THE UNIVERSITY OF CHICAGO

ADAPTATION TO COASTAL ENVIRONMENTS IN SWEDISH *ARABIDOPSIS THALIANA*

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECOLOGY AND EVOLUTION

BY

LAURA ALIDA MERWIN

CHICAGO, ILLINOIS

DECEMBER 2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor, Joy Bergelson. I first spoke to Joy while sitting on a bus in Perth, Australia. She brought me to the University of Chicago and, along with Marcus Kronforst, Jocelyn Malamy, Cathy Pfister, and David Salt, provided the help and guidance without which this thesis would never have been completed. More than that, my committee was incredibly supportive not just of my research, but of my simultaneous pursuit of pedagogical education and experience while in graduate school. Without the encouragement, opportunities to teach, and letters of reference they supplied, I would not have the job I have today, and for that I am immensely grateful.

The Bergelson lab group (past and present) is unique for the incredibly broad range of interests and experience covered by its members; it is also a great place to work. In particular, Ben and Talia answered endless dumb questions about coding and Alice read many drafts of many documents. The lab runs smoothly thanks to the continuous effort of Tim Morton, and, before him, Wayan Mulyati. They are lab managers extraordinaire.

A number of cheerful and hardworking undergraduates and techs contributed many hours of labor to the projects described here. Without Alex, Celia, Kalie, Monica,

Meaghan, Adrianna, Rod, and Bailey, this thesis would literally not have been possible. Many thanks to them, both for their efforts and their camaraderie.

Sandra Suwanski and John Zdenek, faced with the Herculean task of managing a greenhouse used by a dozen groups to study organisms from tobacco plants to butterflies, run a tight ship with grace and good nature. I have yet to make a request that is not met by "Sure, we can do that." Their expertise and care made the experimental portions of this thesis a whole lot easier. In the same way, Bonnie Brown, Connie Homan, and Jeff Wisniewski in the Ecology and Evolution office provide tireless service. Submitting grants, hiring undergraduates, getting reimbursements, registering for conferences, and a thousand other administrative tasks are made painless in their presence.

At the point where acknowledgements turn from the more professional to the more personal, I want to recognize a person who straddles that boundary. Alison Anastasio, a former Bergelson lab member whom I first met at a conference in 2008, shares my interest in plant local adaptation and was the first to introduce me to the Swedish beach populations of *Arabidopsis*. As a scientific mentor, she provided much useful advice and guidance during the planning, experimental design, and data analysis phases of these projects. On a more personal level, she has helped to guide me through the thickets of grad school, offering wisdom, experience, and occasionally a shoulder to cry on. Our common enthusiasm for urban ecology and

sustainable cities will, I hope, continue to draw us together, and I look forward to future collaborations with Alison as both a friend and colleague.

It is impossible to briefly encapsulate the delightful personalities that make up the students of the Darwin cluster. In journal clubs, weekday lunches, and of course at Jimmy's and the Cove, conversations with the Darwinians never fail to be intellectually stimulating, cheerfully combative, and completely enjoyable. (They are also often hilarious.) Throughout, the Darwinians demonstrate their intelligence, thoughtfulness, creativity, and compassion. I am honored to have been part of this group, and I am excited to watch what my colleagues do in the world in the years to come. Special thanks go to my amazing, incredible, stellar cohort: Justin Lemberg, Aaron Olsen, Ben Rubin, Chris Schell, Kristin Voorhies, and Ben Winger, with extra special thanks to Colin Kyle, Tom Stewart, Talia Karasov, Carrie Albertin and Alice MacQueen. In the Darwin auxiliaries, I would also like to thank Justin Heckman for approximately one thousand gin drinks.

Without the support of my family, both now and throughout my life, it goes without saying that I would not have made it this far. Conversations with my dad can range from world history and philosophy of science to technology and global politics; they are always interesting and enlightening. Moreover, his experience as a scientist and manager has helped me to navigate graduate school successfully and with fewer moments of absolute panic. My mother has been the most steady source of support to me throughout the years, heroically available to me day and night (except of

course when a new season of Downton Abbey has just been released). She knows how important this is to me. And my brother, Jack, is one of the few people in the world who has nearly as many thoughts, opinions, and feelings about the Chicago Blackhawks as I do myself.

Last, but not least, I would like to thank the Chicago Blackhawks, who won not one but three Stanley Cups during my graduate school tenure and without which I would not have made it through the long, cold winters.

ABSTRACT

Plants display a variety of specialist adaptations to particular environments. Beach ecotypes are especially common due to a suite of abiotic stressors that act as environmental drivers of adaptation. These stressors include low water availability, salt deposition, low nutrient soil, burial by sand, and strong wind; associated phenotypes include dwarfism, succulence, cuticles and salt excluding glands, rapid growth, light coloration, reflective leaves, transpirational cooling, long hypocotyls, a high density of trichomes, and large seed size. The model organism *Arabidopsis thaliana* is found in a variety of environments across the globe, including coastal environments. As a predominantly selfing species, *A. thaliana* is a good candidate for the production of local ecotypes. Additionally, high-resolution genomic data and ease of phenotyping make *A. thaliana* an attractive choice for the experimental exploration of adaptation to beach environments. I use several complementary approaches to investigate coarse sand beach populations of *A. thaliana* on the Baltic Sea coast of southern Sweden.

In the first chapter, I investigate the genetic architecture of adaptation to beaches in *A. thaliana* using two types of genomic scans. First, I perform an unbiased genome wide scan using Population Branch Statistic (PBS) to identify regions of the genome under selection in the beach population of interest. I apply additional population genetic metrics to the top-scoring regions in order to identify putative genes

involved in adaptation. Second, I focus on three phenotypes that differentiate beach and inland populations and are characteristic of beach plants: hypocotyl length, trichome density, and seed weight. I perform genome wide association mapping using 298 Swedish lines for each phenotype and identify significant associations with each phenotype, including a region of chromosome 4 near the candidate gene *SPA2* significantly associated with hypocotyl length and a region on chromosome 2 strongly associated with trichome density and the candidate genes TCL1/TCL2/ETC2. Combining the two approaches described above, I ask whether the most closely associated SNPs for each phenotype have a higher average PBS value than expected. This test asks whether SNPs associated with phenotypes of interest show evidence of adaptation in the beach population, something I find to be true for hypocotyl length and seed weight, but not trichome density. Thus, hypocotyl length and seed weight appear to be under selection in the beach population.

In the second chapter, I investigate hypocotyl length more deeply. I demonstrate that plants from coarse sand beaches have significantly longer hypocotyls than inland and fine sand beach conspecifics and identify a genetic variant accounting for 18% of the variance in hypocotyl length in the mapping population. I confirm the correlation between this genetic variant and hypocotyl length in two additional populations, a global set of lines from the RegMap panel and in a set of Bay x Sha RIL lines, and demonstrate that this genetic variant increases seedling emergence frequency when seeds are buried. I conclude that this genetic variant is in linkage with a gene involved in hypocotyl elongation, possibly *SPA2* (a suppressor of

phytochrome A important for growth in darkness) and that this gene confers a fitness advantage when seeds are buried.

In the third chapter, I test the hypothesis that beach lines are more tolerant of drought and salt stress than inland conspecifics. I use total seed production as a proxy for lifetime fitness and quantify reductions in fitness under several levels of drought and salt spray stress. I detect no significant response to salt spray. While drought has a negative effect on fecundity, beach lines do not consistently show different responses to drought than inland lines.

This set of complementary studies approaches adaptation to beaches from genotypic, phenotypic, and environmental perspectives. I identify phenotypes and genetic regions important for adaptation in this population of *A. thaliana,* as well as an environmental driver of selection. Additionally, I introduce a new method of combining types of genomic data to yield information about phenotypes under selection in populations of interest.

INTRODUCTION

I.1 The study of adaptation in plants

Humans have been impressed with the matching of organisms to their environments for as long as they have been observing the natural world. Darwin's journals are full of observations of adaptation; such observations led to his formulation of the theory of natural selection (Darwin 1845, 1865). Adaptation has remained a central concern of biologists seeking to understand the past, current, and future occurrence of organisms across various environments. Beginning in the twentieth century, the study of genetics allowed scientists to explicitly investigate mechanisms underlying adaptive evolutionary change. More recently, the emerging field of ecological genomics is concerned specifically with the ability to link genotype with phenotype and environment (Ungerer et al. 2008).

Adaptation of plants to their local environments appears to be common. The term ecotype (Turesson 1922), denoting distinct ecological races of a single species, was first used to compare conspecifics of several plant species growing at coastal and inland sites in Sweden. Early reciprocal transplant experiments in California demonstrated that phenotypic differences associated with altitude in the Sierra Nevada were the product of fixed genetic differences between populations of plants (Clausen, Keck, & Heisey 1940). In the decades since these foundational studies, adaptations to particular sites or environmental conditions have been identified in a range of plant species. These include tolerance to

copper in species located on old mine sites (Gregory & Bradshaw 1965; Allen & Sheppard 1971; Macnair 1983), to serpentine soils and other edaphic factors (Rajakaruna 2010; Sherrard & Maherali 2012; Barry 2013), and to salt spray (Lowry et al. 2008; Lowry et al. 2009). Even when specific selection pressures have not been identified, the correlation between broad-scale genotypic and environmental patterns suggests adaptation, for example, to climate (Fournier-Level 2011; Hancock et al. 2011; Stearns & Fenster 2013; Yoder et al. 2015). Indeed, in a metastudy of reciprocal transplant experiments, both Leimu and Fisher (2008) and Hereford (2009) found local adaptation (when native populations have increased fitness compared to conspecifics from other sites) in 71% of cases. The average magnitude of the fitness advantage was 45% (Hereford 2009). Local adaptation is also possible at very fine scales (Paccard et al. 2013; Richardson et al. 2014).

Aside from the general goal of increasing understanding of the natural world, an understanding of plant adaptation leads to several practical applications. As anthropogenic change alters environments across the globe, scientists, policy makers, and conservation managers increasingly seek to understand how populations have adapted to environmental conditions in the past. Such information can address questions regarding the capacity of ecosystems to adapt to climate change and the likelihood that various protective or restorative measures will be successful (Friesen & von Wettberg 2010; Weinig, Ewers & Welch 2014).

In addition, naturally occurring ecotypes of various plant species demonstrate adaptation to low nutrient soils, reduced water availability, and high salt loads, among other stressors.

Such traits are exactly those required for agricultural production on marginal land, an increasingly necessary practice as human population expansion and anthropogenic change combine to both increase demand and decrease available land for food production (*UN 2010 Revision of the World Population Prospects*; Galpaz & Reymond 2010; Chaves & Davies 2010).

I.2 Adaptation of plants to beach environments

Beaches have been recognized as particularly harsh ecosystems for terrestrial plants due to a suite of co-occurring abiotic stressors: high light; poor quality, fast draining soils; high wind, dramatic diurnal fluctuations in temperature; salt spray; and burial in sand (Maun 1994, McLachlan & Brown 2006, Voronkova et al. 2008; Maun 2009). This marginal environment restricts the type and number of plants that can survive and it produces strong selective forces favoring adaptation to beach conditions. Turesson (1922, 1925) was only the first to identify coastal ecotypes in a number of species. Subsequently, adaptation to coastal sites has been identified in many other species, including *Diodia teres* (Jordan 1992), *Gilia capitata* (Nagy & Rice 1997), *Borrichia aborescens* (Morrison 2002), a set of species from far eastern Russia (Voronkova et al. 2008), *Brassica nigra* (Bischoff & Hurault 2013), and *Arabidopsis thaliana* in Spain (Busoms et al. 2015).

In at least two recent cases, coastal ecotypes have been used to identify genetic regions associated with differential fitness at coastal and inland sites. Verhoeven et al. (2004, 2008) performed reciprocal transplants of wild barley (*Hordeum spontaneum*) at coastal and

inland steppe sites in Israel that differed in water availability and soil quality. They detected a home site advantage and identified a QTL associated with high fitness at the inland site. The QTL was located over a flowering time gene; the authors suggest that rapid flowering provides an advantage in this inhospitable habitat. Lowry et al. (2009) identified a QTL associated with salt spray tolerance in coastal populations of *Mimulus guttatus* in the Pacific Northwest. In addition to salt spray tolerance, *M. guttatus* coastal and inland ecotypes differ in life history (perennial vs. annual), as well as morphological and phenological traits. Subsequent analysis revealed an inversion polymorphism associated both with contrasting life histories and the availability of soil moisture during the summer months (Lowry & Willis 2010).

Certain features tend to be common among beach plants including dwarfism, succulence, cuticles and salt excluding glands, rapid growth, light coloration, reflective leaves, transpirational cooling, long hypocotyls, a high density of trichomes, and large seed size (Turesson 1925; McLachlan & Brown 2006; Voronkova et al. 2008; Maun 2009). Voronkova et al. (2008) noted the similarity in adaptations to deserts and beaches, suggesting again that beach plants are shaped by harsh abiotic stresses.

I.3 *Arabidopsis thaliana* as a model system for studying adaptation

Arabidopsis thaliana is a small annual selfing member of the family Brassicaceae that is broadly distributed across the Northern Hemisphere. As an opportunist, it takes advantage of patchy disturbed habitats, including pastures, cropland, roadside verges, and stone walls,

as well as naturally patchy ecosystems like beach margins. *A. thaliana* displays considerable genetic variation across its range, with a general pattern of isolation by distance (Platt et al. 2010).

Though perhaps better known for its role in molecular and genetic studies, *A. thaliana* has long been touted as a model organism for evolutionary ecology (Mitchell-Olds 2001; Mitchell-Olds & Schmitt 2006; Shindo et al. 2007; Alonso Blanco et al. 2009; Bergelson & Roux 2010; Anderson et al 2011; Trontin et al. 2011). *A. thaliana* is a good candidate for such studies for a number of reasons. First, its global distribution encompasses a wide variety of climates and ecosystems, and there is evidence for adaptation to local environments (Weinig et al. 2003; Fournier-Level 2011; Hancock et al. 2011; Agren & Shemske 2012; Stearns & Fenster 2013). Second, *A. thaliana* displays variation for nearly every phenotype measured (Koornneef et al. 2004). Third, high quality genotype information is available for a large number of accessions from across the globe. Fourth, as a predominantly selfing species, clonal lines can be maintained and phenotyped under various conditions, allowing fixed and plastic effects to be disentangled and high powered association mapping to be carried out without the expense of newly genotyping individuals for each experiment.

Long-established populations of *A. thaliana* have been identified on beaches on the Baltic coast of southern Sweden within a few dozen meters of the water. These plants differ both phenotypically and phenologically different from inland conspecifics. I focused my study on twelve lines from beaches with very coarse sand. In the greenhouse they appear to

consistently produce longer hypocotyls, more trichomes, fewer, larger seeds, shorter inflorescence stems and broader rosettes. My goal was to explore the genetic architecture of adaptation to beaches and to identify both heritable adaptive phenotypes and their underlying genetic variation.

I.4 Approaches to studying adaptation to beaches in *A. thaliana*

Adaptation to a particular environment occurs when selection acts to favor beneficial phenotypes. As a result, allele frequencies at genes underlying these phenotypes are expected to differ between populations of conspecifics adapted to other environments. A complete understanding of adaptation in a specific location would identify (a) the selective pressure(s) in the environment (b) the responding phenotype(s) and (c) the underlying genetic variation.

Approaches to studying local adaptation may begin by focusing on any one of these three components. To completely avoid *a priori* assumptions about selection pressures or adaptive phenotypes, one can search the genome for patterns created by selective events (Tiffen & Ross-Iberra 2014). However, the relationship between regions or genes identified in this manner and phenotype at the organism level is often unclear. Alternatively, examination of environmental and phenotypic differences between populations of conspecifics may suggest putative mechanisms of adaptation (indeed, many populations of interest are chosen for this reason). One can then begin with experiments examining phenotypic variation or the effect of the putative selective pressure in controlled conditions,

leaving the difficult challenge of determining the underlying genetic basis until precise adaptive phenotypes have been defined. Irrespective of the starting point, a complete understanding of an organism's adaptation to a particular place generally requires a set of complementary experiments and analyses (Ungerer et al. 2008; Anderson et al. 2011; Agren & Schemske 2012; Pardo-Diaz et al. 2014; Tiffen & Ross-Iberra 2014).

In this dissertation, I investigated *Arabidopsis thaliana* adaptation to beaches using a number of the approaches discussed above. I used Population Branch Statistic, a metric based on pairwise F_{st} measurements, to perform an unbiased genome wide scan for regions under selection in the beach populations. To complement this analysis, I chose three phenotypes observed to differentiate beach and inland populations, hypocotyl length, trichome density, and seed size. I performed genome wide association studies to identify genetic variation underlying these phenotypes. I combined this GWA data with the PBS data to look for evidence that these phenotypes are under selection in beach populations by asking whether SNPs with high scores in the PBS scan had lower than average GWA p-values. These approaches, separately and in combination, suggest which phenotypes are important in beach plants.

To further investigate a single important phenotype and to link phenotype with potential environmental stressor, I explored hypocotyl length more fully. I characterized the extent of phenotypic differentiation between beach and inland populations and tested the effect of a candidate genetic variant identified using association mapping. I also tested for a fitness

advantage of both phenotype and genotype when plants were subject to a beach stressor, burial in sand.

Lastly, I examined plant response to candidate beach stressors, drought and salt. I used lifetime fecundity measurements to estimate the effect on fitness of these stressors in both beach and inland populations. In this way, I tested for whole plant tolerance to putative beach conditions.

This suite of experiments sought to explore plant adaptation to beaches from phenotypic, genotypic, and environmental starting points in order to identify important characteristics of beach populations of *A. thaliana*.

CHAPTER 1: GENOME-WIDE APPROACHES TO THE STUDY OF ADAPTATION SUGGEST GENOTYPES AND PHENOTYPES UNDER SELECTION ON BEACHES

1.1 Abstract

Both forward and reverse genetic approaches can identify loci important for adaptation. I applied both approaches to identify putative adaptive loci in *Arabidopsis thaliana* from coarse sand beaches in southern Sweden. I used an F_{st} based genome wide scan (population branch statistic, PBS) to identify regions that have undergone selection in my focal population, but not in inland conspecifics. For regions in the top 0.1% of PBS scores, I calculated nucleotide diversity, Tajima's D, and linkage disequilibrium separately for the beach and inland lines. I identified 20 genes with extreme PBS values and extreme differences between the beach and inland lines for the population genetic statistics. As a complementary approach, I chose three phenotypes that differed between beach and inland lines, hypocotyl length, trichome density, and seed weight, and performed genome wide association mapping in a panel of 298 Swedish lines. I identified strong candidate genes for hypocotyl length and trichome density, but not seed weight. Combining data from the two approaches, I tested whether the most strongly associated SNPs in the GWAs had higher PBS values than expected by chance and whether the SNPs with the highest PBS values had lower p-values than expected by chance. I found both to be true for hypocotyl length and seed size, but not trichome density, suggesting that these two phenotypes are

under selection in the beach population. I demonstrate that genomic data can provide information about phenotypes under selection, even when candidate genes underlying these phenotypes cannot be identified.

1.2 Introduction

1.2.1 Background

Understanding how organisms adapt to their environment is a central focus of evolutionary biology. Before the advent of modern molecular techniques, scientists documented correlations between the features of organisms and characteristics of their environments, both within and between species. Reciprocal transplant and common garden experiments were used to demonstrate that observed differences were genetically based and adaptive (for example, Clausen, Keck, and Heisey 1940). More recently, genome sequencing has allowed a closer investigation of the genetic changes that lead to the production of adaptive phenotypes. These new technologies have provided alternative methods for identifying adaptive traits (Pardo-Diaz et al. 2014, de Villemereuil & Gaggiotti 2015).

The availability of high quality genomic data allows for both forward and reverse genetic approaches to identify loci important for adaptation. Genomic scans that search for regions showing high levels of differentiation between populations of interest are usefully agnostic in that they do not require previously selected candidate genes or phenotypes (Stinchcombe & Hoekstra 2008, Pannell & Fields 2013, Tiffen & Ross-Iberra 2014, Pardo-

Diaz et al. 2014, de Villemereuil & Gaggiotti 2015). On the other hand, when candidate phenotypes or environmental characters have been identified, association mapping can identify the genomic regions underlying these putatively adaptive traits (Atwell et al 2010, Bergelson & Roux 2010, Rellstab et al. 2015). A combination of forward and reverse approaches has the potential to augment the identification of adaptive loci where separate approaches fail or where the number of candidates is large (Stinchcombe & Hoekstra 2008, Pardo-Diaz et al. 2014, Habel et al. 2015, de Villemereuil & Gaggiotti 2015)

Two recent studies demonstrate the utility of combining forward and reverse genetic approaches. Fischer et al. (2013) used an F_{st} based analysis to identify highly differentiated SNPs in a set of alpine *Arabidopsis halleri* populations. They then examined associations between these SNPs and a set of topo-climatic variables to identify 175 candidate genes for local adaptation, including four genes known to be associated with site water balance and solar radiation. Nadeau et al. (2014) performed F_{st} outlier analysis and association mapping for wing color patterning in *Heliconius* hybrid zones in South America. Association mapping for color pattern identified many genes known from experimental crosses; there was considerable overlap between these results and genomic regions identified as F_{st} outliers.

Here, I used forward and reverse genetic approaches separately and in combination to identify putative adaptive genotypes and phenotypes in a beach population of *Arabidopsis thaliana.* Beach sites can select for locally adapted plant species and ecotypes (Maun 1994, McLachlan & Brown 2006, Voronkova et al. 2008, Maun 2009) due to their numerous

abiotic stressors. Most obvious among these are low water and nutrient availability, high light and temperature, salt spray, wind, and blowing sand. Phenotypically and genetically differentiated beach ecotypes have been identified in a number of species, including a set of Swedish species (Turesson 1922, 1925) *Diodia teres* (Jordan 1992), *Gilia capitata* (Nagy & Rice 1997), *Borrichia aborescens* (Morrison 2002), a set of species from far eastern Russia (Voronkova et al. 2008), *Brassica nigra* (Bischoff & Hurault 2013), and *Arabidopsis thaliana* in Spain (Busoms et al. 2015). A number of adaptive traits are common in beach plants, including dwarfism, succulence, cuticles and salt excluding glands, rapid growth, light coloration, reflective leaves, transpirational cooling, long hypocotyls, a high density of trichomes, and large seed size (Turesson 1925; McLachlan & Brown 2006; Voronkova et al. 2008; Maun 2009).

I performed a genome wide scan for regions under selection as well as association mapping for three putatively adaptive traits, hypocotyl length, trichome density, and seed size. I show that both forward and reverse approaches yield strong candidate genes, but that strong candidate loci are not detected for every putatively adaptive phenotype tested. I define a methodology for combining information from the forward and reverse approaches that can identify phenotypes under selection using genomic data.

1.2.2 Approach

I sought to identify genetic regions putatively associated with adaptation to beaches in a southern Swedish population of *Arabidopsis thaliana* using two approaches. First, I used

population branch statistic (PBS), an agnostic scan based on pairwise F_{st}, to identify regions that appear to have undergone selection in beach lineages, but not in closely related inland populations. Second, I used genome wide association mapping to identify genomic regions associated with three phenotypes: hypocotyl length, trichome density, and seed size. These phenotypes are thought to be important in beach populations, and I had observed them to be extreme in the Swedish beach populations of interest. I asked:

(a) Which genomic regions appear important for adaptation in beach populations?

(b) Do candidate phenotypes show strong association with specific genomic regions?

To demonstrate that combining the data from these two approaches can yield additional information about phenotypes under selection in this population, I asked:

(c) Do genomic regions associated with putatively adaptive traits show a significant signature of adaptation using the PBS scan results?

1.2.3 Population branch statistic

PBS is based on comparing per locus estimates of divergence times.

$PBS_{Beach} = (T^{B,SS} + T^{B,Out} - T^{SS,Out})/2$

Where:

T = divergence time between given populations; $-\log(1-F_{ST}^{B,SS})$

B = Beach

SS = Southern Sweden (inland)

Out = Outgroup

PBS differs from pairwise F_{st} statistics in its use of an outgroup to identify sites that have undergone selection in a focal population but not in a closely related sister group. Using pairwise comparisons between the focal population and the sister group, and between each of these and the outgroup, PBS gives the apparent length of the branch leading to the population of interest since its divergence from the sister group (Figure 1.1). PBS values at loci that have undergone selection in the population of interest will be large relative to loci evolving due to drift.

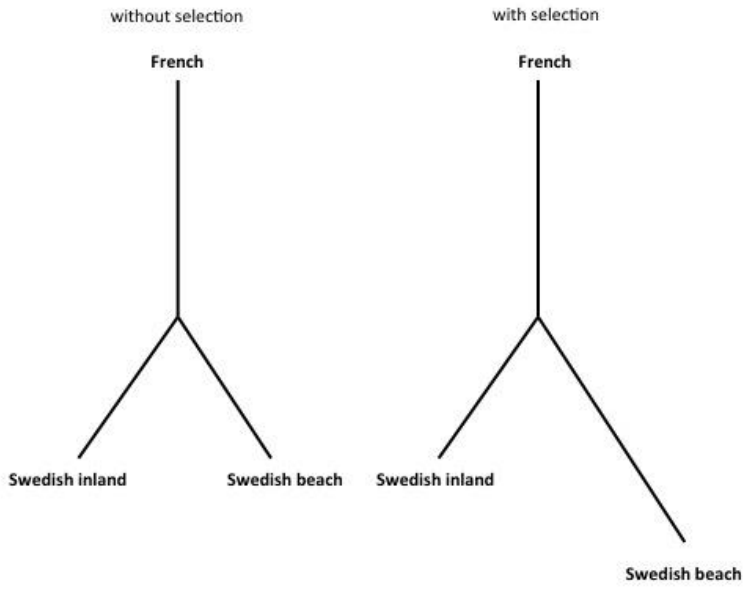

Figure 1.1 Graphic depiction of PBS. Loci that have undergone selection will appear to have a longer time since divergence from the sister group. The asymmetrical nature of the scan (only loci that have undergone selection in the beach population will display this long branch) is achieved by using a third population as an outgroup (French). This figure is adapted from Figure S11 in Huerta-Sanchez et al. (2013).

The PBS method has power to detect both complete and incomplete selective sweeps acting on standing variation or on de novo mutations (Yi et al. 2010). PBS was first used by Yi et al. (2010) and Huerta-Sanchez et al. (2013) in order to understand adaptation to high altitude

in Tibetan and Ethiopian human populations. SNPs with high PBS values were located in genes known to be involved in response to hypoxia. More recently, Zhan et al. (2014) used PBS in combination with other population genetic metrics to identify a collagen subunit differentiating migratory and non-migratory populations of monarch butterflies (*Danaus plexippus*).

1.2.4 Genome wide associations for hypocotyl length, trichome density, and seed weight

I chose three phenotypes that I had noted were consistently different between beach and inland lines when grown in the greenhouse and that are noted in the literature as common adaptive traits in beach plants. I measured these three phenotypes in a set of Swedish accessions and conducted genome wide association mapping.

Hypocotyl length

Hypocotyl elongation is the mechanism by which seedlings emerge from sand burial, a potential hazard on beaches (Maun 2009). Tolerance to burial is a determinant of plant community zonation in dune systems (Van der Valk 1974, Maun and Perumal 1999, Wilson and Sykes 1999, Dech and Maun 2005). A positive relationship between hypocotyl length and burial depth was observed after experimental burial of seeds of *Cirsium pitcheri* (Chen & Maun 1999), and rapid emergence from burial conveys a fitness advantage over inter- and intraspecific competitors (Zhang & Maun 1990, Maun 2009, Weis et al. 2015).

Plant responses to darkness and to burial are similar (Sykes & Wilson 1990). Shade avoidance, a plastic response to low light conditions resulting in morphological changes including elongated hypocotyls, internodes, and petioles and upward bending of the leaves, is also beneficial during burial, allowing for extension of leaves toward light and allowing sand to slide off vertically oriented leaves (Sykes & Wilson 1990). Though shade avoidance is a plastic response, variation in plasticity itself can be an adaptive trait, as loss of plasticity is associated with lower fitness under some light conditions (Schmitt et al. 1995, Ganade & Westoby 1999, Pigliucci & Schmitt 1999, Schmitt et al. 2003, Pierek & de Wit 2013).

Mutants that display hypocotyl length phenotypes inappropriate to the light environment have proven valuable for investigating light sensitive signaling pathways. The large set of genes involved in these pathways and contributing to constitutive and induced hypocotyl length phenotypes includes the five phytochromes (PHYA-E) and regulators such as PIFs (phytochrome interacting factors), the SPA (suppressor of phytochrome A), HY, and COP families, and DELLA proteins, that work to suppress photomorphogenesis in the dark and promote it in light (reviewed in Vandenbussche et al. 2005 and Wang & Wang 2015). In addition, Filiault and Maloof (2012) assembled *a priori* and *de novo* candidate genes associated with hypocotyl length for a mapping study of natural variation in hypocotyl length in *A. thaliana.* Appendix A, Table 1 lists candidate genes from these studies.

Trichome density

Dense trichomes are often noted as characteristic of beach plants (Turesson 1925, McLachlan & Brown 2006, Voronkova et al. 2008), and are believed to be protective against herbivory, high light, aridity, and salt spray (Mauricio 1998, Morrison 2002, Voronkova et al. 2008, Steets et al. 2010, Bloomer et al. 2014). Observed natural variation in trichome density appears to correlate with selection pressure in local populations. Sharma (1984) found higher trichome numbers on leaves from a coastal population of *Salvia lyrata* when compared to an inland population, and Morrison (2002) found a correlation between the ratio of pubescent and glabrous ecotypes of the sea daisy *Borrichia aborescens* and distance to the coast in the Bahamas. Both authors speculated that trichomes protected leaves from salt spray at coastal sites (but see Cartica & Quinn 1980). Karkkaninen et al. (2004) measured variation in trichome density in *Arabidopsis lyrata*, the perennial sister species of *A. thaliana*, in eastern Sweden and found evidence for selection potentially related to herbivory. Mauricio and Rausher (1997) used manipulative field experiments to demonstrate that *A. thaliana* trichome density responds to herbivory related selection pressures.

The genes underlying trichome development have been well characterized, and several studies have identified genomic regions associated with 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, Bloomer et al. 2014). Key proteins include TRANSPARENT TESTA GLABRA (TTG1), GLABRA3 (GL3), ENHANCER OF GLABRA 3 (EGL3) R2R3 MYB GLABRA 1

(GL1), and TT8, which complex to initiate trichome development, and a set of R3 MYBs (including TCL1, TCL2, and ETC2) that suppress initiation in surrounding cells to ensure correct patterning across the leaf. Appendix A, Table 2 lists candidate genes associated with trichome density.

Seed weight

Large seeds have been associated with beaches (Jordan 1992, Maun 1994, Zhang 1996) and with low resource environments generally, especially those that experience low water availability (Baker 1972 for 2500 species, Lowry et al. 2014 for *Panicum hallii*, but see Manzan-Piedras et al. 2014). Moreover, a relationship between hypocotyl length and seed size has been documented in many species and systems (van der Valk 1974, Maun & Riach 1981, Barbour 1985, Maun & LaPierre 1986 (within and between species); Yanful & Maun 1996, Ganade & Westoby 1999, but also see Chen & Maun 1999).

Seed size is a complex polygenic trait that is also affected strongly by environment (Kesavan et al. 2012, Van Daele et al. 2012, Lowry et al. 2014). Genes regulating the development of the embryo, the endosperm, and the seed coat all affect seed size. Key genes known to be involved in these processes include HAIKU1 and 2, MINISEED3, APETALA2, TRANSPARENT TESTA GLABRA2, AUXIN RESPONSE FACTOR2, and AINTEGUMENTA; additionally, several groups have performed QTL mapping studies on natural variation in *A. thaliana* seed size (Alonso-Blanco 1999, Simon et al. 2008, Herridge et al. 2011, Joosen et al. 2012, Kesavan et al. 2012, Moore et al. 2012, Van Daele et al. 2012),

identifying further candidate genes, but the function of many of these genes is not yet known. Appendix A, Table 3 lists candidate genes for seed size.

1.2.5 Combining the two approaches

Peaks identified using a PBS scan represent regions of the genome under current or recent selection in the population of interest. They may represent few or many phenotypes. On the other hand, peaks identified using GWA represent regions of the genome associated with a particular phenotype in the mapping population. This phenotype may or may not be under selection in the population of interest.

A strong signal over the same genomic location in both scans suggests the identification of a phenotype and genotype under selection in a focal population. Such a relationship between selection, phenotype, and genotype, once validated, demonstrates adaptation in a particular environment. However, in the absence of strong and congruent peaks in PBS and GWAS analysis, one can still probe for evidence of selection on adaptive traits. Many genomic regions combine to shape a quantitative trait. Therefore, for complex, putatively adaptive traits, it might not be realistic to expect to find a large, single GWA peak that is congruent with evidence of selection. Instead, one can ask whether SNPs with low p-values in a particular GWA have a higher average PBS value than a random SNP set of the same size. Such a result would suggest that the phenotype mapped in the GWA is under selection in the population of interest. This conclusion rests on the assumption that the set of SNPs

with the lowest p-values represents true associations with the phenotype in question, rather than false positives due to incomplete accounting for population structure.

Similarly, one can ask whether genomic regions showing evidence of selection contribute strongly (but not completely) to a putatively adaptive trait. To do so requires testing whether the set of SNPs with the highest PBS values has a lower than expected average p-value using a particular GWA result. This would suggest that selection is occurring on the phenotype in question. Moreover, it would suggest that the phenotype is relatively important for adaptation in the population of interest. To see why, imagine a situation where many phenotypes are under selection in the population of interest. In this case, the SNP sets associated with all these phenotypes would display higher than average PBS values, but the high-PBS SNP set will exhibit lower than average p-values only for the phenotypes under the very strongest selection.

1.3 Methods

1.3.1 Population branch statistic

I used population branch statistic (PBS) to conduct an unbiased genome wide search in *Arabidopsis thaliana* for loci that appear to be under selection in beach populations, but not in inland conspecifics. Using the measure of F_{st} developed by Reynolds et al. (1983) and the 250K SNP set (Kim et al. 2007), I calculated per locus pairwise F_{st} between beach and inland Swedish populations and between each of these and a French outgroup (Appendix B

Table 1). I used these values to calculate a per locus value of PBS, then averaged PBS values across 20,000 bp sliding windows with a step length of 1000 bp, a window size smaller than the distance at which linkage disequilibrium in *A. thaliana* is expected to decay (within 50 kb, Nordborg et al. 2005), but large enough to yield clear peaks. Windows containing only one SNP were discarded. I extracted the windows with the top 0.1% of PBS values (n=116). I grouped overlapping windows from this set together to identify high-PBS peaks in the genome and identified the genes under these peaks using the TAIR10 database (www.arabidopsis.org).

To search for additional signals of selection within these peaks and to narrow the candidate gene list, I calculated three population genetic metrics for these peaks using a higher density SNP set from the 1001 Genomes Project (1001genomes.org) and the R packages pegas (Paradis et al. 2010) and adegenet (Jombart et al. 2008). In particular, I measured nucleotide diversity, Tajima's D, and linkage disequilibrium (LD) for the set of top peaks identified in the PBS scan.

Several population genetic methods test for signatures of selection by comparing observed statistics to those expected under a model of neutral evolution. Genomic regions that have undergone recent selection are expected to exhibit low sequence diversity. Nucleotide diversity, the average per site pairwise nucleotide differences (Nei & Li 1979), is one such measure of sequence diversity. Another, Tajima's D (Tajima 1989), compares two measures of nucleotide diversity, one of which is sensitive to sample genealogy. If a site is evolving neutrally, the two measures will be equal, causing the test statistic, D, to equal 0. A negative

value of D indicates that purifying selection has operated at the site in question, while a positive value of D suggests that the site is undergoing balancing selection. Linkage disequilibrium, a measurement of the correlation between a pair of loci, is expected to drop rapidly with genomic distance if sites are evolving neutrally. Sequences that have recently undergone a selective sweep, however, may maintain high linkage disequilibrium across many thousands of base pairs.

I recalculated PBS within each peak using the higher density SNP set. Then I removed the French outgroup lines from the data set and separated the Swedish lines into two sets: coarse sand beach lines (n=12) and inland conspecifics (n=135). To look for differences in signatures of selection between the sister sets of lines, I calculated nucleotide diversity, Tajima's D, and linkage disequilibrium (LD) within each peak at a higher resolution than the original scan, using sliding windows of 2000 bp with a step size of 250 bp. For each window, I also calculated the absolute value of the difference between the metrics for the beach and inland sets. Within each peak, among the 2000 bp windows, I looked for windows with a difference between the beach and inland lines in the top 1% for at least two of the three population genetic statistics that also fell in the top 1% of the PBS scores. I then narrowed the candidate gene list to genes that fell under these windows.

To test how extreme the beach population is compared to subsets of the inland accessions of the same size (n=12), I randomly sampled 12 lines from the inland set 100 times. For each sample set, I calculated the metrics for sliding windows of 2000 bp with a step size of

250 bp. For the windows identified above, I asked if the beach population fell in the 1% tail of the distribution of these sample sets.

1.3.2 Genome wide association mapping

For genome wide association mapping, I used a set of 298 Swedish lines (Appendix B Table 2).

Hypocotyl length

Five experimental blocks were planted, one on each of five successive days. Each experimental block consisted of one sample from each line on each of the following substrates: soil (50:50 Fafard C-2:Metro Mix 200), sand (Vigoro Torpedo) and pebbles (Vigoro Pea Pebbles, ~0.5 – 1 cm diameter). Seeds were planted on the moist substrate surface in 98 cell flats, skipping the top and bottom rows to prevent edge effects caused by uneven watering. Top and bottom rows were filled with substrate, but free of plants. For ease of planting, seeds were planted alphabetically by line name. Planted flats were covered and placed in a 4 °C cold room for seven days for stratification, then moved to a 20 °C, 16-hour day greenhouse. Plants were watered as needed and fertilized twice weekly. After one week, plants were thinned to one seedling per cell. On day ten, seedlings were harvested and hypocotyl length (defined as the stem between the base of the cotyledons and the root:shoot junction) was measured by hand.

I performed GWA using both univariate and multivariate linear mixed models in GEMMA (Genome-wide Efficient Mixed Model Association, Zhou & Stephens 2012, 2014) with the 250K SNP set (Kim et al. 2007). GEMMA tests each SNP for association with either a univariate phenotype (hypocotyl length in each growth substrate tested separately) or a multivariate phenotype of the form (hypocotyl length in soil, hypocotyl length in sand, hypocotyl length in pebbles; Stephens 2013). Because the association scenario for a set of phenotypes is not known *a priori,* it is not clear which method has optimal power. Generally, a multivariate model increases the power of the association test, even when only one of the phenotypes tested is driving the association with a variant. However, univariate tests do perform better than multivariate when a single phenotype is directly associated with a variant, and the remaining phenotypes are only indirectly associated (i.e. the phenotypes are correlated). I assessed SNPs with p-values below 5×10^{-6} for nearby candidate genes. This threshold is less stringent than that required by a strict Bonferroni correction for multiple testing. However, the Bonferroni threshold is extremely conservative because it treats each SNP as independent, even though this is not the case. A threshold of 5×10^{-6} captures peaks that fall below the threshold but that are still likely to represent real associations.

Trichome density

Three replicates for each of the 298 lines were planted on moist soil (50:50 Fafard C-2:Metro Mix 200) in 36-cell pull-apart flats. Cells were then separated and replicates were randomized across 25 flats. Flats were covered and stratified at 4 °C for seven days, then

moved to a 20 °C 16-hour day greenhouse, where they were watered as needed and fertilized twice weekly. Plants were thinned to one per cell on day six. On days 24 and 25, rosettes were harvested. Because trichome density differs between leaves that are produced early and later in a plant's life, as well as between juvenile and adult leaves (Larkin et al. 1996, Telfer et al. 1997, Mauricio 2005), I sampled the same leaves (identified by production order) at the same time post-germination. Leaves 1-6 were discarded, and hole punches were taken from the center tip of leaves 7, 8, and 9. Punches were stored in moist 98-well PCR plates. For each punch, I counted trichome number on the upper surface of the leaf under a dissection scope. The average number of trichomes per punch for each line was used as the phenotype in the GWA analysis. GWA analyses were performed in R, using EMMAX, a mixed linear model approach with a correction for population structure (Kang et al. 2008, 2010) and the 250K SNP set described in Kim et al. (2007); SNPs with a p-value below 5×10^{-6} were assessed for nearby candidate genes.

Seed weight

Using a seed counter (Elmor C3), 500 seeds from stock tubes were counted for each line and then weighed. Two separate 500 seed aliquots were weighed for each line (n=2 was considered sufficient due to the high replicability between aliquots). Average weight per 500 was used as the phenotype in the GWA analysis. As above, GWA analyses were performed in R, using EMMAX (Kang et al. 2008, 2010) and the 250K SNP set (Kim et al. 2007); SNPs with a p-value below 5×10^{-6} were assessed for nearby candidate genes.

1.3.3 Combining PBS and GWA results

Selecting GWA results with increasingly stringent p-values should result in increasingly smaller SNP sets that are enriched for true associations with the phenotype. If the phenotype being mapped is under selection, then these progressively smaller SNP sets should also exhibit a higher mean PBS value. For each set of GWA results, I plotted the average PBS value using increasingly stringent p-value cutoffs, from one including nearly half of the 250K SNPs to a conservative threshold close to that required by a strict Bonferroni correction (for the hypocotyl length GWA, I used the multivariate results). A similar procedure was used to plot the average GWAS p-value of SNP sets using increasingly restrictive PBS cutoffs, ranging from the top 50% of PBS values to the top 0.0001%.

Next, I compared the real mean PBS and GWAS p-values of my top SNP sets under one specific set of thresholds to a null distribution of means of random SNP sets. For the PBS scan, I defined the highest scoring 0.1% as top SNPs (n=192; corresponding to PBS values over 1.31; note this uses individual SNP PBS values and not window averages as were used in defining top PBS peaks above). For the GWA studies, I used a threshold p-value of 0.0005 (n=235, 246, and 290 for hypocotyl length, trichome density, and seed weight respectively). I constructed null distributions of PBS values or p-values for random samples of SNPs of the same size as the top SNP sets in question and compared the real values to this distribution.

SNPs meeting the threshold criteria for either type of scan are not distributed evenly across the genome, but tend to cluster as peaks. Due to linkage disequilibrium, the PBS or p-value of SNPs close to each other in the genome are not independent; I maintained this structure in the null PBS and p-value sets. To do so, I separated the vector of PBS values by chromosome and reordered the chromosomes randomly, with each chromosome either inverted or not. Next, I shifted the list of p-values by a random number of positions (with values at the top of the list moving to the bottom, as if the vector had been circularized). The resulting vector was matched with the original ordered SNP list. Thus, each SNP id is matched with a random PBS value, but neighboring SNPs are matched with neighboring PBS values.

The SNP ids corresponding to the top set from each GWA (now matched with random PBS values) were extracted and the average PBS calculated. This process was repeated 10,000 times for each set of top SNPs to create a null distribution of mean PBS values. An analogous approach was taken using the p-values from each GWA and the top SNPs from the PBS scan.

A note on p-values: Strict frequentists object to the use of the magnitude of p-values as a measure of significance or method of prioritizing candidates. However, Tiffen & Ross-Iberra (2014) support ranking candidate adaptive loci based on the magnitude of p-values. They assert that "the use of ranks may capture functionally important loci that would not be captured if formal probability-based rejection of a null model is required", though they also note that this approach is prone to high levels of false positives. Two studies that have

used p-values in this way are Fournier-Level et al. (2011), who ranked candidate loci from a GWA for field fitness performed on large *A. thaliana* reciprocal transplant experiment, and Yoder et al. (2014), who used a similar protocol when testing for association with climate parameters in *Medicago truncatula* populations.

1.4. Results

1.4.1 Population branch statistic

The top 0.1% of windows (n=116) had mean PBS values over 0.57. They fell into twelve distinct peaks (Figure 1.2) encompassing 195 genes. For 2000 bp windows in these peaks, I calculated nucleotide diversity, Tajima's D, and linkage disequilibrium in beach and inland populations separately and recalculated PBS; I identified 99 windows in nine of the twelve peaks that fell within both the 1% tail of recalculated PBS scores and the 1% tail of differences between the beach and inland lines for at least two of the three population genetic statistics (Figure 1.3). These windows generally fell in the 1% tail of the distribution of randomly sampled (n=12) inland subsets, showing that the beach population is extreme in these genomic regions. There were 20 candidate genes that fell in these windows, including, notably, three genes involved in plant response to light, two of which are explicitly linked to hypocotyl elongation. (Table 1.1).

