis vinifera*). *Annals of Botany, 98*(1), 257-265.

Littell, R. C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D. and O. Schabenberger (2006). *SAS for mixed models* (Second ed.). Cary: SAS Institute Inc.

Lowry, D. B., Sheng, C. C., Lasky, J. R., & Willis, J. H. (2012). Five anthocyanin polymorphisms are associated with an R2R3-MYB cluster in *Mimulus Guttatus* (Phrymaceae). *American Journal of Botany, 99*(1), 82-91.

Lu, J., Li, J. N., Lei, B., Wang, S. G., & Chai, Y. R. (2009). Molecular cloning and characterization of two *Brassica napus TTG1* genes reveal genus-specific nucleotide preference, extreme protein-level conservation and fast divergence of organ-specificity. *Genes & Genomics, 31*(2), 129-142.

Lynch, V. J., & Wagner, G. P. (2008). Resurrecting the role of transcription factor change in developmental evolution. *Evolution, 62*(9), 2131-2154.

Maes, L., Inze, D., & Goossens, A. (2008). Functional specialization of the *TRANSPARENT TESTA GLABRA1* network allows differential hormonal control of laminal and marginal trichome initiation in *Arabidopsis* rosette leaves. *Plant Physiology, 148*(3), 1453-1464.

Malmberg, R. L., Held, S., Waits, A., & Mauricio, R. (2005). Epistasis for fitness-related quantitative traits in *Arabidopsis thaliana* grown in the field and in the greenhouse. *Genetics, 171*(4), 2013-2027.

Marks, M. D., & Feldmann, K. A. (1989). Trichome development in *Arabidopsis thaliana* .1. T-DNA tagging of the *GLABROUS1* gene. *Plant Cell, 1*(11), 1043-1050.

Martin, C., Ellis, N., & Rook, F. (2010). Do transcription factors play special roles in adaptive variation? *Plant Physiology, 154*(2), 506-511.

Martin, C., & Glover, B. J. (2007). Functional aspects of cell patterning in aerial epidermis. *Current Opinion in Plant Biology, 10*(1), 70-82.

Martin, N. H., Sapir, Y., & Arnold, M. L. (2008). The genetic architecture of reproductive isolation in louisiana irises: Pollination syndromes and pollinator preferences. *Evolution, 62*(4), 740-752.

Matsui, K., Umemura, Y., & Ohme-Takagi, M. (2008). *AtMYBL2*, a protein with a single MYB domain, acts as a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. *Plant Journal, 55*(6), 954-967.

Mauricio, R. (1998). Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. *American Naturalist, 151*(1), 20-28.

Mauricio, R. (2005). Ontogenetics of QTL: the genetic architecture of trichome density over time in *Arabidopsis thaliana*. *Genetica, 123*(1-2), 75-85.

Mauricio, R., & Rausher, M. D. (1997). Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defense. *Evolution, 51*(5), 1435-1444.

Mauricio, R., Stahl, E. A., Korves, T., Tian, D. C., Kreitman, M., & Bergelson, J. (2003). Natural selection for polymorphism in the disease resistance gene *RPS2* of *Arabidopsis thaliana*. *Genetics, 163*(2), 735-746.

McKay, J. K., Richards, J. H., Nemali, K. S., Sen, S., Mitchell-Olds, T., Boles, S., et al. (2008). Genetics of drought adaptation in *Arabidopsis thaliana* II. QTL analysis of a new mapping population, Kas-1 x Tsu-1. *Evolution, 62*(12), 3014-3026.

Meinhard, H., & Gierer, A. (1974). Applications of a theory of biological pattern formation based on lateral inhibition. *Journal of Cell Science, 15*(2), 321-346.

Michaels, S. D., He, Y. H., Scortecci, K. C., & Amasino, R. M. (2003). Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America, 100*(17), 10102-10107.

Morohashi, K., & Grotewold, E. (2009). A systems approach reveals regulatory circuitry for *Arabidopsis* trichome initiation by the *GL3* and *GL1* selectors. *Plos Genetics, 5*(2).