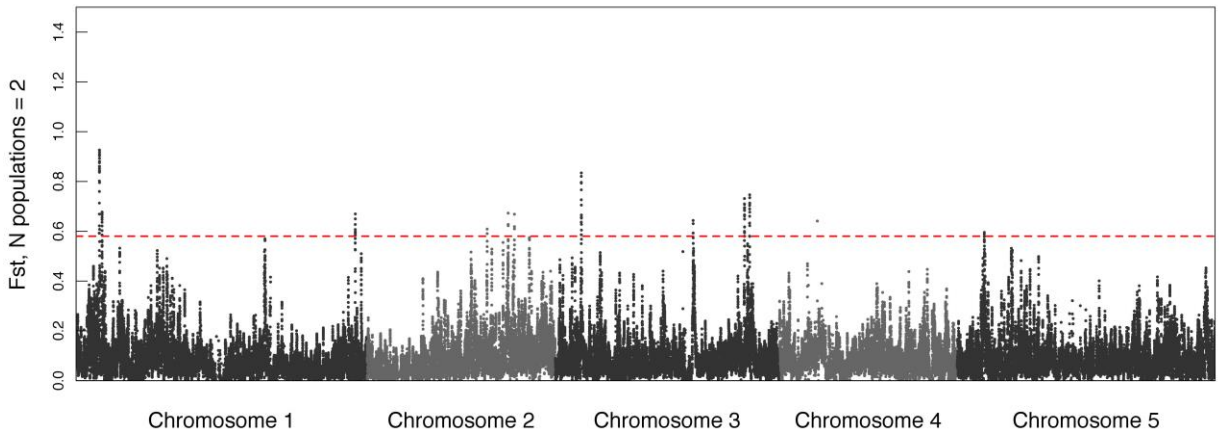

Figure 1.2 Manhattan plot of PBS results. Each dot represents the average PBS value for a 20,000 bp window (n=116,525). The red line represents the cutoff for the top 0.1% of windows.

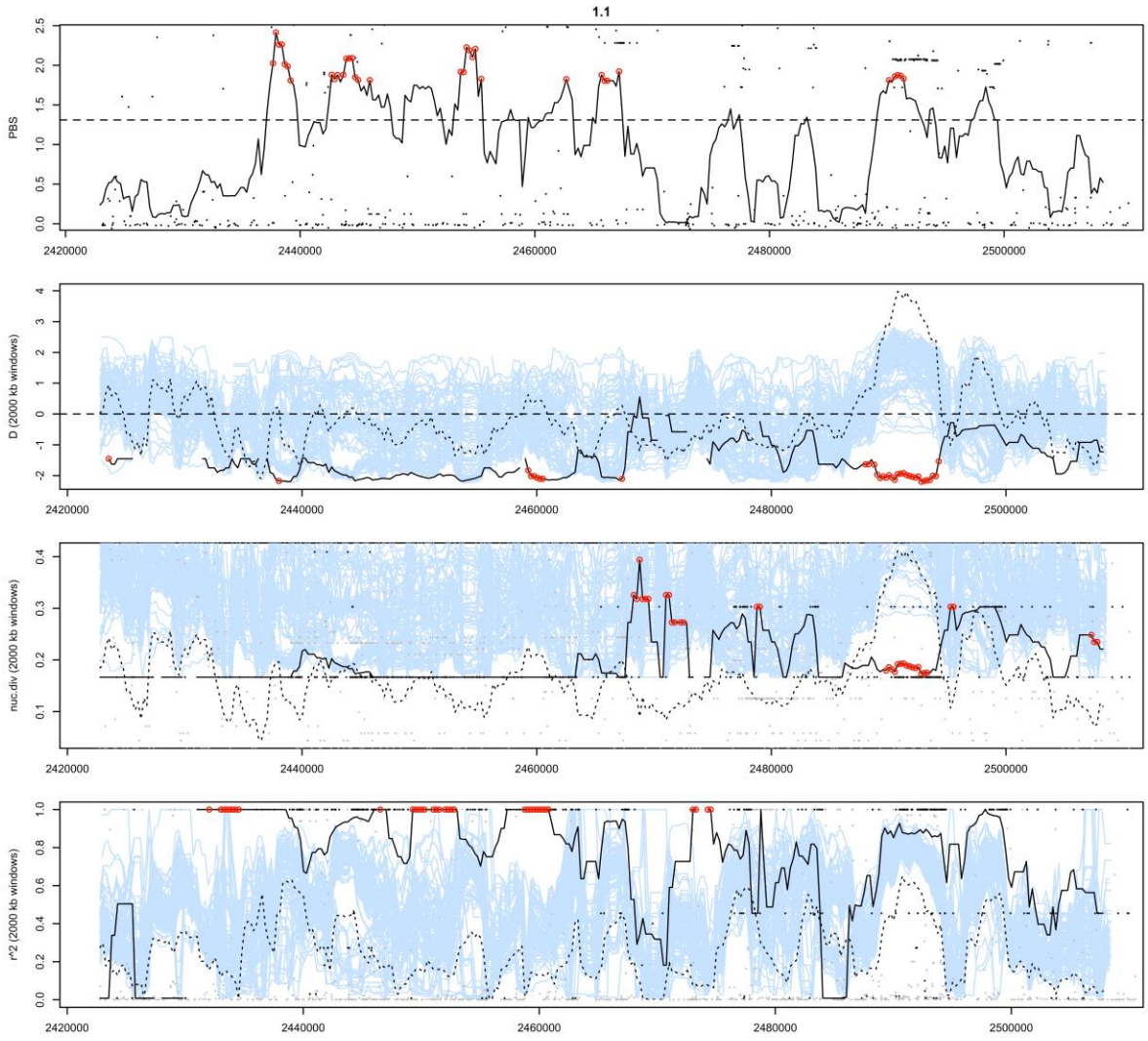

Figure 1.3 Example of population genetic analyses carried out on chromosome 1 peak identified from top 0.1% of original PBS scan as described in Methods. (a) PBS reanalysis with higher density SNP set (b) Tajima's D (c) nucleotide diversity (d) linkage disequilibrium. Dotted line is inland lines, solid is beach. Each blue line represents a single (n=12) random sample of inland lines. Red dots are top 1% (of values for PBS; of difference between beach and inland for rest). Peak at ~ 2495000 falls over the gene TPST, a tyrosylprotein sulfotransferase implicated in many processes including hypocotyl elongation.

Table 1.1 Candidate genes from PBS

chr	start	stop	locus	name	description
1	2491809	2493250	AT1G08030	ACTIVE QUIESCENT CENTER1, AQC1, HPS7, HYPERSENSITIVE TO PI STARVATION 7, TPST, TYROSYLPROTEIN SULFOTRANSFERASE	Tyrosylprotein sulfotransferase (TPST), acts in the auxin pathway to maintain postembryonic root stem cell niche by defining the expression of the PLETHORA stem cell transcription factor genes. Loss-of-function mutant has marked dwarf phenotype, stunted roots, pale green leaves, reduction in higher order veins, early senescence, reduced number of flowers and siliques.
1	29189984	29190975	AT1G77670		Pyridoxal phosphate (PLP)-dependent transferases superfamily protein; involved in 1-aminocyclopropane-1-carboxylate synthase activity, pyridoxal phosphate binding, the asparagine catabolic process, glutamate catabolic process to oxaloacetate, aspartate transamidation.
2	12564431	12564664	AT2G29220	L-TYPE LECTIN RECEPTOR KINASE III.1, LECRK-III.1	Concanavalin A-like lectin protein kinase family protein.
2	14760539	14756711	AT2G35020	GLCNA.UT2, GLCNAC1PUT2, N-ACETYLGLUCOSAMINE-1-PHOSPHATE URIDYLYLTRANSFERASE 2	N-acetylglucosamine-1-phosphate uridylyltransferase, catalyzes formation of UDP-N-acetylglucosamine (UDP-GlcNAc), an essential precursor for glycolipid and glycoprotein synthesis, also used for regulatory protein modification in signaling pathways; can also catalyze the reverse reaction using both UDP-GlcNAc and UDP-N-acetylgalactosamine; can also act on glucose-1-phosphate to produce UDP-glucose.
2	14760539	14763077	AT2G35030	COD1, CYTOCHROME C OXIDASE DEFICIENT 1	Pentatricopeptide repeat (PPR) superfamily protein
2	16969405	16965933	AT2G40660		Nucleic acid-binding, OB-fold-like protein; involved in tRNA aminoacylation for protein translation.
2	16969405	16970230	AT2G40670	ARR16, RESPONSE REGULATOR 16, RR16	Part of cytokinin-activated signaling pathway, involved in cytokinin induced senescence, red light sensing, circadian rhythm, phosphorelay signal transduction system.
3	14416678	14416062	AT3G42255	transposable element gene	Pseudogene, similar to putative Ac-like transposases.
3	14431053	14432589	AT3G42256	transposable element gene	Gypsy-like retrotransposon family (Athila)
3	14436678	14438817	AT3G42258	transposable element gene	Gypsy-like retrotransposon family (Athila)
3	14454553	14454332	AT3G42300	transposable element gene	
3	19787274	19784502	AT3G53360		Tetratricopeptide repeat (TPR)-like superfamily protein.
3	19787274	19788843	AT3G53370		S1FA-like DNA-binding protein.
3	20339277	20338522	AT3G54880		Unknown protein

Table 1.1 continued

chr	start	stop	gene	name	description
3	20339277	20341103	AT3G54890	LHCA1, PHOTOSYSTEM I LIGHT HARVESTING COMPLEX GENE 1	Encodes a component of the light harvesting complex associated with photosystem I.
5	2836454	2838090	AT5G08710	RCC1/UVR8/GEF-LIKE 1, RUG1	Regulator of chromosome condensation (RCC1) family protein.
5	2844954	2847415	AT5G08730	ARABIDOPSIS ARIADNE 16, ARI16, ARIADNE 16, ATARI16	Specific, to siliques; functions in zinc ion and nucleic acid binding, protein ubiquitination.
5	2844954	2845381	AT5G08720		Unknown protein
5	2858454	2860262	AT5G08790	ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN 81, ATAF2, ANAC081	Induced by wounding, belongs to a large family of putative transcriptional activators with NAC domain; associated with stresses related to pathogens, cold/light, arsenic. Functions in multicellular organismal development, regulation of cell size, regulation of transcription, defense responses, response to light stimulus, response to sucrose.
5	2858454	2856740	AT5G08780		Winged-helix DNA-binding transcription factor family protein; involved in DNA binding and nucleosome assembly.

ARR16 (AT2G40670, Figure 1.4) is part of a set of *Arabidopsis response regulator* genes involved in negative regulation of cytokinin signaling (Ren et al. 2009). ARR16 has been implicated in root growth, shoot regeneration, and leaf senescence (Ren et al. 2009). It has also been identified as a potential binding target of LONG HYPOCOTYL5 (HY5), a transcription factor that positively controls photomorphogenesis and root development in *A. thaliana* (Song et al. 2008). HY5 knockout mutants fail to inhibit hypocotyl elongation in light conditions. Downregulation of ARR16 is also part of cytokinin control of the shade response (Boonman et al. 2007).

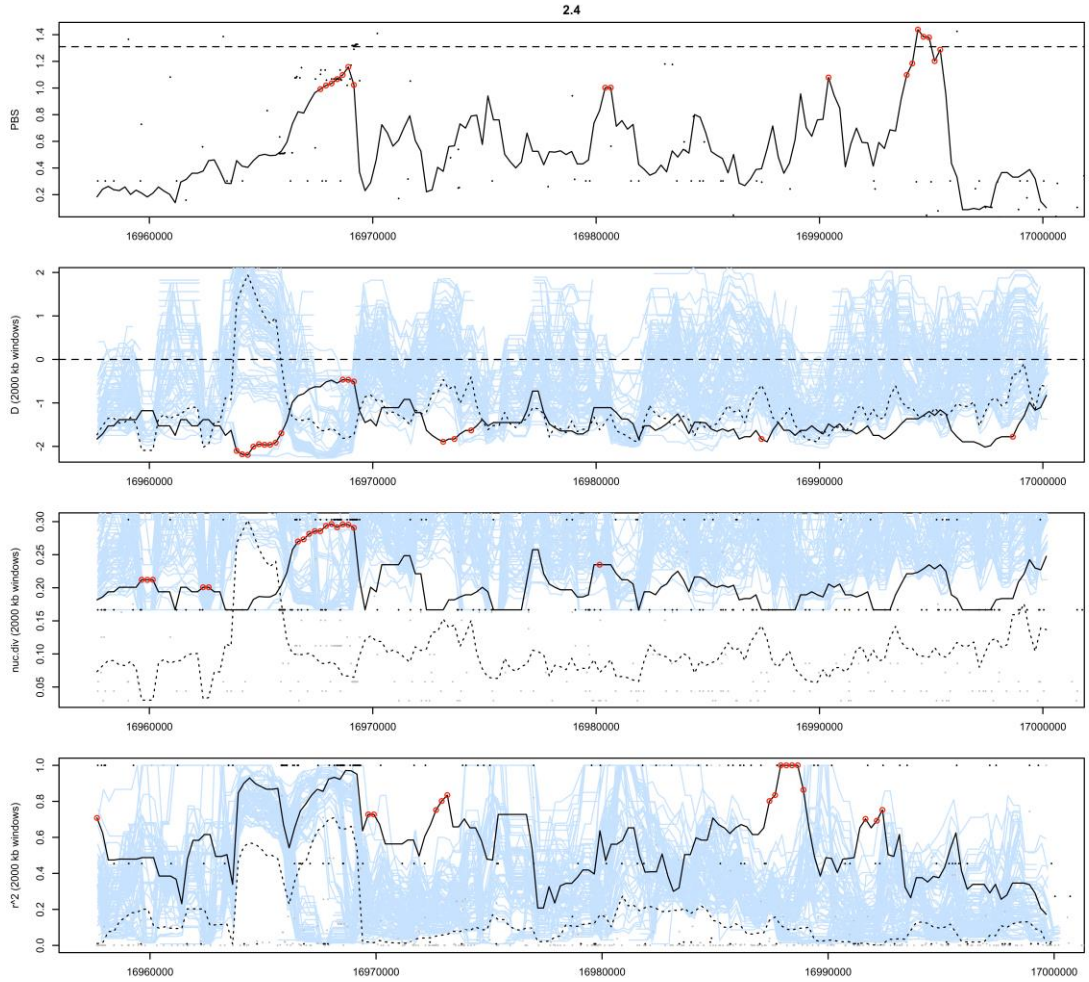

Figure 1.4 Population genetic analyses carried out on chromosome 2 peak identified from top 0.1% of original PBS scan as described in Methods. (a) PBS reanalysis with higher density SNP set (b) Tajima's D (c) nucleotide diversity (d) linkage disequilibrium. Dotted line is inland lines, solid is beach. Each blue line represents a single (n=12) random sample of inland lines. Red dots are top 1% (of values for PBS; of difference between beach and inland for rest). Peak at ~16970000 falls over the gene ARR16, a response regulator involved in negative regulation of cytokinin signaling and implicated in many processes including photomorphogenesis and the shade response.

Posttranslational tyrosine sulfation of the peptide hormones PSY1 and PSK by the tyrosylprotein sulfotransferase (TPST, AT1G08030; Figure 1.3; Komori et al. 2009) is associated with many diverse plant processes, including defense signaling (Igarashi et al. 2012, Mosher 2013), root stem cell maintenance (Matsuzaki et al. 2010, Zhou et al. 2010),

pollen tube growth (Stuhrwohldt et al. 2015), and response to phosphate and copper deficiencies (Kang et al. 2014, Wu et al. 2015). However, TPST has also been associated with hypocotyl elongation through PSK. PSK deficient cells have shorter hypocotyls due to reduced cell elongation; a similar phenotype was seen in TPST knockouts (Stuhrwohldt et al. 2011).

A third gene involved in light related processes in *A. thaliana*, photosystem I light harvesting complex 1 (LHCA1, AT3G54890; Figure 1.5), was also identified in this candidate list.

Additionally, one gene of unknown function (AT1G77670; Figure 1.6) was a candidate gene in a trichome density QTL study using the St-0 x Sf-2 RIL family (Bloomer et al. 2014).

Thus, although the PBS scan required no *a priori* assumptions about phenotypes under selection in the beach population, the candidate gene list generated contains genes linked to putatively adaptive phenotypes. Obvious follow-up candidates emerge, especially when the candidate list is coupled with the results of combining PBS and GWA data discussed below.

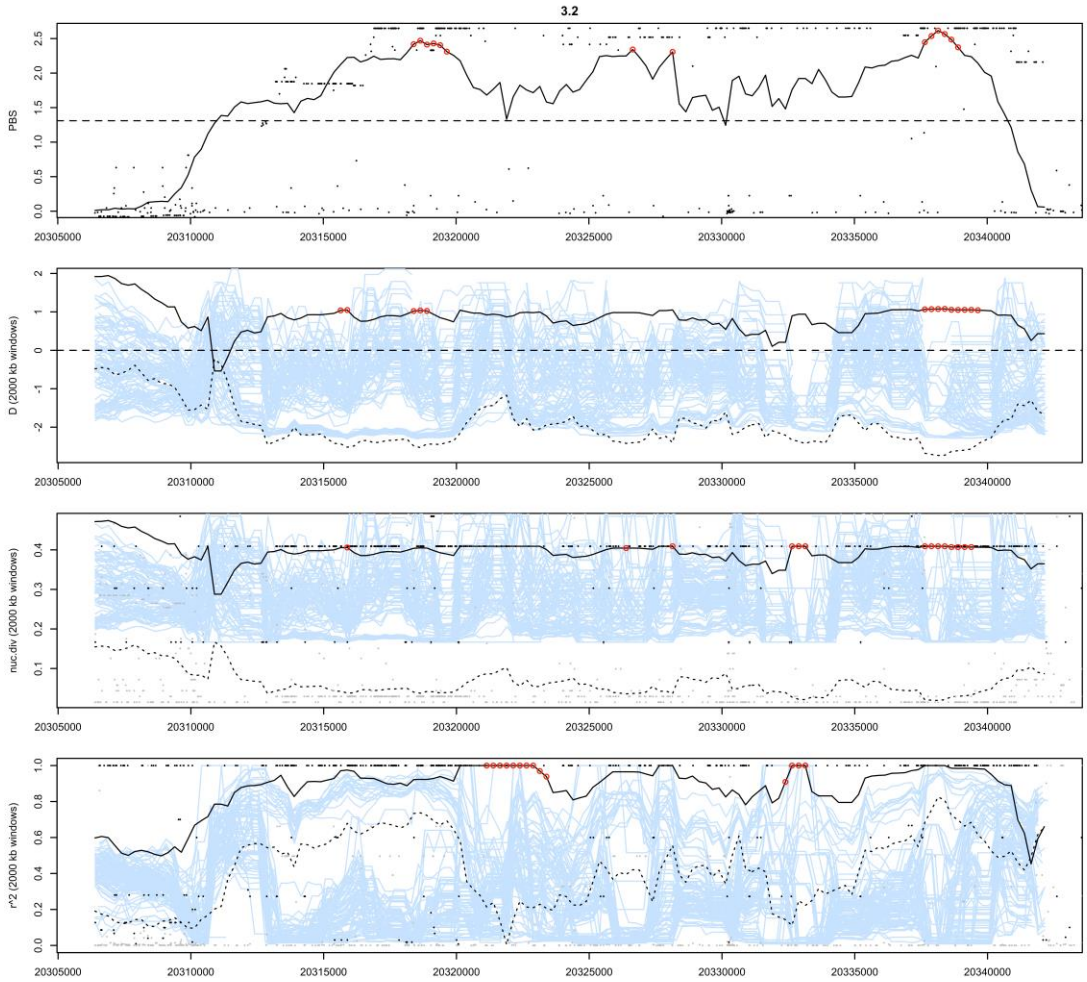

Figure 1.5 Population genetic analyses carried out on chromosome 3 peak identified from top 0.1% of original PBS scan as described in Methods. (a) PBS reanalysis with higher density SNP set (b) Tajima's D (c) nucleotide diversity (d) linkage disequilibrium. Dotted line is inland lines, solid is beach. Each blue line represents a single (n=12) random sample of inland lines. Red dots are top 1% (of values for PBS; of difference between beach and inland for rest). Peak at ~20340000 falls over the gene LHCA1, a component of the photosystem I light harvesting complex.

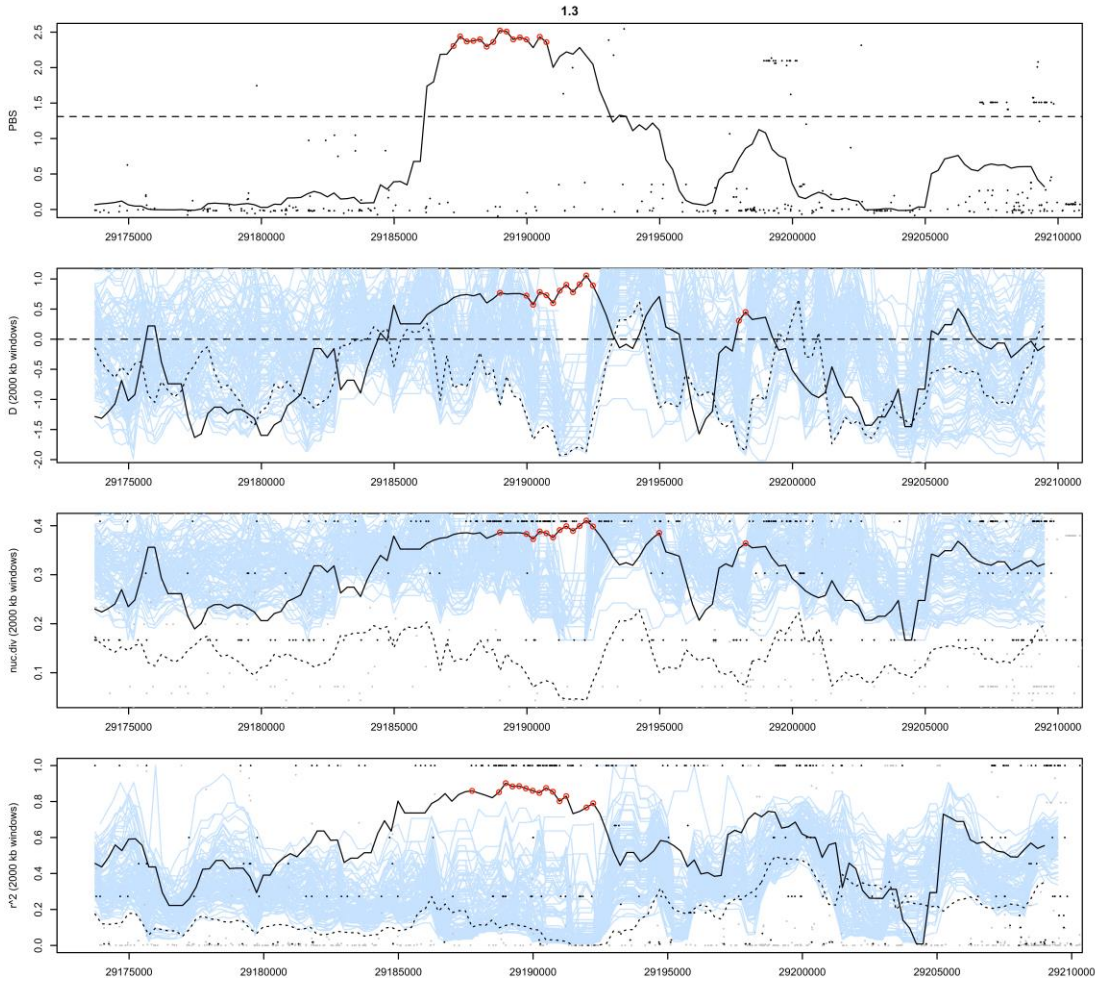

Figure 1.6 Population genetic analyses carried out on chromosome 1 peak identified from top 0.1% of original PBS scan as described in Methods. (a) PBS reanalysis with higher density SNP set (b) Tajima's D (c) nucleotide diversity (d) linkage disequilibrium. Dotted line is inland lines, solid is beach. Each blue line represents a single (n=12) random sample of inland lines. Red dots are top 1% (of values for PBS; of difference between beach and inland for rest). Peak at ~16970000 falls over the gene AT1G77670, encoding a protein of unknown function. This gene was also identified in a study mapping trichome density QTL (Bloomer et al. 2014).

1.4.2 Genome wide association studies

Hypocotyl length

As expected, hypocotyl length was extreme in the twelve beach lines. The average hypocotyl length across all substrates for these twelve lines was 5.4 mm (sd = 2.8 mm), compared to a mean of 3.7 mm (sd = 1.9 mm) for the remainder of the mapping lines (Wilcoxon rank sum test p-value < 0.0001; Figure 1.7). The beach lines accounted for seven of the twenty lines with the longest hypocotyls.

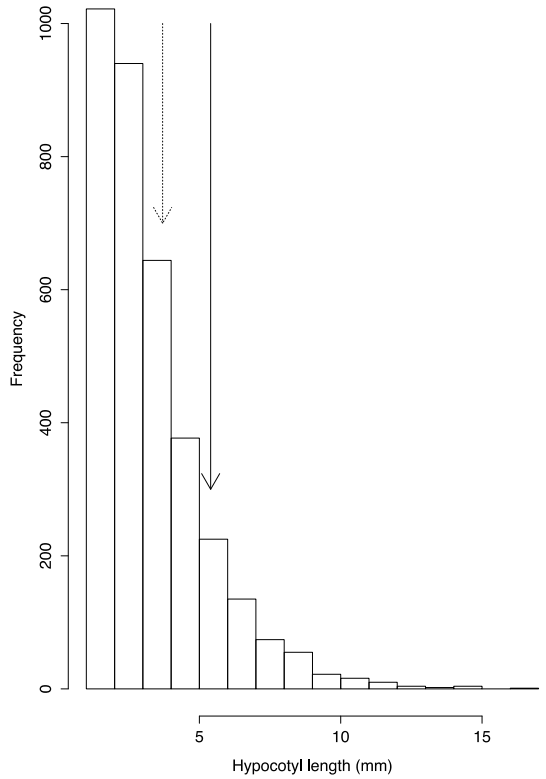

Figure 1.7 Histogram of hypocotyl lengths for 298 Swedish mapping lines (n=3531). Solid arrow indicates mean of coarse sand beach lines, dotted arrow indicates mean of other lines.

Multivariate GWA mapping for hypocotyl length resulted in several SNPs meeting the 5 x 10^{-6} threshold (Figure 1.8). In addition, a peak on chromosome 2 was identified in the univariate analysis for hypocotyl length on soil, a peak on chromosome 4 was identified in the univariate analyses for hypocotyl length on both soil and sand, and two SNPs on chromosome 5 were identified in the univariate analysis for hypocotyl length on sand (Figure 1.9). Each univariate analysis also had SNPs meeting the threshold that were not identified in the multivariate analysis.

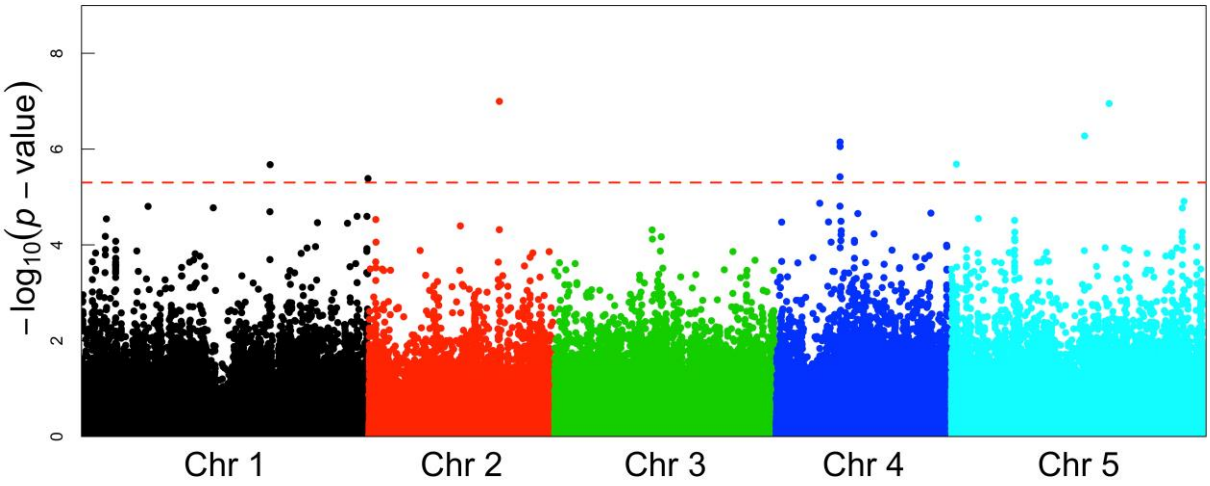

Figure 1.8 Manhattan plot of multivariate GWA results for the hypocotyl length phenotype. Dotted red line is Bonferroni adjusted significance threshold

Appendix A Table 4 presents candidate genes found in 20 kb windows centered on each of the SNPs that reached the threshold. The most promising candidate gene identified in this scan is *SPA2*, located under the chromosome 4 peak seen in the multivariate analysis and the univariate analyses for hypocotyl length in soil and sand. Five of the twelve accessions from the focal beach population bear the minor allele at the most highly associated SNP, as well as another line found near the water in northern Sweden.

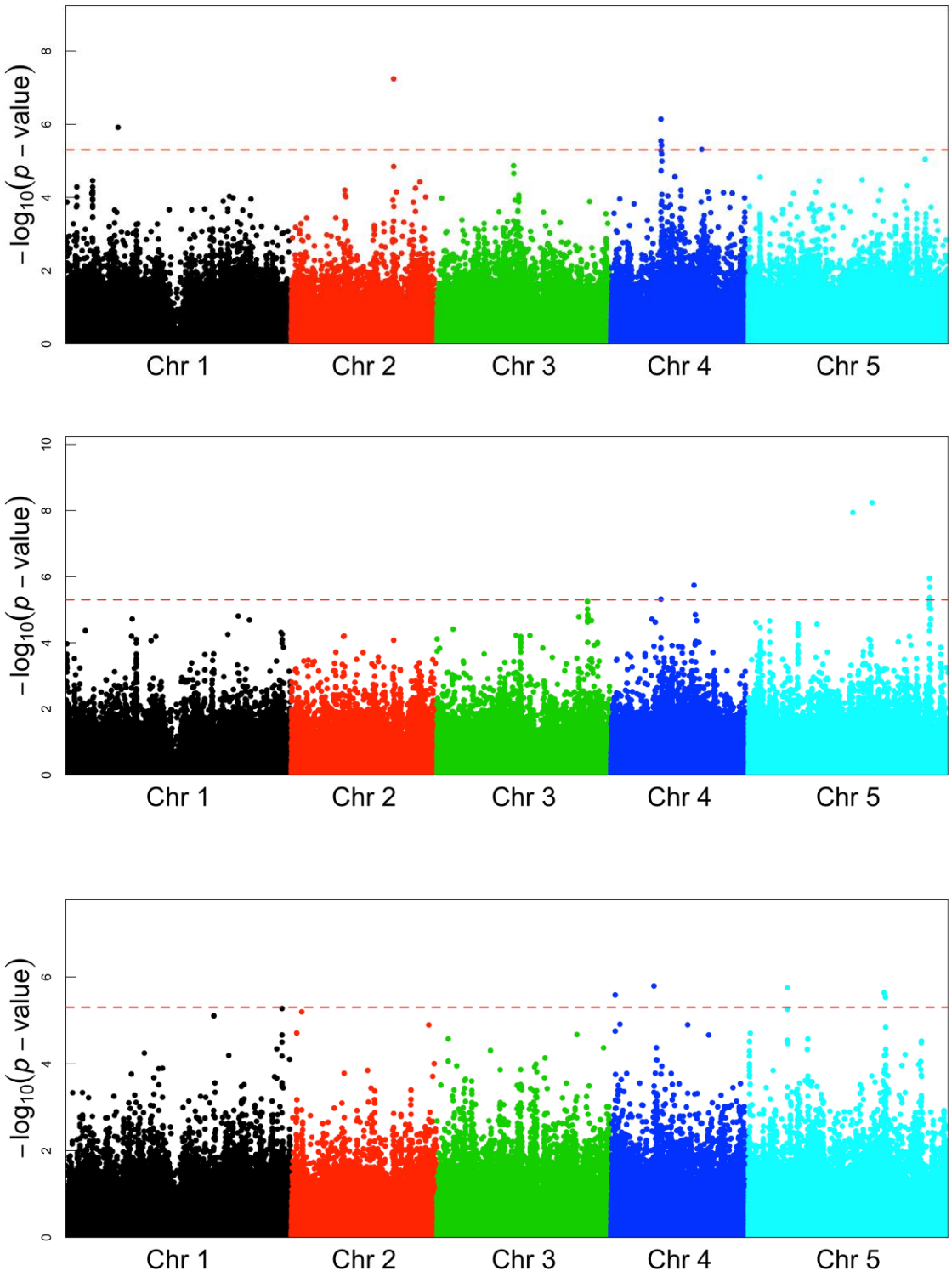

Figure 1.9 Univariate GWA results (a) hypocotyl length in soil (b) hypocotyl length in sand (c) hypocotyl length in pebbles. Dotted line is Bonferroni adjusted significance threshold.

SPA2 (suppressor of phytochrome A) is implicated in the repression of photomorphogenesis in darkness (Laubinger & Hoecker 2003, Laubinger et al. 2004, Fittinghoff et al. 2006, Roulauffs et al. 2012). One of four *SPA* family proteins that complex with COP1, SPA2 is the only one to be almost completely inactivated in light conditions (Balcerowicz et al. 2011). Plants bearing knockout mutations at multiple *SPA* loci exhibit constitutive photomorphogenesis in darkness (Laubinger et al. 2004). Coarse sand beach plants, in contrast, do not differ from fine sand beach or inland lines when grown in darkness, but produce long hypocotyls in light conditions. Therefore, it is possible that an overactive variant of SPA2 could contribute to the observed phenotype. Further exploration of this peak is described in Chapter 2. No complementary peak was found in the PBS scan in this region, possibly because the beach population is itself polymorphic at the site.

Trichome density

Despite my earlier anecdotal observations, trichome density did not differ dramatically between coarse sand beach and other populations. Beach lines had a mean number of trichomes per punch of 31.3 (sd = 15.3), while the remainder of the lines had a mean trichome number of 24.9 (sd = 14.5). While this is a significant difference (Wilcoxon rank sum test p-value = 0.0006), the beach line mean does not fall into the tail of the distribution of trichome numbers (Figure 1.10).

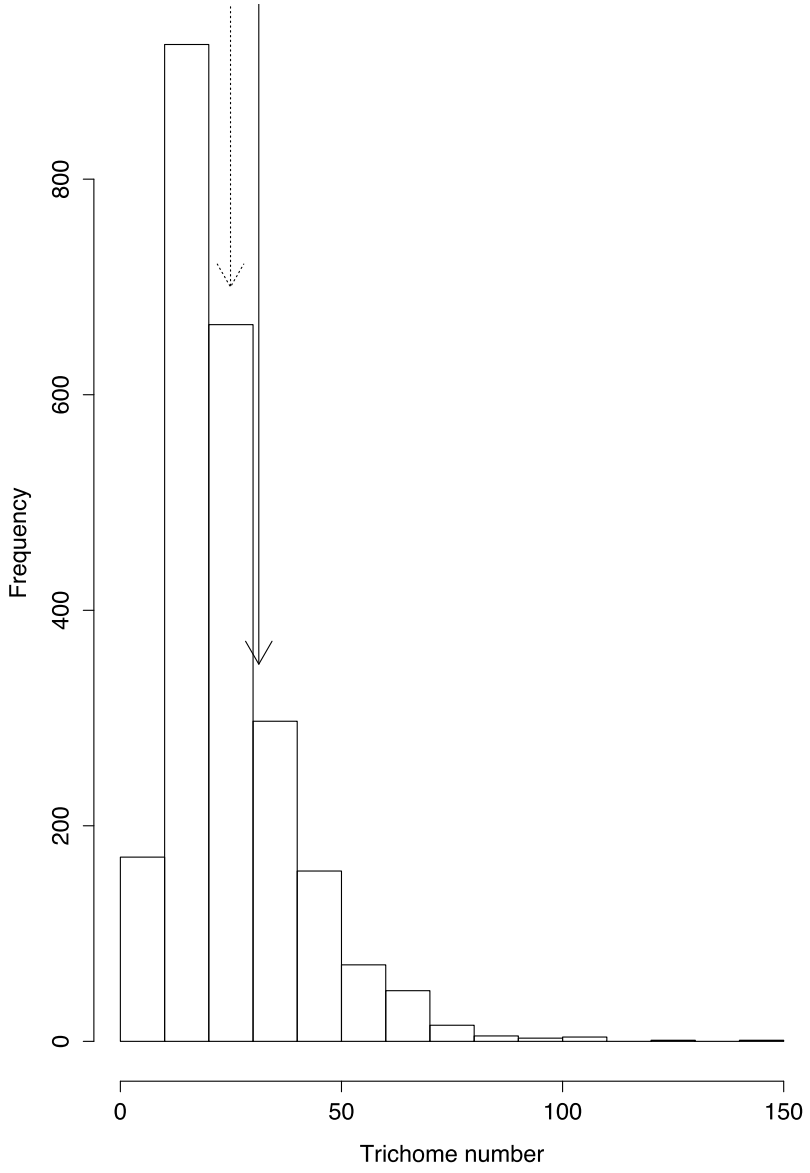

Figure 1.10 Histogram of trichome number for 294 Swedish mapping lines (n=2363). Solid arrow indicates mean of coarse sand beach lines, dotted arrow indicates mean of other lines.

GWA mapping for trichome density resulted in a clear peak on chromosome 2, as well as SNPs that reached the 5 x 10^{-6} threshold on chromosomes 1 and 4 (Figure 1.11). Candidate genes in a 20kb window around the SNPs that reached the threshold (Appendix A Table 5)

include *TCL1* and *ETC2*. The identification of *TCL1/ETC2* as a candidate gene for trichome density corresponds to the results of mapping studies in other *A. thaliana* genotypes (Hilscher et al. 2009, Bloomer et al. 2014), suggesting that the genetic architecture underlying trichome density is shared across many accessions. However, both coarse sand beach and other lines display polymorphism at this locus. Moreover, it is not clear why quantitative results did not match my initial observations of large trichome density differences between beach and inland lines; below I discuss two possible contributing factors.

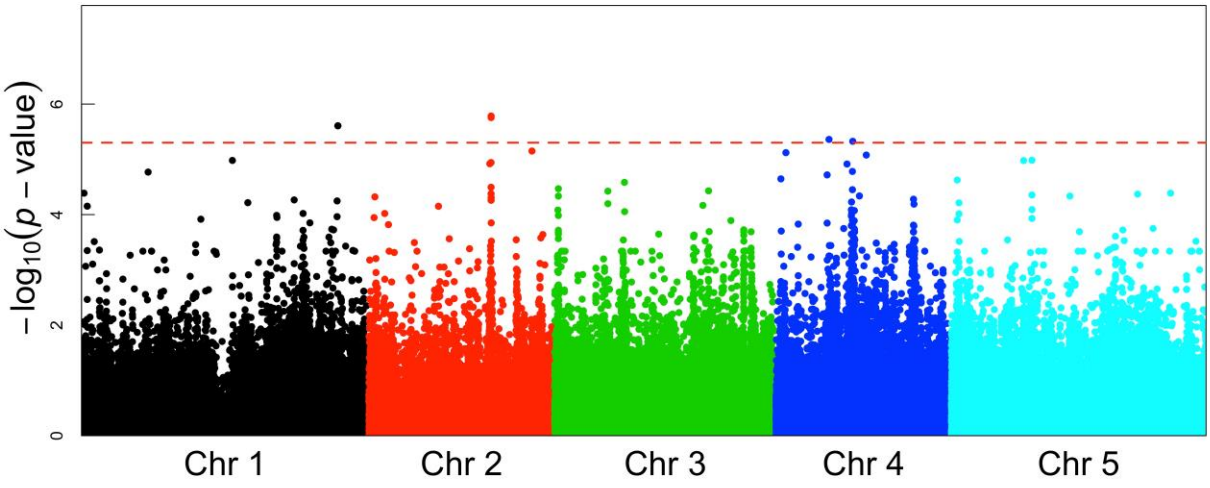

Figure 1.11 Manhattan plot of GWA results for trichome density phenotype. Dotted red line is Bonferroni adjusted significance threshold.

Trichome density is dependent on leaf identity (early leaves have none or few trichomes; Telfer et al. 1997,) and leaf age (juvenile leaves have higher density than fully expanded adult leaves; Larkin et al. 1996, Mauricio 2005). Coarse sand beach plants differ from inland conspecifics in developmental timing, with leaf production lagging. Leaf size and shape also differ across accessions; early leaves of coarse sand beach plants are much

narrower than inland lines. All these factors make it difficult to choose times and leaf identities that reflect developmentally equivalent leaves in all accessions. I chose to sample the 7^{th}, 8^{th} and 9^{th} leaves produced, a fixed time after germination, but it is not clear how alternative choices would affect trichome density measurements.

Another possible explanation for the discrepancy between observation and measurement is differences in trichome size. Larger trichomes may lend beach plants a hairier, more silvery appearance than inland conspecifics, leading to an overestimation of the difference in densities. I did not measure trichome size, but I did observe variation within the mapping population.

Seed weight

As expected, seed weight was extreme in beach populations (Figure 1.12). Beach lines had an average seed weight per 500 of 16.3 mg (sd = 2.5 mg) while the remainder had an average weight per 500 of 11.8 mg (sd = 1.8 mg; Wilcoxon rank sum test p-value < 0.0001).

Associations reaching the 5×10^{-6} threshold were detected in a GWA for seed weight on chromosomes 1, 2, and 3 (Figure 1.13). Appendix A Table 6 shows candidate genes in a 20 kb window around each SNP, but there are no candidates among them previously associated with seed size phenotypes.

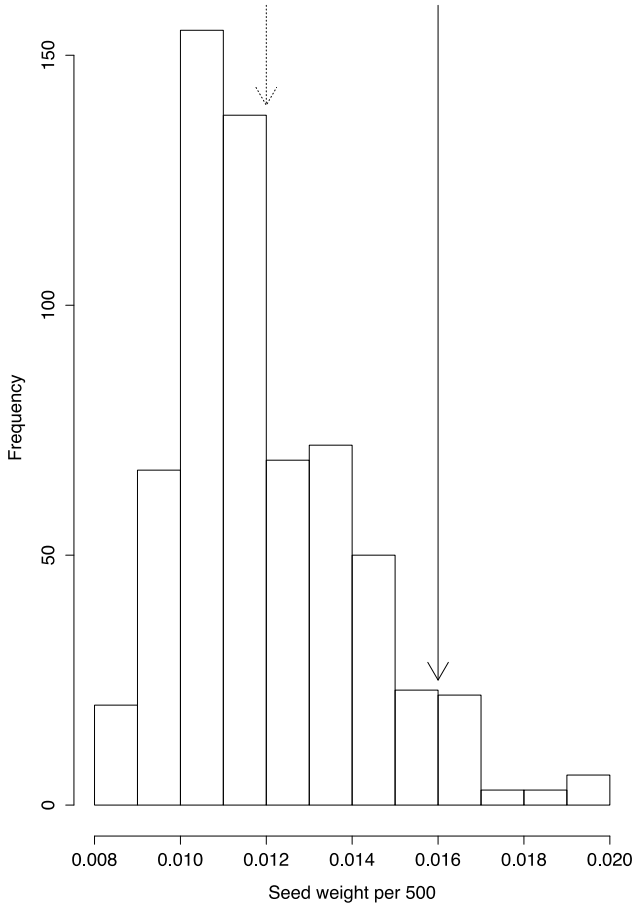

Figure 1.12 Histogram of seed weight per 500 for 298 Swedish mapping lines (n=628). Solid arrow indicates mean of coarse sand beach lines, dotted arrow indicates mean of other lines.

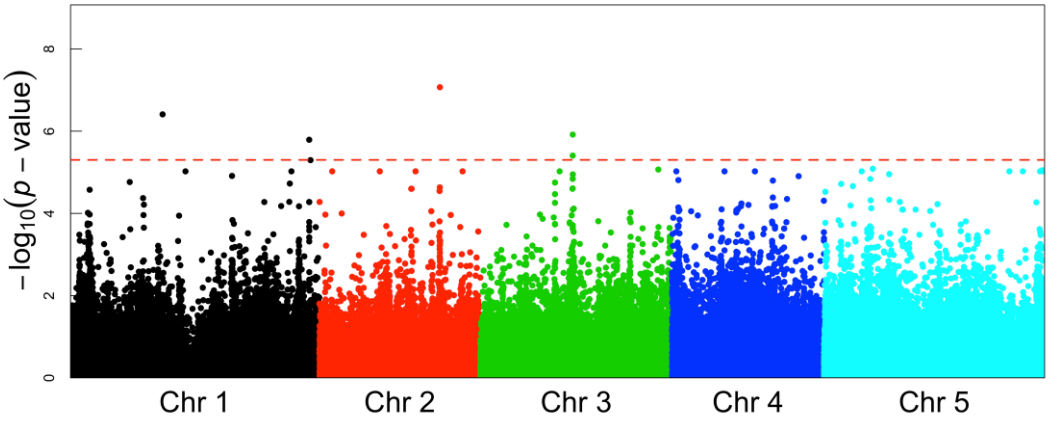

Figure 1.13 Manhattan plot of GWA results for seed weight per 500 phenotype. Dotted red line is Bonferroni adjusted significance threshold.

These GWA results indicate that seed weight is a complex polygenic trait; thus they present no clear picture of the underlying genetic architecture. However, the observation that coarse sand beach lines were differentiated from fine sand beach and inland lines by seed weight is suggestive of adaptive differences in seed size associated with the beach. Thus, combining data from PBS and GWA scans provides another approach by which to query putative adaptation in beach plants.

1.4.3 Combining PBS and GWA results

If a mapped phenotype is under selection, mean PBS should increase as GWA results are restricted to SNPs meeting more conservative significance thresholds. Plots of mean PBS show that SNP sets associated with hypocotyl length and seed weight exhibit higher PBS values with increasingly stringent p-value cutoffs (Figure 1.14a). Sets associated with trichome density, however, show only a very modest increase in mean PBS with

increasingly stringent p-value cutoffs. Even the set with the most conservative threshold has a mean PBS only marginally higher than the mean PBS of the whole genome.

Plots of mean p-value of SNP sets with various PBS cutoffs show the same pattern (Figure 1.14b). As the value of the PBS cutoff increased, the average p-value of the set of SNPs above the threshold decreased when using data from hypocotyl length or seed weight GWA, but not when using data from trichome density GWA. These two results suggest that hypocotyl length and seed weight are important adaptive phenotypes in the focal beach population, but that trichome density is not.

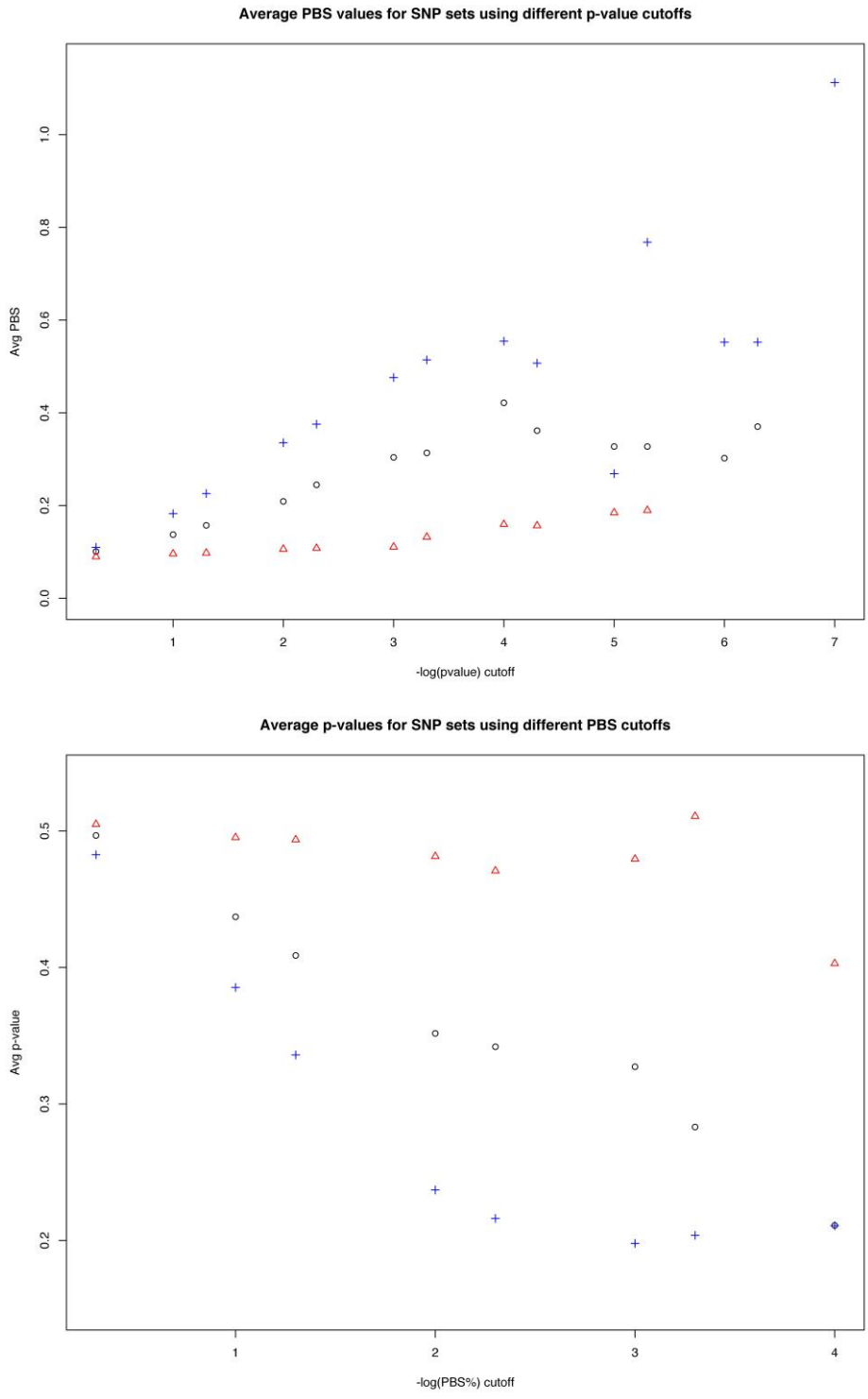

Figure 1.14 (a) Average PBS value increases in SNP sets with increasingly stringent p-value cutoffs for hypocotyl length (circles) and seed weight (crosses) but not trichome density (triangles). (b) Average p-value decreases in SNP sets with increasingly stringent PBS cutoffs for hypocotyl length (circles) and seed weight (crosses) but not trichome density (triangles).

I compared the observed mean PBS and p-values using a single set of thresholds (top 0.1% PBS; p-value<0.0005) to null distributions of random means calculated as described in the Methods. The observed mean PBS of the top hypocotyl length and seed weight SNPs was outside of the range described by the null distribution. That is, none of the random sets of SNPs had a mean as extreme as the true value. The observed mean PBS of the top trichome density SNPs was in the 0.01% tail of the null distribution (Table 1.2). Similar patterns were observed for the mean p-values of the top PBS SNPs. The observed mean p-values were outside the null distributions for hypocotyl length and seed weight, whereas the observed mean p-value for the top PBS SNPs for trichome density was not significantly different from the null average (Table 1.3). Again, this suggests that hypocotyl length and seed size are important adaptive phenotypes in the population of interest, while trichome density is not.

	Full data set		**GWA w/o beach**	
	# SNPS <0.0005	Mean PBS	# SNPS <0.0005	Mean PBS
Hypocotyl length	235	0.31	240	0.14
Trichome density	246	0.13	258	0.089
Seed weight/500	290	0.51	252	0.16

Table 1.2 Real mean PBS values of GWA SNPs with p-values below 0.0005 using full data set and GWA is performed without focal beach population (see text).

	Full data set	GWA w/o beach
	Mean p-value	Mean p-value
Hypocotyl length	0.33	0.51
Trichome density	0.49	0.53
Seed weight/500	0.20	0.39

Table 1.3 Real mean p-values of SNPs in the top 0.1% of PBS values using full data set (n=192) and when GWA is performed without focal population (see text).

Distributions of average PBS values calculated with random SNP sets for hypocotyl, trichome density, and seed weight were very similar, each having a mean of 0.089, (Figure 1.15), even though the underlying structure of these SNP sets varied (number of peaks, distribution across the genome). The similarity of the null distributions suggests that there was no effect of this structure. P-value null distributions were also very similar to each other (Figure 1.16).

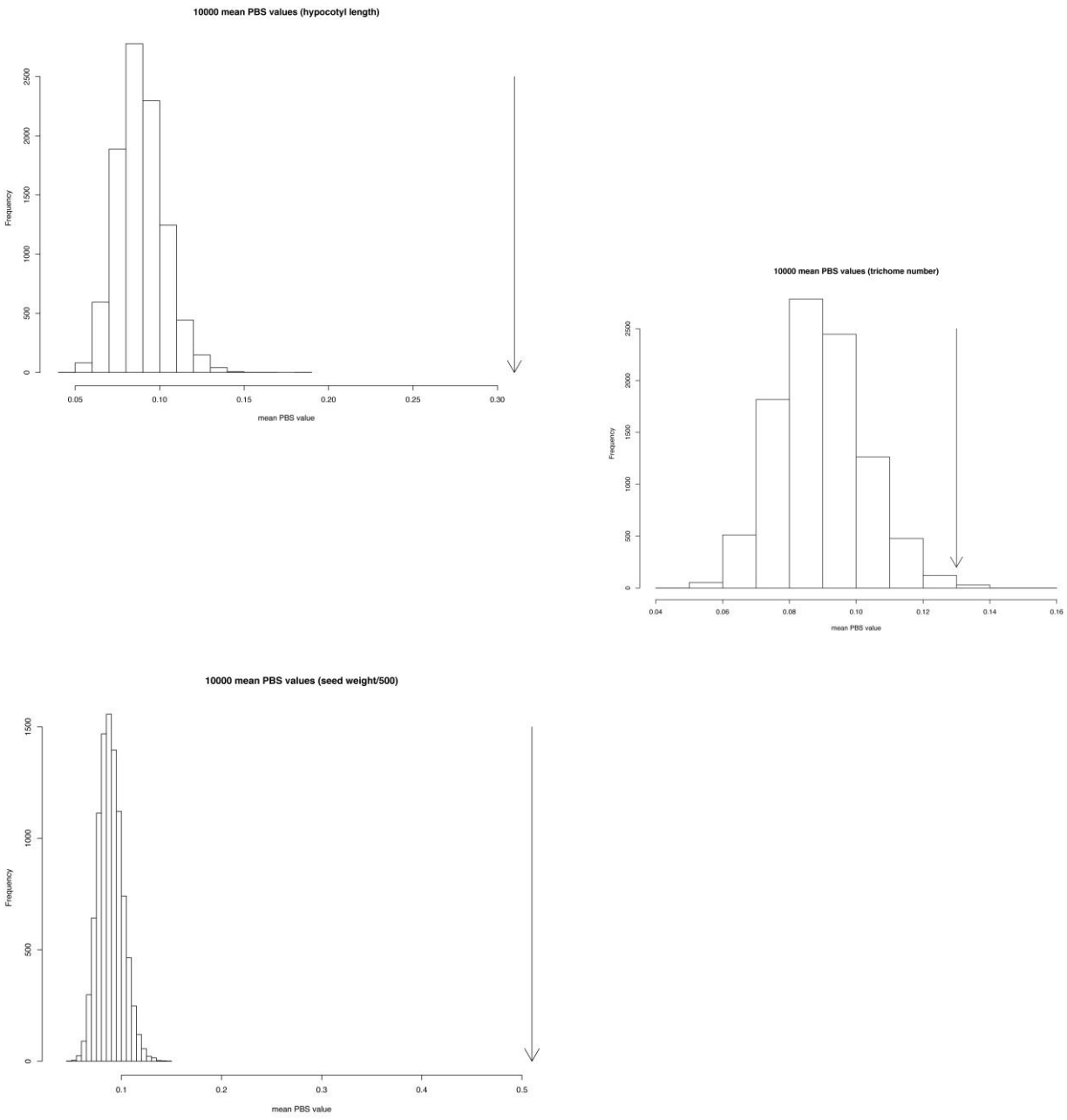

Figure 1.15 Distributions of mean PBS values for random SNP sets generated as described in methods using (a) SNPs from hypocotyl length GWA (n=235) (b) SNPs from trichome density GWA (n=246) (c) SNPs from seed weight GWA (n=290). Arrows are real mean PBS values of the top SNP set for each phenotype.

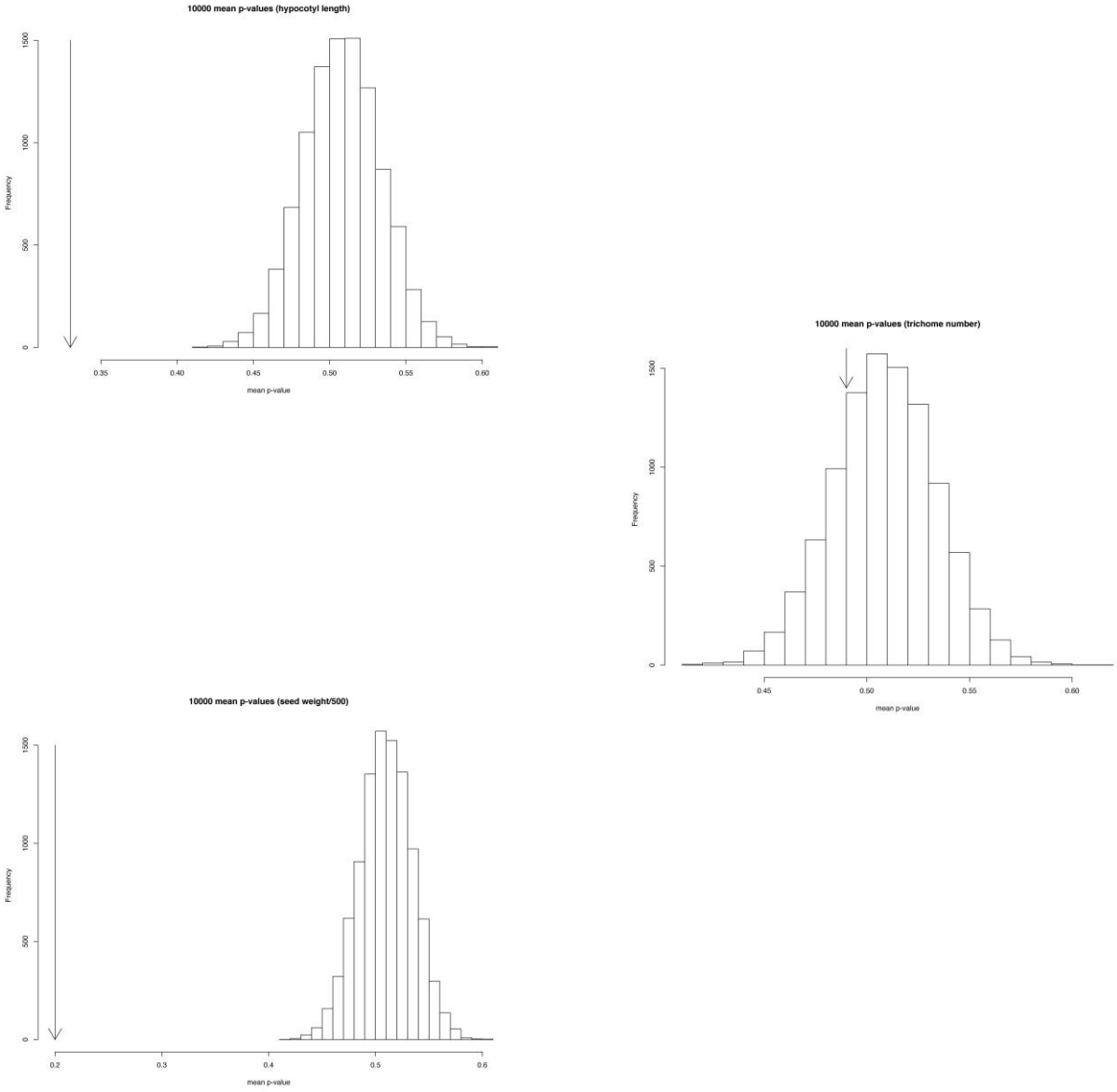

Figure 1.16 Distributions of mean p-values for random SNP sets (n=192) generated as described in methods using (a) p-values from hypocotyl length GWA (b) p-values from trichome density GWA (c) p-values from seed weight GWA. Arrows are real mean p-values of top PBS SNP set using each GWA result.

Hypocotyl length, trichome density, and seed weight were chosen because they appeared to differentiate the population of interest from closely related conspecifics. However, phenotypes that are strongly differentiated between populations have the potential to produce false positives in GWA analysis, because alleles private to one population will be correlated with the phenotype even if there is no true causal association. The population structure correction applied by EMMAX is based on genome-wide average levels of relatedness, but specific loci may display a greater signature of population structure, such that corrected results are imperfect. Furthermore, it seems likely that false positives induced by population structure in the GWA results would also have high PBS values, since PBS values are the result of tests for regions that (a) are shared among individuals from the population of interest and (b) differentiate the population from its sister group. Such a phenomenon could explain the different responses of the phenotypes; since trichome density was not found to differ dramatically between coarse sand beach and other lines whereas both hypocotyl length and seed size did.

In order to test whether uncorrected population structure is the cause of the significant results in the combined GWA and PBS data, I removed the focal population from the GWA mapping panel and repeated the GWA analysis. Since the second set of GWA do not contain the focal population, genotypes shared exclusively by these accessions (likely to have high PBS scores) cannot be causing spurious results. Indeed, this is an overly conservative approach, since rare genotypes that are truly associated with the traits of interest but are restricted to beach populations have been excluded.

Conclusions drawn from the reanalysis are qualitatively similar to the original analysis for seed size and trichome density (Figure 1.17 -1.19, Table 1.2-1.3). The observed mean PBS value of the top GWA SNPs and the observed mean p-value of the top PBS SNPs still fell in the 1% tail of the null distribution for seed weight, while there was no significant difference between the observed mean PBS or p-value and the null distribution mean for trichome density.

The observed mean PBS value of the top hypocotyl length GWA SNPs still fell in the 1% tail of the null distribution (Figure 1.18, Table 1.2), however, there was no significant difference between the observed mean p-value of the top PBS SNPs and the null distribution mean (Figure 1.19, Table 1.3). That is, high PBS SNPs had lower than average p-values in the hypocotyl length GWA in the full analysis, but not in the reanalysis. This result could reflect false positives; alternatively, it could represent SNPs genuinely associated with hypocotyl length that have private alleles in the beach population. Such SNPs would be excluded from a reanalysis excluding the beach population, because they are monomorphic in the remaining accessions. There is some evidence to suggest that this is the case. For example, SNPs in the peak on chromosome 4 in the full GWA carry the minor allele only in beach populations, and this locus has been demonstrated to be genuinely associated with hypocotyl length (see Chapter 2). In addition, SNPs highly associated with hypocotyl length in the reanalysis still have higher than average PBS values.

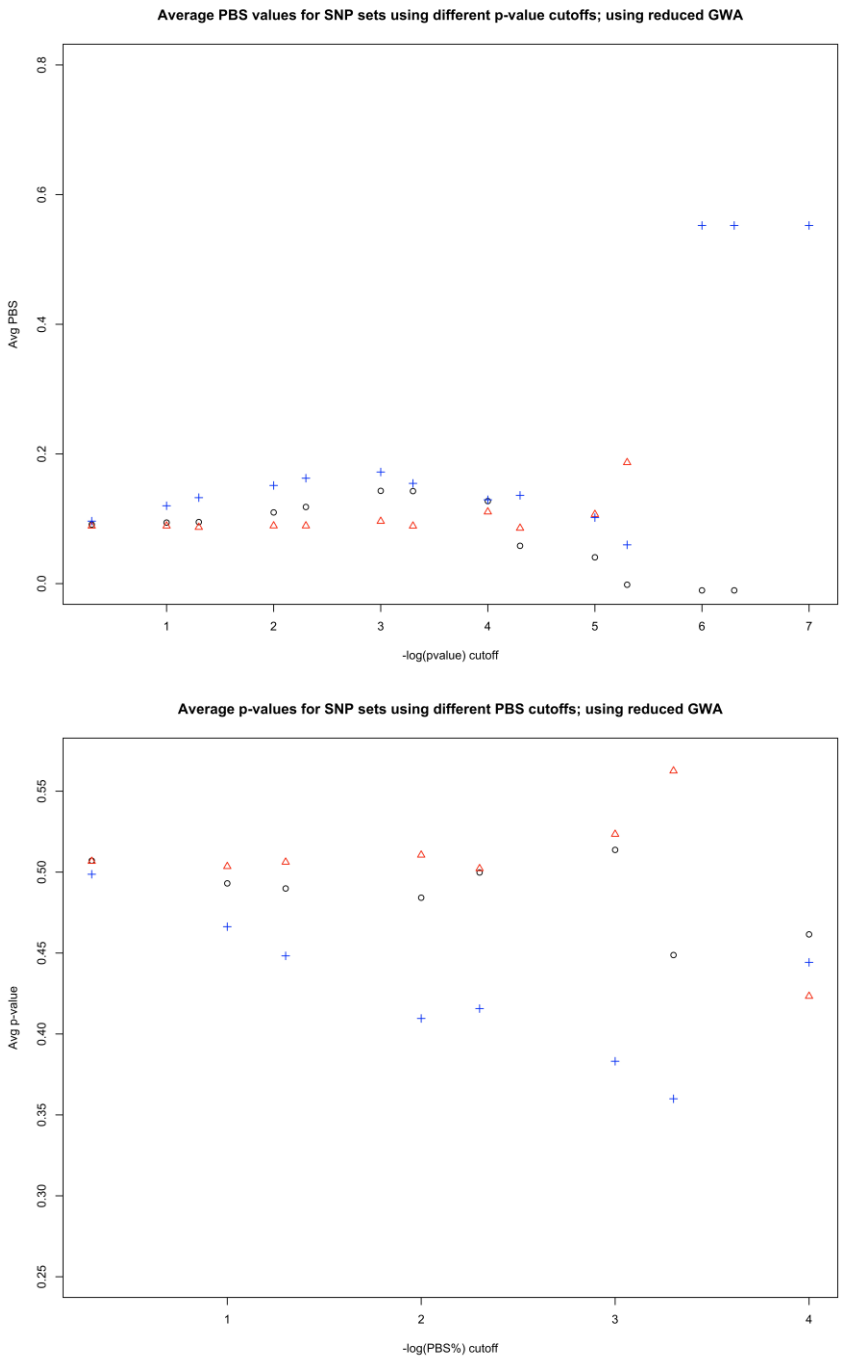

Figure 1.17 Analogous to Figure 1.14 above, but with GWA performed without the focal beach population. (a) Average PBS value does not dramatically increase in SNP sets with increasingly stringent p-value cutoffs for hypocotyl length (circles), trichome density (triangles), or seed weight (crosses), except for with very small (n=2) seed weight SNP sets. (b) Average p-value decreases in SNP sets with increasingly stringent PBS cutoffs for hypocotyl length (circles) and seed weight (crosses) but not trichome density (triangles), except for when using the most stringent cutoff.

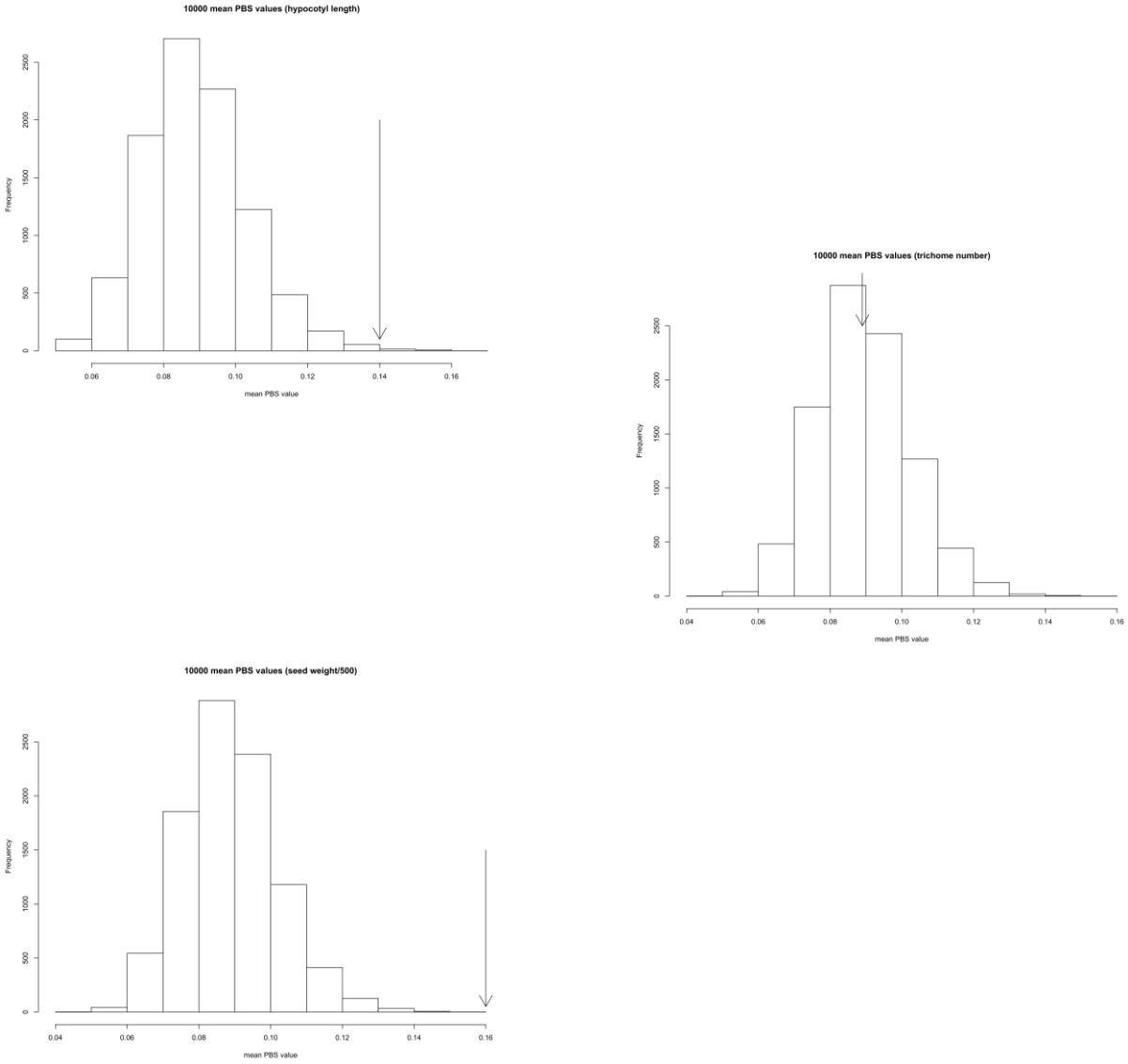

Figure 1.18 Performing GWA without the focal beach population as described in Methods, distributions of mean PBS values for random SNP sets generated as described in methods using (a) SNPs from hypocotyl length GWA (n=240) (b) SNPs from trichome density GWA (n=258) (c) SNPs from seed weight GWA (n=252). Arrows are observed mean PBS values of the top SNP set for each phenotype.

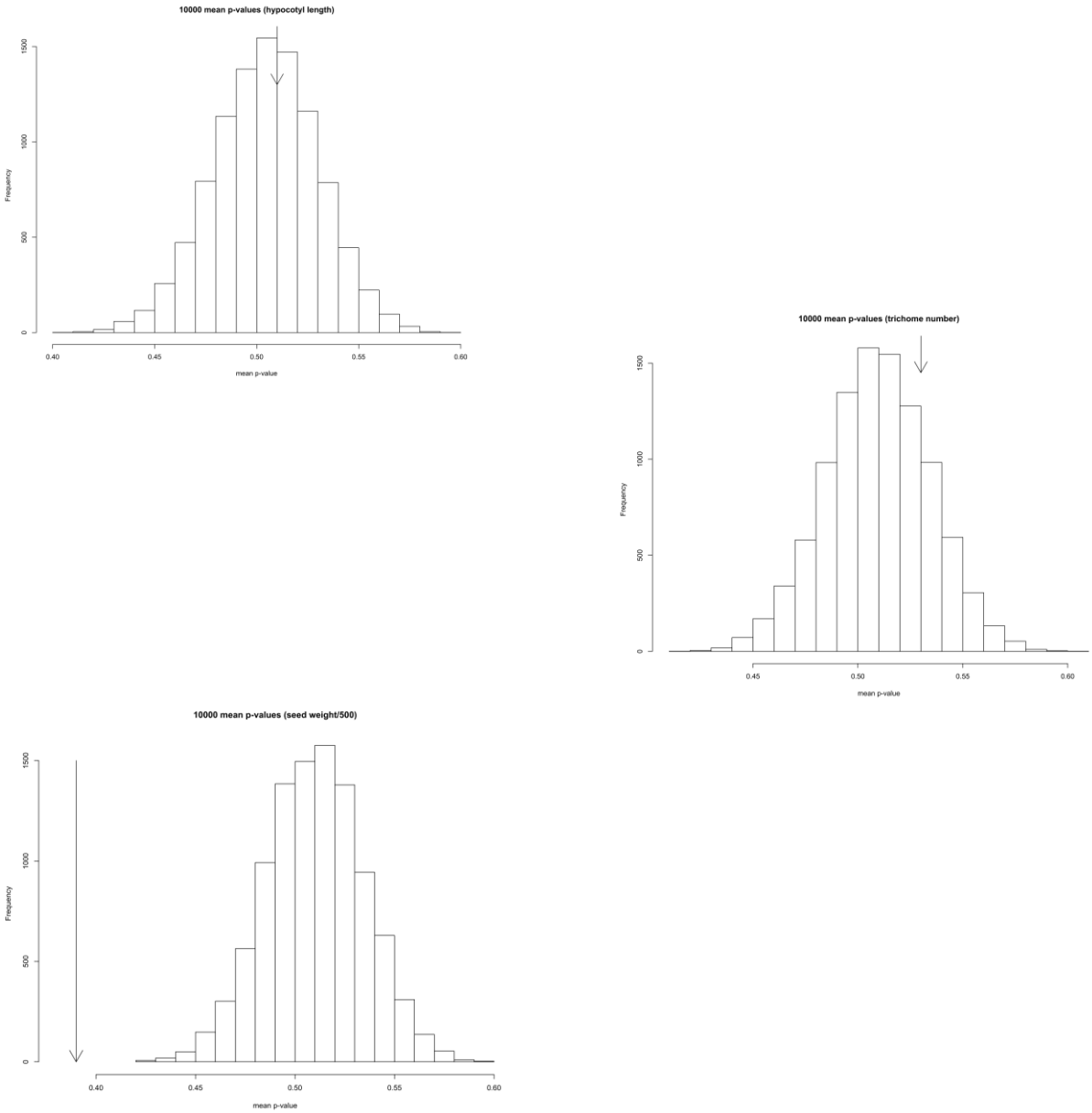

Figure 1.19 Performing GWA without the focal beach population as described in Methods, distributions of mean p-values for random SNP sets (n=199) generated as described in methods using (a) p-values from hypocotyl length GWA (b) p-values from trichome density GWA (c) p-values from seed weight GWA. Arrows are observed mean p-values of top PBS SNP set using each GWA result.

Thus, the results confirm that genetic variation associated with hypocotyl length and seed size appears to be under selection in the population of interest. It is interesting to note that for both phenotypes, the observed mean PBS values were less extreme than with the full GWA. That is, some SNPs with high PBS values had low p-values in the full GWA, but not in the reduced GWA. It seems likely that the beach population has private alleles at these SNPs and that they are unassociated with either phenotype (that is, in the full GWA they show false association with the phenotype due to uncorrected population structure). Thus, the reanalysis shows that uncorrected population structure contributed to the extremity of the original results, but that removing its effect still yields significant evidence of selection on hypocotyl length and seed weight.

1.5 Discussion

1.5.1 Adaptation in an *A. thaliana* beach population

I used a combination of methodologies to explore adaptation to coarse sand beaches in a Swedish population of *Arabidopsis thaliana.*

Of the three putative phenotypes identified, two appear to be adaptively important in this beach population. The dramatic increase in mean PBS and dramatic decrease in mean p-value in increasingly restrictive sets of GWA (and PBS) results for hypocotyl length and seed size, as well as the extreme mean PBS and p-values compared to null distributions, suggest that hypocotyl length and seed size are phenotypes under selection in the beach

population. In contrast, while trichome density may be under some level of selection, as suggested by a slight increase in average PBS value when restricting GWA results and a mean PBS value in the far right tail of the null distribution, it does not appear to be an important adaptive trait in these populations, as it shows no decrease in p-value with restriction of PBS results and no significant difference between mean p-value and mean of the null distribution.

Strong candidate genes were identified for two of the three phenotypes. *SPA2,* a suppressor of phytochrome A, falls under a clear peak in GWA mapping of hypocotyl length. This gene encodes a protein shown to be important for growth in darkness. For beach plants whose seeds are smaller than the coarse sand grains of the substrate, early hypocotyl elongation in both light and dark conditions might be necessary for seedlings to emerge from the sand. Interestingly, candidate genes from the agnostic genome wide PBS scan also yielded candidate genes for hypocotyl length. ARR16 and TPST have both been linked to hypocotyl elongation phenotypes. Thus, a combination of approaches suggests the strong importance of hypocotyl elongation for adaptation to this beach.

TCL1/ETC2, genes identified in previous studies of trichome density (Hilscher et al. 2009, Bloomer et al. 2014), fall under a peak on chromosome 2 identified in the GWA for trichome density. However, beach and inland lines both display polymorphism in this region, suggesting that while these genes are determinants of trichome density in these populations, they do not differentiate beach and inland lines. This corresponds to the observations that trichome density does not appear to be under selection in the beach

population and that trichome density in beach lines is not dramatically greater than in inland lines.

Strong candidate genes were not identified for seed size. This is perhaps not surprising considering the complex polygenic nature of the seed size trait. Nevertheless, the combination of PBS and GWA data does suggest that seed size is an important trait in this population.

The combination of reverse and forward approaches applied here, as well as the synthesis of data from both approaches, provides insight into the phenotypes and genotypes important to *A. thaliana* on coarse sand beaches and the environmental selective drivers likely acting in this population, such as burial.

1.5.2 Combining PBS and GWA results provides new information

PBS and GWA scans are two different approaches to the problem of investigating adaptation to a particular environment. While PBS requires no *a priori* assumptions about what genes or phenotypes are important and is thus less susceptible to bias, it is often unclear how to define thresholds of significance, and choosing genes for follow up is a largely subjective process, especially if many candidate genes are poorly annotated. Thus, even when candidate regions are clearly differentiated between focal and sister populations, the identity and function of the genes driving the adaptive differentiation may remain unclear.

GWA, on the other hand, utilizes knowledge about the organism or environment to choose candidate phenotypes, leveraging existing information from observation or previous studies. However, while GWA can clearly show the underlying genetic architecture of relatively simple traits, the results of mapping complex polygenic traits are typically much less easily interpretable. Moreover, in neither case can GWA provide information about whether a particular variant is or has been under selection in a population of interest.

Here, I show that combining PBS and GWA results can provide information about phenotypes under selection in a population of interest that neither approach can provide alone. In some cases, this method confirms results from other experiments. For example, here I show that hypocotyl length is extreme in beach populations and identify a strong peak in the GWA results. Beach and inland populations are differentiated at the top SNPs in this peak, which falls over a strong candidate gene, *SPA2.* In Chapter 2, I show that hypocotyl length and this genetic variant in particular confer fitness advantages in a beach-like environment. That the combination of PBS and GWA data suggest that there is selection on hypocotyl length on the beach confirms conclusions drawn from these data. However, the GWA peak over *SPA2* is not found in the PBS results, suggesting that additional genetic variation underlying hypocotyl length is under selection. Thus, even in this case, the synthesis of GWA and PBS results provides additional insight into the selective process in question.

In other cases, the combination of PBS and GWA results provides support for a hypothesis that is not well supported by other types of evidence. Combining PBS and GWA data in this case provided evidence that seed size is under selection in the beach population, a conclusion that could not have been reached with either PBS data alone, because top PBS SNPs were not closely linked to genes annotated as important for seed size, or GWA data alone, because the complex polygenic nature of the trait made it impossible to identify strong candidate genes. Support for the hypothesis of selection on seed size not only helps more fully describe the suite of phenotypic characteristics important for life on a beach, it also provides a rationale for additional mapping experiments with the goal of identifying the genetic variants underlying seed size differences in beach and inland lines.

1.5.3 Potential applications of this methodology

These results suggest that the combination of F_{st} based statistics and association mapping results can be fruitful even when phenotypic data is unavailable for the population of interest. As long as the genetic basis of the trait is shared – as is likely if association mapping was conducted in closely related lines – this method has the ability to identify phenotypes under selection in a population of interest. With the increasing availability of genomic data, phenotyping is often the most difficult and time-consuming step in an investigation. While this method cannot obviate the need for phenotyping altogether, it may provide a way to screen candidate phenotypes *a priori* or generate hypotheses about important drivers of local adaptation at sites of interest.

For example, Atwell et al. (2010) compiled a dataset of 107 phenotypes related to flowering, resistance, development, and ionomics and performed GWA analysis in a worldwide set of 96 accessions, plus an additional 95 accessions for which flowering phenotypes were available. For lines within this geographic range for which genomic data is available, whether or not they were included in the original phenotyping set, this method could suggest which traits are under selection. For researchers interested in identifying drivers of local adaptation, a genomic approach may allow for selection of good candidates prior to phenotyping. A small set of candidate phenotypes could then be assessed in the field or greenhouse.

Environmental data have been used as if they were phenotypes in association mapping studies to find genetic variants associated with particular climatological factors (Hancock et al. 2011). Such factors are likely to be common drivers of local adaptation. Thus, where climate, soil, or environmental data is available at an appropriate grain size, association mapping and PBS data could be used to suggest environmental factors linked to adaptive pressures in local populations.

Here, using a Swedish beach population of *Arabidopsis thaliana,* I have demonstrated that a combination of genome wide approaches to exploring adaptation, as well as a synthesis of data from these approaches, can be used to confirm hypotheses about phenotypes under selection and to generate candidate loci for follow up. I show that the synthesis of PBS and GWA data is a useful approach to screen phenotypes or environmental drivers thought to

be important to adaptation to a particular site and can provide evidence of selection on a phenotype even when mapping approaches fail to identify gene candidates.

CHAPTER 2: HYPOCOTYL LENGTH AND ASSOCIATED GENOTYPES ARE LINKED TO FITNESS ON BEACHES

2.1 Abstract

Due to the possibility of burial by sand, elongated hypocotyls are considered an important trait for seedling survival on beaches. I noticed that *Arabidopsis thaliana* plants from coarse sand beaches had consistently longer hypocotyls than lines from inland sites or fine sand beaches when grown in soil in the greenhouse. To demonstrate that hypocotyl length is a genetically based adaptive trait in these populations, I confirmed that hypocotyls from coarse sand beach lines were significantly longer than inland and fine sand conspecifics across a variety of growth conditions. I performed a genome wide association study on the hypocotyl length phenotype using a set of lines from Sweden and identified a peak on chromosome 4 near the candidate gene *SPA2*. I examined the genetic variants under this peak and found a correlation between the most highly associated SNP in the peak and hypocotyl length in a global set of accessions and in a set of RIL lines. I also demonstrated that the variant is associated with increased seedling emergence from burial in the original panel of lines from Swedish beaches as well as in the global panel. I conclude that this variant is associated with a genetic region involved in hypocotyl elongation and that this region confers a fitness advantage related to seed burial in beach populations. I speculate that the gene responsible is *SPA2*, a known regulator of growth in darkness.

2.2 Introduction

2.2.1 Beaches as candidate sites for local adaptation

Naturalists and evolutionary biologists have long been fascinated by organisms' adaptations to specific environments. Such adaptations are evident both at the species level and in ecotypes within species. Within-species variation of this sort has provided insight into the selective forces characteristic of various ecosystem types and led to the identification of the genetic architecture underlying adaptive traits. Examples in plants include tolerance to heavy metals (Gregory & Bradshaw 1965, Allen & Sheppard 1971, Macnair 1983), to serpentine soils and other edaphic factors (Rajakaruna 2010, Sherrard & Maherali 2012, Barry 2013), and to coastal environments (Turesson 1922, 1925; Jordan 1992; Nagy & Rice 1997, Morrison 2002; Voronkova et al. 2008; Lowry et al. 2009, Lowry & Willis 2010, Bischoff & Hurault 2013; Busoms et al. 2015).

Coastal ecotypes are common in plants and were among the first noted by ecologists (Turesson 1922, 1925). Beaches are good candidates for the production of locally adapted phenotypes due to a suite of abiotic stressors including low-nutrient, sandy soils, salt spray, high winds, and dramatic temperature fluctuations (Maun 1994, McLachlan & Brown 2006, Voronkova et al. 2008, Maun 2009). As a result, coastal populations often possess phenotypes distinct from inland conspecifics, including dwarfism, succulence, cuticles and salt excluding glands, rapid growth, light coloration, reflective leaves, transpirational cooling, long hypocotyls, a high density of trichomes, and large seed size (Turesson 1922,

1925; Morrison 2002; Voronkova et al. 2008; Lowry et al. 2008). Moreover, coastal phenotypes are demonstrably associated with fitness benefits on beaches or in beach-like environments. Reciprocal transplant experiments commonly reveal a home site advantage for coastal populations (Jordan 1992, Nagy & Rice 1997, Lowry et al. 2008, Bischoff & Hurault 2013), and coastal plants exhibit a greater tolerance to stressors like drought and salt in controlled common garden experiments (Lowry et al. 2009, Busoms et al. 2015).

Beaches and dunes are one of many global ecosystems occupied by the model plant *Arabidopsis thaliana.* Several studies have suggested that local adaptation in *A. thaliana* is common (Weinig et al. 2003, Fournier-Level 2011, Hancock et al. 2011, Kniskern et al. 2011, Agren & Shemske 2012, Stearns & Fenster 2013), and local adaptation to beaches has been shown in two populations. Arany et al. (2009) described home site advantage associated with reduced herbivory of dune *A. thaliana* in the Netherlands, while Busoms et al. (2015) attributed local adaptation in Spanish coastal accessions to salinity tolerance. On the Baltic coast of southern Sweden, long established populations of *A. thaliana* (Figure 2.1) appear phenotypically and phenologically different than inland conspecifics, both in the field and in the greenhouse. Notable phenotypes include features common to beach plants such as larger seeds, dense trichomes, and long hypocotyls. However, the adaptive advantage conferred by these phenotypes has not been experimentally demonstrated.

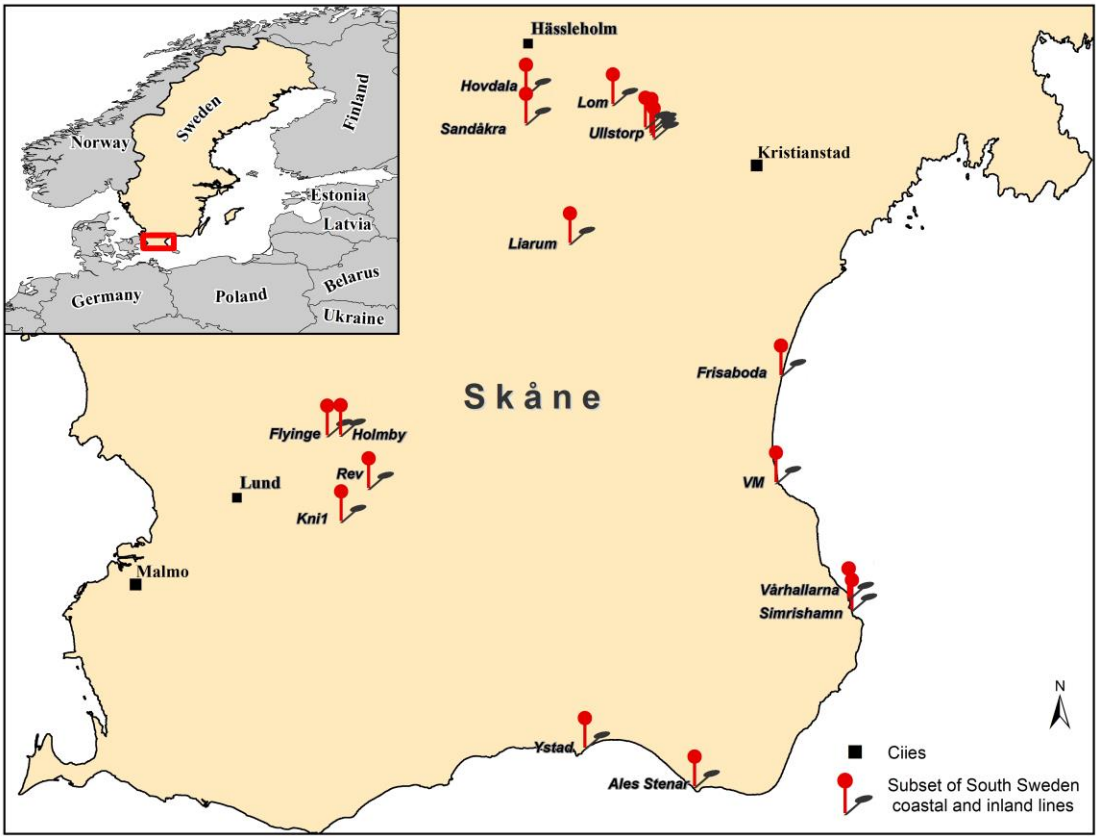

Figure 2.1 Location of beach and inland *A. thaliana* populations in southern Sweden used in the experiments below.