Morohashi, K., Zhao, M. Z., Yang, M. L., Read, B., Lloyd, A., Lamb, R., et al. (2007). Participation of the *Arabidopsis* bHLH factor *GL3* in trichome initiation regulatory events. *Plant Physiology, 145*(3), 736-746.

Morse, A. M., Whetten, R. W., Dubos, C., & Campbell, M. M. (2009). Post-translational modification of an R2R3-MYB transcription factor by a MAP Kinase during xylem development. *New Phytologist, 183*(4), 1001-1013.

Moyano, E., MartinezGarcia, J. F., & Martin, C. (1996). Apparent redundancy in Myb gene function provides gearing for the control of flavonoid biosynthesis in *Antirrhinum* flowers. *Plant Cell, 8*(9), 1519-1532.

Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M., & Lepiniec, L. (2000). The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques. *Plant Cell, 12*(10), 1863-1878.

Nesi, N., Jond, C., Debeaujon, I., Caboche, M., & Lepiniec, L. (2001). The *Arabidopsis TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell, 13*(9), 2099-2114.

Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H. G., et al. (2005). The pattern of polymorphism in *Arabidopsis thaliana*. *Plos Biology, 3*(7), 1289-1299.

O'Neill, C. M., Morgan, C., Kirby, J., Tschoep, H., Deng, P. X., Brennan, M., et al. (2008). Six new recombinant inbred populations for the study of quantitative traits in *Arabidopsis thaliana*. *Theoretical and Applied Genetics, 116*(5), 623-634.

Oppenheimer, D. G., Herman, P. L., Sivakumaran, S., Esch, J., & Marks, M. D. (1991). A MYB gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell, 67*(3), 483-493.

Palaniswamy, S. K., James, S., Sun, H., Lamb, R. S., Davuluri, R. V., & Grotewold, E. (2006). AGRIS and AtRegNet. A platform to link cis-regulatory elements and transcription factors into regulatory networks. *Plant Physiology, 140*(3), 818-829.

Pang, Y. Z., Wenger, J. P., Saathoff, K., Peel, G. J., Wen, J. Q., Huhman, D., et al. (2009). A WD40 repeat protein from *Medicago truncatula* is necessary for tissue-specific anthocyanin and proanthocyanidin biosynthesis but not for trichome development. *Plant Physiology, 151*(3), 1114-1129.

Payne, C. T., Zhang, F., & Lloyd, A. M. (2000). *GL3* encodes a bHLH protein that regulates trichome development in *Arabidopsis* through interaction with *GL1* and *TTG1*. *Genetics, 156*(3), 1349-1362.

Perazza, D., Vachon, G., & Herzog, M. (1998). Gibberellins promote trichome formation by up-regulating *GLABROUS1* in *Arabidopsis*. *Plant Physiology, 117*(2), 375-383.

Pesch, M., & Hulskamp, M. (2009). One, two, three ... models for trichome patterning in *Arabidopsis*? *Current Opinion in Plant Biology, 12*(5), 587-592.

Pfalz, M., Vogel, H., Mitchell-Olds, T., & Kroymann, J. (2007). Mapping of QTL for resistance against the crucifer specialist herbivore *Pieris brassicae* in a new *Arabidopsis* Inbred Line population, Da(1)-12xEi-2. *Plos One, 2*(6).

Pyhajarvi, T., Garcia-Gil, M. R., Knurr, T., Mikkonen, M., Wachowiak, W., & Savolainen, O. (2007). Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics, 177*(3), 1713-1724.

Qi, T. C., Song, S. S., Ren, Q. C., Wu, D. W., Huang, H., Chen, Y., et al. (2011). The Jasmonate-ZIM-Domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*. *Plant Cell, 23*(5), 1795-1814.

Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J., et al. (1999). Molecular analysis of the anthocyanin2 gene of petunia and its role in the evolution of flower color. *Plant Cell, 11*(8), 1433-1444.