2.2.2 The ecological importance of hypocotyl elongation

Long hypocotyls are an important characteristic of beach plants (Maun 2009). Tolerance to sand burial is a major determinant of plant community zonation on dunes (Van der Valk 1974, Maun and Perumal 1999, Wilson and Sykes 1999, Dech and Maun 2005), and this tolerance is mediated in part by the capacity for hypocotyl elongation (Maun 2009). For example, Chen and Maun (1999) observed a positive relationship between hypocotyl length and burial depth in experimentally buried *Cirsium pitcheri*. Some beach species appear to require burial for optimal germination; burial may allow anchoring to soil or

protect seeds from desiccation on the surface, from extremes of temperature or light, or from insects (Maun & Riach 1981, Maun & LaPierre 1986, Maun 1994). The ability of seedlings to emerge rapidly from sand before inter- or intraspecific competitors also appears to convey a fitness advantage (Zhang & Maun 1990, Maun 2009, Weis et al. 2015).

Sykes & Wilson (1990) measured the length of time mature plants of several species survived when subject either to total darkness or to burial and found a correlation between survival times in the two conditions. They observed that hypocotyl elongation varied in response to both light and burial treatments. Dark grown seedlings of *A. thaliana* exhibit a characteristic etiolated morphology with an elongated hypocotyl, reduced root system, and closed, pale, and underdeveloped cotyledons (also called skotomorphogenic). In contrast, exposure to light results in a short, thick hypocotyl; open, green cotyledons; and a more extensive root system (Boron & Vissenberg 2014).

Mutants that display constitutive or condition-inappropriate skoto- or photomorphogenesis, often measured using hypocotyl length, led to the identification of genes involved in light sensitive signaling pathways. Further work has identified five light sensitive phytochromes (A through E) and a large set of players involved in downstream signaling governing photomorphogenic growth, germination, circadian rhythms, and flowering time (Vandenbussche et al. 2005, Wang & Wang 2015). These include proteins that act in the dark to repress photomorphogenesis and promote hypocotyl elongation (Wang & Wang 2015).

While natural variation in constitutive and light responsive changes in hypocotyl length has been documented in *A. thaliana* (Maloof et al. 2001, Borevitz et al. 2002, Filiault & Maloof 2012), few studies have examined hypocotyl length in an ecological context. Those that do have tended to focus on the plastic shade response exhibited by *A. thaliana* and other plants (Devlin et al. 1999; Callahan & Pigliucci 2002; Schmitt et al. 2003; Pigliucci & Schmitt 1999, 2004; Sessa et al. 2011; Wang et al. 2011). Though shade avoidance has not been explicitly studied in beach plants, insights about plant responses to shade conditions are likely to be relevant to understanding plant responses to other conditions in which light availability is reduced, like burial. Shade avoidance is characterized by morphological changes including elongated hypocotyls, internodes, and petioles and upward bending of the leaves when plants sense shading. This suite of phenotypes is also beneficial during burial, enabling a plant to extend leaves toward the light and allowing sand to slide off vertically oriented leaves (Sykes & Wilson 1990). Reduction of photosynthetically active radiation or the red:far red light ratio associated with canopy cover can trigger a shade avoidance response (Pierik & de Wit 2013).

Although shade avoidance is a plastic response to an environmental condition, plants bearing mutations in phytochromes and associated signaling proteins demonstrate genetically based variation in plasticity. Moreover, loss of plasticity is associated with lower fitness under some experimental light conditions, demonstrating that plasticity itself can be an adaptive trait (Schmitt et al. 1995, Ganade & Westoby 1999, Pigliucci & Schmitt 1999, Schmitt et al. 2003, Pierek & de Wit 2013). Despite this, field trials attempting to measure the fitness effects of both fixed differences in shade tolerance and shade

avoidance plasticity have garnered mixed results. While Schmitt et al. (2003) found evidence for adaptive divergence in shade response between sun and shade habitats in *Impatiens capensis*, and McIntyre and Strauss (2013) found a positive relationship between plasticity and fitness in *Claytonia perfoliata*, Callahan and Pigliucci (2002) found no effect on fitness in *A. thaliana.*

2.2.3 Experimental approach

I observed long hypocotyls in the greenhouse in *A. thaliana* accessions collected from coarse sand beaches on the Baltic coast of southern Sweden (Figure 2.1). When seeds were planted on the soil surface, seedlings at the cotyledon stage were noticeably taller than inland conspecifics. By the two- to four-leaf stage, the weight of the growing rosette often caused these seedlings to lie along the surface of the soil rather than stand upright. Similar behavior was not noted in inland and fine sand beach conspecifics from southern Sweden.

Statistically significant phenotypic differences between conspecifics from different subpopulations may arise as the result of several processes (Merila & Hendry 2013). First, drift may have occurred in one or both populations since their split, resulting in divergent phenotypes that are nonetheless neutral with respect to the environment. Alternatively, divergent phenotypes may represent a plastic response to environmental conditions. In this situation, though a correlation with environmental factors is observed, no genetic differences underlie phenotype differences. From the perspective of an evolutionary

biologist, a more interesting scenario occurs when differing selection regimes have led to genetic changes underlying adaptive phenotypes in one or both populations.

I hypothesized an adaptive role for the hypocotyl phenotype in my focal coarse sand beach populations. Although *A. thaliana* generally germinates on the soil surface, where it is exposed to light, populations on beaches may be more subject to burial. However, the Baltic beaches on which the focal populations are found are not dune systems with the associated high rates of fine sand accumulation. Instead, a layer of coarse sand lies on a mostly flat beach. The grain size of this coarse sand layer is large enough to be much less disturbed by wind and is considerably larger than an *A. thaliana* seed. Therefore, while traditional models of burial may be inappropriate, it may still be realistic to think about a seed falling between large grains, experiencing low light conditions, and emerging via hypocotyl elongation.

Linking a phenotype with a fitness advantage under specific conditions and, if possible, with underlying genetic variants, sheds light on the genetic and physiological mechanisms of adaptation to a particular environment. Moreover, connections between phenotype, genotype, and environment provide insight into the ecological and evolutionary processes of ecotype production. To elucidate one facet of plant local adaptation to beach environments, I sought to determine whether the observed long hypocotyl phenotype represents a heritable adaptive trait in beach *A. thaliana* from southern Sweden. To demonstrate this, I asked:

(1) Do beach plants have significantly longer hypocotyls than inland conspecifics, and are differences in hypocotyl length fixed or plastic?

(2) What genetic regions are associated with natural variation in this phenotype?

(3) Are genetic variants identified in (2) also associated with the long hypocotyl phenotype in other *A. thaliana* accessions?

(4) Do the long hypocotyl phenotype and associated genetic variants found in the Swedish beach population correlate with increased emergence of *A. thaliana* seedlings when seeds are buried, both for this set of lines and for a larger, worldwide set of accessions?

2.3 Methods

2.3.1 Hypocotyl length in plants from coarse sand beaches

To confirm my observation that lines from coarse sand beaches have longer hypocotyls than inland or fine sand beach conspecifics, I measured hypocotyl length in a panel of 27 *A. thaliana* accessions from southern Sweden plus Columbia (Appendix B Table 3). Lines came from a range of environments, including coarse sand beaches, fine sand beaches, and sand and soil sites located inland. I measured hypocotyl length when plants were grown in a variety of substrates to explore fixed and plastic differences between accessions.

Three or four replicates per line were grown in each of four substrate types: soil (50:50 Fafard C-2:Metro Mix 200), fine sand (Vigoro Torpedo), clay particles (Turface Athletics),

and pebbles (Vigoro Pea Pebbles, ~0.5 – 1 cm diameter). Seeds were planted on the substrate surface and watered, then stratified at 4°C for seven days to break dormancy and synchronize germination. Trays were moved to the greenhouse (20°C, 16-hour day) after stratification, watered as needed, and fertilized twice weekly. Seedlings were harvested at eight and eleven days after removal to the greenhouse, and hypocotyl length was measured from the bottom of the petioles to the root-shoot junction.

I removed lines with poor germination from the experiment (<10/14 individuals). To account for variance within each source (coarse sand beach, fine sand beach, soil inland, sand inland) due to random effects of accession, I performed separate linear regressions of the formula (Hypocotyl length ~ Line) for each source. I then used the residuals from these regressions in a linear model of the formula (Hypocotyl length ~ Source*Substrate) to test for effects of source, growth substrate, and their interaction on hypocotyl length.

2.3.2 Genome wide association mapping of hypocotyl length

To further explore natural variation in hypocotyl length and to test for genetic associations, a panel of 298 accessions sampled from across Sweden (Appendix B Table 2) was planted on soil. One block was planted on each of five successive days for a total of five blocks. Each block consisted of one representative of each accession. Seeds were stratified and grown in the greenhouse as above. On the tenth day, hypocotyl length was measured and these values were used as the phenotype in a genome wide association study. GWA was performed in R, using EMMAX, a mixed linear model approach with a correction for

population structure (Kang et al. 2008, 2010) and the 250k SNP set described in Kim et al. (2007).

2.3.3 Hypocotyl length in additional accessions

To test the relationship between HYPO5, the genetic variant of interest identified in Part 2, and the long hypocotyl phenotype in other populations, I used lines from the RegMap panel (Horton et al. 2012) and lines from the Bay x Sha RIL family. From the RegMap, I chose 83 lines with the G allele and 78 lines with the T allele at HYPO5. Lines with contrasting alleles were chosen from France, Germany, Spain, Sweden, the UK, the US, the Netherlands, and Russia (Appendix B Table 4). For the Bay x Sha family (Loudet et al. 2002), I identified the closest markers flanking the genetic variant of interest and selected 24 lines with Bay-0 alleles at both markers (assuming that the intervening sequence is also Bay-0) and 24 lines with Sha alleles at both markers (Appendix B Table 5).

I stratified each set of seeds in water in the refrigerator for one week, planted them on moist soil in 288-cell flats, and placed the flats in a growth room (20°C, 16 hour day). Eight seeds per line were planted for the RegMap accessions and six seeds per line were planted for the Bay x Sha RILs. Hypocotyl length was measured on the tenth day after flats were moved to the growth room.

2.3.4 Seedling emergence and hypocotyl length

To look for a fitness advantage associated with the long hypocotyl phenotype, I tested for relationships between hypocotyl length and both the genetic variant of interest and frequency of emergence from burial. I designed an assay similar to Maun & Riach (1981) to test for the ability of seeds to emerge when buried in conditions mimicking coarse sand beaches. In pilot experiments (summarized in Appendix C) unburied controls exhibited near 100% germination; therefore, my final assay did not include an unburied control. Between-line differences in frequency of emergence were attributed to variation in the ability to emerge from burial.

I used twelve lines from coarse sand beaches and from the Angso site, six with the G allele and six with the T allele at HYPO5 (Appendix B Table 6). Seventy-two seeds per line were stratified in water in the refrigerator for seven days. Cheesecloth was sandwiched between pairs of 288-cell flats to prevent loss of sand from drainage holes in the bottom of the flat. Three flats were half filled with very fine sand (Vigoro Torpedo, portion sieved through 6 micron mesh), moistened, and then one seed was planted per cell. Two flats were misted with fertilizer, then covered with coarse sand to a depth of ~1cm (Vigoro Torpedo, remnant portion after sieving through 2 mm mesh). To assess whether clay particles yielded similar results to coarse sand, I covered one additional flat with ~1 cm clay particles (Turface Athletics). Flats were watered lightly daily and fertilized twice weekly. Emergence of germinants was recorded daily for two weeks.

To test for a relationship between the genetic variant of interest and frequency of emergence in a set of 421 accessions (161 RegMap accessions plus 260 additional Swedish accessions; Tables 2.2 and 2.3; Bay x Sha RIL lines were not used due to poor germination rates), I used the above assay with some amendments. Because the larger set of accessions had not been screened for germination rates or poor performance on sand, I included unburied controls. I assigned cells in 48-cell flats in treatment/controls pairs. Each accession was represented by four treatment/control pair replicates, randomized across all flats. As above, cells were lined with cheesecloth and half filled with very fine sand (Vigoro Torpedo, portion sieved through 6 micron mesh). Twenty seeds (previously stratified in water for seven days at 4°C) were planted in each cell. In the beach line burial assay, emergence from clay particles was later and lower than emergence from coarse sand, but the difference in emergence between G and T bearing lines was similar between substrates. Thus, I used clay particles in the expansion of this assay as it reduced the experimental preparation time. Flats were misted with fertilizer, then treatment cells were buried with ~1 cm of clay particles (Turface Athletics). Flats were watered daily and fertilized twice weekly.

Emerged germinants were counted one and two weeks after planting. Before data analysis, I removed treatment-control pairs that had no emerged seedlings as well as those with final control germinant counts less than or equal to five seedlings. Such pairs either reflected poor germination rates of the seed stock, errors in planting (i.e. an empty control cell), or poor conditions in the control cell due to experimental error (i.e. excessive dryness). I also removed pairs where planting errors had been noted. The remaining data

set included 1603 treatment-control pairs from 421 accessions. Emergence for each time point was calculated as treatment seedlings emerged/control seedlings germinated for each pair.

I used treatment count as the response variable in a generalized linear mixed model with a Poisson error distribution (R package lme4; Bates et al. 2015). I tested separately for an effect of HYPO5 and an effect of the average per line hypocotyl length (as measured in Parts 2 and 3). These analyses were separated due to the correlation between SNP and hypocotyl length. Control count, identity of the counter, identity of the planter, flat, cell position, and accession were included as random effects.

2.4 Results

2.4.1 Hypocotyl length in plants from coarse sand beaches

After removing one line with poor germination (Tjo 10), I analyzed hypocotyl measurements of 356 individuals from 27 accessions (Figure 2.2). Hypocotyl length in these 27 accessions demonstrated both fixed and plastic differences (Table 2.1). There was no significant difference in the length of hypocotyls harvested on day 8 (mean = 4.2 mm) and day 11 (mean = 4.6 mm, Wilcoxon rank sum test p-value = 0.19). Day 11 data is reported below, in Figures 2.3 and 2.4, and in Tables 2.6 and 2.7.

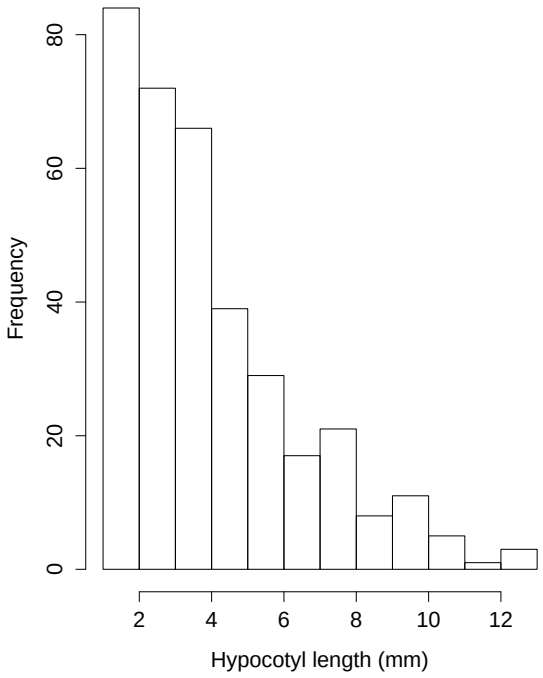

Figure 2.2 Hypocotyl length measurements for a panel of 27 southern Swedish accessions from coarse sand beaches, fine sand beaches, and inland sand and soil sites, grown on soil, fine sand, clay particles, and pebbles (n=356).

	Soil	Sand	Clay particles	Pebbles	All substrates
Coarse sand beach	6.8	4.1	5.8	8.5	6.3
Inland soil	5.1	2.4	3.8	5.1	4.1
Fine sand beach	4.1	2.1	2.5	5.1	3.5
Inland sand	3.9	2.0	3.1	5.0	6.5
All sources	5.2	2.9	4.1	6.1	

Table 2.1 Mean hypocotyl length (mm) of lines from different source populations in different substrates, day 11 (n=202).

Growth substrate had an effect on hypocotyl length (Table 2.2, Figure 2.3a). Accessions grew the longest hypocotyls on pebbles (mean of all lines 6.1 mm), although this was not significantly longer than hypocotyls grown on soil (5.2 mm). Accessions grew shorter hypocotyls on fine sand (2.9 mm) and clay particles (4.1 mm).

Effect	DF	Sum Sq.	Mean Sq.	F	p-value
Source	3	919.3	306.5	136.6	<0.0001***
Substrate	3	330.8	110.3	49.2	<0.0001***
Source:Substrate	9	20.8	2.3	1.0	0.42

Table 2.2 ANOVA table for linear model of the formula (Hypocotyl length ~ Source*Substrate).

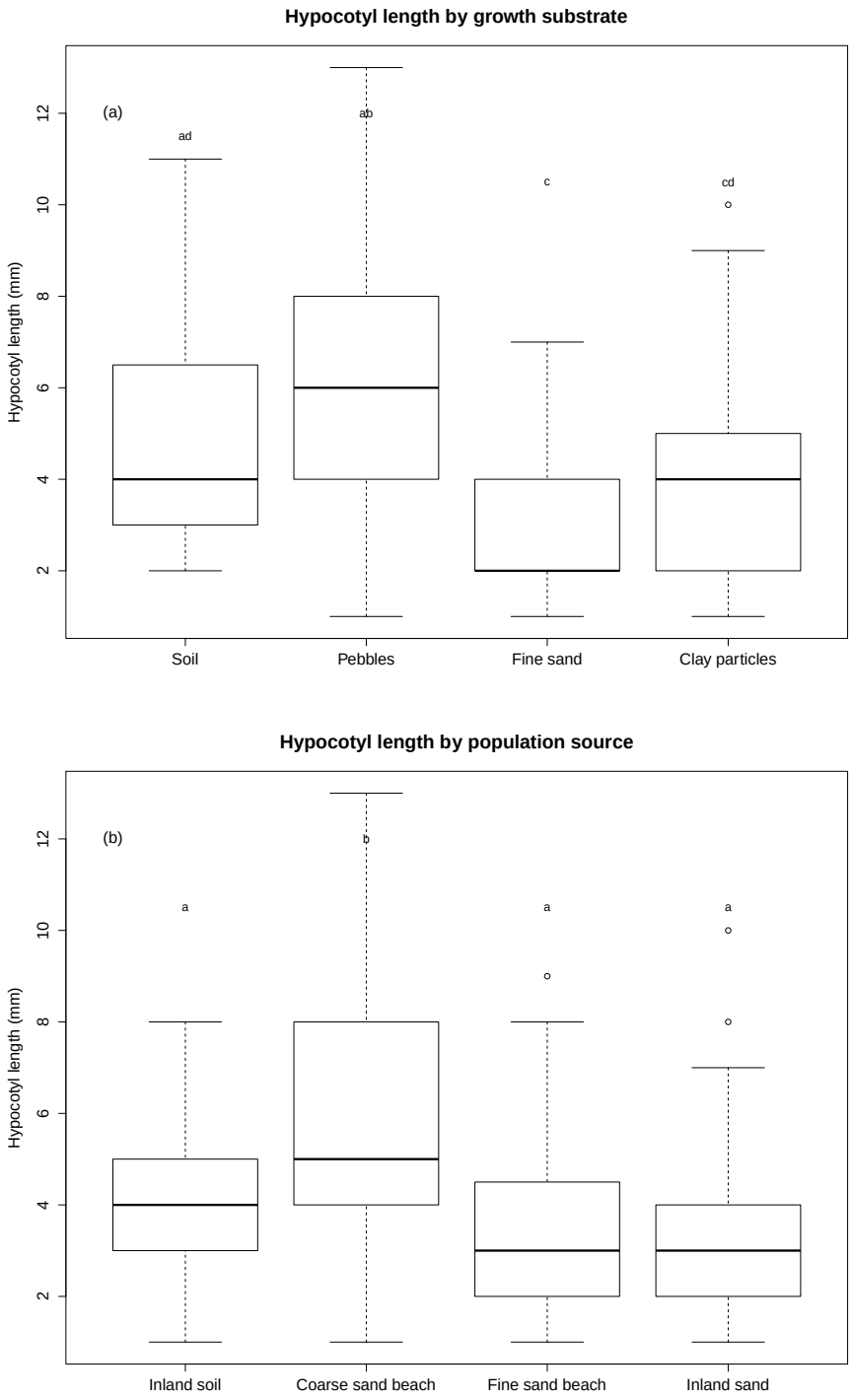

Figure 2.3 (a) Hypocotyl length in accessions grown in soil, pebbles, fine sand, and clay particles, measured on day 11. Accessions grew the longest hypocotyls on pebbles, although this was not significantly longer than hypocotyls grown on soil. (b) Hypocotyl length in accessions from inland soil sites, coarse sand and fine sand beaches, and inland sand sites, measured on day 11 (n=202). Accessions from coarse sand beaches displayed significantly longer hypocotyls than fine sand beach and non-beach accessions.

Population source also had an effect on hypocotyl length (Table 2.2, Figure 2.3b). Accessions from coarse sand beaches displayed significantly longer hypocotyls than fine sand beach and non-beach accessions, with an average length of 6.3 mm (across all substrates) in comparison to fine sand beach lines (3.5 mm), inland sand lines (3.5 mm) and inland soil lines (4.1 mm).

No interaction between source and substrate was detected (Table 2.2). That is, lines from different source populations exhibited similar degrees and directions of plasticity to the growth substrates (Figure 2.4).

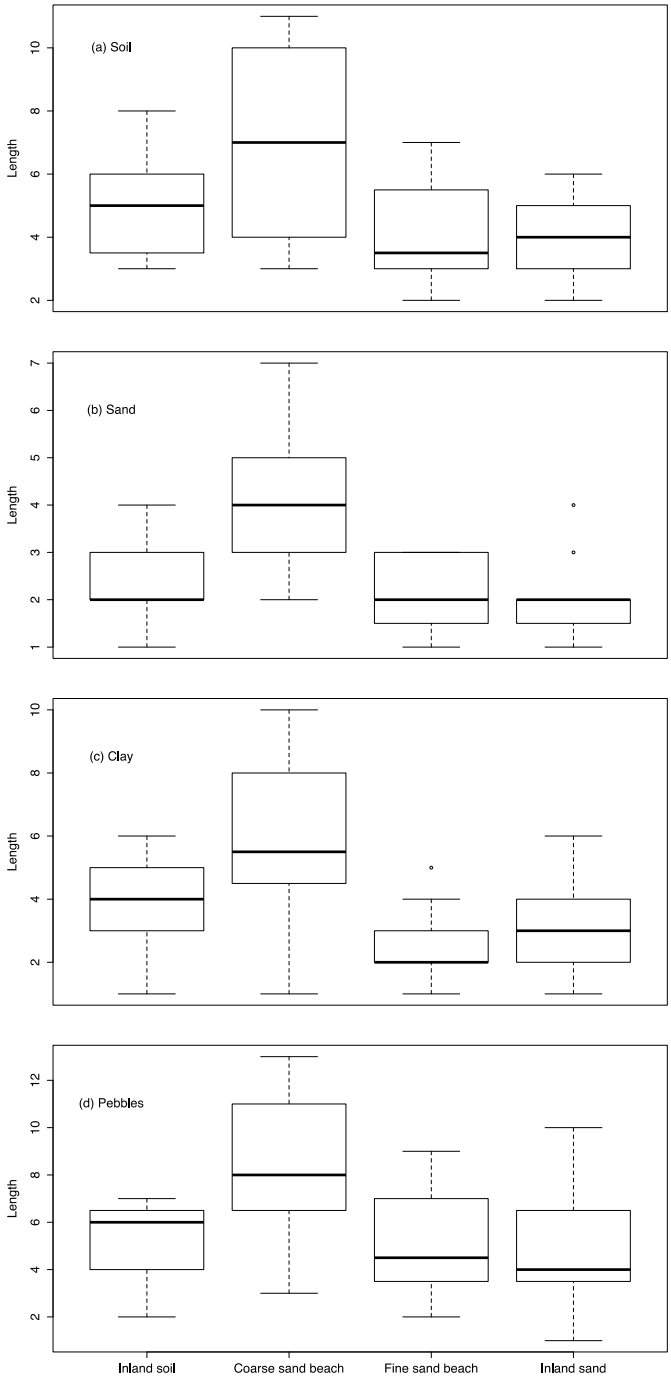

Figure 2.4 Hypocotyl length of accessions from inland soil and sand sites and coarse and fine sand beaches grown in (a) soil (b) fine sand (c) clay particles (d) pebbles on day 11 (n=202).

2.4.2 Genome wide association mapping of hypocotyl length

As in the smaller panel of lines, fixed differences in hypocotyl length among the 298 accessions were clear, with a broad sense heritability of 0.30. The twelve lines from coarse sand beaches had an average hypocotyl length of 5.5 mm. The remaining lines exhibited a significantly shorter average hypocotyl length, 3.6 mm (Wilcoxon rank sum test p-value < 0.0001, Figure 2.5).

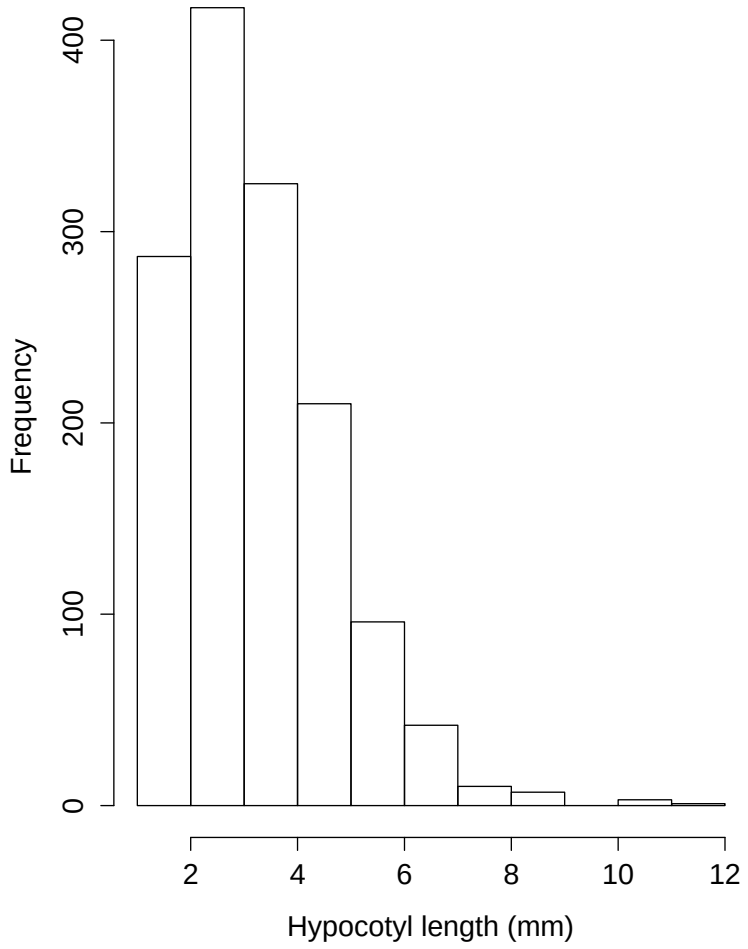

Figure 2.5 Hypocotyl length distribution for 298 lines in regional Swedish mapping panel (n=1398)

I used the per line average hypocotyl length to perform a GWA. I identified two peaks, one on chromosome 2 and one on chromosome 4 (Figure 2.6). Regions 20kb up and downstream from the most significantly associated SNP in each peak were searched for candidate genes (Appendix A Table 7). A strong candidate gene, *SPA2* (suppressor of phytochrome A), falls under the chromosome 4 peak, but no strong candidates were found near the chromosome 2 peak. For this reason, I focused on the chromosome 4 peak. Five of the twelve coarse sand beach lines bear the minor allele at the top five SNPs with the strongest association. In addition, a single line from northern Sweden, Angso 80, bears the minor allele (T) at one of these SNPs (Table 2.3). In the following assays, I focus on this SNP, HYPO5, as the variant of interest and include Angso lines with and without the minor allele to investigate the effect of the SNP in multiple genetic backgrounds.

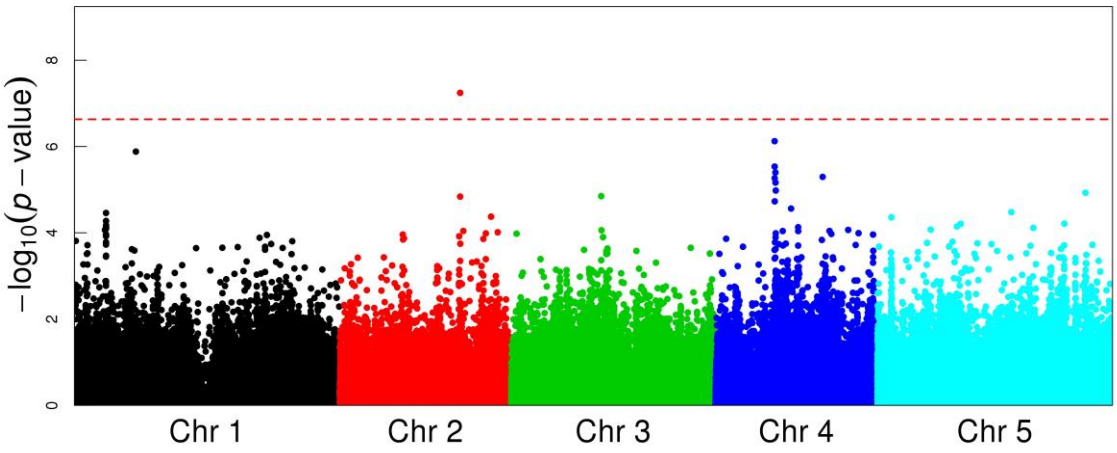

Figure 2.6 Manhattan plot of GWA results. Response phenotype was mean per line hypocotyl length for the 298 Swedish mapping panel; dotted red line is Bonferroni adjusted significance threshold.

SNP	**6775136**	**6778523**	**6781439**	**6867260**	**6784304**
beta	-0.85	-1.08	-1.08	-0.67	-0.95
p-value	**7.52 x 10-7**	**2.92 x 10-6**	**2.92 x 10-6**	**3.99 x 10-6**	**5.43 x 10-6**
Angso-80	C	C	G	G	T
Sim 5	G	A	A	C	T
TV-22	G	A	A	C	T
TV-30	G	A	A	C	T
TV-4	G	A	A	C	T
TV-7	G	A	A	C	T
Sim 1	C	C	G	G	G
Sim 2	C	C	G	G	G
TV-10	C	C	G	G	G
TV-38	C	C	G	C	G
Var A1	C	C	G	G	G
Var 2-1	C	C	G	G	G
Var 2-6	C	C	G	G	G

Table 2.3 Top SNPs in the chromosome 4 peak in the twelve beach lines plus Angso 80.

2.4.3 Hypocotyl length in additional accessions

The T allele at HYPO5 was associated with significantly longer hypocotyls in both the RegMap lines and the Bay x Sha RIL lines. RegMap lines with the T allele had longer hypocotyls (mean = 4.96 mm) than lines with the G allele (mean = 4.51; Wilcoxon rank sum test p value < 0.0001; Figure 2.7). Bay x Sha RIL lines with the T allele also had longer hypocotyls (mean = 3.78) than lines with the G allele (mean = 3.28; Wilcoxon rank sum test p-value = 0.03; Figure 2.8).

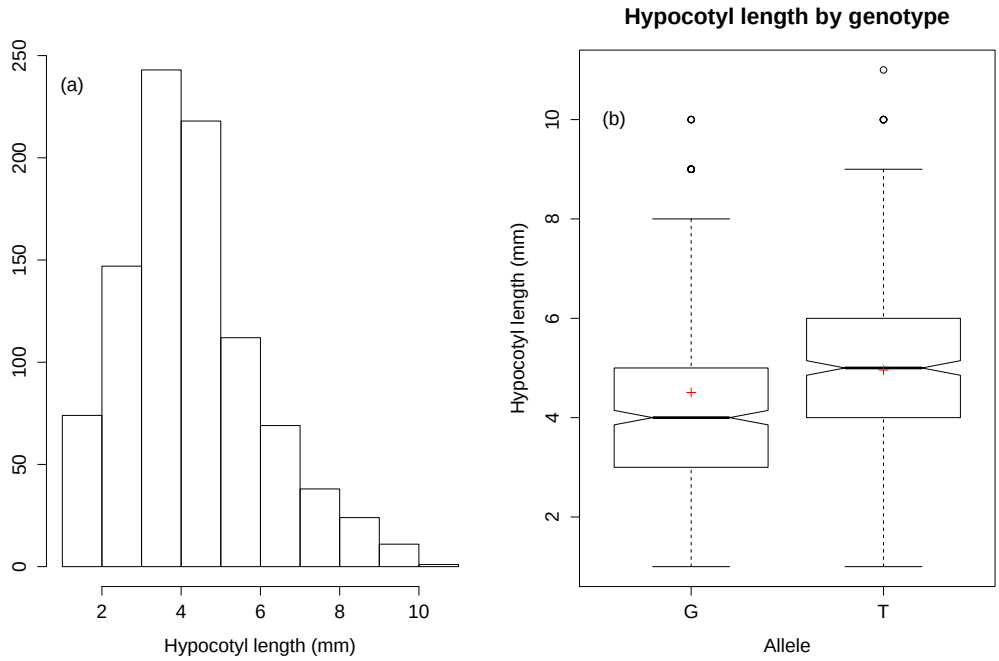

Figure 2.7 (a) Distribution of hypocotyl lengths for 149 RegMap lines (n=937) (b) Hypocotyl length of 149 RegMap lines by allele at HYPO5. Red crosses are the mean values (mean G = 4.51 mm, mean T = 4.96 mm; Wilcoxon rank sum test p value < 0.0001)

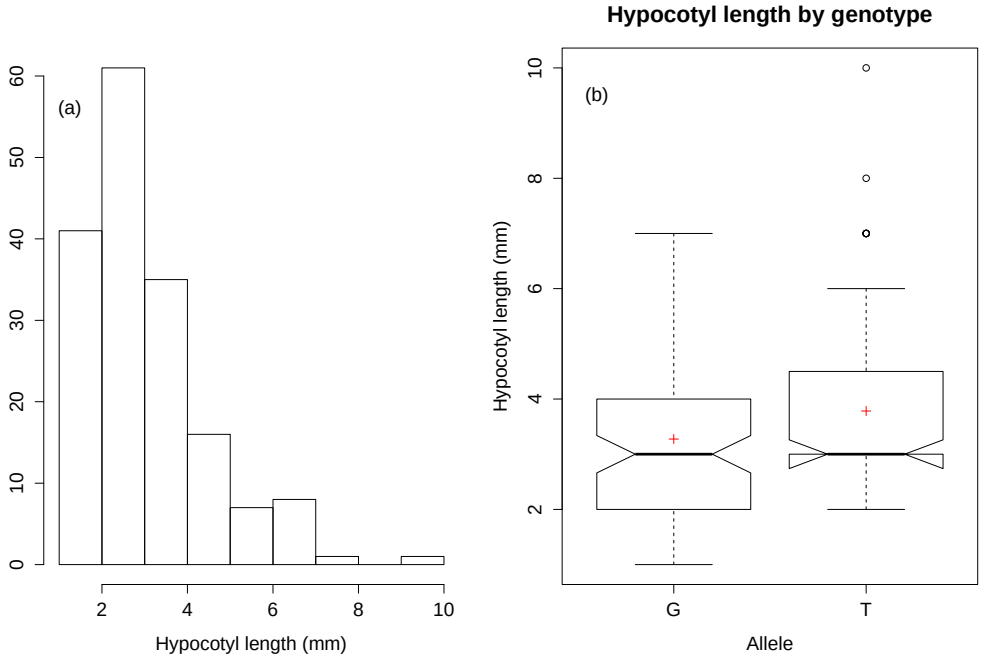

Figure 2.8 (a) Distribution of hypocotyl lengths for 39 Bay x Sha RIL lines(n=170) (b) Hypocotyl length of 39 Bay x Sha RIL lines by allele at HYPO5. Red crosses are the mean values (mean G = 3.28 mm, mean T = 3.78 mm; Wilcoxon rank sum test p value < 0.03).

2.4.4 Seedling emergence and hypocotyl length

In the beach accessions, lines with the T allele at HYPO5 emerged from burial more frequently than lines with the G allele (Table 2.4; chi-square test of independence; p-value = 0.001). Nevertheless, exceptions exist. Sim 1, with a G allele, exhibited emergence rates more similar to T bearing TV lines, while its close geographic neighbor Sim 5 (T allele) exhibited much lower emergence rates. Angso 80 (T), behaved much more similarly to other Angso lines (G) than to other T lines (Figure 2.9).

	G (frequency)	T (frequency)
Emerged	256 (.59)	302 (0.70)
Failed	176 (.41)	129 (.30)

Table 2.4 Number and frequency of emerged and failed seedlings from the burial assay. Chi-square test for independence p-value = 0.001.

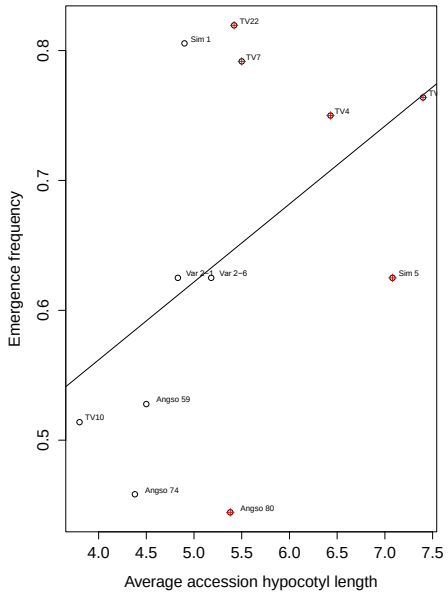

Figure 2.9 Emergence frequency of twelve beach lines as a function of average hypocotyl length on soil. Lines marked with a red cross have the T allele at HYPO5 (n = 863). Line represents slope from linear model of the form (Emergence frequency ~ average accession hypocotyl length).

For the burial assay with the larger set of accessions, at both time points, lines with the T allele had a higher average emergence rate than lines with the G allele at HYPO5 (Table 2.5; Figure 2.10). Generalized linear models indicated a significant effect of both SNP and hypocotyl length at both time points (Table 2.6).

	G	T
Emergence @ Time point 1	0.43	0.50
Emergence @ Time point 2	0.54	0.62

Table 2.5 Emergence frequency differences between lines with G and T at the SNP of interest.

Phenotype	Fixed effect	Estimate	Standard error	p-value
Treatment count (Time point 1)	SNP	0.19	0.07	0.01*
Treatment count (Time point 2)	SNP	0.017	0.06	0.008**
Treatment count (Time point 1)	Hypocotyl	0.08	0.03	0.003**
Treatment count (Time point 2)	Hypocotyl	0.06	0.02	0.01*

Table 2.6 Results a generalized linear model of burial assay with a Poisson error distribution. Random effects were control count, counter, planter, flat, cell position, and ecotype id (n=1603).

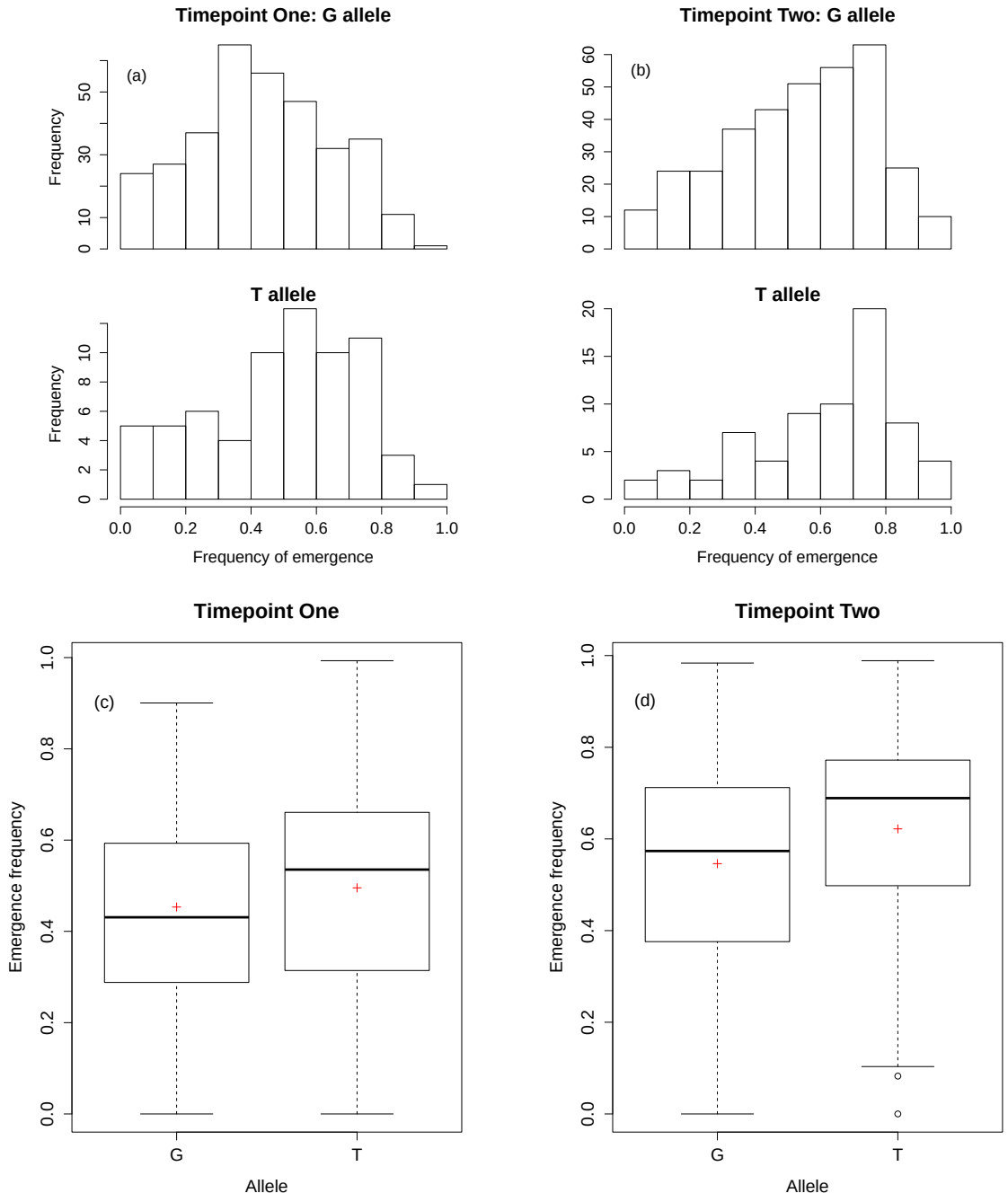

Figure 2.10 (a) Histogram of average per line emergence frequencies at time point one for both alleles at HYPO5 (b) Histogram of average per line emergence frequencies at time point two for both alleles at HYPO5 (c) Average per line emergence frequencies at time point one for both alleles at HYPO5 (mean G = 0.46; mean T = 0.51) (d) Average per line emergence frequencies at time point two for both alleles at HYPO5 (mean G = 0.55; mean T = 0.63). For better visualization of all plots above, lines with emergence frequencies above one have been omitted (n=5 for time point one, n=6 for time point 2).

2.5 Discussion

2.5.1 Coarse sand beach lines have longer hypocotyls than inland and fine sand beach lines

Studies of beach plant responses to burial largely focus on dune sites where sand accumulation due to wind is the main source of burial. However, in my panel lines from beaches with fine sand do not have hypocotyls significantly different in length than lines from inland sites. On the other hand, lines from beaches with coarse sand exhibit long hypocotyls despite the fact that the substrate is less susceptible to movement by wind.

What accounts for this counterintuitive result? As discussed in the introduction, the grain size of the substrate on the coarse sand beaches is larger than an *A. thaliana* seed, suggesting that long hypocotyls may be necessary for seedlings to reach the surface. The fine sand beaches represented here are not traditional dune communities. Plants were sampled from flat sandy areas above the tideline where they may not be subject to heavy sand accumulation. Moreover, wind events tend to be seasonally distributed on most temperate beaches (Maun & Perumal 1999). If sand accumulation due to wind does not coincide with a sensitive part of the *A. thaliana* life cycle, selection pressure for burial tolerance may be weak or absent in these lines.

2.5.2 Identification of a candidate locus for long hypocotyls in the beach lines

Standard GWA protocols usually involve the exclusion of SNPs with minor allele frequencies below a specific threshold in order to reduce the chances that rare alleles (often private to a particular subpopulation) show a chance association with the phenotype. This is true even when using a protocol like EMMAX that corrects for population structure, because this correction reflects genome-wide average levels of relatedness while specific regions of the genome may exhibit a greater or lesser signature of population structure. However, despite the fact that only six of 298 accessions bear the minor allele at HYPO5, there are good reasons to believe the result is not spurious. First, although five of the six lines are closely related, the sixth accession is geographically and phylogenetically separate from the others, and the SNP appears in a different background (Table 2.3), but the accession is found close to the water and exhibits an extreme hypocotyl phenotype. Second, the association between a low frequency allele and an extreme phenotype that is geographically restricted is expected in a scenario where a small population has experienced adaptation to a harsh environment. Third, a strong candidate gene, *SPA2* (suppressor of phytochrome A), falls near this peak. Lastly, follow up experiments confirmed that this variant was associated with hypocotyl length in two independent panels of accessions.

SPA2 is one of four *SPA* family proteins that complex with COP1 and interact with phytochrome A to repress growth in darkness and to control the shade avoidance response (Laubinger & Hoecker 2003, Laubinger et al. 2004, Fittinghoff et al. 2006, Roulauffs et al.

2012). Upon exposure to light, SPA2 (unlike other members of the *SPA* family) is inactivated and degraded (Balcerowicz et al. 2011). Dark grown *SPA* double and triple mutants exhibit constitutive photomorphogenesis (failure to etiolate, open cotyledons, etc.; Laubinger et al. 2004). In contrast, beach line seedlings produce long hypocotyls in light conditions (though they do not display other features of etiolated plants, such as the apical hook), while dark grown seedlings exhibit normal dark-growth phenotypes and do not have significantly longer hypocotyls than inland or fine sand beach conspecifics (Appendix D). Thus, a strong or overactive variant of SPA2 could potentially be responsible for the observed phenotypes.

Using full sequence data for 209 accessions (1001genomes.org) aligned with bwa2, I identified of genetic variants within the *SPA2* coding region and in 1 kb flanking regions on either side of the gene, as well as 20kb window centered on HYPO5. Testing for association of the phenotype with genetic variants in *SPA2* does not identify any SNPs strongly associated with the long hypocotyl phenotype, with the highest R^2 value only 0.03 (Figure 2.11). R^2 values are higher for SNPS in the intergenic region immediately surrounding HYPO5, but even here associations are not particularly strong (max $R^2 = 0.2$). However, such a test would fail if the causative variation was an indel or if multiple SNPs acted together to produce the phenotype.

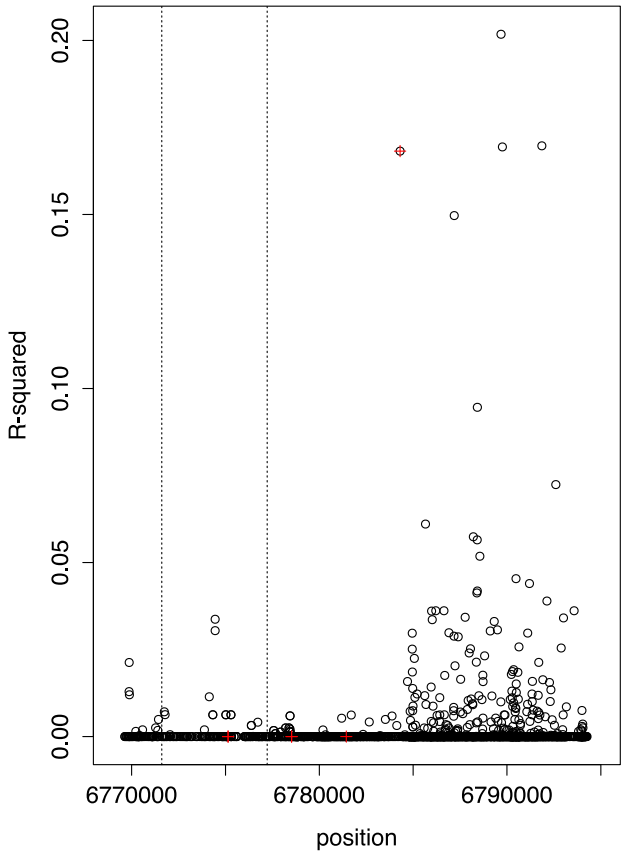

Figure 2.11 R^2 values for SNPs in a 20kb window around HYPO5. Upstream window has been extended to include the entire coding region of SPA2 and 1kb flanking window. Dotted lines indicate the ORF of SPA2.

If a mutation in *SPA2* is indeed the variant responsible for the long hypocotyl phenotype, a simple knockout of *SPA2* will not be sufficient to demonstrate causality. Because the beach allele appears to be overactive, a strategy for validation involves the introduction of the beach allele into *SPA* knockout mutants. One could then test for an increase in hypocotyl length or improved frequency of emergence when buried associated with the presence of the beach allele in the F2 generation. Complementary experiments could determine whether the SPA2 protein is present (and not degraded) in light conditions in accessions

with the long hypocotyl phenotype. Once validation of the gene is achieved, further experiments could pinpoint the mutation(s) responsible for the phenotype change and lead to an understanding of the mechanism of hypocotyl elongation at a molecular level.

2.5.3 Fitness benefits conferred by the genetic variant of interest

HYPO5 is significantly associated with increased emergence from burial in the 421 lines studied, as was the average accession hypocotyl length. However, SNP allele/hypocotyl length explain very little of the variation in emergence frequencies (based on linear models using emergence as the response variable: R^2 = 0.7-1.4% for SNP allele; 2.5-3.4% for hypocotyl length). Apart from the fact that plant characteristics other than hypocotyl length doubtless contribute to emergence success, there is good reason to believe that the emergence assay may underestimate the fitness benefit of this genetic variant in the field.

Many studies have demonstrated that early emergence confers a significant fitness advantage over conspecifics, in some cases a many-fold increase in seed production (Zhang & Maun 1990, Maun 2009, Weis et al. 2015). The burial assay used here emphasized maximum or total possible emergence as the response phenotype, but daily data from the 12-line assay show that lines with the T allele generally emerge earlier than lines with the G allele (Figure 2.12; data on emergence timing is unavailable for the larger experiment due to the infeasibility of monitoring some 1700 treatment/control pairs daily). For example, even though the line Sim 1 (G allele) has an overall emergence frequency similar to lines TV4, 7, 22, and 30 (T allele), the four T lines outpace Sim 1 for the entire first week

of the experiment (Figure 2.12). In direct competition, this could dramatically decrease Sim 1 success. When unburied on sand or soil, lines with the T allele do not germinate earlier than lines with the G allele; thus, emergence advantage is limited to buried seeds. Further experiments confirming competitive disadvantage of late emergence could result in a more accurate assessment of the importance of early emergence.

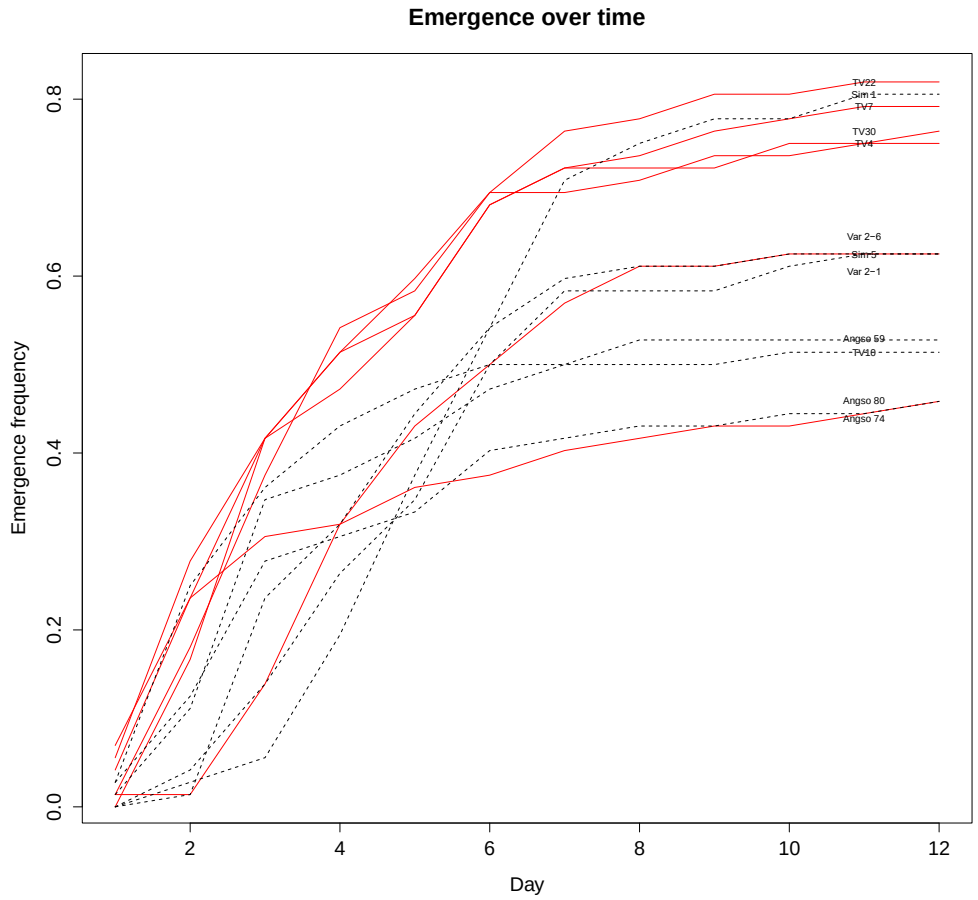

Figure 2.12 Proportion of buried seedlings emerged by day (cumulative) for twelve beach lines with the T allele (red lines) and G allele (dashed lines).

In addition, while differences in emergence represent one effect of HYPO5 on fitness, they might not represent the entire effect of the SNP. Hypocotyl elongation and etiolation have been shown to involve tradeoffs with root system growth and stem thickness, potentially

increasing fragility or susceptibility to pathogen attack (Ganade & Westoby 1999, Boron & Vissenberg 2014). The effects of the tradeoffs necessary to emerge may be less mitigated in G plants and these effects might be expressed at a later stage (Jordan 1992) or under a different or new stressor. Specific measures of tradeoffs – reduction in leaf area, root system growth, stem thickness, growth rate, and eventually flower and fruit production – could give insight into the whole-plant effect of emergence from burial.

2.6 Conclusion

In this chapter, I have demonstrated that *Arabidopsis thaliana* plants from coarse sand beaches:

(a) have longer hypocotyls than fine sand beach and inland conspecifics.

(b) show an association between hypocotyl length and a genetic variant on chromosome 4.

(c) exhibit variation in seedling emergence from burial that is associated with hypocotyl length and the genetic variant.

These results have been confirmed in a regional set of Swedish accessions and in a global set of RegMap accessions. Thus, this work links an adaptive phenotype and genotype with a particular environmental stressor.

Not all coarse sand beach lines possess the T allele, despite the demonstrated effect on fitness. This could reflect the ongoing process of a favored genetic variant spreading through a population. Although there is a relationship between hypocotyl length and this

SNP within the beach lines, beach lines with both G and T alleles exhibit longer hypocotyls than inland and fine sand beach conspecifics. Hypocotyl length is clearly a polygenic trait, and selection for increased hypocotyl length on coarse sand beaches may affect other loci.

Field experiments and competition experiments could lead to a more sophisticated understanding of the fitness benefits conveyed by long hypocotyls. Because fitness at one life stage has effects on fitness at later stages, longer experiments could measure lifetime reproductive fitness differences associated with early life survival and competitive success due to hypocotyl length.

SPA2 is a strong candidate for the causal gene underlying this phenotype. Burial and darkness have been shown to evoke similar phenotypic responses; thus, linking mutations in light sensing pathways to burial related fitness is logical. Work on the phytochromes and associated signaling pathways has revealed the ability of plants to tailor their growth to the ambient light environment using both fixed genetic variation and plastic responses to environmental signals.

SPA2 awaits confirmation as the causal gene. Ongoing experiments involve crossing *SPA* knockout mutants with coarse sand beach lines and phenotyping F2 plants for hypocotyl length. If longer hypocotyls are associated with the beach allele of *SPA2* and shorter hypocotyls are associated with the knockout *SPA2* allele, *SPA2* will be validated as the causal gene.

Elongated hypocotyls allow seedlings to emerge from a coarse-grained substrate and represent a phenotype important for survival on coarse sand beaches. Other putatively adaptive phenotypes that differentiate beach and inland plants along with hypocotyl length constitute a suite of characteristics that define a unique coarse sand beach ecotype. Future work on this ecotype would involve exploration of other adaptive phenotypes.

CHAPTER 3: SWEDISH BEACH *ARABIDOPSIS THALIANA* DO NOT DISPLAY ADAPTATION TO DROUGHT OR SALT

3.1 Abstract

Numerous examples of phenotypically and genotypically differentiated beach and inland ecotypes have been identified. Commonly, beach plants must cope with salt spray on their tissues and reduced water availability; reciprocal transplants suggest that inland ecotypes tend to suffer greater fitness consequences than beach ecotypes when subject to beach stressors, suggesting that many beach plants have evolved adaptations to these stressors. *Arabidopsis thaliana* populations found on southern Swedish beaches are good candidates for this type of adaptation given the persistence of the populations and the considerable natural variation exhibited in this species. I tested for tolerance to salt and drought in 34 accessions of *A. thaliana* from southern Sweden by measuring the reduction in fecundity when subject to these stressors. Salt spray did not reduce fecundity in either beach or non-beach populations. This suggests that salt spray is not a major selection pressure on Swedish beaches, perhaps because of the low salinity of the Baltic or perhaps because major wind events do not coincide with sensitive growth periods of *A. thaliana*. In an initial study using the same 34 accessions, I found that lines from coarse sand beaches had higher fecundity when subject to mild drought stress relative to replete conditions. They also suffered less fecundity loss when subject to harsh drought than inland conspecifics. However, in an experiment including 298 Swedish accessions, neither trend was repeated.

Coarse sand beach plants produced fewer seeds than inland conspecifics, but did not respond differentially to drought. GWA mapping results for fecundity changes under drought did not show enrichment for genes identified as drought responders *a priori.*

3.2 Introduction

3.2.1 Background

Salt spray and drought are two stressors commonly faced by beach plants (Maun 1994, McLachlan & Brown 2006, Maun 2009). While beaches may not receive less rainfall than inland areas, sandy soils tend to drain quickly, reducing water availability at any given time. At the same time, salt spray leads to deposition of salt on plant tissues and increased salt loads in the soil. Salt on exposed plant tissue can lead to tissue necrosis, dropping of leaves, and asymmetric growth, and may be partially responsible for dwarfism in coastal ecotypes (Boyce 1954). In the soil, salt buildup creates osmotic stress, mimicking and augmenting conditions of low water availability (Boyce 1954, Maun 2009).

Salt spray tolerance is likely responsible for zonation of plant communities in dune ecosystems (Wilson & Sykes 1999). Coastal ecotypes often possess salt spray tolerance mechanisms absent in inland conspecifics (Morrison 2002). The genetic basis of salt spray tolerance has been detected in Pacific coast *Mimulus guttatus* populations, where coastal alleles at tolerance QTL conferred fitness advantages at coastal sites (Lowry et al. 2009). In the same species, alternative life history strategies in coastal and inland populations were

attributed to the selective pressure of drought, though in this case inland populations are subject to more arid conditions (Lowry & Willis 2010).

Though not considered a halophyte, *A. thaliana* demonstrates natural variation in salt tolerance. In European populations of *A. thaliana,* higher leaf sodium content is associated with a particular allele of HKT1, a sodium transporter, and correlates with proximity to the coast and to saline soils (Rus et al. 2006, Baxter et al. 2010). Natural variation for salinity tolerance both during germination and vegetative growth in *A. thaliana* appears to be mediated by Na^+ transporters like HKT1 (Quesada et al. 2002, Katori et al. 2010). Recently, Busoms et al. (2015) detected local adaptation to coastal sites in Spanish populations of *A. thaliana* and attributed this to elevated salinity tolerance of the coastal populations.

A. thaliana demonstrates natural variation in drought tolerance, with significant genotype by environment effects documented in drought response studies (Bouchabke et al. 2008, Verslues & Juenger 2011, Juenger 2013, El-Soda et al. 2015). *A. thaliana* accessions also vary in drought tolerance strategy. McKay et al. (2003) demonstrated trade-offs between drought avoiding plants that flower early to escape major drought effects, and drought resisting plants that maintain metabolic function during periods of low water availability through stomatal closure, increased root:shoot ratio, solute accumulation, and other physiological mechanisms.

Many studies have compiled lists of candidate genes for drought and salt responses in *A. thaliana.* Responses to drought and salt stress in plants overlap considerably, as the initial

reaction to increased salinity in the soil is an osmotic response to low effective soil water potential (Munns 2011, Yadav et al. 2011). For example, in a screen of 1,700 rice genes, 56 of 57 genes with increased expression after salt exposure were also induced by drought stress (Rabbani et al. 2003, Hadiarto & Tran 2011). I hoped to leverage the power of these extensive candidate lists by using them to parse the results of genome wide association mapping studies conducted in a regional Swedish mapping population to identify genes responsible for adaptation to salt and water stress in a southern Swedish beach population of *A. thaliana.*

3.2.2 Approach

In its northern European native range, *Arabidopsis thaliana* occupies a variety of habitats, including agricultural fields, roadsides, quarries, and beaches. Long established beach populations appear phenotypically and phenologically distinct from inland conspecifics in the field (A. Anastasio, personal communication) and in the greenhouse (personal observation). Rosettes are larger, inflorescence stems are shorter, and plants have longer hypocotyls, more trichomes, and fewer, larger seeds. These differences suggest that these accessions are the product of selection for a coastal ecotype. Thus, I hypothesized that beach populations of *A. thaliana* would exhibit tolerance phenotypes when exposed to the well-known beach stressors, drought and salt.

Plants have evolved multiple strategies to cope with environmental stressors. Even within a single species, multiple pathways may be used to cope with stress depending on the

ecotype or specific environmental condition. Thus, though many physiological mechanisms ameliorate drought and salt stress, measurements of any one of these may fail to capture important responses in a specific plant. The ultimate measure of an organism's success in coping with environmental stresses is the organism's ability to survive and reproduce. For this reason, I chose to use total seed production, a catchall fitness proxy, as a tolerance phenotype. This phenotype integrates all physiological responses into a single measure of the plant's ability to successfully reproduce in various conditions. Because *A. thaliana* is an annual, seed production represents lifetime reproductive fitness for the individual. Because seed size varied between accessions, I used total seed weight rather than number as my response phenotype. Thus, this method is sensitive to both changes in seed number and seed size in response to stress.

I first performed pilot experiments in a panel of 34 beach and inland lines. I measured fecundity changes under various levels of salt and drought stressors. I hypothesized that lines from beaches would exhibit a smaller reduction in fecundity in response to stressors than inland conspecifics. When beach lines showed differential responses to stressors in the pilot experiment, I repeated the tolerance assays in a panel of 298 accessions in order to conduct genome wide association mapping.

3.3 Salt Spray Pilot

3.3.1 Methods

Pilot experiments were conducted using a panel of 34 accessions of *A. thaliana* from southern Sweden. Lines were originally sampled from both coarse and fine sand beaches, as well as from inland sites featuring either soil or sand (Table 3.1).

Seeds were sown on moist soil (50:50 Fafard C-2:Metro Mix 200) in every other cell of 36-cell flats (so that expanded rosettes would not shade each other). To break dormancy and synchronize germination, flats were covered and placed in a 5 °C cold room for 7-8 days. Flats were then moved to a 20 °C, 16-hour day greenhouse. A randomized block design was used; three replicates of each line were randomized within each of fifteen blocks. Each block was composed of a single treatment group (for ease of treatment application) with three blocks for each treatment (Table 3.2). After one week, plants were thinned to one per cell. Treatments began after two weeks, when the plants had about five true leaves. Plants were fertilized once a week (Scott's Pro Peat Lite Special 15-16-17 fertilizer, 200 ppm N) and watered every other day or as needed.

Line name	ID number	Latitude	Longitude	Home environment
Ale41-1	991	55.3833	14.05	Fine sand beach
Ale44-4	992	55.3833	14.05	Fine sand beach
FlyA3	9380	55.7488	13.3742	Inland sand
Fri 1	9381	55.81064	14.20914	Fine sand beach
Fri2	9382	55.81064	14.20914	Fine sand beach
Fri 3	9383	55.81064	14.20914	Fine sand beach
HolA1 2	9405	55.7491	13.399	Inland soil
HolA2 2	9407	55.7491	13.399	Inland soil
Hov2-1	8423	56.1	13.74	Inland soil
Hov 3-2	6036	56.1053	13.7132	Inland sand
Hov3-5	6038	56.1	13.74	Inland sand
Hov 4-1	8306	56.1	13.74	Inland sand
Hovdala 2	6039	56.1053	13.7132	Inland sand
Kni 1	6040	55.66	13.4	Inland soil
Liarum	8241	56.0328	14.775	Inland soil
Rev 1	8369	55.6942	13.4504	Inland soil
Rev 2	6076	62.8	18.2	Inland soil
Rev 3	6077	62.8	18.2	Inland soil
San 2	8247	56.07	13.74	Inland sand
Sim 1	9442	55.5678	14.3398	Coarse sand beach
TV4	6252	55.5796	14.3336	Coarse sand beach
TV7	6255	55.5796	14.3336	Coarse sand beach
TV10	6258	55.5796	14.3336	Coarse sand beach
TV22	6268	55.5796	14.3336	Coarse sand beach
TV30	6276	55.5796	14.3336	Coarse sand beach
Ull 2-3	6973	56.0648	13.9707	Inland soil
Ull2-5	6974	56.0648	13.9707	Inland soil
Ull3-4	6413	56.0648	13.9707	Inland soil
Var A1	9476	55.58	14.334	Coarse sand beach
Var 2-1	7516	55.58	14.334	Coarse sand beach
Var 2-6	7517	55.58	14.334	Coarse sand beach
VM12		55.7	14.2	Fine sand beach
Yst 1	9481	55.42422	13.84837	Fine sand beach
Yst 2	9482	55.42422	13.84837	Fine sand beach

Table 3.1 Lines used in pilot salt and drought stress experiments.

Treatment	Days between spray	Spray intensity
Control	2	Heavy
Mild	4	Light
Moderate (2-day)	2	Light
Moderate (4-day)	4	Heavy
Harsh	2	Heavy

Table 3.2 Salt spray treatment regimen. Leaves were misted with either water (Control) or 7.5 ppt solution of Instant Ocean®. This salinity level mimics that of the Baltic Sea off the coast of southeastern Sweden (helcom.fi).

Most accessions from Sweden require overwintering to stimulate flowering. 23 days after placement in the greenhouse, flats were covered and placed in a 5 °C, 8-hour day cold room for eight weeks. During this time, flats were watered once a week or as needed to keep soil moist. After eight weeks, flats were returned to the greenhouse (20 °C, 16 hour day) and treatments resumed as above. When plants flowered and siliques began to dry, plants were bagged in order to collect seeds. Plants were considered dead when no green tissue remained. Dead plants were moved from flats and allowed to dry completely. Seeds were collected from dried plants and the seed output for each individual plant weighed.

Because accessions varied in seed production under replete conditions, I used the proportional change in seed production when treated as the response phenotype. Because control and treatment plants were not paired in the experiment, I used two different approaches to calculate proportional differences and to test for effects of treatment, source (coarse sand beach, fine sand beach, inland sand, inland soil) and treatment:source interactions. A significant treatment:source interaction would suggest that plants from different source populations respond differently to stressors – an indication of site-specific adaptation.

In the first approach (Method A), I used the average seed production for each line in each treatment to calculate the proportional difference in seed production between treatment and control: (Average Treatment – Average Control) / Average Control. Thus, a line with a positive value experiences an increase in fecundity when treated and a line with a negative value experiences a decrease in fecundity when treated. I used a box cox transformation to normalize the data and then tested for effects of treatment and source using a linear model with the formula: (Proportional difference ~ Treatment + Source + Treatment:Source).

In the second approach (Method B), for each treatment level, I randomly matched each individual treatment plant with an individual control plant from the same line (without replacement) and calculated the proportional difference in seed production for each pair: (Treatment – Control) / Control. I applied a box-cox transformation to this set of proportional differences and used the linear model (Proportional difference ~ Treatment + Source + Treatment:Source) to test for effects of Treatment, Source, and Treatment:Source. I stored the p-values for each effect, repeated this exercise for a total of 1000 datasets constructed using random pairs as described above and plotted the distribution of p-values for each effect.

3.3.2 Results

Methods A and B both suggest significant effects of treatment and source, but no significant effect of the interaction between treatment and source (Table 3.3, Figure 3.1).

Method	Effect	p-value
A	Treatment	<0.0001***
	Source	0.019*
	Treatment:Source	0.55
B	Treatment	0.0002***
	Source	0.006**
	Treatment:Source	0.47

Table 3.3 Model results for Methods A and B using salt spray pilot data. P-values reported for Method B are averages of 1000 models (see text and Figure 3.1)

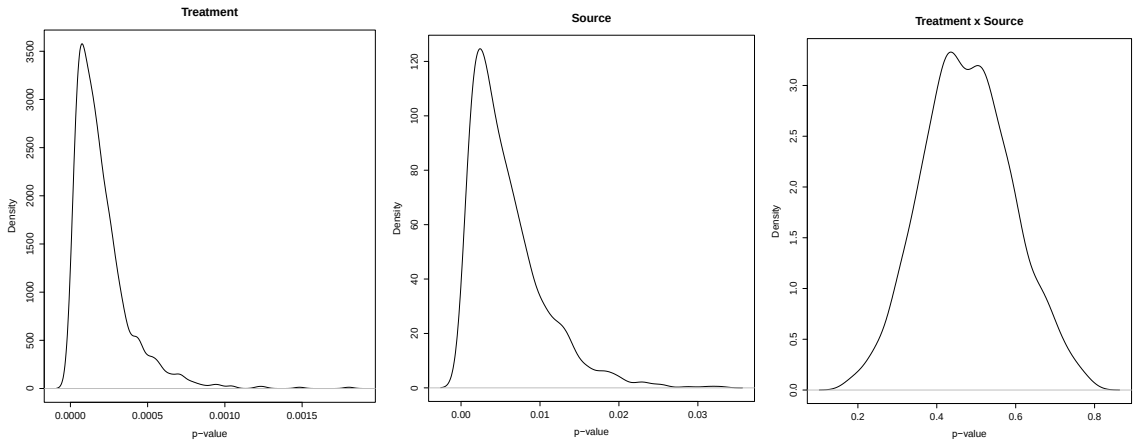

Figure 3.1 Density plots of p-values produced from 1000 ANOVAs as described in the text. (a) Treatment (b) Source (c) Treatment:Source

I expected to see a significant decrease in fecundity with increasing intensity of salt spray treatments. However, the observed treatment effect does not suggest a decline in fecundity when plants are treated with salt spray (Figure 3.2). Instead, the significant effect of treatment appears to be the result of a small increase in fecundity exhibited by plants sprayed heavily every four days.

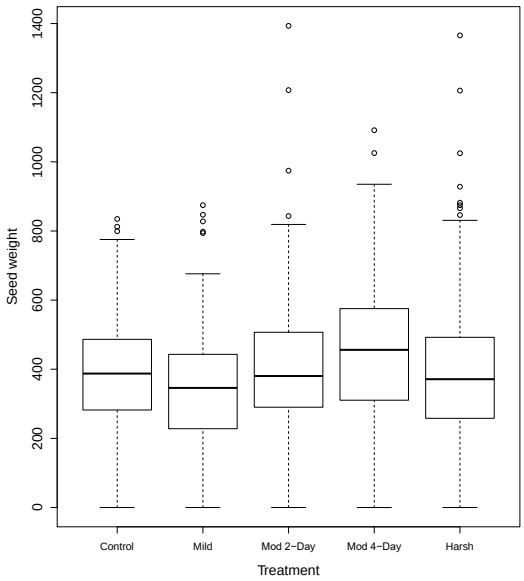

Figure 3.2 Seed production (by total weight) of all lines by salt spray treatment level.

Plants from coarse sand beaches had lower fecundity overall (Figure 3.3a). In addition, the significant effect of source on proportional seed production appears to be the result of two coarse sand beach lines (Var A1 and TV30) in the Moderate-4-Day treatment (Figure 3.3b). However, there is no significant treatment by source effect; coarse sand beach lines do not appear to react differently to salt spray than inland or fine sand beach lines (Figure 3.4). Rather, lines from different sources demonstrate fixed differences in fecundity; these differences are unaffected by salt spray treatment. For these reasons, I conducted no further testing of salt spray tolerance. I suggest potential reasons that I failed to detect tolerance to salt spray in the Discussion below.

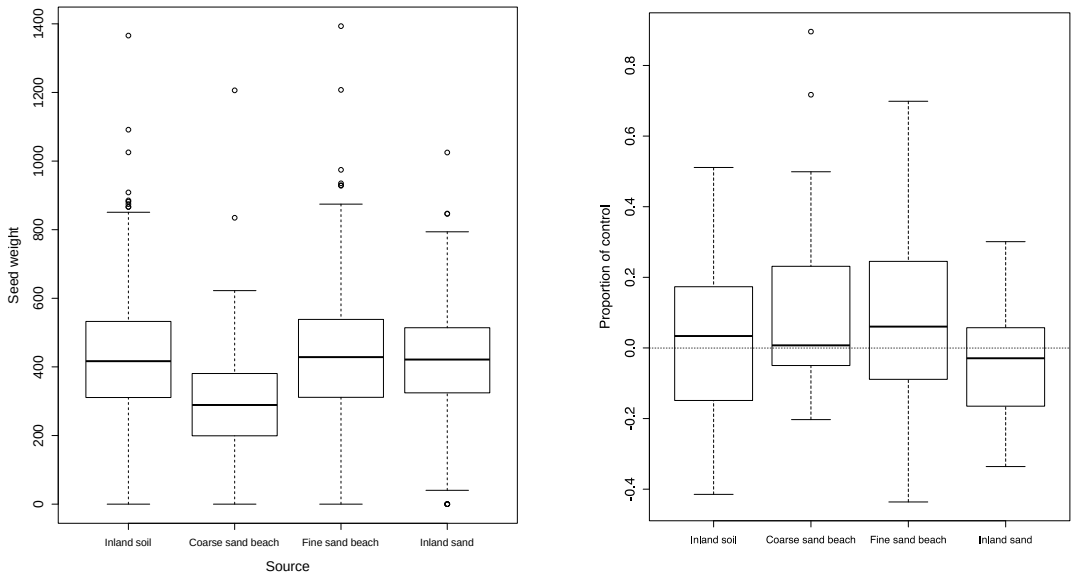

Figure 3.3 (a) Seed production (by total weight) by source (including all treatment levels). (b) Proportion of control seed produced by source (including all treatment levels).

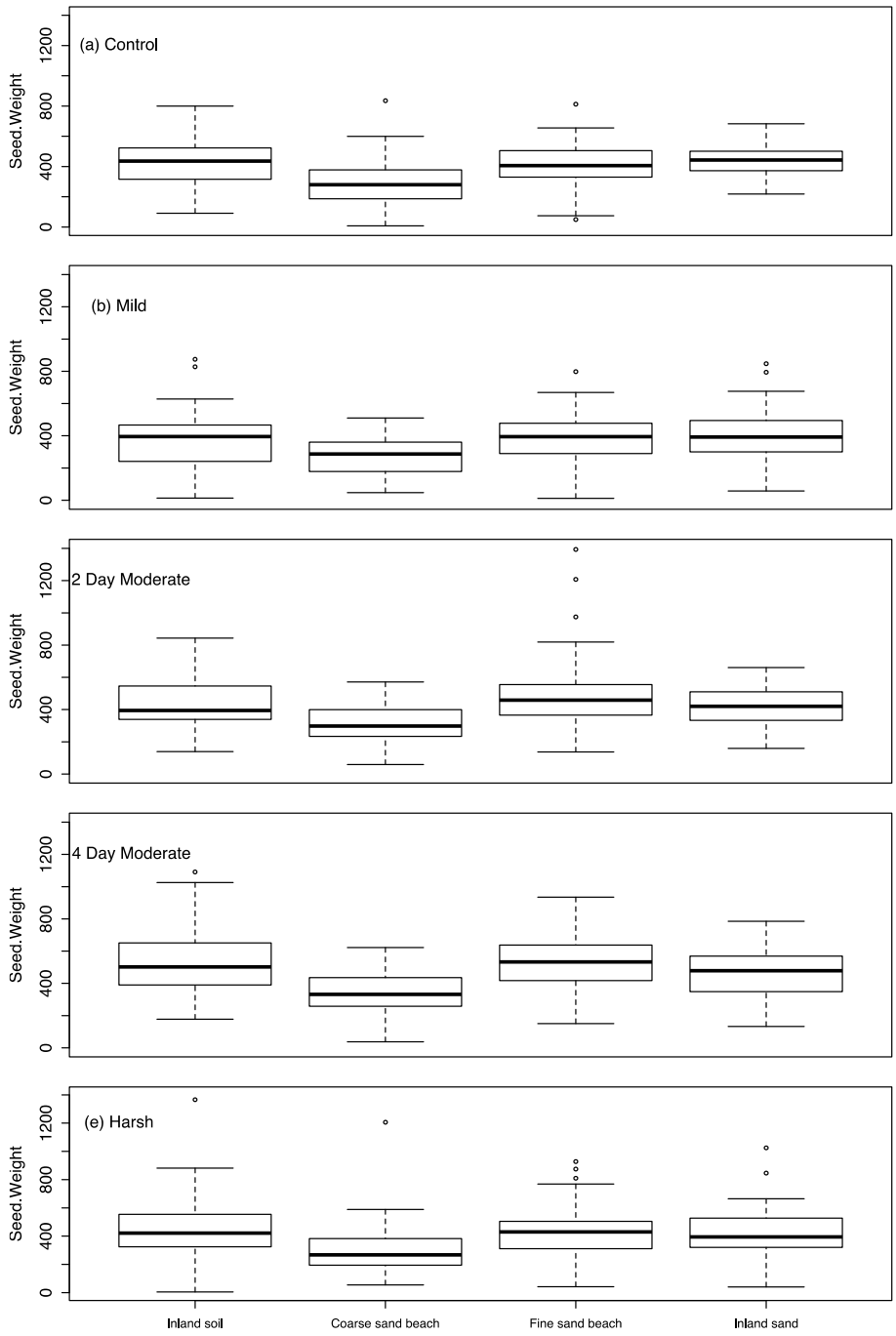

Figure 3.4 Seed production by source for each treatment level (a) Control (b) Mild (c) Moderate 2-Day (d) Moderate 4-Day (e) Harsh.

3.4 Drought Pilot

3.4.1 Methods

Drought pilot experiments were conducted using the same 34 accession panel as the salt spray pilot (Table 3.1).

As above, seeds were sown on moist soil (50:50 Fafard C-2:Metro Mix 200) in every other cell of 36-cell flats. To break dormancy and synchronize germination, flats were covered and placed in a 5 °C cold room for 7-8 days. Flats were then moved to a 20 °C, 16 hour day greenhouse. As above, I used a randomized block design with three replicates per line randomized in each of three blocks of each treatment (twelve total, Table 3.4). After one week, plants were thinned to one per cell. Treatments began after two weeks, when the plants had about five true leaves. Plants were fertilized once a week (Scott's Pro Peat Lite Special 15-16-17 fertilizer, 200 ppm N).

Treatment	Days between watering
Control	2
Mild	3
Moderate	4
Harsh	5

Table 3.4 Drought treatment regimen. Plants were watered to soil saturation. Treatment levels were chosen such that treated plants were visibly stressed but did not exhibit significant mortality (n=75 dead out of 6104 germinants).

As above, after 23 days of growth, flats were vernalized for eight weeks in a 5 °C, 8 hour day cold room, after which treatment was resumed in the greenhouse. Plants were bagged

when siliques began to dry and removed from treatment when completely dead. Seeds were collected from dried plants and the seed output for each individual plant weighed.

Analysis of the proportional change in seed production under drought stress was conducted using the two methods presented above in the Salt Spray Methods section.

3.4.2 Results

Overall, plants that were mildly drought treated did not differ in seed production from control plants, but moderate and harsh drought treatment plants produced fewer seeds (Figure 3.5). These results suggest that treatment levels were appropriately chosen. As in the salt spray experiment, plants from coarse sand beaches produced fewer seeds than plants from fine sand beaches and inland sites across all treatments (Figure 3.6a). Both approaches to the data described in the Methods section showed significant effects of treatment, source, and treatment:source interaction in the pilot panel of 34 lines (Table 3.5, Figure 3.7).

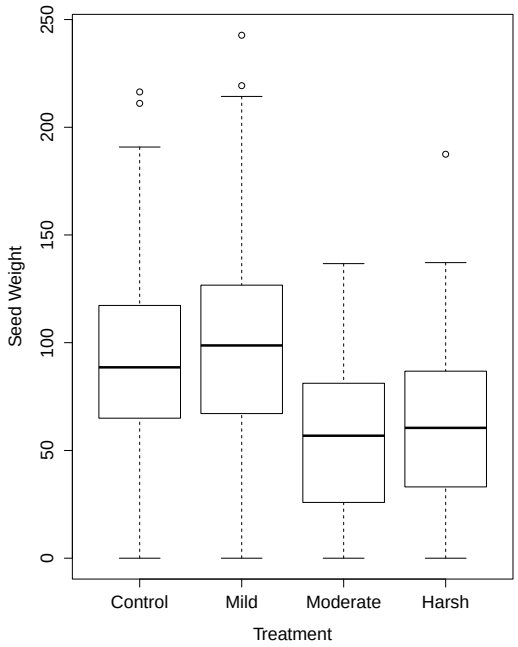

Figure 3.5 Seed production (by total weight) of all lines by treatment level.

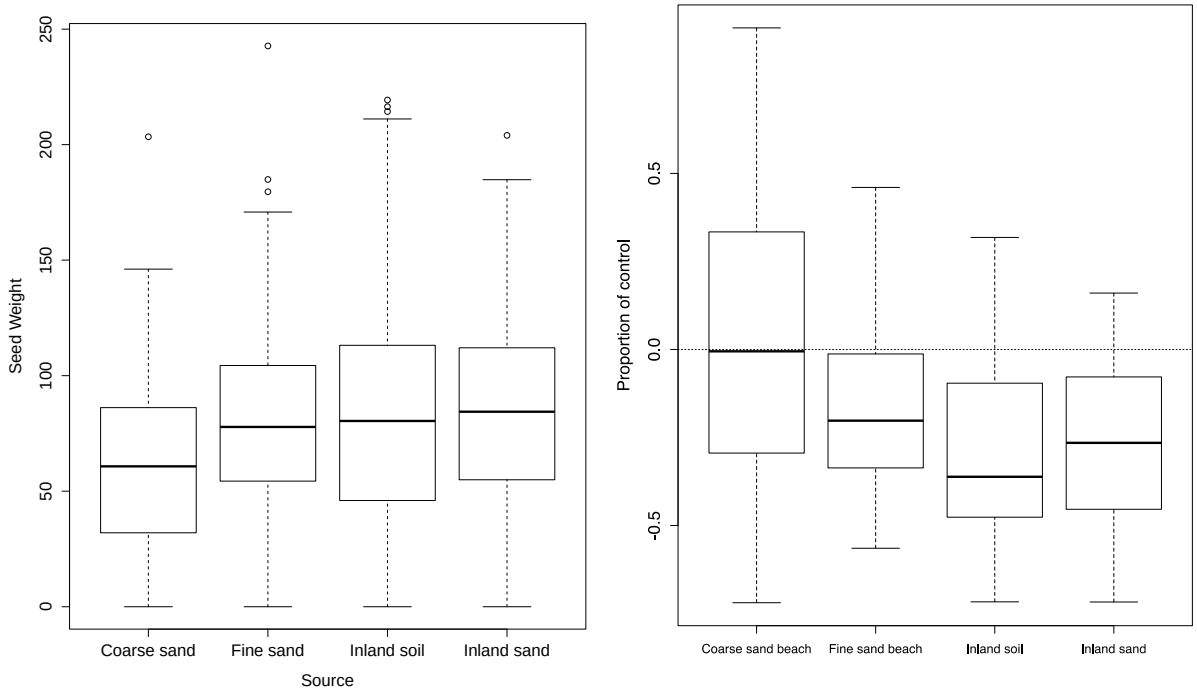

Figure 3.6 (a) Seed production (by total weight) by source (including all treatment levels). (b) Proportion of control seed produced by source across all treatments.

Method	Effect	p-value
A	Treatment	<0.0001***
	Source	<0.0001***
	Treatment:Source	0.0005**
B	Treatment	<0.0001***
	Source	<0.0001***
	Treatment:Source	<0.0001***

Table 3.5 Model results for Methods A and B using drought pilot data. P-values reported for Method B are averages of 1000 models (see text and Figure 3.5)

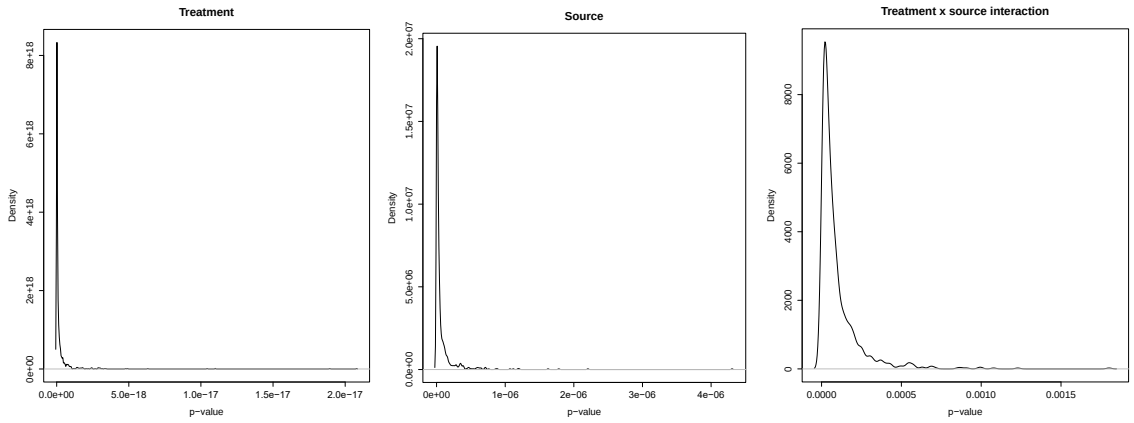

Figure 3.7 Density plots of p-values produced from 1000 ANOVAs as described in the text. (a) Treatment (b) Source (c) Treatment:Source

Plants from different sources exhibited different responses to drought treatment, as predicted by likely water regimes in beach and inland environments. Across all treatments, plants from fine sand beaches and inland sites experienced an average decrease in fecundity compared to control, while coarse sand beach plants did not (Figure 3.6b). I noted two separate effects that contributed both to the significant effect of source and to the significant effect of treatment x source on proportional seed production. First, while mildly drought treated inland and fine sand beach plants did not significantly differ in seed production from control plants, coarse sand beach plants produced on average 47% more

seed in mild drought conditions (Table 3.6, Figure 3.8, 3.9). This increased performance in drier conditions could reflect local adaptation to a constitutively drier beach environment.

	Coarse sand beach lines	Inland and fine sand beach lines
Mild	0.47	0.028
Moderate	-0.41	-0.37
Harsh	0.0031	-0.37

Table 3.6. Average proportional change in total seed produced by weight for each of three treatment levels for lines from coarse sand beaches and conspecific lines from inland sites and fine sand beaches.

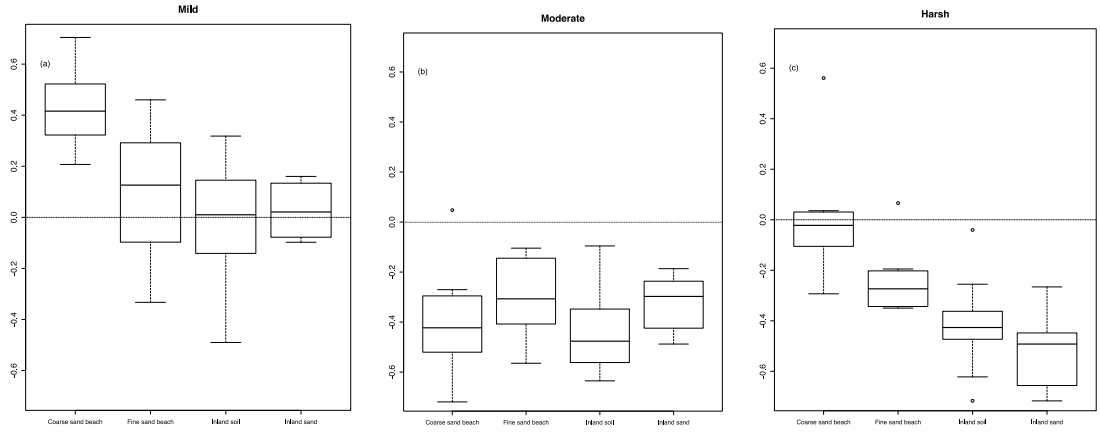

Figure 3.8 Proportion of control seed production by source for each treatment level (a) mild (b) moderate (c) harsh.