Rose, L. E., Bittner-Eddy, P. D., Langley, C. H., Holub, E. B., Michelmore, R. W., & Beynon, J. L. (2004). The maintenance of extreme amino acid diversity at the disease resistance gene, *RPP13*, in *Arabidopsis thaliana*. *Genetics, 166*(3), 1517-1527.

Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X., & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics, 19*(18), 2496-2497.

Sanyal, A., & Linder, C. R. (2012). Quantitative trait loci involved in regulating seed oil composition in *Arabidopsis thaliana* and their evolutionary implications. *Theoretical and Applied Genetics, 124*(4), 723-738.

Schellmann, S., Schnittger, A., Kirik, V., Wada, T., Okada, K., Beermann, A., et al. (2002). *TRIPTYCHON* and *CAPRICE* mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. *Embo Journal, 21*(19), 5036-5046.

Schlichting, C. D., & Pigliucci, M. (1993). Control of phenotypic plasticity via regulatory genes. *American Naturalist, 142*(2), 366-370.

Schmid, K. J., Ramos-Onsins, S., Ringys-Beckstein, H., Weisshaar, B., & Mitchell-Olds, T. (2005). A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics, 169*(3), 1601-1615.

Schubert, D., Lechtenberg, B., Forsbach, A., Gils, M., Bahadur, S., & Schmidt, R. (2004). Silencing in *Arabidopsis* T-DNA transformants: The predominant role of a gene-specific RNA sensing mechanism versus position effects. *Plant Cell, 16*(10), 2561-2572.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology, 18*(2), 233-234.

Schwartz, C., Balasubramanian, S., Warthmann, N., Michael, T. P., Lempe, J., Sureshkumar, S., et al. (2009). Cis-regulatory changes at *FLOWERING LOCUS T* mediate natural variation in flowering responses of *Arabidopsis thaliana*. *Genetics, 183*(2), 723-732.

Schwinn, K., Venail, J., Shang, Y. J., Mackay, S., Alm, V., Butelli, E., et al. (2006). A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell, 18*(4), 831-851.

Shimizu, H., Maruoka, M., Ichikawa, N., Baruah, A. R., Uwatoko, N., Sano, Y., et al. (2010). Genetic control of phenotypic plasticity in Asian cultivated and wild rice in response to nutrient and density changes. *Genome, 53*(3), 211-223.

Shindo, C., Bernasconi, G., & Hardtke, C. S. (2007). Natural genetic variation in *Arabidopsis*: Tools, traits and prospects for evolutionary ecology. *Annals of Botany, 99*(6), 1043-1054.

Simon, M., Loudet, O., Durand, S., Berard, A., Brunel, D., Sennesal, F. X., et al. (2008). Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. *Genetics, 178*(4), 2253-2264.

Stamp, N. (2003). Theory of plant defensive level: example of process and pitfalls in development of ecological theory. *Oikos, 102*(3), 672-678.

Steets, J. A., Takebayashi, N., Byrnes, J. M., & Wolf, D. E. (2010). Heterogenous selection on trichome production in Alaskan *Arabidopsis kamchatica* (Brassicaceae). *American Journal of Botany, 97*(7), 1098-1108.

Stern, D. L., & Orgogozo, V. (2009). Is genetic evolution predictable? *Science, 323*(5915), 746-751.

Stinchcombe, J. R., Weinig, C., Ungerer, M., Olsen, K. M., Mays, C., Halldorsdottir, S. S., et al. (2004). A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proceedings of the National Academy of Sciences of the United States of America, 101*(13), 4712-4717.

Stracke, R., Werber, M., & Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology, 4*(5), 447-456.

Streisfeld, M. A., Liu, D., & Rausher, M. D. (2011). Predictable patterns of constraint among anthocyanin-regulating transcription factors in *Ipomoea*. *New Phytologist, 191*(1), 264-274.

Symonds, V. V., Godoy, A. V., Alconada, T., Botto, J. F., Juenger, T. E., Casal, J. J., et al. (2005). Mapping quantitative trait loci in multiple populations of *Arabidopsis thaliana* identifies natural allelic variation for trichome density. *Genetics, 169*(3), 1649-1658.