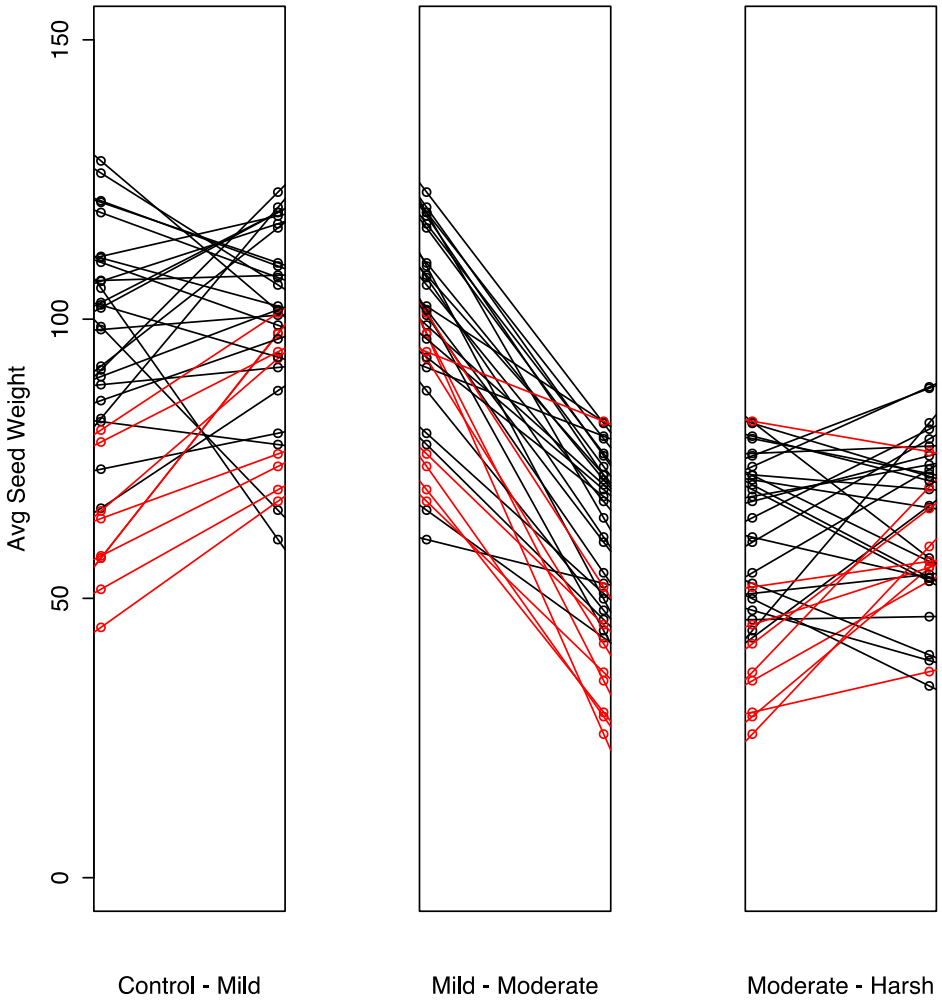

Figure 3.9 Seed production across all treatments. Each accession is represented by a line. Inland and fine sand accessions are in black, coarse sand beach lines are in red.

Secondly, under moderate and harsh drought stress, fine sand beach and inland plants reduced their fecundity by 37%. Coarse sand beach plants suffered similar reductions in fecundity when moderately stressed (41%), but under harsh conditions produced as much seed as coarse sand lines in control conditions (+0.3% difference). This pattern could result from physiological responses active only when the plant is subject to more extreme water deficits.

With this evidence of differential drought response in inland and beach lines, I sought to explore the genetic variation underlying these patterns by repeating the tolerance assay in a larger regional mapping panel.

3.5 Drought Mapping

3.5.1 Methods

To generate drought phenotypes for GWA mapping, I repeated the drought tolerance experiment described above using a panel of 298 Swedish accessions (Appendix B Table 2). I used the same methodology as the pilot with the following alterations. Instead of a randomized block design, I used a fully randomized design. I randomized six replicates per line for each treatment across 100 flats (again using every other cell of 36 cell flats), then randomized all 400 flats across the greenhouse. To minimize the effect of position in the cold room and the greenhouse, I shifted plants to new positions every 4-5 days until they flowered. Plants flowering in the first three weeks, before overwintering, were placed in new flats and not subject to vernalization.

Data analysis

First, I removed plants that flowered before vernalization from the analysis. For lines where more than half of the individuals flowered before vernalization, all individuals from

that line were removed. Other plants that flowered before vernalization were also removed, but the vernalized individuals from these lines remained in the data set.

Next, I repeated the analyses described above for 32 of the 34 lines also used in the pilot experiment (Hovdala-2 and Ull 2-3 flowered before vernalization and were removed; their inclusion does not change the results described below). I compared the results to the pilot data in order to see if the effects detected there were repeated in the larger experiment.

GWA mapping

I performed GWA using both univariate and multivariate linear mixed models in GEMMA (Genome-wide Efficient Mixed Model Association, Zhou & Stephens 2012, 2014) with the 250K SNP set (Kim et al. 2007). GEMMA tests each SNP for association with either a univariate phenotype, in this case, proportional seed production in mild conditions, proportional seed production in moderate conditions, and proportional seed production in harsh conditions were each tested separately, or a multivariate phenotype (Stephens 2013); in this case, each accession had a multivariate phenotype of the form (proportional seed production in mild conditions, proportional seed production in moderate conditions, proportional seed production in harsh conditions). Both methods were used because different models have optimal power under different association scenarios, but the association scenario is unknown *a priori*. Testing multivariate phenotypes generally increases the power of the association test, even when only one of the phenotypes tested is driving the association with a variant. However, univariate tests perform the best when one

of the phenotypes tested is directly associated with a variant and the other phenotypes are indirectly associated (i.e. the phenotypes are correlated).

I used a list of SNPs ranked by p-value to create a ranked list of genes found within 20 kb of the SNPs, then applied GORILLA (Gene Ontology enRIchment anaLysis and visuaLizAtion tool) to test for enrichment of function categories at the top of the list (Eden et al. 2007, 2009) without selection of a particular threshold of significance.

Using PANTHER (Thomas 2003; Mi et al. 2012, 2013), I tested for statistical overrepresentation of Gene Ontology functional categories in genes within 20 kb of SNPs with p-values smaller than 5×10^{-6} (a threshold less stringent than that demanded by the overly conservative Bonferroni correction chosen to capture all apparent peaks in the data). I also used the p-values associated with these SNPs to test for significant p-value enrichment in any functional category.

3.5.2 Results

I removed all individuals from 22 lines where more than half of the individuals flowered before vernalization, as well as 73 other individuals (from 22 different lines, Appendix B Table 8). The resulting data set consisted of 5575 individuals from 276 lines.

Treatment effects seen in the pilot were recapitulated in the mapping panel. Control and mild treatment groups did not differ from each other in seed production. Moderate and

harsh groups also did not differ from each other, but produced less seed than control and mild plants (Figure 3.10).

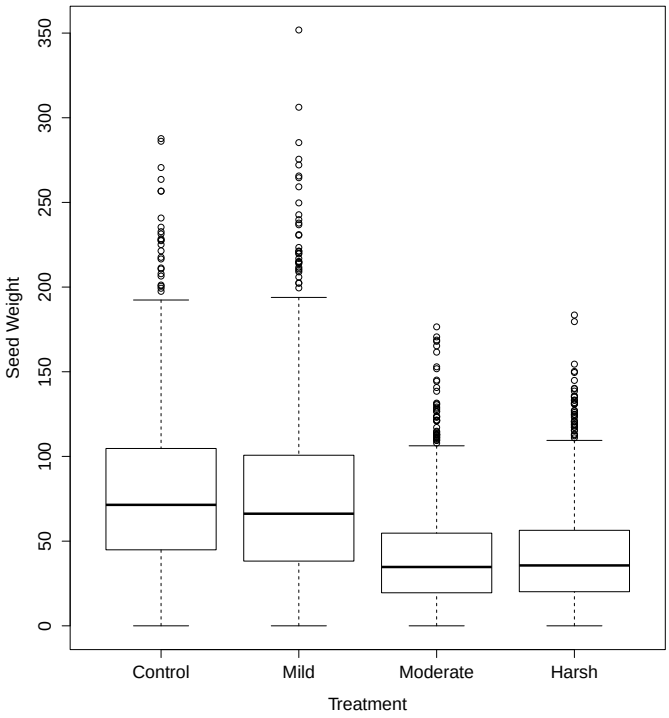

Figure 3.10 Seed production (by total weight) of all lines by treatment level.

Before examining the data for all 276 accessions, I extracted the results for 32 of the 34 lines used in the pilot experiment and repeated the analyses described above. As expected, there was a significant effect of treatment (Table 3.7, Figure 3.11a, 3.12). I saw a significant effect of source with Method B, but not with Method A (Table 3.7, Figure 3.11b, 3.13). This is likely due to the large variance in coarse sand proportional seed production, resulting in no significant effect when line averages were used (Method A) but an average significant effect when randomly created datasets were used (Method B). The effect of source did not

recapitulate results from the pilot experiment; coarse sand beach plants overall did not differ from fine sand and inland plants in proportional seed production.

Method	Effect	p-value
A	Treatment	<0.0001***
	Source	0.57
	Treatment:Source	0.90
B	Treatment	<0.0001***
	Source	0.006*
	Treatment:Source	0.62

Table 3.7 Model results for Methods A and B using drought mapping data for 32 lines also used in the pilot experiment. P-values reported for Method B are averages of 1000 models (see text and Figure 3.11)

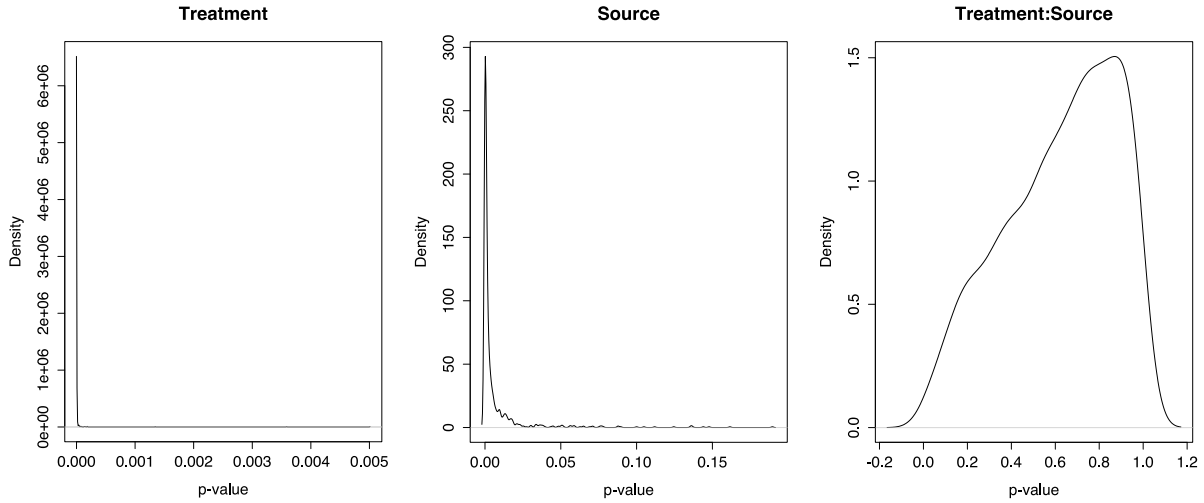

Figure 3.11 Density plots of p-values produced from 1000 ANOVAs as described in the text. (a) Treatment (b) Source (c) Treatment:Source

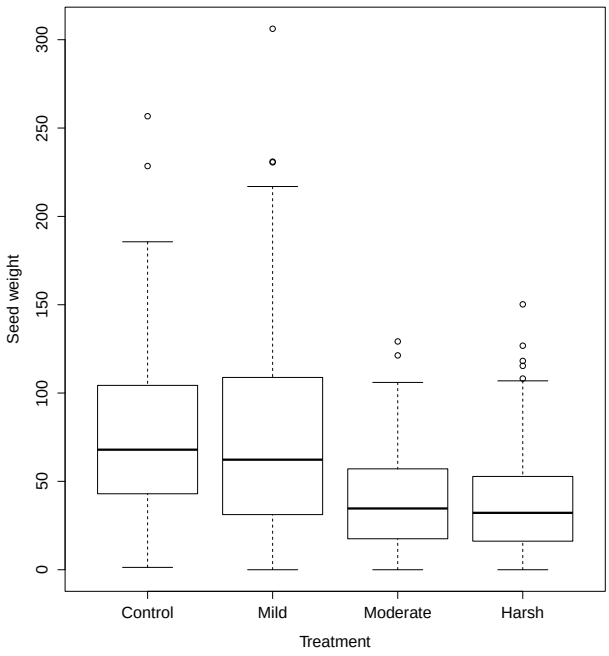

Figure 3.12 Seed production (by total weight) of 33 lines also used in the pilot experiment lines by treatment level.

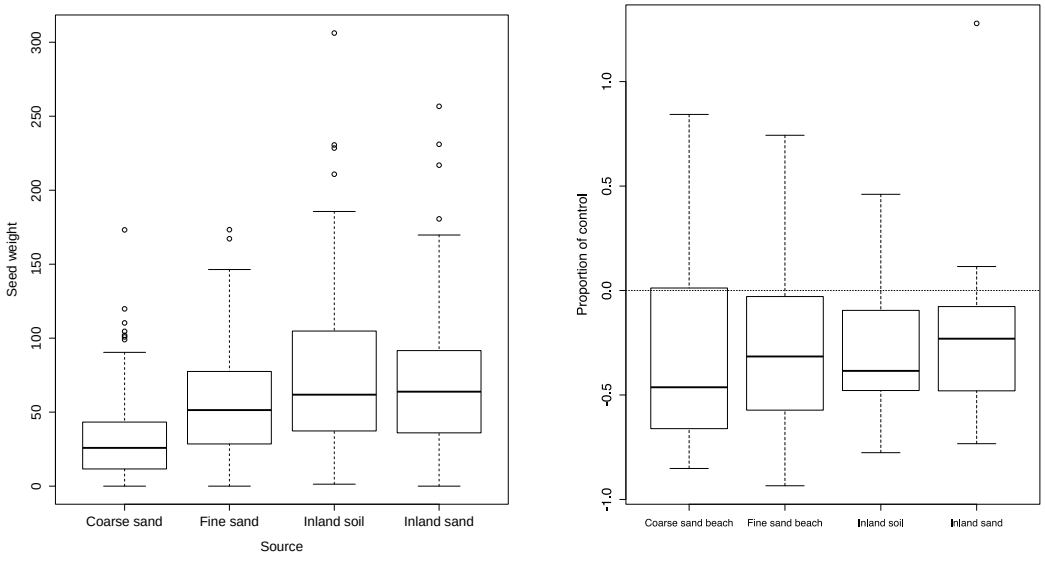

Figure 3.13 (a) Seed production (by total weight) by source (including all treatment levels) for the 33 lines also used in the pilot experiment. (b) Proportion of control seed produced by source across all treatments.

Moreover, I also failed to see significant treatment:source interactions (Table 3.7, Table 3.8, Figure 3.11c, 3.14). Both effects seen in the pilot experiment were no longer detected. When subject to mild drought conditions, coarse sand beach plants did not perform better than when in control conditions, experiencing a non-significant 3% decrease in seed production. When subject to harsh drought conditions, coarse sand plants did not perform differently than in moderate drought conditions, suffering 43 and 50% losses in fecundity (Figure 3.14). Inland and fine sand beach lines experienced similar responses to drought as in the pilot experiment (Table 3.8).

	Coarse sand lines		Inland and fine sand beach lines	
	Pilot	**Mapping**	**Pilot**	**Mapping**
Mild	0.47	-0.029	0.028	0.0076
Moderate	-0.41	-0.43	-0.37	-0.46
Harsh	0.0031	-0.50	-0.37	-0.43

Table 3.8 Comparison between experiments of average proportional change in total seed produced by weight for each of three treatment levels for lines from coarse sand beaches and conspecific lines from inland sites and fine sand beaches.

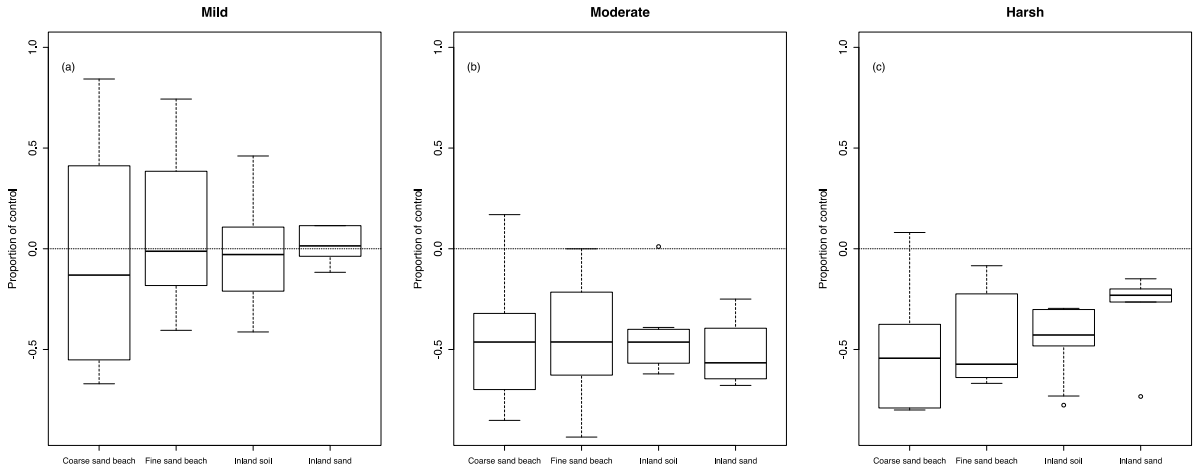

Figure 3.14 For 33 lines also used in the pilot experiment, proportion of control seed production by source for each treatment level (a) mild (b) moderate (c) harsh.

Figure 3.15 shows the pattern of drought response in the 32 lines and, in contrast to Figure 3.9, illustrates that the coarse sand plants did not respond differently from conspecifics (though as in other experiments, their fecundity was lower in all treatments).

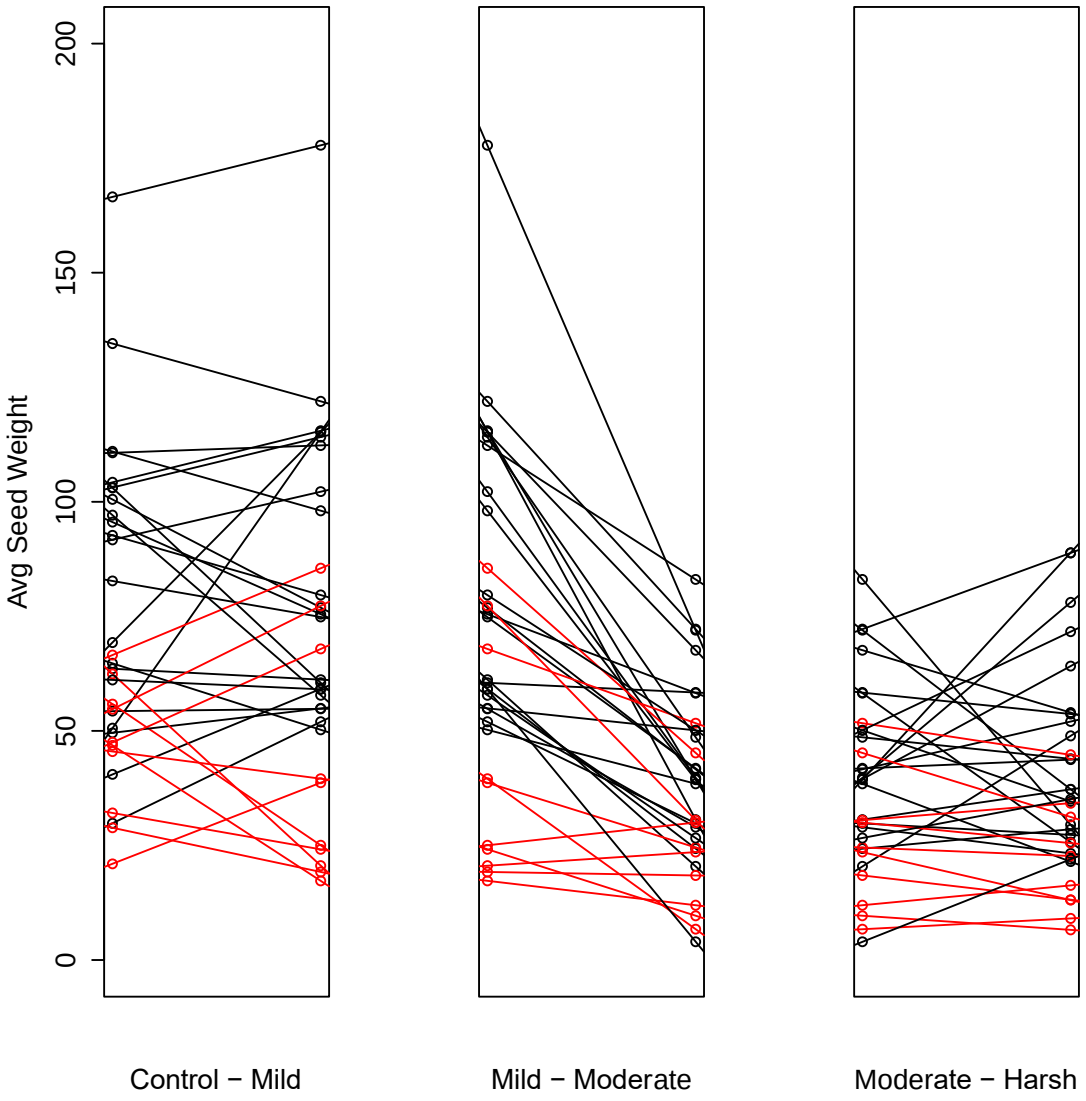

Figure 3.15 Seed production across all treatments. Each accession is represented by a line. Inland and fine sand accessions are in black, coarse sand beach lines are in red.

The discrepancy in results between these two experiments is discussed below.

GWA mapping

Association testing using both univariate and multivariate models resulted in very few peaks (Figure 3.16). For the univariate analysis of fecundity reduction under moderate drought stress, PANTHER detected no overrepresentation or enrichment for any GO functional categories. For the multivariate analysis and for the univariate analysis of mild drought stress, kinase activators were enriched in the set of genes within 20 kb of the top SNPs. For the univariate harsh analysis, genes associated with top SNPs were enriched for transmembrane transporter activity and for involvement in mitosis and cytokinesis. However, in all cases, most genes identified were located in tandem repeats, meaning that they did not occur independently of one another. Moreover, the frequency of genes annotated as involved in the osmotic stress response in the set associated with all top SNPs was not significantly different than their frequency in the genome as a whole.

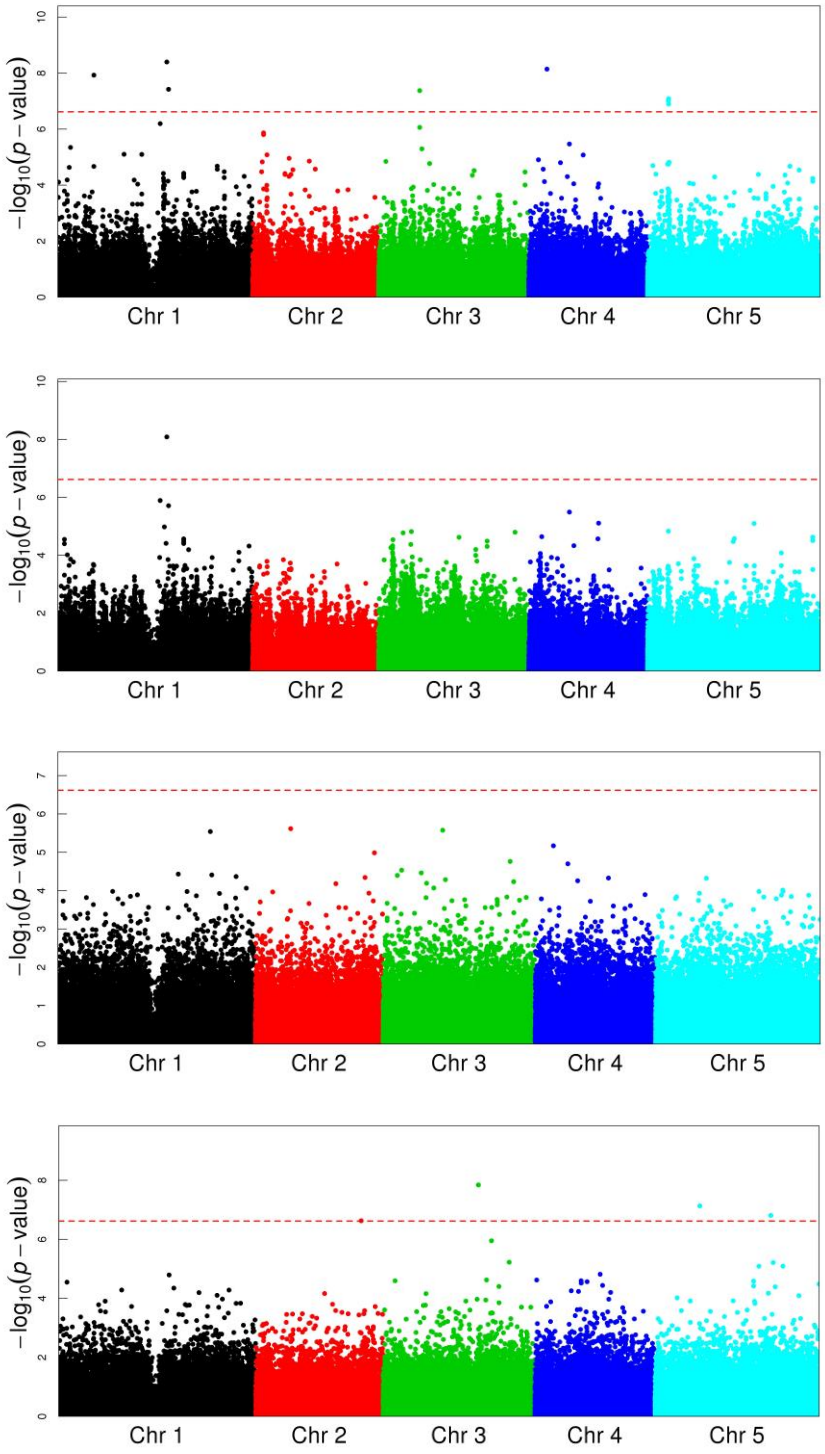

Figure 3.16 Manhattan plot of GWA results for multivariate and univariate models using the 298 Swedish mapping panel. (a) Multivariate drought response phenotype (proportional seed production in mild, moderate, and harsh conditions) (b) univariate mild (c) univariate moderate and (d) univariate harsh drought response. Dotted red line is Bonferroni adjusted significance threshold.

3.6 Discussion

3.6.1 A. thaliana from southern Swedish beaches do not exhibit a response to salt spray

There are several potential reasons for the failure to detect a response to salt spray in Swedish *A. thaliana.*

First, I may have failed to consistently detect tolerance phenotypes because they do not exist. The assumption that beach plants are subject to salt stress was made based on general characteristics of beaches. However, it is possible that *A. thaliana* on Swedish beaches are not salt stressed, or experience the stressor infrequently enough that it does not represent a major selective force. In support of this hypothesis is the fact that the Baltic Sea off the coast of southeastern Sweden, though saline, has much lower levels of salt than the open ocean. All published studies of salt spray have dealt with plants on sea coasts with much higher salinity levels than the Baltic (helcom.fi). In addition, Maun and Perumal (1999) note that in temperate regions with cold winters, major wind and storm events with significant salt spray often occur largely in the autumn and winter when plants are dormant and less subject to damage. Not all coastal ecotypes are salt tolerant. Locally adapted populations of *Brassica nigra* on the coast appeared to avoid high salinity soils, causing the authors to speculate that climate, not salinity, is driving adaptation to the coast in this species (Bischoff & Hurault 2013).

This possibility could be addressed by collecting additional data about salt load on the beaches in question. Salt deposition on plant tissues can be quantified by use of salt spray traps on beaches (Boyce 1954, Wilson & Sykes 1999). Soil salinity measurements also provide data about salt loads experienced by beach plants. Such data could suggest that plants rarely face salt stress on Baltic beaches. If salt does appear to be present at levels likely to be harmful to *A. thaliana*, greenhouse experiments could be designed to more closely replicate beach conditions and test for differences in tolerance between beach and inland lines.

A second possibility is that *A. thaliana* only exhibits a reduction in fecundity in response to salt spray if plants experience the stressor at a particular developmental period (for example, at the seedling stage, not tested here) or in conjunction with other stressors (for example, low water availability or pathogen damage). The obvious advantage of controlled greenhouse experiments is their ability to test single controlled factors, but at the expense of the reality of field conditions. If stressors applied in the greenhouse fail to adequately replicate the conditions or combinations of field stressors, the experimental results have no bearing on natural plant behavior. Thus, meaningful tolerance phenotypes may exist, but were not detected under the conditions I used.

If candidate stressors were identified, further experimentation could reveal the effect of multiple interacting stressors. However, without any additional information about the nature or timing of such stressors, it is difficult to know what combinations to test. Field observations could potentially give clues – for example, if periods of acute water stress and

heavy salt spray coincided. Additional experimentation could then lead to a deeper understanding of abiotic pressures faced by plants on the Baltic coast.

3.6.2 GWA results are not enriched for drought/stress response candidate genes

Neither GWA results alone nor tests for enrichment of GO functional categories among genes near the most strongly associated SNPs suggested interesting avenues of follow-up for drought response in this mapping population. Given the lack of clear GWA results, this is perhaps unsurprising. The response phenotype used, changes in total lifetime reproductive fitness with respect to a control group, is a polygenic trait influenced by many environmental factors. Though it captures all effects of drought in an evolutionarily relevant measure, its complexity makes it difficult to parse the underlying genetic architecture. In future studies of this population, mapping specific physical or physiological phenotypes correlated with drought response could provide a route to understanding mechanisms involved in response to drought. However, if beach populations are not strongly differentiated in drought response (as they were not in this GWA study), such experiments will not provide insight into beach-specific adaptation to drought.

3.6.3 Drought responses in *A. thaliana* lines from coarse sand beaches are inconsistent across experiments

What explains the discrepancy between the results for the 32 lines in the pilot and mapping experiments? The reduction in number of replicates per line from nine to six in the second

experiment could have potentially have reduced the experimental power such that the effect of interest was no longer significant. However, in this case, I would still expect to see similar patterns in the data, with coarse sand plants performing better in mild than in control conditions and increasing their seed production in harsh compared to moderate conditions. Instead, coarse sand beach lines do not appear to behave differently from other lines in any condition.

A potential explanation for the failure of the mapping experiment to detect differential effects of drought is a change in experimental conditions. There may have been an undetected change in the application of the drought stress or an additional unknown stressor such as a pest or pathogen present in the greenhouse during the second but not the first experiment. Overall, control plants in the second experiment produced 15% less seed than in the pilot; this decreased fecundity was seen across lines and treatments, suggesting that some additional factor suppressed fecundity in these plants. A differential response to this unknown factor in lines from different sources could have swamped the differential drought response.

One possible stressor present in the mapping population and not in the pilot involved early drought stress. After seed planting but before stratification, planted cells were randomized across flats. To perform this three-day randomization, flats were spread out uncovered in the greenhouse; thus, soil was subject to drying, and cells randomized later were subject to greater soil drying than cells randomized earlier. Because plants were randomized one treatment group at a time in order of intensity, early soil drying corresponded with later

drought treatment. Moderate and harsh treatment groups did suffer significantly higher rates of germination failure than control and mild treatment groups (Figure 3.17), presumably as a result of soil drying (all failed germinants were removed from the data set). Though unplanned, however, this stressor represents an increase in treatment intensity rather than a confounding factor. It could only be responsible for the discrepancy in results between the pilot and mapping experiments if early drought stress eliminated or mitigated the drought response of coarse sand beach plants seen in the previous experiment. While not impossible, this seems unlikely.

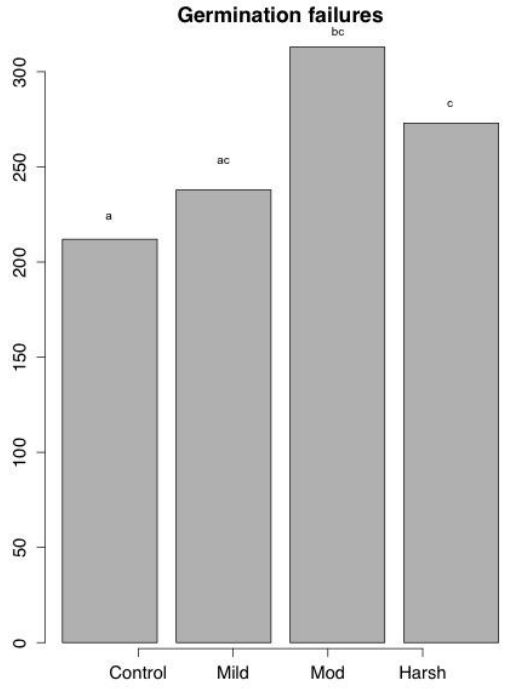

Figure 3.17 Germination failures by treatment level for mapping experiment.

If the pilot experiment was a false positive, we must consider, as with salt spray, whether the beach population is subject to drought stress with a great enough intensity and frequency to select for a different response. The coarse sand beach population in southern

Sweden does experience slightly less annual rainfall than inland lines (47.6 mm compared to 55.6 mm), but fine sand beach lines experience only 49.25 mm and were not differentiated in drought response from inland lines (Hancock et al. 2011). The ability of soil to hold water also has a large effect on water availability, but detailed data on substrate type and quality is lacking for the accessions in question. It is possible that beach lines rarely suffer from reduced water availability. If this is true, then some unidentified factor led to the effects seen in coarse sand plants in the pilot experiment, and these results do not speak to drought response in these lines.

As with salt spray, further data about conditions at the beach would allow for more carefully designed studies mimicking beach conditions. Additionally, a new experimental design with increased power is necessary to settle this question. An improved design would also eliminate early soil drying by moving the randomization step to the pre-planting stage.

3.7 Conclusion

In this chapter, I measured fecundity changes in response to salt spray and drought stress in populations of *Arabidopsis thaliana* from beach and inland sites. I found that no consistent decrease in fecundity occurred when plants were subject to salt spray. I found that lines from coarse sand beaches responded differently to drought stress than fine sand beach and inland conspecifics, producing the most seed under mild drought stress and

reducing fecundity less under harsh drought conditions. However, I was unable to replicate this effect.

Both sets of experiments would have benefited from additional data on field conditions at beach sites in Sweden. Detailed information on salt loads, timing of salt deposition, timing of precipitation events, and soil water capacity would have informed experimental design so that greenhouse conditions more closely mimicked field conditions faced by plants.

Combined with adequately powered experiments, such measures would allow a more decisive conclusion regarding tolerance to salt and drought in southern Swedish beach lines. Nevertheless, the current data does suggest some tentative conclusions. First, the lack of fecundity response in salt spray treated plants, even those in the harshest treatment, suggests that the low salinity of the Baltic sea may mean that coastal plants do not suffer from salt spray stress. Indeed, an increase in fecundity was seen in moderately treated plants.

Though the drought experiments remain inconclusive, the pilot experiment suggests that differential responses may occur in coarse sand beach plants and fine sand beach and inland conspecifics. If confirmed, mapping experiments could indicate genetic regions associated with maintenance of fecundity, and measurements of physiological responses to drought could indicate the mechanisms by which plants maintain fecundity under harsh drought.

CONCLUSION

Studying the process and outcome of adaptation to a variety of environments is a central preoccupation at the intersection of ecology and evolution. Connecting phenotypes important for fitness to the underlying genetic basis and the relevant environmental drivers is crucial to characterizing adaptation of a particular population to a particular environment. Here, I have explored characteristics important for adaptation to coarse sand beaches in southern Swedish *Arabidopsis thaliana.* To attempt to identify the three features described above, I approached this question from several angles, searching for genomic regions associated with selection with an agnostic scan, as well as investigating phenotypes and environmental stressors believed to be associated with beach plants generally and with the focal population in particular. I found evidence that specific phenotypes as well as a set of genomic regions are important for adaptation to beaches, and in one case also identified the associated environmental stressor.

In the focal beach population of *A. thaliana,* the hypocotyl length phenotype and an associated genetic variant differentiated the population from inland conspecifics and were shown experimentally to convey a fitness advantages when seedlings were required to emerge from burial in order to survive. In addition, the agnostic genomic scan identified genes implicated in hypocotyl elongation. The combination of PBS and GWA data also suggested that multiple regions in the genome contribute to adaptive variation in this phenotype. Thus, in the case of hypocotyl length, connections between phenotype, genotype, and environmental stressor were all identified.

Seed size was also identified as a phenotype important to beach plants, but the complex polygenic nature of this trait did not allow for confirmation of specific genetic variants associated with the phenotype. Additionally, greenhouse experiments showed no germination or seedling survival advantage associated with larger seeds. Large seeds are generally believed to be important in harsh environments like beaches, but neither a specific environmental stressor nor a fitness advantage could be experimentally demonstrated here. Work to address these outstanding questions is discussed below.

Additionally, a set of genomic regions that appear to be under selection were identified, but the nature and specific role of the genes driving the signal in these regions are unknown.

On the other hand, specific environmental stressors common to beaches showed no evidence of being adaptive drivers in this population. Salt spray at a concentration similar to the Baltic Sea did not decrease the total reproductive fitness of *A. thaliana* lines from either beach or inland sites, suggesting that the mild salinity of the Baltic is not sufficient to cause stress. Reduced water availability did negatively impact fecundity in *A. thaliana,* but no compelling evidence demonstrated a difference in response between beach and inland lines, as would be expected if beach lines were adapted to drier conditions.

These results present a partial picture of the characteristics of beach *A. thaliana.* They confirm initial observations of phenotypic differentiation between beach and inland lines and indicate that coarse sand beach plants have undergone selection due to a unique suite

of environmental drivers, arriving at a phenotypic and genotypic identity that differentiates them from inland and fine sand beach conspecifics and that includes long hypocotyls and large seeds, presumably among other unidentified adaptive phenotypes.

In addition, the results demonstrate that a combination of approaches is beneficial to address adaptive questions. Here, each approach provided a unique insight into the population in question; for example, hypocotyl elongation genes identified in the PBS scan were not identified in the hypocotyl GWA (presumably because they represent loci of small effect), and evidence for selection on seed weight was garnered from the combination of PBS and GWA data, but not from either approach independently.

However, some approaches were more fruitful than others. Prior identification of environmental drivers proved useful in thinking about adaptive reasons for elongated hypocotyls, a phenotype already noted in beach plants. However, investigations of drought and salt, common candidate drivers of adaptive differentiation on beaches, were not informative. No adaptive responses were detected in experiments using these two stressors. Unlike the case of hypocotyl length, no specific phenotypes were identified *a priori* as associated with drought and salt stress. In studying the effects of these stressors, I used fecundity as a catch-all phenotype, integrating all physiological responses into a single evolutionarily relevant measure. Comparing the relative success of the hypocotyl length and drought and salt studies suggests that positive results may be more likely when using more narrowly defined phenotypes (discussed further below). Using a complex, polygenic

and environmentally responsive phenotype like fecundity may not be a reliable approach to investigating adaptation.

A revised approach to this project would maintain a multi-pronged approach, but would adopt a set of field experiments Field experiments early in the process would provide important data, serving several functions. First, reciprocal transplants would provide rigorous quantitative evidence for local adaptation in beach populations, a feature absent from this work. Though phenotypic differences and the presence of unique environmental drivers, as well as the commonness of coastal ecotypes, strongly suggest that beach populations are locally adaptive, reciprocal transplants are the only true gold standard.

Secondly, field experiments would ground greenhouse experiments more firmly in reality by defining experimental parameters, for example, amount and timing of water availability. Field observations would likely also suggest additional important phenotypes and environmental drivers. As discussed above, using total reproductive fitness as a response phenotype is problematic; it is possible that drought and salt spray could have been more effectively studied by using a more narrowly defined, easily measureable phenotype, clearly associated with the environmental driver in question, for example, visible tissue damage as a result of salt spray. While this will not help if adaptation to the driver is not present (if Baltic salt spray really is too weak to cause damage), if phenotype effects are subtle, a precise approach could identify effects of the driver of interest.

Thirdly, effects of substrate clearly play a large role in these populations. The phenotypic differences between fine and coarse sand beaches are as dramatic as those between coarse sand and inland plants. Detailed information on the physical and chemical composition of soils from the beach and inland sites where plants were sampled would improve the realism of greenhouse experiments and potentially lead to the formation of new hypotheses.

The current results also suggest a variety of approaches to more completely describe a coarse sand beach *A. thaliana.* First, several lines of inquiry would lead to a deeper understanding of the important phenotypes already identified. Phenotyping and emergence experiments in F2 plants from crosses with different alleles at the SNP of interest and in F2 plants from crosses between beach plants and *SPA* knockout mutants could confirm that the long hypocotyl phenotype and associated fitness advantages are a result of a strong *SPA2* allele. I have suggested in Chapter 2 that an overactive SPA2 protein, not degraded in light, is responsible for the phenotype; assays for protein levels in long and short hypocotyl lines could shed light on the molecular mechanism involved. Moving away from SPA2, the hypocotyl elongation genes identified in the PBS scan could also be explored for association with fitness advantages during burial. A complex polygenic trait under selection is likely to show allele frequency changes at multiple sites in the genome.

While there is evidence for adaptation related to seed size, the nature of the environmental driver and the underlying genetic variation remain unknown. Field experiments could reveal germination or seedling survival advantages of larger seeded lines that are not seen

in milder greenhouse conditions. Though GWA peaks did not occur over genes known to be involved in seed size, genes falling under these peaks may well contribute to this complex polygenic trait, and may be worth following up. Ranking genes from GWA peaks on the basis of PBS score may be a valuable approach to prioritizing candidates.

More broadly, the PBS results could suggest more genes of interest. Here, I have applied specific thresholds to identify twenty genes extreme in several population genetic measures associated with selection; these are certainly not the only twenty genes implicated in selection in the genome. Exploring a longer list or applying different thresholds will likely reveal more interesting genes.

Lastly, the approach to combining GWA and PBS data that I have described in Chapter 1 immediately suggests an approach to generating more candidate phenotypes. Beginning with the Atwell et al. (2010) phenotype data from a global set of accessions and potentially including other published phenotype sets from global or European *A. thaliana,* testing for an increase in average PBS value with restricted subsets of GWA data will likely yield additional phenotypes that appear to be under selection in the beach population. These phenotypes can then be dissected with a suite of approaches similar to those described in this thesis.

Here, I have described a complementary set of approaches and used them to explore adaptation to beaches in a particular population of southern Swedish *A. thaliana.* I have identified genetic and phenotypic variation important to survival on coarse sand beaches,

and I have demonstrated how various approaches provide different types of information. I have also described a new methodology for combining genomic data from different types of scans to provide information on phenotypes under selection that cannot be gleaned from either scan alone. The data and methods included here contribute to the broad ecological study of organisms' adaptation to varied environments.

REFERENCES

Adamski, N.M., Anastasiou, E., Eriksson, S., O'Neill, C.M. & Lenhard, M. 2009. Local maternal control of seed size by KLUH/CYP78A5-dependent growth signaling. *Proceedings of the National Academy of Sciences USA,* **106**, 20115–20120.

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

Allen, W.R. & Sheppard, P.M. 1971. Copper tolerance in some Californian populations of the monkey flower, *Mimulus guttatus*. *Proceedings of the Royal Society B*, **177**, 177-196.

Alonso-Blanco, C., Blankestijn-De Vries, H., Hanhart, C.J. & Koornneef, M. 1999. Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana. Proceedings of the National Academy of Sciences USA*, **96,** 4710-4717.

Alonso-Blanco, C., Aarts, M.G.M., Bentsin, L., Keurentjes, J.J.B., Reymond, M., Vreugdenhil, D. & Koornneef, M. 2009. What has natural variation taught us about plant development, physiology, and adaptation? *The Plant Cell,* **21,** 1877-1896.

Anderson, J.T., Willis, J.H. & Mitchell-Olds, T. 2011. Evolutionary genetics of plant adaptation. *Trends in Genetics*, **27**, 258-266.

Arany A.M., de Jong, T.J. & van der Meijden, E. 2009. Herbivory and local genetic differentiation in natural populations of *Arabidopsis thaliana* (Brassicaceae). *Plant Ecology*, **201**, 651–659

Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y.,…Nordberg, M. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature,* **465,** 627-631.

Baker, H.G. 1972. Seed weight in relation to environmental conditions in California. *Ecology,* **53,** 997-1010.

Balcerowicz, M., Fittinghoff, K., Wirthmueller, L, Maier, A., Fackendahl, P.,…Hoecker, U. 2011. Light exposure of *Arabidopsis* seedlings causes rapid de-stabilization as well as selective post-translational inactivation of the repressor of photomorphogenesis SPA2. *The Plant Journal,* **65,** 712-723.

Barbour, M.G. & DeJong, T.M. 1985. Response of west coast beach taxa to salt spray, seawater inundation, and soil salinity. *Bulletin of the Torrey Botanical Club,* **104,** 29-34.

Barry, T. 2013. Local adaptation of two cryptic species, *Lasthenia californica* and *Lasthenia gracilis*, to distinct regions within a serpentine outcrop. *Master's Theses,* Paper 4258.

Bates D, Maechler M, Bolker B and Walker S (2015). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-8, http://CRAN.R-project.org/package=lme4.