Symonds, V. V., Hatlestad, G., & Lloyd, A. M. (2011). Natural allelic variation defines a role for *ATMYC1*: Trichome cell fate determination. *Plos Genetics, 7*(6).

Symonds, V. V., & Lloyd, A. M. (2003). An analysis of microsatellite loci in *Arabidopsis thaliana*: Mutational dynamics and application. *Genetics, 165*(3), 1475-1488.

Symonds, V. V., & Lloyd, A. M. (2004). A simple and inexpensive method for producing fluorescently labelled size standard. *Molecular Ecology Notes, 4*(4), 768-771.

Szymanski, D. B., Jilk, R. A., Pollock, S. M., & Marks, M. D. (1998). Control of *GL2* expression in *Arabidopsis* leaves and trichomes. *Development, 125*(7), 1161-1171.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*(3), 585-595.

Tetard-Jones, C., Kertesz, M. A., & Preziosi, R. F. (2011). Quantitative trait loci mapping of phenotypic plasticity and genotype-environment interactions in plant and insect performance. *Philosophical Transactions of the Royal Society B-Biological Sciences, 366*(1569), 1368-1379.

Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M., & Bergelson, J. (2003). Fitness costs of *R*-gene-mediated resistance in *Arabidopsis thaliana*. *Nature, 423*(6935), 74-77.

Tian, D. C., Araki, H., Stahl, E., Bergelson, J., & Kreitman, M. (2002). Signature of balancing selection in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America, 99*(17), 11525-11530.

Toledo-Ortiz, G., Huq, E., & Quail, P. H. (2003). The *Arabidopsis* basic/helix-loop-helix transcription factor family. *Plant Cell, 15*(8), 1749-1770.

Tominaga-Wada, R., & Nukumizu, Y. (2012). Expression analysis of an R3-Type MYB transcription factor *CPC-LIKE MYB4* (*TRICHOMELESS2*) and *CPL4*-related transcripts in *Arabidopsis*. *International Journal of Molecular Sciences, 13*(3), 3478-3491.

Tonsor, S. J., Alonso-Blanco, C., & Koornneef, M. (2005). Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. *Plant Cell and Environment, 28*(1), 2-20.

Traw, M. B. (2002). Is induction response negatively correlated with constitutive resistance in black mustard? *Evolution, 56*(11), 2196-2205.

Traw, M. B., & Bergelson, J. (2003). Interactive effects of jasmonic acid, salicylic acid, and gibberellin on induction of trichomes in *Arabidopsis*. *Plant Physiology, 133*(3), 1367-1375.

Ungerer, M. C., Halldorsdottir, S. S., Purugganan, M. A., & Mackay, T. F. C. (2003). Genotype-environment interactions at quantitative trait loci affecting inflorescence development in *Arabidopsis thaliana*. *Genetics, 165*(1), 353-365.

Valverde, P. L., Fornoni, J., & Nunez-Farfan, J. (2001). Defensive role of leaf trichomes in resistance to herbivorous insects in *Datura stramonium*. *Journal of Evolutionary Biology, 14*(3), 424-432.

van Ooijen, J. W. V., R. E. (2001). *JoinMap 3.0. Software for the calculation of genetic linkage maps.* Wageningen: Plant Research International.

Walker, A. R., Davison, P. A., Bolognesi-Winfield, A. C., James, C. M., Srinivasan, N., Blundell, T. L., et al. (1999). The *TRANSPARENT TESTA GLABRA1* locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell, 11*(7), 1337-1349.

Wang, S., Wang, J. W., Yu, N., Li, C. H., Luo, B., Gou, J. Y., et al. (2004). Control of plant trichome development by a cotton fiber MYB gene. *Plant Cell, 16*(9), 2323-2334.

Wang, S. C., & Chen, J. G. (2008). *Arabidopsis* transient expression analysis reveals that activation of *GLABRA2* may require concurrent binding of *GLABRA1* and *GLABRA3* to the promoter of *GLABRA2*. *Plant and Cell Physiology, 49*(12), 1792-1804.

Wang, S. C., Hubbard, L., Chang, Y., Guo, J. J., Schiefelbein, J., & Chen, J. G. (2008). Comprehensive analysis of single-repeat R3 MYB proteins in epidermal cell patterning and their transcriptional regulation in *Arabidopsis*. *BMC Plant Biology, 8*, 13.

Wang, S. C., Kwak, S. H., Zeng, Q. N., Ellis, B. E., Chen, X. Y., Schiefelbein, J., et al. (2007). *TRICHOMELESS1* regulates trichome patterning by suppressing *GLABRA1* in *Arabidopsis*. *Development, 134*(21), 3873-3882.

Werker, E. (2000). Trichome diversity and development. *Advances in Botanical Research, 31*, 1-35.

Wester, K., Digiuni, S., Geier, F., Timmer, J., Fleck, C., & Hulskamp, M. (2009). Functional diversity of R3 single-repeat genes in trichome development. *Development, 136*(9), 1487-1496.

Wright, S. I., Bi, I. V., Schroeder, S. G., Yamasaki, M., Doebley, J. F., McMullen, M. D., et al. (2005). The effects of artificial selection of the maize genome. *Science, 308*(5726), 1310-1314.

Xia, Y., Yu, K. S., Navarre, D., Seebold, K., Kachroo, A., & Kachroo, P. (2010). The *glabra1* mutation affects cuticle formation and plant responses to microbes. *Plant Physiology, 154*(2), 833-846.

Yin, P., Kang, J. Q., He, F., Qu, L. J., & Gu, H. Y. (2010). The origin of populations of *Arabidopsis thaliana* in China, based on the chloroplast DNA sequences. *Bmc Plant Biology, 10*.

Yoshida, Y., Sano, R., Wada, T., Takabayashi, J., & Okada, K. (2009). Jasmonic acid control of *GLABRA3* links inducible defense and trichome patterning in *Arabidopsis*. *Development, 136*(6), 1039-1048.

Yu, J. M., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics, 38*(2), 203-208.

Zhang, F., Gonzalez, A., Zhao, M. Z., Payne, C. T., & Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. *Development, 130*(20), 4859-4869.

Zhao, H. T., Wang, X. X., Zhu, D. D., Cui, S. J., Li, X., Cao, Y., et al. (2012). A single amino acid substitution in IIIf subfamily of basic helix-loop-helix transcription factor *AtMYC1* leads to trichome and root hair patterning defects by abolishing its interaction with partner proteins in *Arabidopsis*. *Journal of Biological Chemistry, 287*(17), 14109-14121.

Zhao, K. Y., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C. L., et al. (2007). An *Arabidopsis* example of association mapping in structured samples. *Plos Genetics, 3*(1).

Zhao, M., Morohashi, K., Hatlestad, G., Grotewold, E., & Lloyd, A. (2008). The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. *Development, 135*(11), 1991-1999.

Zhen, Y., & Ungerer, M. C. (2008). Relaxed selection on the CBF/DREB1 regulatory genes and reduced freezing tolerance in the southern range of *Arabidopsis thaliana*. *Molecular Biology and Evolution, 25*(12), 2547-2555.

Zhu, H. F., Fitzsimmons, K., Khandelwal, A., & Kranz, R. G. (2009). *CPC*, a single-repeat R3 MYB, is a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. *Molecular Plant, 2*(4), 790-802.

Zimmermann, I. M., Heim, M. A., Weisshaar, B., & Uhrig, J. F. (2004). Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like bHLH proteins. *Plant Journal, 40*(1), 22-34.

Zust, T., Joseph, B., Shimizu, K. K., Kliebenstein, D. J., & Turnbull, L. A. (2011). Using knockout mutants to reveal the growth costs of defensive traits. *Proceedings of the Royal Society B-Biological Sciences, 278*(1718), 2598-2603.