Baxter, I., Brazelton, J.N., Yu, D., Huang, Y.S., Lahner, B., Yakubova, E.,…Salt, D.E. 2010. A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1;1. *PLoS Genetics,* **6,** 1-8.

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

Bischoff, A. & Hurault, B. 2013. Scales and drivers of local adaptation in *Brassica nigra* (Brassicaceae) populations. *American Journal of Botany,* **100**, 1162-1170.

Bloomer, R.H., Lloyd, A.M. & Symonds, V.V. 2014. The genetic architecture of constitutive and induced trichome density in two new recombinant inbred line populations of *Arabidopsis thaliana*: phenotypic plasticity, epistasis, and bidirectional leaf damage response. *BMC Plant Biology,* **14,** 10.1186/1471-2229-14-119.

Boonman, A., Prinsen, E., Gilmer, F., Schurr, U., Peeters, A.J.M., Voesenek, L.A.C.J., & Pons, T.L. 2007. Cytokinin import rate as a signal for photosynthetic acclimation to canopy light gradients. *Plant Physiology,* **143**, 1841-1852.

Borevitz, J., Maloof, J.N., Lutes, J., Dabi, T., Redfern, J.L.,…Chory, J. 2002. Quantitative traid loci controlling light and hormone response in two accessions of *Arabidopsis thaliana. Genetics,* **160,** 683-696.

Boron, A.K. & Vissenberg, K. 2014. The *Arabidopsis thaliana* hypocotyl, a model to identify and study control mechanisms of cellular expansion. *Plant Cell Reports,* **33,** 697-706.

Bouchabke, O., Chang, F., Simon, M., Voisin, R., Pelletier, G. & Durand-Tardif, M. 2008. Natural variation in *Arabidopsis thaliana* as a tool for highlighting differential drought responses. *PLoS ONE,* **3,** 1-8.

Boyce, S.G. 1954. The salt spray community. *Ecological Monographs,* **24,** 29-67.

Busoms, S., Teres, J., Huang, X., Bomblies, K., Danku, J.,…Salt, D.E. 2015. Salinity is an agent of divergent selection driving local adaptation of *Arabidopsis thaliana* to coastal habitats. *Plant Physiology,* **163,** 915-929.

Callahan, H.S. & Pigliucci, M. 2002. Shade induced plasticity and its ecological significance in wild populations of *Arabidopsis thaliana. Ecology,* **83,** 1965-1980.

Cartica, R.J. & Quinn, J.A. 1980. Responses of populations of *Solidago sempervirens* (Compositae) to salt spray across a barrier beach. *American Journal of Botany,* **67,** 1236-1242.

Chaves, M. & Davies, B. 2010. Drought effects and water use efficiency: Improving crop production in dry environments. *Functional Plant Biology,* **37,** iii-iv.

Chen, H. & Maun, M.A. 1999. Effects of sand burial depth on seed germination and seedling emergence of *Cirsium pitcheri. Plant Ecology,* **140,** 53-60.

Clausen, J., Keck, D.D. & Hiesey, W. 1940. *Experimental studies on the nature of species I. Effect of varied environments on western North American plants.* Carnegie Institute of Washington 520.

Darwin, C.R. 1845. *Journal of researches into the natural history and geology of the countries visited during the voyage of H.M.S. Beagle round the world, under the Command of Capt. Fitz Roy, R.N.* (Second ed.), London: John Murray.

Darwin, C.R. 1869. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.

Dech, J.P. & Maun, M.A. 2005. Zonation of vegetation along a burial gradient on the leeward slopes of Lake Huron sand dunes. *Canadian Journal of Botany,* **83,** 227-236.

Devlin, P.F., Robson, P.R.H., Patel, S.R., Goosey, L., Sharrock, R.A. & Whitelam, G.C. 1999. Phytochrome D acts in the shade-avoidance syndrome in *Arabidopsis* by controlling elongation growth and flowering time. *Plant Physiology,* **119,** 909-915.

Dilkes, B.P., Spielman, M., Weizbauer, R,, Watson, B., Burkart-Waco, D., Scott, R.J. & Comai, L. 2008. The maternally expressed WRKY transcription factor TTG2 controls lethality in interploidy crosses of *Arabidopsis*. *PLoS Biology,* **6**, 2707–2720.

Eden, E., Lipson D., Yogev, S. & Yakhini, Z. 2007. Discovering Motifs in Ranked Lists of DNA sequences. *PLoS Computational Biology*, **3,** e39.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. 2009. GOrilla: A Tool For Discovery And Visualization of Enriched GO Terms in Ranked Gene Lists. *BMC Bioinformatics*, **10**, 48.

El-Soda, M., Kruijer, W., Malosetti, M., Koornneef, M. & Aarts, M.G.M. 2015. Quantitative trait loci and candidate genes underlying genotype by environment interaction in the response of *Arabidopsis thaliana* to drought. *Plant, Cell & Environment,* **38,** 585-599.

Filiault, D.L. & Maloof, J.N. 2012. A genome-wide association study identifies variants underlying the *Arabidopsis thaliana* shade avoidance response. *PLoS Genetics,* **8,** e1002589.

Fischer, M.C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F.,…Widmer, A. 2013. Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology,* **22,** 5594-5607.

Fittinghoff, K., Laubinger, S., Nixdorf, M., Fackendahl, P., Baumgardt, R-L., Batschauer, A. & Hoecker U. 2006. Functional and expression analysis of Arabidopsis *SPA* genes during seedling photomorphogenesis and adult growth. *The Plant Journal,* **47,** 577-590.

Fournier-Level, A., Korte, A., Cooper, M.D., Nordborg, M., Schmitt, J. & Wilczek, A.M. 2011. A map of local adaptation in *Arabidopsis thaliana. Science,* **334,** 86-89.

Friesen, M.L. & von Wettberg, E.J. 2010. Adapting genomics to study the evolution and ecology of agricultural systems. *Current Opinion in Plant Biology,* **13,** 119-125.

Galpaz, N. & Reymond, M. 2010. Natural variation in *Arabidopsis thaliana* revealed a genetic network controlling germination under salt stress. *PLoS ONE*, **5**, 1-10.

Ganade, G. & Westoby, M. 1999. Seed mass and the evolution of early-seedling etiolation. *The American Naturalist,* **154,** 469-480.

Gregory, R.P.G. & Bradshaw, A.D. 1965. Heavy metal tolerance in populations of *Agrostis tenuis* Sibth. and other grasses. *New Phytologist,* **64,** 131-143.

Habel, J.C., Zachos, F.E., Dapporto, L., Rodder, D., Radespiel, U.,…Schmitt, T. 2015. Population genetics revisited – towards a multidisciplinary research field. *Biological Journal of the Linnean Society,* **115,** 1-12.

Hadiarto, T. & Tran, L-S.P. 2011. Progress studies of drought-responsive genes in rice. *Plant Cell Reports,* **30,** 297-310.

Hancock, A.M., Brachi, B., Faure, N., Horton, M.W., Jarymowycz, L.B.,…Bergelson, J. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science,* **334,** 83-86.

Hereford, J. 2009. A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist,* **173,** 579-88.

Herridge, R.P., Day, R.C., Baldwin, S. & Macknight, R.C. 2011. Rapid analysis of seed size in *Arabidopsis* for mutant and QTL discovery. *Plant Methods,* **7**:3.

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

Horton, M.W., Hancock, A.M., Huang, Y.S., Toomajian, C., Atwell, S.,…Bergelson, J. 2012. Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nature Genetics,* **44,** 212-216.

Huerta-Sanchez, E., DeGiorgio, M., Pagani, L., Tarekegn, A., Ekong, R.,…Nielsen, R. 2013. Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Molecular Biology and Evolution,* **30,** 1877-1888.

Igarashi, D., Tsuda, K. & Katagiri, F. 2012. The peptide growth factor, phytosulfokine, attenuates pattern-triggered immunity. *The Plant Journal,* **71,** 194-204.

Jofuku, K.D., Omidyar, P.K., Gee, Z. & Okamuro, J.K. 2005. Control of seed mass and seed yield by the floral homeotic gene APETALA2. *Proceedings of the National Academy of the Sciences USA*, **102**, 3117–3122.

Johnson, C.S., Kolevski, B. & Smyth, D.R. 2002. TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor. *Plant Cell*, **14**, 1359–1375.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* **24**, 1403-1405.

Joosen, R.V.L., Arends, D., Willems, L.A.J., Ligterink, W., Jansen, R.C. & Hilhorst, H.W.M. 2012. Visualizing the genetic landscape of *Arabidopsis* seed performance. *Plant Physiology,* **158,** 570-589.

Jordan, N. 1992. Path analysis of local adaptation in two ecotypes of the annual plant *Diodia teres* Walt. (Rubiaceae). *The American Naturalist,* **140,** 149-165.

Juenger, T.E. 2013. Natural variation and genetic constraints on drought tolerance. *Current Opinion in Plant Biology,* **16,** 274-281.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S., Freimer, N.B.,…Eskin, E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics,* **42,** 348-354.

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

Kang, J., Yu, H., Tian, C., Zhou, W., Li, C., Jiao, Y. & Liu, D. 2014. Suppression of photosynthetic gene expression in roots is required for sustained root growth under phosphate deficiency. *Plant Physiology,* **165,** 1156-1170.

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

Katori, T., Ikeda, A., Iuchi, S., Kobayashi, M., Shinozaki, K., Maehashi, K.,…Taji, T. 2010. Dissecting the genetic control of natural variation in salt tolerance of *Arabidopsis thaliana* accessions. *Journal of Experimental Botany,* **61,** 1125-1138.

Kesavan, M., Song, J.T. & Seo, H.S. 2012. Seed size: A priority trait in cereal crops. *Physiologia Plantarum,* **147,** 113-120.

Kim, S., Plagnol, V., Hu, T.T., Toomajian, C., Clark, R.M., Ossowski, S.,…Nordberg, M. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana. Nature Genetics,* **39**, 1151-1155.

Kniskern, J.M., Barrett, L.G. & Bergelson, J. 2011. Maladaptation in wild populations of the generalist plant pathogen *Pseudomonas syringae. Evolution,* **65,** 818-830.

Kohler, C., Page, D.R., Gagliardini, V. & Grossniklaus, U. 2005. The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nature Genetics*, **37**, 28–30.

Komori, R., Amano, Y., Ogawa-Ohnishi, M. & Matsubayashi, Y. 2009. Identification of tyrosylprotein sulfotransferase in *Arabidopsis. Proceedings of the National Academy of Sciences USA,* **106,** 15067-15072.

Koornneef, M., Alonso-Blanco, C. & Vreugdenhil, D. 2004. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology,* **55,** 141-72.

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

Laubinger, S., Fittinghoff, K. & Hoecker, U. 2004. The SPA quartet: A family of WD-repeat proteins with a central role in suppression of photomorphogenesis in *Arabidopsis. The Plant Cell,* **16,** 2293-2306.

Laubinger, S. & Hoecker, U. 2003. The SPA1-like proteins SPA3 and SPA4 repress photomorphogenesis in the light. *The Plant Journal,* **35,** 373-385.

Leimu, R. & Fischer, M. 2008. A meta-analysis of local adaptation in plants. *PLoS ONE,* **3,** 1-8.

Li, Y.H., Zheng, L.Y., Corke, F., Smith, C. & Bevan, M.W. 2008. Control of final seed and organ size by the DA1 gene family in *Arabidopsis thaliana*. *Genes and Development,* **22**,1331 – 1336.

Loudet, O., Chaillou, S., Camilleri, C., Bouchez, D. & Daniel-Vedele, F. 2002. Bay-0 x Shahdara recombinant inbred line population: A powerful tool for the genetic dissection of complex traits in *Arabidopsis. Theoretical and Applied Genetics,* **104,** 1173-1184.

Lowry, D.B., Hall, M.C., Salt, D.E. & Willis, J.H. 2009. Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus. New Phytologist,* **183,** 776-88.

Lowry, D.B., Hernandez, K., Taylor, S.H., Meyer, E., Logan, T.L.,…Juenger, T.E. 2014. The genetics of divergence and reproductive isolation between ecotypes of *Panicum hallii. New Phytologist,* **205,** 402-414.

Lowry, D.B., Rockwood, R.C. & Willis, J.H. 2008. Ecological reproductive isolation of coast and inland races of *Mimulus guttatus. Evolution,* **62,** 2196-2214.

Lowry, D.B. & Willis, J.H. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology,* **8,** 1-11.

Luo, M., Dennis, E., Berger, F., Peacock, W.J. & Chaudhury, A. 2005. MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*. *Proceedings of the National Academy of the Sciences USA*, **102**, 17531–17536.

Macnair, M.R. & Watkins, A.D. 1983. The fitness of the copper tolerance gene of *Mimulus guttatus* in uncontaminated soil. *New Phytologist,* **95,** 133-137.

Maloof, J.N., Borevitz, J.O., Dabi, T., Lutes, J., Nehring, R.B.,…Chory, J. 2001. Natural variation in light sensitivity of *Arabidopsis. Nature Genetics,* **29,** 441-446.

Manzan-Piedras, E., Marcer, A., Alonso-Blanco, C. & Pico, F.X. 2014. Deciphering the adjustment between environment and life history in annuals: Lessons from a geographically explicit approach in *Arabidopsis thaliana. PLoS ONE,* **9,** e87836.

Matsuzaki, Y., Ogawa-Ohnishi, M., Mori, A. & Matsubayashi, Y. 2010. Secreted peptide signals required for maintenance of root stem cell niche in *Arabidopsis. Science,* **329,** 1065-1067.

Maun, M.A. 1994. Adaptations enhancing survival and establishment of seedlings on coastal dune systems. *Vegetatio,* **111,** 59-70.

Maun, M.A. 2009. *The biology of coastal sand dunes.* Oxford University Press, Oxford.

Maun, M.A. & LaPierre, J. 1986. Effects of burial by sand on seed germination and seedling emergence of four dune species. *American Journal of Botany,* **73,** 450-455.

Maun, M.A. & Riach, S. 1981. Morphology of caryopses, seedlings and seedling emergence of the grass *Calamovilfa longifolia* from various depths in sand. *Oecologia,* **49,** 137-142.

Maun, M.A. & Perumal, J. 1999. Zonation of vegetation on lacustrine coastal dunes: effects of burial by sand. *Ecology Letters,* **2,** 14-18.

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

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

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

McIntyre, P.J. & Strauss, S.Y. 2013. Phenotypic and transgenerational plasticity promote local adaptation to sun and shade environments. *Evolutionary Ecology,* **28,** 229-246.

McKay, J.K., Richards, J.H. & Mitchell-Olds, T. 2003. Genetics of drought adaptation in *Arabidopsis thaliana:* I. Pleiotropy contributes to genetic correlations among ecological traits. *Molecular Ecology,* **12,** 1137-51.

McLachlan, A. & Brown, A. 2006. *The ecology of sandy shores.* 2nd ed. Academic Press, Burlington.

Merila, J. & Hendry, A.P. 2013. Climate change, adapatation, and phenotypic plasticity: The problem and the evidence. *Evolutionary Applications,* **7,** 1-14.

Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S. & Thomas, P.D. 2012. PANTHER version 7: Improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Research,* **38,** D204-D210.

Mi, H., Muruganujan, A., Casagrande, J.T. & Thomas, P.D. 2013. Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols,* **8,** 1551-1566.

Mitchell-Olds, T. 2001. *Arabidopsis thaliana* and its wild relatives: A model system for ecology and evolution. *TRENDS in Ecology & Evolution,* **16,** 693-700.

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

Moore, C.R., Gronwall, D.S., Miller, N.D. & Spalding, E.P. 2012. Mapping quantitative trait loci affecting *Arabidopsis thaliana* seed morphology features extracted computationally from images. *G3: Genes Genomes Genetics,* **3,** 109-118.

Morrison, L.W. 2002. The geographic distribution of pubescence in the sea daisy, *Borrichia aborescens,* on Bahamian Islands. *Global Ecology & Biogeography,* **11,** 247-252.

Mosher, S., Sybold, H., Rodriguez, P., Stahl, M., Davies, K.A.,…Kemmerling, B. 2013. The tyroside-sulfated peptide receptors PSKR1 and PSY1R modify the immunity of *Arabidopsis* to biotrophic and necrotropic pathogens in an antagonistic manner. *The Plant Journal,* **73,** 469-482.

Munns, R. 2011. Plant adaptations to salt and water stress: Differences and commonalities. *Advances in Botanical Research,* **57,** 1-32.

Nadeau, N.J., Ruiz, M., Salazar, P., Counterman, B., Medina, J.A.,…Papa, R. 2014. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato. Genome Research,* **24,** 1316-1333.

Nagy, E.S. & Rice, K.J. 1997. Local adaptation in two subspecies of an annual plant: Implications for migration and gene flow. *Evolution,* **51,** 1079-1089.

Nei, M. & Li, W-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Acadamy of the Sciences USA,* **76,** 5269-5273.

Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C,…Bergelson, J. 2005. The pattern of polymorphism in *Arabidopsis thaliana. PLoS Biology*, **3**, e196.

Ohto, M., Fischer, R.L., Goldberg, R.B., Nakamura, K. & Harada, J.J. 2005. Control of seed mass by APETALA2. *Proceedings of the National Academy of the Sciences USA*, **102**, 3123–3128.

Paccard, A., Fruleux, A. & Willi, Y. 2013. Latitudinal trait variation and responses to drought in *Arabidopsis lyrata. Oecologia,* **175,** 577-587.

Pannell, J.R. & Fields, P.D. 2013. Evolution in subdivided plant populations: Concepts, recent advances and future directions. *New Phytologist,* **201,** 417-432.

Paradis, E. 2010. Pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics,* **26,** 419-420.

Pardo-Diaz, C. Salazar, C. & Jiggens, C.D. 2014. Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution,* **6,** 445-464.

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)-12 x Ei-2. *PLoS ONE,* **6,** e578.

Pierik, R. & de Wit, M. 2013. Shade avoidance: Phytocrhome signalling and other aboveground neighbor detection cues. *Journal of Experimental Botany,* **11,** 2815-2824.

Pigliucci, M. & Schmitt, J. 1999. Genes affecting phenotypic plasticity in *Arabiodopsis:* pleiotropic effects and reproductive fitness of photomorphogenic mutants. *Journal of Evolutionary Biology,* **12,** 551-562.

Pigliucci, M. & Schmitt, J. 2004. Phenotypic plasticity in response to foliar and neutral shade in geibberellin mutants of *Arabidopsis thaliana. Evolutionary Ecology Research,* **6,** 243-259.

Platt, A., Horton, M., Huang, Y.S., Li, Y., Anastasio, A.E., Mulyati, N.W.,…Borevitz, J. 2010. The scale of population structure in *Arabidopsis thaliana. PLoS Genetics,* **6,** 1-8.

Quesada, V., García-Martínez, S., Piqueras, P., Ponce, M.R. & Micol, J.L. 2002. Genetic architecture of NaCl tolerance in *Arabidopsis. Plant Physiology,* **130,** 951-63.

Rabbani, M.A., Maruyama, K., Abe, H., Khan, M.A., Katsura, K., Ito, Y.,…Yamaguchi-Shinozaki, K. 2003. Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant Physiology,* **133,** 1755-1767.

Rajakaruna, N. 2004. The edaphic factor in the origin of plant species. *International Geology Review,* **46,** 471-8.

Rellstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M., & Holderegger, R. 2015. A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology,* in press.

Ren, B., Liang, Y., Deng, Y., Chen, Q., Zhang, J.,…Zua, J. 2009. Genome-wide comparative analysis of type-A *Arabidopsis response regulator* genes by overexpression studies reveals their diverse roles and regulatory mechanisms in cytokinin signalling. *Cell Research,* **19,** 1178-1190.

Reynolds, J., Weir, B.S. & Cockerham, C.C. 1983. Estiation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics,* **105,** 767-779.

Richardson, J.L., Urban, M.C., Bolnick, D.I. & Skelly, D.K. 2014. Microgeographic adptation and the spatial scale of evolution. *TRENDS in Ecology & Evolution,* **29,** 165-176.

Riefler, M., Novak, O., Strnad, M. & Schmulling, T. 2006. *Arabidopsis* cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin metabolism. *Plant Cell,* **18**, 40–54.

Rolauffs, S., Fackendahl, P., Sahm, J., Fiene, G. & Hoecker, U. 2012. *Arabidopsis COP1* and *SPA* genes are essential for plant elongation but not for acceleration of flowering time in response to a low red light to a far-red light ratio. *Plant Physiology,* **160,** 2015-2027.

Rus, A., Baxter, I., Muthukumar, B., Gustin, J., Lahner, B., Yakubova, E. & Salt, D.E. 2006. Natural variants of At*HKT1* enhance Na^+ accumulation in two wild populations of *Arabidsopsis. PLoS Genetics,* **2,** 1964-73.

Schmitt, J., McCormac, A.C. & Smith, H. 1995. A test of the adaptive plasticity hypothesis using transgenic and mutant plants disabled in phytochrom-mediated elongation responses to neighbors. *The American Naturalist,* **146,** 937-953.

Schmitt, J., Stinchcombe, J.R., Heschel, M.S. & Huber, H. 2003. The adaptive evolution of plasticity: Phytochrome-mediated shade avoidance responses. *Integrative Computational Biology,* **43,** 459-469.

Schruff, M.C., Spielman, M., Tiwari, S., Adams, S., Fenby, N. & Scott, R.J. 2005. The AUXIN RESPONSE FACTOR 2 gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. *Development,* **133**, 251 – 261.

Sessa, G., Carabelli, M., Sassi, M., Ciolfi, A., Possenti, M., Mittempergher, F.,…Ruberti, I. 2005. A dynamic balance between gene activation and repression regulates the shade avoidance response in *Arabidopsis. Genes & Development,* **19,** 2811-2815.

Sharma, G.K. 1984. Leaf surface effects of environmental pollution on sugar maple (*Acer saccharum*) in Montreal. *Montreal Candadian Journal of Botany,* **53,** 2312-2314.

Sherrard, M.E. & Maherali, H. 2012. Local adaptation across a fertility gradient is influcenced by soil biota in the invasive grass, *Bromus inermis. Evolutionary Ecology,* **26,** 529-544.

Shindo, C., Bernasconi, G. & Hardtke, C.S. 2007. Natural genetic variation in *Arabidopsis:* Tools, traits and prospects for evolutionary ecology. *Annals of Botany,* **99,** 1043-54.

Simon, M., Loudet, O., Durand, S., Berard, A., Brunel, D.,…Camilleri, C. 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,** 2253-2264.

Song, Y.H., Yoo, C.M., Hong, A.P., Kim, S.H., Jeong, H.J.,…Hong, J.C. 2008. DNA-binding study identifies C-box and hybrid C/G-box or C/A-box motifs as high-affinity binding sites for STF1 and LONG HYPOCOTYL5 proteins. *Plant Physiology,* **146,** 1862-1877.

Stearns, F.W. & Fenster, C.B. 2013. Evidence for parallel adaptation to climate across the natural range of *Arabidopsis thaliana. Ecology and Evolution,* **3,** 2241-2250.

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,** 1098-1108.

Stephens, M. 2013. A unified framework for association analysis with multiple related phenotypes. *PLoS ONE*, **8**, e65245.

Stinchcombe, J.R. & Hoekstra, H.E. 2008. Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity,* **100,** 158-70.

Stuhrwohldt, N., Dahkle, R.I., Kutschmar, A., Peng, X., Sun, M-X. & Sauter, M. 2015. Phytosulfokine peptide signaling controls pollen tube growth and funicular pollen tube guidance in *Arabidopsis thaliana. Physiologia Plantarum,* **153,** 643-653.

Sturwohldt, N., Dahlke, R.I., Steffens, B., Johnson, A. & Sauter, M. 2011. Phytosulfokine-α controls hypocotyl length and cell expansion in *Arabidopsis thaliana* through phytosulfokine receptor 1. *PLoS ONE,* **6,** e21054.

Sykes, M.T. & Wilson, J.B. 1990. An experimental investigation in to the response of New Zealand sand dune species to different depths of burial by sand. *Acta Botanica Neerlandica,* **39,** 171-181.

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

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

Telfer, A., Bollman, K.M. & Poethig, R.S. 1997. Phase change and the regulation of trichome distribution in *Arabidopsis thaliana. Development,* **124,** 645-654.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Kariak, B.,…Narechania, A. 2003. PANTHER: A library of protein families and subfamilies indexed by function. *Genome Research,* **13,** 2129-2141.

Tiffen, P. & Ross-Iberra, J. 2014. Advances and limits of using population genetics to understand local adaptation. *TRENDS in Ecology and Evolution,* **29,** 673-680.

Trontin, C., Tisné, S., Bach, L. & Loudet, O. 2011. What does *Arabidopsis* natural variation teach us (and does not teach us) about adaptation in plants? *Current Opinion in Plant Biology,* **14,** 225-231.

Turesson, G. 1922. The species and the variety as ecological units. *Hereditas,* **3,** 100-113.

Turesson, G. 1925. The plant species in relation to habitat and climate: Contributions to the knowledge of genecological units. *Hereditas,* **6,** 147-236.

Ungerer M.C., Johnson, L.C. & Herman, M.A. 2008. Ecological genomics: Understanding gene and genome funciton in the natural environment. *Heredity,* **100,** 178-183.

United Nations, Department of Economic and Social Affairs, Population Division (2011). World Population Prospects: The 2010 Revision.

Van Daele I., Gonzalez, N., Vercauteren, I., de Smet, L, Inze, D.,…Vuylsteke, M. 2012. A comparative study of seed yield parameters in *Arabidopsis thaliana* mutants and transgenics. *Plant Biotechnology Journal,* **10,** 488-500.

Van der Valk, A.G. 1974. Environmental factors controlling the distribution of forbs on coastal foredunes in Cape Hatteras National Seashore. *Canadian Journal of Botany,* **52,** 1057-1073.

Vandenbussche, F., Verbelen, J-P. & Van Der Straeten, J. 2005. Of light and length: regulation of hypocotyl growth in *Arabidopsis. BioEssays* **27,** 275-284.

Verhoeven, K.J.F., Poorter, H., Nevo, E. & Biere, A. 2004. Habitat-specific natural selection at a flowering-time QTL is a main driver of local adaptation in two wild barley populations. *Molecular Ecology,* **17,** 3416-3424.

Verhoeven, K.J.F., Poorter, H., Nevo, E. & Biere, A. 2008. Habitat-specific natural selection at a flowering-time QTL is a main driver of local adaptation in two wild barley populations. *Molecular Ecology,* **17,** 3416-24.

Verslues, P.E. & Juenger, T.E. 2011. Drought, metabolites, and *Arabidopsis* natural variation: A promising combination for understanding adaptation to water-limited environments. *Current Opinion in Plant Biology,* **14,** 1-6.

de Villemereuil, P. & Gaggiotti, O.E. 2015. A new F_{ST}-based method to uncover local adaptation using environmental variables. *Methods in Eoclogy and Evolution,* in press.

Voronkova, N.M., Burkovskaya, E.V., Bezdeleva, T.A. & Burundukova, O.L. 2008. Morphological and biological features of plants related to their adaptation to coastal habitats. *Russian Journal of Ecology,* **39,** 1-7.

Wang, A., Garcia, D., Zhang, H., Feng, K., Chaudhury, A.,…Luo, M. 2010. The VQ motif protein IKU1 regulates endosperm growth and seed size in *Arabidopsis*. *Plant Journal*, **63**, 670–679.

Wang, H. & Wang, H. 2015. Phytochrome signaling: Time to tighten up the loose sends. *Molecular Plant,* **8,** 540-551.

Wang, X., Roig-Villanova, I., Khan, S., Shanahan, H., Quail, P., Martinez-Garcia, J.F. & Devlin, P.F. 2011. A novel high-throughput *in vivo* molecular screen for shade avoidance mutants identifies a novel phyA mutation. *Journal of Experimental Biology,* **162,** 2973-2987.

Weinig, C., Dorn, L.A., Kane, N.C., German, Z.M., Halldorsdottir, S.S.,…Schmitt, J. 2003. Heterogenous selection at specific loci in natural environments in *Arabidopsis thaliana. Genetics,* **165,** 32-329.

Weinig, C., Ewers, B.E. & Welch, S.M. 2014. Ecological genomics and process modeling of local adaptation to climate. *Current Opinion in Plant Biology,* **18,** 66-72.

Weis, A.E., Turner, K.M., Petro, B., Austen, E.J. & Wadgymar, S.M. 2015. Hard and soft selection on phenology through seasonal shifts in the general and social environments: A study on plant emergence time. *Evolution,* **69,** 1361-1374.

Werner, T., Motyka, V., Laucou, V., Smets, R., Van Onckelen, H. & Schmulling, T. 2003. Cytokinin-deficient transgenic *Arabidopsis* plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *Plant Cell,* **15**, 2532–2550.

Wilson, J.B. & Sykes, M.T. 1999. Is zonation on coastal sand dunes determined primarily by sand burial or by salt spray? A test in New Zealand dunes. *Ecology Letters,* **2,** 233-236.

Wu, T., Kamiya, T., Yumoto, H., Sotta, N., Katsushi, Y.,…Fujiwara, T. 2015. An *Arabidopsis thaliana* copper-sensitive mutatnt suggests a role of phytosulfokine in ethylene production. *Journal of Experimental Botany,* **66,** 3657-3667.

Xiao, W., Brown, R.C., Lemmon, B.E., Harada, J.J., Goldberg, R.B. & Fischer, R.L. 2006. Regulation of seed size by hypomethylation of maternal and paternal genomes. *Plant Physiology*, **142**, 1160–1168.

Xing, Y. & Zhang, Q. 2010. Genetic and molecular bases of rice yield. *Annual Review of Plant Biology*, **61**, 421–442.

Yadav, R.S., Sehgal, D. & Vadez, V. 2011. Using genetic mapping and genomics approaches in understanding and improving drought tolerance in pearl millet. *Journal of Experimental Botany,* **62,** 397-408.

Yanful, M. & Maun, M.A. 1996. Effects of burial of seeds and seedlings from different seed sizes on the emergence and growth of *Strophostyles helvola. Canadian Journal of Botany,* **74,** 1322-1330.

Yi, X., Liang, Y, Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P.,…Wang, J. 2010. Sequncing of 50 human exomes reveals adaptatin to high altitude. *Science,* **329,** 75-78.

Yoder, J.B., Standon-Geddes, J., Zhou, P., Briskine, R.,Young, N.D. & Tiffen, P. 2015. Genomic signature of adaptation to climate in *Medicago truncatula. Genetics,* **196,** 1263-1275.

Zhan, S., Zhang, W., Niitepold, K., Hsu, J., Haeger, J.F.,…Kronforst, M.R. 2014. The genetics of monarch butterfly migration and warning colouration. *Nature,* **514,** 317-321.

Zhang, J. 1996. Seed mass effects across environments in an annual dune plant. *Annals of Botany,* **77,** 555-563.

Zhang, J. & Maun, M.A. 1990. Effects of sand burial on seed germination, seedling emergence, survival, and growth of *Agropyron psammophilum. Canadian Journal of Botany,* **68,** 304-310.

Zhou, W., Wei, L., Xu, J., Xhai, Q., Jiang, H.,…Li, C. 2010. *Arabidopsis* tyrosylprotein sulfotransferase acts in the Auxin/PLETHORA pathway in regulating postembryonic maintenance of the root stem cell niche. *The Plant Cell,* **22,** 3692-3709.

Zhou, X. & Stephens, M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, **44**, 821-824.

Zhou, X. & Stephens, M. 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods*, **11**, 407-409.

APPENDIX A: CANDIDATE GENE LISTS

Table A.1 Hypocotyl candidate genes from the literature.

chr	locus	start	end	name	function	reference
4	AT4G26200	13275307	13277188	ACS7	ethylene biosynthesis	Filiault & Maloof 2012
				ACT	actin genes	Vandenbussche et al. 2005
1	AT1G48410	17886098	17892586	AGO1	RNA slicer regulating auxin and light signalling	Filiault & Maloof 2012
3	AT3G05800	1727306	1728296	AIF1	BR signalling	Filiault & Maloof 2012
2	AT2G46530	6628068	6633087	ARF11/ IAA22	TCF responding to auxin	Filiault & Maloof 2012
3	AT3G59900	22129726	22130464	ARGOS	auxin regulated; involved in organ size control	Filiault & Maloof 2012
				ARR4	interact with phytochromes	Wang & Wang 2015
5	AT5G43700	17550233	17551581	ATAUX-11/IAA4	auxin inducible, similar to TCF	Filiault & Maloof 2012
4	AT4G16780	9449114	9450743	ATHB2	TCF in phytochrome B signalling	Filiault & Maloof 2012
3	AT3G01220	73481	75538	ATHB20	regulates transcription during germination/early devlopment	Filiault & Maloof 2012
				ATHB23	interact with phytochromes	Wang & Wang 2015
5	AT5G40820	16342745	16353898	ATR	response to DNA damage; cell cycle regulation	Vandenbussche et al. 2005
2	AT2G32410	13757593	13761077	AXL	auxin signalling	Filiault & Maloof 2012
1	AT1G05180	1498114	1501824	AXR1	regulates auxin response	Vandenbussche et al. 2005
1	AT1G19350	6688535	6690411	BES1	BR signalling	Filiault & Maloof 2012
3	AT3G23210	8283090	8285081	bHLH34	TCF, response to UV-B	Filiault & Maloof 2012
4	AT4G39400	18324661	18328826	BRI1	receptor kinase involved in BR signalling	Vandenbussche et al. 2005
3	AT3G19820	6879617	6881652	CBB1/ DIM1/ DWF1	BR biosynthesis; dwarf mutants	Vandenbussche et al. 2005

Table A.1 continued

chr	locus	start	end	name	function	reference
2	AT2G46830	19245672	19248909	CCA	TCF in circadian rhythms	Vandenbussche et al. 2005; Wang & Wang 2015
				CES	cellulose synthetase genes	Vandenbussche et al. 2005
4	AT4G26150	13253210	13254904	CGA1	TCF, responsive to cytokinins and light stimulus	Filiault & Maloof 2012
2	AT2G32950	13977933	13983535	COP1	represses photomorphogenesis/induces skotomorphogenesis in dark	Vandenbussche et al. 2005; Wang & Wang 2015
4	AT4G14110	8132886	8134920	COP9	represses photomorphogenesis/induces skotomorphogenesis in dark	Vandenbussche et al. 2005; Wang & Wang 2015
5	AT5G05690	1702688	1706780	CPD/ CBB3/ DWF3	cytochrome, mutant is constitutively photomorphogenic dwarf	Vandenbussche et al. 2005
4	AT4G08920	5724103	5727253	CRY1	blue light receptor	Vandenbussche et al. 2005
1	AT1G04400	1185556	1188276	CRY2	blue light receptor	Vandenbussche et al. 2005
5	AT5G24850	8535282	8538107	CRY3	probably a photoreceptor	Vandenbussche et al. 2005
				DDB1	complexes with COP and DET	Wang & Wang 2015
4	AT4G10180	6346463	6349276	DET1	repressor of photomorphogenesis	Vandenbussche et al. 2005; Wang & Wang 2015
5	AT5G54510	22131093	22133678	DFL1	hypocotyl growth in light	Vandenbussche et al. 2005
4	AT4G03400	1497536	1499865	DFL2	red light specific hypocotyl length elongation	Vandenbussche et al. 2005
3	AT3G02580	546903	548795	DWF7	brassinosteroid biosynthesis	Vandenbussche et al. 2005
				ELF	circadian rhythm genes	Vandenbussche et al. 2005; Wang & Wang 2015
				FAR1	regulate FHY1/FHL transcription	Wang & Wang 2015

Table A.1 continued

chr	locus	start	end	name	function	reference
				FHL	chaperone protein for nuclear translocation of phyA	Wang & Wang 2015
				FHY1	chaperone protein for nuclear translocation of phyA	Wang & Wang 2015
				FHY3	regulate FHY1/FHL transcription	Wang & Wang 2015
2	AT2G46370	19033741	19036445	FIN219	FR insensitive; jasmonic acid signalling pathway	Vandenbussche et al. 2005
4	AT4G19990	10832303	10835698	FRS1	response to R/FR light	Filiault & Maloof 2012
2	AT2G32250	13693105	13696932	FRS2	response to R/FR light	Filiault & Maloof 2012
2	AT2G26990	11519507	11522587	FUS12/ COP12	represses photomorphogenesis/induces skotomorphogenesis in dark	Filiault & Maloof 2012; Wang & Wang 2015
4	AT4G02780	1237767	1244813	GA1	GA synthesis	Vandenbussche et al. 2005
4	AT4G25420	12990884	12992458	GA20ox 1	GA biosynthesis	Filiault & Maloof 2012
5	AT5G51810	21055188	21056808	GA20ox 2	GA biosynthesis	Filiault & Maloof 2012
1	AT1G14920	5149226	5151354	GAI	DELLA; GA signalling	Vandenbussche et al. 2005
1	AT1G22770	8061844	8067716	GI	circadian rhythms	Vandenbussche et al. 2005; Wang & Wang 2015
				HFR1	promote photomorphogenesis in light	Wang & Wang 2015
3	AT3G09150	2803528	2805564	HY2	photosignalling	Vandenbussche et al. 2005
5	AT5G11260	3593380	3594992	HY5	photosignalling	Vandenbussche et al. 2005; Wang & Wang 2015
				HYH	promote photomorphogenesis in light	Wang & Wang 2015
				IAA	stabilize auxin response factor proteins	Vandenbussche et al. 2005; Wang & Wang 2015

Table A.1 continued

chr	locus	start	end	name	function	reference
3	AT3G15540	5264024	5265678	IAA19	auxin signalling	Filiault & Maloof 2012
2	AT2G32440	13775570	13778002	KAO1 (2)	GA biosynthesis	Filiault & Maloof 2012
5	AT5G49720	20197391	20200288	KOR1	cellulose biosynthesis, cell wall structuring	Vandenbussche et al. 2005
1	AT1G65610	24391679	24395360	KOR2	cellulose biosynthesis, cell wall structuring	Vandenbussche et al. 2005
4	AT4G24260	12577871	12580133	KOR3	cellulose biosynthesis, cell wall structuring	Vandenbussche et al. 2005
				LAF1	promote photomorphogenesis in light	Wang & Wang 2015
2	AT2G35300	14862901	14863387	LEA18	embryo development, response to low water availability	Filiault & Maloof 2012
1	AT1G01060	33666	37840	LHY	MYB TCF involved in circadian rhythms	Vandenbussche et al. 2005; Wang & Wang 2015
				LIP1	circadian rhythms	Wang & Wang 2015
1	AT1G80350	30205315	30208226	LUE	cytoskeleton	Vandenbussche et al. 2005
				LUX	circadian rhythms	Wang & Wang 2015
5	AT5G43900	17656894	17667451	MYA2	vesicle transport	Vandenbussche et al. 2005
2	AT2G32460	13782208	13784471	MYB101	GA signalling	Filiault & Maloof 2012
3	AT3G15570	5269849	5271922		response to light stimulus	Filiault & Maloof 2012
3	AT3G25880	9469713	9470033		response to auxin	Filiault & Maloof 2012
3	AT3G25890	9475575	9477455		TCF, response to ethylene	Filiault & Maloof 2012
3	AT3G59470	21978961	21980544		R/FR response	Filiault & Maloof 2012
5	AT5G01990	377231	379911		auxin transport	Filiault & Maloof 2012
1	AT1G09570	3095256	3100357	PHYA	R/FR light receptor	Vandenbussche et al. 2005; Wang & Wang 2015
2	AT2G18790	8139881	8144430	PHYB	R/FR light receptor	Vandenbussche et al. 2005; Wang & Wang 2015
5	AT5G35840	14007826	14011764	PHYC	R/FR light receptor	Vandenbussche et al. 2005; Wang & Wang 2015

Table A.1 continued

chr	locus	start	end	name	function	reference
4	AT4G16250	9195602	9199486	PHYD	R/FR light receptor	Vandenbussche et al. 2005; Wang & Wang 2015
4	AT4G18130	10042137	10046082	PHYE	R/FR light receptor	Vandenbussche et al. 2005; Wang & Wang 2015
				PIFs	phytochrome interacting factors	Wang & Wang 2015
				PIN1-7	auxin distribution genes	Vandenbussche et al. 2005
4	AT4G31500	15273471	15275310	RED1/SUR2	photosignalling	Vandenbussche et al. 2005
2	AT2G01570	255249	257550	RGA	DELLA, GA signalling	Vandenbussche et al. 2005
1	AT1G66350	24748195	24750043	RGL1	DELLA, GA signalling	Vandenbussche et al. 2005
3	AT3G03450	819337	821406	RGL2	DELLA integrating GA and light signalling	Filiault & Maloof 2012
3	AT3G12830	4012608	4013583	SAP5	positive regulator of stress response	Filiault & Maloof 2012
				SAX		Vandenbussche et al. 2005
1	AT1G69640	26193659	26195695	SBH1	sphingoid base hydroxylase	Filiault & Maloof 2012
1	AT1G04240	1128188	1129551	SHY2/IAA3	auxin signalling; photosignalling	Vandenbussche et al. 2005
5	AT5G04470	1266625	1267416	SIM	endoreduplication, brassinosteroid biosynthesis, trichome differentiation	Vandenbussche et al. 2005
2	AT2G46340	19022173	19027413	SPA1	suppressor of phyA, involved in repression of photomorphogenesis in darkness	Wang & Wang 2015
4	AT4G11110	6771605	6777225	SPA2	suppressor of phyA, involved in repression of photomorphogenesis in darkness	Wang & Wang 2015

Table A.1 continued

chr	locus	start	end	name	function	reference
3	AT3G15354	5169095	5172837	SPA3	suppressor of phyA, involved in repression of photomorphogenesis in darkness	Wang & Wang 2015
1	AT1G53090	19783542	19786902	SPA4	suppressor of phyA, involved in repression of photomorphogenesis in darkness	Wang & Wang 2015
2	AT2G03680	1120800	1122055	SPR1	microtubule associated; cytoskeleton	Vandenbussche et al. 2005
4	AT4G27060	13581401	13585155	SPR2	microtubule associated; cytoskeleton	Vandenbussche et al. 2005
4	AT4G08810	5616052	5617939	SUB1	cryptochrome and phytochrome coaction	Filiault & Maloof 2012
5	AT5G57560	23307130	23308282	TCH4	cell wall structuring	Vandenbussche et al. 2005
				TIC	circadian rhythms	Wang & Wang 2015
3	AT3G62980	23273116	23276375	TIR1	mediates auxin-regulated transcription	Vandenbussche et al. 2005
3	AT3G02260	430984	448489	TIR3	auxin transport	Vandenbussche et al. 2005
5	AT5G20730	7016470	7022113	TIR5/ IAA21/23/35	auxin regulated transcriptional activator; blue light response	Vandenbussche et al. 2005
				XCT	circadian rhythms	Wang & Wang 2015
				XTH	cell wall structuring genes	Vandenbussche et al. 2005
4	AT4G37390	17579657	17582022	YDK	auxin signalling pathways; list says blue and FR elongation but not TAIR	Vandenbussche et al. 2005

Table A.2 Candidate genes for trichome density from the literature.

chr	locus	start	end	name	function	reference
1	AT1G74950	28148575	28150444	JAZ2	ABA, jasmonic acid signalling pathways; jasmonic acid biosynthesis	Bloomer et al 2014
1	At1G77670	29188901	29190975	none	biosynthetic process	Bloomer et al 2014
1	At1G79840	30036956	30041440	GL2	epidermal cell identity	Bloomer et al 2014
2	At2G30420	12960827	12962162	ETC2	TCF, trichome formation	Bloomer et al 2014; Atwell et al. 2010; Hilscher et al. 2009
2	At2G30424	12964506	12965468	TCL2	MYB TCF, negatively regulates trichome formation	Bloomer et al 2014; Atwell et al. 2010
2	At2G30432	12968615	12970209	TCL1	MYB TCF, negatively regulates trichome formation	Bloomer et al 2014; Atwell et al. 2010
2	At2G31660	13464298	13471821	URM9/ SAD2	ABA signalling	Bloomer et al 2014
2	At2G37260	15644840	15647065	TTG2	cell fate determianation in seed integument and endosperm; seed coat dev	Bloomer et al 2014; Symonds et al. 2005; Larkin et al. 1996
3	AT3G27920	10361945	10363506	GL1	MYB TCP, initiates trichomes with JAZ & DELLA; GA signalling	Symonds et al. 2005
4	At4G09820	6182023	6186493	TT8	anthocyanin biosynthesis	Bloomer et al 2014
5	AT5G24520	8370587	8372847	TTG1	epidermal cell fate specification; trichomes and root hairs; axin responsive; anthocyanin accumulation	Symonds et al. 2005
5	AT5G41315	16529455	16532879	GL3	interacts with GL1, trichome dev	Symonds et al. 2005

Table A.3 Candidate genes for seed weight from the literature.

chr	locus	start	end	name	function	reference
1	At1g02580	544783	549202	MEA/ FIS1	Negative regulator of seed development	Johnson et al. (2002) Kohler et al. (2005) Xiao et al. (2006)
1	At1g13710	4702725	4704657	KLU/ KLUH	Increases silique width and regulates seed size	Adamski et al. (2009)
1	At1g19270	6662501	6666028	DA1	Ubiquitin receptor; increases seed size; ABA signaling	Li et al. (2008)
1	At1g27320	9487780	9492572	AHK3	Histadine kinase; increase embryo size; cytokinin metabolism	Riefler et al. (2006)
1	At1g55600	20774046	20776294	MINI3	Member of WRKY TCF group 1; endosperm growth	Luo et al. (2005)
1	At1g65330	24266481	24267320	PHE1	Type 1 MADS-box protein; endosperm dev	Kohler et al. 2005
2	At2g35230	14842293	14844434	HAIKU1 /IKU1	Contains a plant-specific VQ motif; involved in endosperm growth; seed size det.	Luo et al. (2005) Wang et al. (2010)
2	At2g37260	15644840	15647065	TTG2	Encodes protein similar to WRKY; seed length; proanthocynidin synthesis	Johnson et al. (2002) Dilkes et al. (2008)
2	At2g35670	14992565	14996756	FIS2	Negative regulator of seed development	Johnson et al. (2002) Kohler et al. (2005) Xiao et al. (2006)
2	At2g41510	17314532	17317003	CKX1	Increases seed size and weight; cytokinin metabolism	Werner et al. (2003) Xing & Zhang (2010)
2	At2g19500	8444204	8447355	CKX2	Increases seed size and weight; cytokinin metabolism	Werner et al. (2003) Xing & Zhang (2010)
2	At2g01830	362983	368082	AHK4	Histadine kinase; increase embryo size; cytokinin metabolism	Riefler et al. (2006)
3	At3g19700	6843662	6847275	HAIKU2 /IKU2	Encodes leucine rich repeat kinase; endosperm growth	Luo et al. (2005)
3	At3g20740	7248809	7252452	FIE/ FIS3	Regulation of endosperm dev.	Kohler et al. (2005) Xiao et al. (2006)
4	At4g02020	886600	891955	SWN	Endosperm dev	Kohler et al. (2005) Xiao et al. (2006)

Table A.3 continued

chr	locus	start	end	name	function	reference
4	At4g36920	17400847	17403332	AP2	Seed abortion, flower dev; neg reg of seed weight	Jofuku et al. 2005 Ohto et al. 2005
5	At5g49160	19932230	19938369	MET1	Seed size; paternal methylation	Xiao et al. (2006)
5	At5g58230	23556012	23558245	MSI1	Seed dev	Kohler et al. (2005) Xiao et al. (2006)
5	At5g04560	1309202	1318401	DME	Endosperm maternal allele specific hypomethylation; seed size	Gehring et al. (2006)
5	At5g62000	24910358	24915210	ARF2	Seed size and weight	Schruff et al. 2005
5	At5g35750	13911371	13916337	AHK2	Histadine kinase; increase embryo size; cytokinin metabolism	Riefler et al. (2006)

Table A.4 Candidate genes from hypocotyl length GWA (Chapter 1 criteria)

Chr	Locus	Coordinates (bp)	Name	Function
2	AT2G32540	13814686 – 13818289	CSLB04	Encodes a gene similar to cellulose synthase.
2	AT2G32550	13819379 – 13822839		Cell differentiation, Rcd1-like protein.
2	AT2G32560	13824794 - 13826986		F-box family protein.
2	AT2G32580	13827663 - 13829317		Unknown protein.
2	AT2G32590	13829420 - 13833209	EMB2795	Mutant has defective embryo.
2	AT2G32600	13833545 - 13835663		Hydroxyproline-rich glycoprotein family protein.
2	AT2G32610	13836234 - 13839513	CSLB01	Encodes a gene similar to cellulose synthase.
2	AT2G32620	13840744 - 13844324	CSLB02	Encodes a gene similar to cellulose synthase.
2	AT2G32630	13844834 - 13846708		Pentatricopeptide repeat (PPR-like) superfamily protein.
2	AT2G32640	13846919 - 13850909		Lycopene beta/epsilon cyclase protein.
2	AT2G32650	13851062 - 13852080		RmlC-like cupins superfamily protein.
2	AT2G32660	13853758 - 13856112	RLP22	Receptor like protein.
2	AT2G32670	13857877 - 13859774	VAMP725	Member of Synaptobrevi-like protein family.
4	AT4G11060	6754587 - 6756532	MTSSB	Mitochondrially targeted single-stranded DNA binding protein.
4	AT4G11070	6759303 - 6760794	WRKY41	Member of WRKY Transcription Factor; Group III.
4	AT4G11080	6760867 - 6763272	3XHMG-BOX1	Encodes a protein containing three copies of the HMG (high mobility group)-box domain; interacts with mitotic and meiotic chromosomes.
4	AT4G11090	6764537 – 6766268	TBL2	Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene family containing a plant-specific DUF231 (domain of unknown function) domain. Members of this family have been shown to be involved in the synthesis and deposition of secondary wall cellulose, presumably by influencing the esterification state of pectic polymers.
4	AT4G11100	6768614 - 6770184		Unknown protein.

Table A.4 continued

Chr	Locus	Coordinates (bp)	Name	Function
4	AT4G11110	6771605 - 6777225	SPA2	Encodes a member of the SPA (suppressor of phyA-105) protein family (SPA1-SPA4). SPA proteins contain an N-terminal serine/threonine kinase-like motif followed by a coiled-coil structure and a C-terminal WD-repeat domain. SPA proteins function redundantly in suppressing photomorphogenesis in dark- and light-grown seedlings. SPA2 primarily regulates seedling development in darkness and has little function in light-grown seedlings or adult plants.
4	AT4G11120	6777802 - 6780166		Translation elongation factor.
4	AT4G11130	6780522 - 6784440	RDR2; SMD1	Encodes RNA-dependent RNA polymerase, required for endogenous siRNA formation.
4	AT4G11140	6794776 - 6795793	CRF1	Encodes a member of the ERF (ethylene response factor) subfamily B-5 of ERF/AP2 transcription factor family. The protein contains one AP2 domain.
5	AT5G35945	14093675 - 14094277		Unknown protein.
5	AT5G35950	14104943 - 14106764		Mannose-binding lectin superfamily protein
5	AT5G35960	14108524 - 14110536		Protein kinase
5	AT5G35965	14113311 - 14117927		copia-like retrotransposon family
5	AT5G35970	14118876 - 14123340		P-loop containing nucleoside triphosphate hydrolases superfamily protein
5	AT5G35980	14127984 - 14136324	YAK1	protein serine/threonine kinase
5	AT5G41730	16684914 - 16687145		Protein kinase
5	AT5G41740	16688626 - 16693119		Disease resistance protein
5	AT5G41750	16693909 - 16698943		Disease resistance protein
5	AT5G41755	16700028 - 16705892		non-LTR retrotransposon family
5	AT5G41760	16706880 - 16709484		Nucleotide-sugar transporter family protein
5	AT5G41765	16716552 - 16717064		DNA-binding storekeeper protein-related transcriptional regulator
5	AT5G41770	16717996 - 16721222		crooked neck protein, putative / cell cycle protein
5	AT5G41780	16723075 - 16724833		myosin heavy chain-related

Table A.5 Candidate genes from trichome density GWA

Chr	Locus	Coordinates (bp)	Name	Function
2	AT2G30380	12948284 – 12950573		Unknown
2	AT2G30390	12951062 – 12954114	FC2	Ferrochelatase; speculated to operate in photosynthetic cytochromes.
2	AT2G30395	12954174 – 12955089	OFP17	Unknown; ovate protein family.
2	AT2G30400	12956561 – 12957554	OFP2	Unknown; ovate protein family.
2	AT2G30410	12959289 – 12960810	KIESEL, KIS, TFCA	Tubulin Folding Cofactor A
2	AT2G30420	12960827 – 12962162	ETC2	Negative regulation of trichome patterning; in a tandem repeat with AT2G30424 and AT2G30432.
2	AT2G30430	12968176 – 12968412		Unknown
2	AT2G30432		TCL1	A single-repeat MYB-type transcription factor that negatively regulates trichome formation by suppressing GL1 (GLABRA1); in a tandem repeat with AT2g30424 and AT2g30420.
2	AT2G30440	12972877 – 12975497	PLSP2B, TPP	Encodes a thylakoidal processing peptidase that removes signal sequences from proteins synthesized in the cytoplasm and transported into the thylakoid lumen.
2	AT2G30450	12975948 – 12976018		pre-tRNA-Gly (anticodon: GCC)
2	AT2G30460	12976146 – 12978868	UTX2	UXT2 is a member of the NST-KT subfamily of nucleotide/sugar transporters; localized to the golgi, functions as a UDP-Xyl transporter.
2	AT2G30470	12980507 – 12985043	HSI2, VAL1, VP1/ABI3-LIKE 1	An active repressor of the Spo minimal promoter (derived from a gene for sweet potato sporamin A1) through the EAR motif.
2	AT2G30480	12987876 – 12993403		Unknown

Table A.6 Candidate genes from seed weight GWA

Chr	Locus	Coordinates (bp)	Name	Function
2	AT2G34920	14728027 – 14731156	EDA18	Embryo sac development, pollen development.
2	AT2G34930	14737066 – 14739904		Disease resistance family protein.
2	AT2G34940	14740468 – 14743417	BP80-3;2, VSR3;2, VSR5	Protein targeting to vacuole.
2	AT2G34950			pre-tRNA-Thr (anticodon: AGT)
2	AT2G34960	14744037 – 14745961	CAT5	Encodes a member of the cationic amino acid transporter (CAT) subfamily of amino acid polyamine choline transporters. Mediates efficient uptake of Lys, Arg and Glu in yeast; localized to plasma membrane.
2	AT2G34970	14746307 – 14748781		Trimeric LpxA-like enzyme.
2	AT2G34980	14748967 – 14749878	SETH1	Encodes a putative phosphatidylinositol-glycan synthase subunit C gene involved in the first step of the glycosylphosphtidyl-inositol (GPI) biosynthetic pathway.
2	AT2G34985	14750064 – 14750134		tRNA-Thr (anticodon: CGT)
2	AT2G34990	14750260 – 14751168		RING/U-box superfamily protein.
2	AT2G35000	14751574 – 14753319	ATL9	E3 ligase-like protein induced by chitin oligomers.
2	AT2G35010	14754271 – 14756129	TO1	Thioredoxin O1.
2	AT2G35020	14756711 - 14760706	GLCNA.UT2, GLCNAC1PUT2	Encodes a protein that functions as an N-acetylglucosamine-1-phosphate uridylyltransferase, catalyzes formation of UDP-N-acetylglucosamine (UDP-GlcNAc), an essential precursor for glycolipid and glycoprotein synthesis also used for regulatory protein modification in signaling pathways. Also catalyzes the reverse reaction;can act on glucose-1-phosphate to produce UDP-glucose.
2	AT2G35030	14760869 – 14763077	COD1	Pentatricopeptide repeat (PPR) superfamily protein.
2	AT2G35035	14763382 – 14765045	URED	Encodes a urease accessory protein essential for activation of urease.
2	AT2G35040	14765160 – 14768288		AICARFT/IMPCHase bi enzyme family.
2	AT2G35050	14768724 - 14775035		Protein kinase superfamily protein with octicosapeptide/Phox/Bem1p domain.

Table A.7 Candidate genes for hypocotyl length GWA (Chapter 2 criteria)

Chr	Locus	Coordinates (bp)	Name	Function
2	AT2G32540	13814686 – 13818289	CSLB04	Encodes a gene similar to cellulose synthase.
2	AT2G32550	13819379 – 13822839		Cell differentiation, Rcd1-like protein.
2	AT2G32560	13824794 – 13826986		F-box family protein.
2	AT2G32580	13827663 – 13829317		Unknown protein.
2	AT2G32590	13829420 – 13833209	EMB2795	
2	AT2G32600	13833545 - 13835663		Hydroxyproline-rich glycoprotein family protein.
2	AT2G32610	13836234 – 13839513	CSLB01	Encodes a gene similar to cellulose synthase.
2	AT2G32620	13840744 – 13844324	CSLB02	Encodes a gene similar to cellulose synthase.
2	AT2G32630	13844834 – 13846708		Pentatricopeptide repeat (PPR-like) superfamily protein.
2	AT2G32640	13846919 – 13850909		Lycopene beta/epsilon cyclase protein.
2	AT2G32650	13851062 – 13852080		RmlC-like cupins superfamily protein.
2	AT2G32660	13853758 – 13856112	RLP22	Receptor like protein.
2	AT2G32670	13857877 – 13859774	VAMP72	Member of Synaptobrevi-like protein family.
4	AT4G11060	6754587 – 6756532	MTSSB	Mitochondrially targeted single-stranded DNA binding protein.
4	AT4G11070	6759303 – 6760794	WRKY41	Member of WRKY Transcription Factor; Group III.
4	AT4G11080	6760867 – 6763272	3XHMG-BOX1	Encodes a protein containing three copies of the HMG (high mobility group)-box domain; interacts with mitotic and meiotic chromosomes.
4	AT4G11090	6764537 - 6766268	TBL23	Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene family containing a plant-specific DUF231 (domain of unknown function) domain. Members of this family have been shown to be involved in the synthesis and deposition of secondary wall cellulose, presumably by influencing the esterification state of pectic polymers.
4	AT4G11100	6768614 – 6770184		Unknown protein.

Table A.7 continued

Chr	Locus	Coordinates (bp)	Name	Function
4	AT4G11110	6771605 – 6777225	SPA2	Encodes a member of the SPA (suppressor of phyA-105) protein family (SPA1-SPA4). SPA proteins contain an N-terminal serine/threonine kinase-like motif followed by a coiled-coil structure and a C-terminal WD-repeat domain. SPA proteins function redundantly in suppressing photomorphogenesis in dark- and light-grown seedlings. SPA2 primarily regulates seedling development in darkness and has little function in light-grown seedlings or adult plants.
4	AT4G11120	6777802 – 6780166		Translation elongation factor.
4	AT4G11130	6780522 – 6784440	RDR2; SMD1	Encodes RNA-dependent RNA polymerase, required for endogenous siRNA formation.
4	AT4G11140	6794776 – 6795793	CRF1	Encodes a member of the ERF (ethylene response factor) subfamily B-5 of ERF/AP2 transcription factor family; contains one AP2 domain.

APPENDIX B: ACCESSION LISTS

Table B.1 Accessions used for PBS scan.

Line name	Line ID	Population
Sim 1	9442	Coarse sand beach
Sim 2	9443	Coarse sand beach
Sim 5	9446	Coarse sand beach
TV-10	6258	Coarse sand beach
TV-22	6268	Coarse sand beach
TV-30	6276	Coarse sand beach
TV-38	6284	Coarse sand beach
TV-4	6252	Coarse sand beach
TV-7	6255	Coarse sand beach
Var2-1	7516	Coarse sand beach
Var2-6	7517	Coarse sand beach
Var A1	9476	Coarse sand beach
App1-12	5830	Inland southern Sweden
App1-14	5831	Inland southern Sweden
App1-16	5832	Inland southern Sweden
Ba4-1	8258	Inland southern Sweden
Ba5-1	8259	Inland southern Sweden
Boo2-1	8266	Inland southern Sweden
Boo2-3	5836	Inland southern Sweden
Bro1-6	8231	Inland southern Sweden
Brosarp-11-135	1061	Inland southern Sweden
Brosarp-21-140	1063	Inland southern Sweden
Brosarp-25-142	1064	Inland southern Sweden
Brosarp-51-157	1072	Inland southern Sweden
Brosarp-53-159	1073	Inland southern Sweden
Dra1-4	5865	Inland southern Sweden
Dra2-1	5867	Inland southern Sweden
Dra3-1	8283	Inland southern Sweden
Dra3-9	5870	Inland southern Sweden
Fja1-1	8422	Inland southern Sweden
Fja1-2	6019	Inland southern Sweden
Fja1-5	6020	Inland southern Sweden
Fja2-4	6021	Inland southern Sweden
Fja2-6	6022	Inland southern Sweden
Fly2-1	6023	Inland southern Sweden
Fly2-2	6024	Inland southern Sweden
FlyA 3	9380	Inland southern Sweden
Ham 1	9399	Inland southern Sweden
HolA1 1	9404	Inland southern Sweden
HolA1 2	9405	Inland southern Sweden

Table B.1 continued

Line name	Line ID	Population
HolA2 2	9407	Inland southern Sweden
Hov1-10	6035	Inland southern Sweden
Hov1-7	6034	Inland southern Sweden
Hov3-2	6036	Inland southern Sweden
Hov3-5	6038	Inland southern Sweden
Hov4-1	8306	Inland southern Sweden
Hovdala-2	6039	Inland southern Sweden
Hovdala-6	8307	Inland southern Sweden
Kal 1	9408	Inland southern Sweden
Kavlinge-1	8237	Inland southern Sweden
Kia 1	9409	Inland southern Sweden
Kni-1	6040	Inland southern Sweden
Kulturen-1	8240	Inland southern Sweden
Lan 1	9421	Inland southern Sweden
Lis-2	8222	Inland southern Sweden
Lis-3	6041	Inland southern Sweden
Lu-1	8334	Inland southern Sweden
Lund	8335	Inland southern Sweden
Or-1	6074	Inland southern Sweden
Rev-1	8369	Inland southern Sweden
Rev-2	6076	Inland southern Sweden
Rev-3	6077	Inland southern Sweden
San-2	8247	Inland southern Sweden
Sparta-1	6085	Inland southern Sweden
Stu-2	6087	Inland southern Sweden
T1000	6090	Inland southern Sweden
T1010	6091	Inland southern Sweden
T1020	6092	Inland southern Sweden
T1030	6093	Inland southern Sweden
T1040	6094	Inland southern Sweden
T1050	6095	Inland southern Sweden
T1060	6096	Inland southern Sweden
T1070	6097	Inland southern Sweden
T1080	6098	Inland southern Sweden
T1090	6099	Inland southern Sweden
T1110	6100	Inland southern Sweden
T1120	6101	Inland southern Sweden
T1130	6102	Inland southern Sweden
T1150	6103	Inland southern Sweden
T1160	6104	Inland southern Sweden
T460	6106	Inland southern Sweden
T470	6107	Inland southern Sweden

Table B.1 continued

Line name	Line ID	Population
T510	6109	Inland southern Sweden
T520	6110	Inland southern Sweden
T530	6111	Inland southern Sweden
T540	6112	Inland southern Sweden
T570	6114	Inland southern Sweden
T580	6115	Inland southern Sweden
T590	6116	Inland southern Sweden
T610	6118	Inland southern Sweden
T620	6119	Inland southern Sweden
T630	6120	Inland southern Sweden
T640	6121	Inland southern Sweden
T670	6122	Inland southern Sweden
T680	6123	Inland southern Sweden
T690	6124	Inland southern Sweden
T710	6125	Inland southern Sweden
T720	6126	Inland southern Sweden
T730	6127	Inland southern Sweden
T740	6128	Inland southern Sweden
T750	6129	Inland southern Sweden
T760	8225	Inland southern Sweden
T780	6131	Inland southern Sweden
T790	6132	Inland southern Sweden
T800	6133	Inland southern Sweden
T810	6134	Inland southern Sweden
T840	6136	Inland southern Sweden
T850	6137	Inland southern Sweden
T860	6138	Inland southern Sweden
T880	6140	Inland southern Sweden
T890	6141	Inland southern Sweden
T900	6142	Inland southern Sweden
T920	6144	Inland southern Sweden
T930	6145	Inland southern Sweden
T940	6146	Inland southern Sweden
T950	6147	Inland southern Sweden
T960	6148	Inland southern Sweden
T970	6149	Inland southern Sweden
T980	6150	Inland southern Sweden
T990	6151	Inland southern Sweden
TDr-1	6188	Inland southern Sweden
TDr-11	6197	Inland southern Sweden
TDr-13	6198	Inland southern Sweden
TDr-14	6199	Inland southern Sweden

Table B.1 continued

Line name	Line ID	Population
TDr-15	6200	Inland southern Sweden
TDr-16	6201	Inland southern Sweden
TDr-17	6202	Inland southern Sweden
TDr-18	6203	Inland southern Sweden
TDr-2	6189	Inland southern Sweden
TDr-22	6207	Inland southern Sweden
TDr-3	6190	Inland southern Sweden
TDr-4	6191	Inland southern Sweden
TDr-5	6192	Inland southern Sweden
TDr-7	6193	Inland southern Sweden
TDr-8	6194	Inland southern Sweden
TDr-9	6195	Inland southern Sweden
Tomegap-2	6242	Inland southern Sweden
Tottarp-2	6243	Inland southern Sweden
Ull1-1	8426	Inland southern Sweden
Ull1-10	6411	Inland southern Sweden
Ull2-13	8427	Inland southern Sweden
Ull2-3	6973	Inland southern Sweden
Ull2-5	6974	Inland southern Sweden
Ull2-7	6412	Inland southern Sweden
UllA 1	9471	Inland southern Sweden
UllA 2	9472	Inland southern Sweden
Vinslov	9057	Inland southern Sweden
ALL1-3	2	France
ALL1-6	5	France
ALL1-7	6	France
ALL1-8	7	France
ALL1-9	8	France
ALL2-4	12	France
CAM-16	23	France
CAM-25	32	France
CAM-42	45	France
CAM-45	48	France
CAM-48	51	France
CAM-56	60	France
CAM-58	62	France
CAM-61	66	France
CAM-64	69	France
CAM-7	74	France
CUR-10	79	France
CUR-2	80	France

Table B.1 continued

Line name	Line ID	Population
CUR-3	81	France
CUR-5	83	France
CUR-6	84	France
CUR-7	85	France
CUR-8	86	France
CUR-9	87	France
CYR	88	France
DAM1	89	France
LAC-3	94	France
LAC-5	96	France
LDV-14	104	France
LDV-16	106	France
LDV-25	116	France
LDV-3	121	France
LDV-30	122	France
LDV-31	123	France
LDV-34	126	France
LDV-53	146	France
LDV-57	148	France
LDV-58	149	France
LDV-6	151	France
LDV-68	153	France
LDV-8	157	France
MAR2-3	159	France
MIB-1	160	France
MIB-100	162	France
MIB-11	163	France
MIB-14	165	France
MIB-15	166	France
MIB-16	167	France
MIB-17	168	France
MIB-19	169	France
MIB-2	170	France
MIB-20	171	France
MIB-22	173	France
MIB-23	174	France
MIB-24	175	France
MIB-28	178	France
MIB-29	179	France
MIB-32	183	France
MIB-33	184	France

Table B.1 continued

Line name	Line ID	Population
MIB-34	185	France
MIB-35	186	France
MIB-36	187	France
MIB-37	188	France
MIB-39	190	France
MIB-40	191	France
MIB-43	194	France
MIB-47	196	France
MIB-50	198	France
MIB-54	200	France
MIB-55	201	France
MIB-57	202	France
MIB-58	203	France
MIB-60	204	France
MIB-61	205	France
MIB-62	206	France
MIB-63	207	France
MIB-64	208	France
MIB-67	210	France
MIB-69	212	France
MIB-70	213	France
MIB-73	214	France
MIB-75	215	France
MIB-76	216	France
MIB-77	217	France
MIB-80	219	France
MIB-83	222	France
MIB-84	223	France
MIB-86	224	France
MIB-87	225	France
MIB-89	227	France
MIB-9	228	France
MIB-90	229	France
MIB-92	230	France
MIB-93	231	France
MOG-11	236	France
MOG-12	237	France
MOG-37	242	France
MOG-40	244	France
MOG-55	252	France
PAR-10	257	France

Table B.1 continued

Line name	Line ID	Population
PAR-3	258	France
PAR-4	259	France
PAR-6	261	France
PAR-8	262	France
PAR-9	263	France
PON	264	France
ROM-1	267	France
ROM-2	268	France
ROM-9	269	France
TOU-A1-105	273	France
TOU-A1-109	275	France
TOU-A1-111	277	France
TOU-A1-112	278	France
TOU-A1-114	280	France
TOU-A1-115	281	France
TOU-A1-116	282	France
TOU-A1-117	283	France
TOU-A1-120	287	France
TOU-A1-122	288	France
TOU-A1-124	290	France
TOU-A1-125	291	France
TOU-A1-128	292	France
TOU-A1-129	293	France
TOU-A1-131	295	France
TOU-A1-133	296	France
TOU-A1-134	297	France
TOU-A1-137	298	France
TOU-A1-143	302	France
TOU-A1-18	306	France
TOU-A1-23	309	France
TOU-A1-26	310	France
TOU-A1-27	311	France
TOU-A1-33	314	France
TOU-A1-37	316	France
TOU-A1-39	318	France
TOU-A1-41	320	France
TOU-A1-43	321	France
TOU-A1-47	323	France
TOU-A1-60	326	France
TOU-A1-61	327	France
TOU-A1-62	328	France

Table B.1 continued

Line name	Line ID	Population
TOU-A1-63	329	France
TOU-A1-65	331	France
TOU-A1-66	332	France
TOU-A1-69	335	France
TOU-A1-73	338	France
TOU-A1-74	339	France
TOU-A1-75	340	France
TOU-A1-77	341	France
TOU-A1-79	343	France
TOU-A1-81	346	France
TOU-A1-82	347	France
TOU-A1-84	348	France
TOU-A1-85	349	France
TOU-A1-93	355	France
TOU-A1-96	357	France
TOU-A1-98	359	France
TOU-C-1	360	France
TOU-C-2	361	France
TOU-D-1	363	France
TOU-D-5	364	France
TOU-E-11	366	France
TOU-E-2	367	France
TOU-E-7	368	France
TOU-G-1	371	France
TOU-H-10	372	France
TOU-H-12	373	France
TOU-H-13	374	France
TOU-H-3	375	France
TOU-I-1	377	France
TOU-I-17	378	France
TOU-I-2	379	France
TOU-I-6	380	France
TOU-J-3	383	France
TOU-K-2	385	France
TOU-K-3	386	France
TOU-L-10	387	France
TOU-L-17	388	France
TOU-L-5	389	France
VOU-1	390	France
VOU-10	391	France
VOU-2	392	France

Table B.1 continued

Line name	Line ID	Population
VOU-4	393	France
VOU-6	395	France
VOU-7	396	France
VOU-8	397	France
Ag-0	6897	France
Lz-0	6936	France
Ra-0	6958	France
Ren-1	6959	France
Ren-11	6960	France
Ann-1	6994	France
Chat-1	7071	France
Cit-0	7075	France
Com-1	7092	France
Di-1	7098	France
Pn-0	7307	France
Nc-1	7430	France
PHW-35	7506	France
PHW-36	7507	France
PHW-37	7508	France
Gy-0	8214	France
Cen-0	8275	France
Lm-2	8329	France

Table B.2 Swedish mapping panel used in GWA

Line name	Line ID
Adal 1	9321
Adal 3	9323
Ale-Stenar-41-1	991
Ale-Stenar-44-4	992
Ale-Stenar-50-11	996
Ale-Stenar-56-14	997
Ale-Stenar-57-16	998
Ale-Stenar-59-18	999
Ale-Stenar-64-24	1002
Ale-Stenar-77-31	1006
Aledal-1-34	1153
Aledal-11-63	1163
Aledal-14-73	1166
Aledal-17-82	1169
Aledal-6-49	1158
Algutsrum	8230
Angso-57-419	1312
Angso-59-422	1313
Angso-74-430	1317
Angso-80-432	1318
App1-12	5830

Table B.2 continued

Line name	Line ID
App1-14	5831
App1-16	5832
Arby-1	6998
Ba4-1	8258
Ba5-1	8259
Bar 1	9332
Bil-3	5835
Bil-5	6900
Bil-7	6901
Bon 1	9336
Boo2-1	8266
Boo2-3	5836
Bot 1	9339
Bot 4	9342
Bro1-6	8231
Brosarp-11-135	1061
Brosarp-21-140	1063
Brosarp-25-142	1064
Brosarp-51-157	1072
Brosarp-53-159	1073
Dja 1	9343
Dja 2	9344
Djk 3	9349
Dod 1	9351
Dod 2	9352
Dod 3	9353
Dor-10	5856
Dra-3	5860
Dra1-4	5865
Dra2-1	5867
Dra3-1	8283
Dra3-9	5870
Eden 15	9354
Eden 16	9355
Eden 17	9356
Eden-1	6009
Eden-2	6913
Eden-4	8218
Eden-5	6010
Eden-6	6011

Table B.2 continued

Line name	Line ID
Eden-7	6012
Eden-9	6013
EdJ 2	9363
Eds-1	6016
Eds-9	6017
EkN 3	9367
EkS 2	9369
Fab-2	6917
FaL 1	9371
FaU 4	9378
Fja1-1	8422
Fja1-2	6019
Fja1-5	6020
Fja2-4	6021
Fja2-6	6022
Fly2-1	6023
Fly2-2	6024
FlyA 3	9380
Fri 1	9381
Fri 2	9382
Fri 3	9383
Fro 1	9384
Fro 3	9385
Gro-3	6025
Gron 12	9386
Gron 14	9388
Gron-5	6030
Had 1	9390
Had 2	9391
Had 3	9392
Ham 1	9399
Ham-10-239	1366
Ham-2-228	1360
Ham-27-256	1374
Ham-6-232	1362
Ham-7-233	1363
Hel 3	9402
Hen-16-268	1585
HolA1 1	9404
HolA1 2	9405

Table B.2 continued

Line name	Line ID
HolA2 2	9407
Hov1-10	6035
Hov1-7	6034
Hov2-1	8423
Hov3-2	6036
Hov3-5	6038
Hov4-1	8306
Hovdala-2	6039
Hovdala-6	8307
Kal 1	9408
Kavlinge-1	8237
Kia 1	9409
Kni-1	6040
Kor 1	9410
Kor 2	9411
Kor 3	9412
Kor 4	9413
Kru 3	9416
Kulturen-1	8240
Kva 2	9418
Lag 1	9419
Lan 1	9421
Liarum	8241
Lillo-1	8242
Lis-2	8222
Lis-3	6041
Lov-1	6043
Lov-5	6046
Lu-1	8334
Lund	8335
Nas 2	9427
Nyl 13	9433
Nyl-2	6064
Nyl-7	6069
Ode 2	9434
Omn-5	6071
OMo1-7	6073
Omo2-1	7518
Omo2-3	7519
Or-1	6074

Table B.2 continued

Line name	Line ID
Ost-0	8351
Puk 2	9437
Rev-1	8369
Rev-2	6076
Rev-3	6077
Rod-17-319	1435
San-2	8247
Sanna-2	8376
Sim 1	9442
Sim 2	9443
Sim 5	9446
Sku-30	1552
Sparta-1	6085
Spr1-6	6965
Spro 1	9450
Spro 2	9451
Spro 3	9452
Sr:3	6086
Stabby-13	1391
Stabby-26	1404
Ste 2	9453
Ste 3	9454
Stu-2	6087
T1000	6090
T1010	6091
T1020	6092
T1030	6093
T1040	6094
T1050	6095
T1060	6096
T1070	6097
T1080	6098
T1090	6099
T1110	6100
T1120	6101
T1130	6102
T1150	6103
T1160	6104
T460	6106
T470	6107

Table B.2 continued

Line name	Line ID
T510	6109
T520	6110
T530	6111
T540	6112
T570	6114
T580	6115
T590	6116
T610	6118
T620	6119
T630	6120
T640	6121
T670	6122
T680	6123
T690	6124
T710	6125
T720	6126
T730	6127
T740	6128
T750	6129
T760	8225
T780	6131
T790	6132
T800	6133
T810	6134
T840	6136
T850	6137
T860	6138
T880	6140
T890	6141
T900	6142
T920	6144
T930	6145
T940	6146
T950	6147
T960	6148
T970	6149
T980	6150
T990	6151
TAA 04	6154
TAA 14	6163

Table B.2 continued

Line name	Line ID
TAA 17	6166
TAD 01	6169
TAD 02	6170
TAD 03	6171
TAD 04	6172
TAD 05	6173
TAD 06	6174
TAL 03	6177
TAL 07	6180
TBO 01	6184
TDr-1	6188
TDr-11	6197
TDr-13	6198
TDr-14	6199
TDr-15	6200
TDr-16	6201
TDr-17	6202
TDr-18	6203
TDr-2	6189
TDr-22	6207
TDr-3	6190
TDr-4	6191
TDr-5	6192
TDr-7	6193
TDr-8	6194
TDr-9	6195
TEDEN 02	6209
TEDEN 03	6210
TFA 02	6212
TFA 04	6214
TFA 05	6215
TFA 06	6216
TFA 07	6217
TFA 08	6218
TGR 01	6220
TGR 02	6221
THO 03	8227
THO 08	6226
TNY 04	6231
TOM 01	6235

Table B.2 continued

Line name	Line ID
TOM 02	6236
TOM 03	6237
TOM 04	6238
TOM 06	6240
TOM 07	6241
Tomegap-2	6242
Tos-31-374	1247
Tos-82-387	1254
Tos-93-391	1256
Tos-95-393	1257
Tottarp-2	6243
TRA 01	6244
Tur 3	9469
Tur 4	9470
TV-10	6258
TV-22	6268
TV-30	6276
TV-38	6284
TV-4	6252
TV-7	6255
Ull1-1	8426
Ull1-10	6411
Ull2-13	8427
Ull2-3	6973
Ull2-5	6974
Ull2-7	6412
Ull3-4	6413
UllA 1	9471
UllA 2	9472
Var A1	9476
Var2-1	7516
Var2-6	7517
Vimmerby	8249
Vinslov	9057
VM-12	1026
Yst 1	9481
Yst 2	9482

Table B.3 Accessions used in initial hypocotyl length phenotyping experiment.

Line name	ID number	Latitude	Longitude	Home environment
Col-0	-	-	-	Lab
Dra 2-1	5867	55.76	14.12	Inland sand
FlyA3	9380	55.7488	13.3742	Inland sand
Fri 1	9381	55.81064	14.20914	Fine sand beach
HolA2 2	9407	55.7491	13.399	Inland soil
Hov 3-2	6036	56.1053	13.7132	Inland sand
Hovdala 2	6039	56.1053	13.7132	Inland sand
Kia 1	9409	56.05727	14.30201	Inland sand
Kni 1	6040	55.66	13.4	Inland soil
Kor 1	9410	57.2746	16.1494	Inland soil
Lom 1-1		56.09	13.9	Inland sand
Rev 1	8369	55.6942	13.4504	Inland soil
Sim 1	9442	55.5678	14.3398	Coarse sand beach
Sim 2	9443	55.5678	14.3398	Coarse sand beach
Spro 1	9450	57.2545	18.2109	Inland soil
Ste 2	9453	57.8009	18.5162	Inland soil
Tjo 1	-	55.6041	14.3023	Coarse sand beach
Tjo 10	-	55.6041	14.3023	Coarse sand beach
TV4	6252	55.5796	14.3336	Coarse sand beach
TV10	6258	55.5796	14.3336	Coarse sand beach
TV22	6268	55.5796	14.3336	Coarse sand beach
Ull 2-3	6973	56.0648	13.9707	Inland soil
Var A1	9476	55.58	14.334	Coarse sand beach
Var 2-1	7516	55.58	14.334	Coarse sand beach
Var 2-6	7517	55.58	14.334	Coarse sand beach
VM2	-	55.7	14.2	Fine sand beach
VM11	-	55.7	14.2	Fine sand beach
Yst 1	9481	55.42422	13.84837	Fine sand beach

Table B.4 Lines from the RegMap used to test for an effect of HYPO5 on hypocotyl length.

Line name	Line ID	Country of origin	Latitude	Longitude	Allele
CYR	88	France	47.4	0.683333	T
DAM1	89	France	48.7	1.98333	G
MAR2-3	159	France	47.35	3.93333	T
MIB-1	160	France	47.3833	5.31667	G
MIB-100	162	France	47.3833	5.31667	T
MIB-11	163	France	47.3833	5.31667	T
MIB-14	165	France	47.3833	5.31667	G
MIB-15	166	France	47.3833	5.31667	G
MIB-16	167	France	47.3833	5.31667	T
MIB-17	168	France	47.3833	5.31667	G
MIB-19	169	France	47.3833	5.31667	T
MIB-2	170	France	47.3833	5.31667	G
MIB-20	171	France	47.3833	5.31667	T
MIB-22	173	France	47.3833	5.31667	G
MIB-23	174	France	47.3833	5.31667	T
MIB-24	175	France	47.3833	5.31667	T
MIB-28	178	France	47.3833	5.31667	T
MIB-29	179	France	47.3833	5.31667	T
MIB-32	183	France	47.3833	5.31667	T
MIB-33	184	France	47.3833	5.31667	G
MIB-34	185	France	47.3833	5.31667	T
MIB-35	186	France	47.3833	5.31667	G
MIB-36	187	France	47.3833	5.31667	G
MIB-37	188	France	47.3833	5.31667	G
MIB-39	190	France	47.3833	5.31667	G
MIB-40	191	France	47.3833	5.31667	T
MIB-43	194	France	47.3833	5.31667	G
MIB-47	196	France	47.3833	5.31667	T
MIB-50	198	France	47.3833	5.31667	G
MIB-54	200	France	47.3833	5.31667	T
MIB-55	201	France	47.3833	5.31667	T
MIB-57	202	France	47.3833	5.31667	T
MIB-58	203	France	47.3833	5.31667	T
MIB-60	204	France	47.3833	5.31667	G
MIB-61	205	France	47.3833	5.31667	T
MIB-62	206	France	47.3833	5.31667	T
MIB-63	207	France	47.3833	5.31667	T
MIB-64	208	France	47.3833	5.31667	G
MIB-67	210	France	47.3833	5.31667	G
MIB-69	212	France	47.3833	5.31667	G
MIB-70	213	France	47.3833	5.31667	T

Table B.4 continued

Line name	Line ID	Country of origin	Latitude	Longitude	Allele
MIB-73	214	France	47.3833	5.31667	G
MIB-75	215	France	47.3833	5.31667	T
MIB-76	216	France	47.3833	5.31667	G
MIB-77	217	France	47.3833	5.31667	T
MIB-80	219	France	47.3833	5.31667	G
MIB-83	222	France	47.3833	5.31667	G
MIB-84	223	France	47.3833	5.31667	G
MIB-86	224	France	47.3833	5.31667	G
MIB-87	225	France	47.3833	5.31667	G
MIB-89	227	France	47.3833	5.31667	G
MIB-9	228	France	47.3833	5.31667	T
MIB-90	229	France	47.3833	5.31667	G
MIB-92	230	France	47.3833	5.31667	G
MIB-93	231	France	47.3833	5.31667	T
MOG-11	236	France	48.6667	-4.06667	G
MOG-12	237	France	48.6667	-4.06667	G
MOG-37	242	France	48.6667	-4.06667	G
MOG-40	244	France	48.6667	-4.06667	G
MOG-55	252	France	48.6667	-4.06667	G
PAR-10	257	France	46.65	-0.25	G
PAR-3	258	France	46.65	-0.25	G
PAR-4	259	France	46.65	-0.25	G
PAR-6	261	France	46.65	-0.25	G
PAR-8	262	France	46.65	-0.25	G
PAR-9	263	France	46.65	-0.25	G
PON	264	France	45.75	4.71667	G
	266	France	NA	NA	T
ROM-1	267	France	45.5333	4.85	G
ROM-2	268	France	45.5333	4.85	G
ROM-9	269	France	45.5333	4.85	G
TOU-A1-105	273	France	46.6667	4.11667	G
TOU-A1-109	275	France	46.6667	4.11667	T
TOU-A1-111	277	France	46.6667	4.11667	T
TOU-A1-112	278	France	46.6667	4.11667	G
TOU-A1-114	280	France	46.6667	4.11667	G
TOU-A1-117	283	France	46.6667	4.11667	T
TOU-A1-134	297	France	46.6667	4.11667	T
TOU-A1-27	311	France	46.6667	4.11667	T
TOU-A1-39	318	France	46.6667	4.11667	T
	322	France	NA	NA	T
TOU-A1-47	323	France	46.6667	4.11667	T

Table B.4 continued

Line name	Line ID	Country of origin	Latitude	Longitude	Allele
TOU-A1-61	327	France	46.6667	4.11667	T
TOU-A1-74	339	France	46.6667	4.11667	T
TOU-A1-75	340	France	46.6667	4.11667	T
TOU-A1-77	341	France	46.6667	4.11667	G
TOU-A1-81	346	France	46.6667	4.11667	T
TOU-A1-82	347	France	46.6667	4.11667	T
TOU-A1-84	348	France	46.6667	4.11667	T
TOU-A1-85	349	France	46.6667	4.11667	T
TOU-A1-93	355	France	46.6667	4.11667	T
TOU-A1-96	357	France	46.6667	4.11667	G
TOU-E-2	367	France	46.6667	4.11667	T
TOU-H-13	374	France	46.6667	4.11667	T
TOU-H-3	375	France	46.6667	4.11667	T
TOU-I-17	378	France	46.6667	4.11667	T
TOU-I-2	379	France	46.6667	4.11667	T
TOU-J-3	383	France	46.6667	4.11667	G
TOU-K-3	386	France	46.6667	4.11667	T
Nc-1	7430	France	48.6167	6.25	T
PHW-37	7508	France	48.6103	2.3086	T
Cen-0	8275	France	49	0.5	T
Bay-0	6899	Germany	49	11	T
Ei-2	6915	Germany	50.3	6.3	T
Ler-1	6932	Germany	NA	NA	T
Ak-1	6987	Germany	48.0683	7.62551	G
Ei-4	7113	Germany	NA	NA	T
Kelsterbach-2	7188	Germany	50.0667	8.5333	G
Kl-1	7195	Germany	50.95	6.9666	G
Kr-0	7201	Germany	51.3317	6.55934	T
Krot-2	7205	Germany	49.631	11.5722	G
Li-5:2	7227	Germany	50.3833	8.0666	G
Li-5:3	7228	Germany	50.3833	8.0666	T
Ma-2	7246	Germany	NA	NA	T
Is-0	8312	Germany	50.5	7.5	T
S96	7472	Netherlands	NA	NA	T
Rsch-4	8374	Russia	NA	NA	T
Bla-1	7015	Spain	41.6833	2.8	T
Bla-11	7017	Spain	41.6833	2.8	G
Bla-1	8264	Spain	41.6833	2.8	T
Can-0	8274	Spain	29.2144	-13.4811	G
Angso-80-432	1318	Sweden	59.5667	16.8667	T
TV-4	6252	Sweden	55.5796	14.3336	T

Table B.4 continued

Line name	Line ID	Country of origin	Latitude	Longitude	Allele
TV-7	6255	Sweden	55.5796	14.3336	T
TV-10	6258	Sweden	55.5796	14.3336	G
TV-22	6268	Sweden	55.5796	14.3336	T
TV-30	6276	Sweden	55.5796	14.3336	T
TV-38	6284	Sweden	55.5796	14.3336	G
Var2-1	7516	Sweden	55.58	14.334	G
Var2-6	7517	Sweden	55.58	14.334	G
Sim 5	9446	Sweden	55.5678	14.3398	T
VarA 1	9476	Sweden	55.5796	14.3336	G
Kent	8238	UK	NA	NA	T
UKSW06-220	4820	United Kingdom	50.4	-4.9	G
UKSW06-227	4827	United Kingdom	50.4	-4.9	G
UKSW06-262	4862	United Kingdom	50.3	-4.9	T
UKSW06-280	4879	United Kingdom	50.3	-4.9	T
UKSW06-329	4927	United Kingdom	50.3	-4.8	G
UKSW06-337	4935	United Kingdom	50.3	-4.6	G
UKSE06-032	4980	United Kingdom	51.3	0.5	G
UKSE06-062	4997	United Kingdom	51.3	0.5	G
UKSE06-192	5056	United Kingdom	51.3	0.5	G
UKSE06-278	5122	United Kingdom	51.3	0.4	G
UKSE06-640	5350	United Kingdom	51.1	0.4	T
UKNW06-019	5364	United Kingdom	54.4	-3	G
UKNW06-079	5386	United Kingdom	54.4	-3	T
UKNW06-101	5394	United Kingdom	54.4	-3	G
UKNW06-105	5398	United Kingdom	54.4	-3	G
UKNW06-197	5461	United Kingdom	54.4	-3	G
CIBC-5	6730	United Kingdom	51.4083	-0.6383	T
LI-OF-002	607	USA	40.7777	-72.9069	G
LI-OF-066	631	USA	40.7777	-72.9069	G
LI-SET-009	719	USA	40.9352	-73.114	G
KYF-21	827	USA	38.0786	-84.473	G
MAW-7	895	USA	42.5319	-72.5595	T
CON-6	946	USA	42.4682	-71.4158	G

Table B.5 Lines from the Bay x Sha RIL family used to test for an effect of HYPO5 on hypocotyl length.

Line number	Genotype at HYPO5	Allele
3	Shahdara	G
4	Shahdara	G
6	Shahdara	G
7	Shahdara	G
8	Shahdara	G
10	Shahdara	G
12	Bay-0	T
13	Shahdara	G
14	Bay-0	T
15	Shahdara	G
16	Bay-0	T
18	Shahdara	G
19	Bay-0	T
21	Shahdara	G
22	Bay-0	T
24	Shahdara	G
27	Shahdara	G
30	Bay-0	T
31	Shahdara	G
33	Bay-0	T
34	Shahdara	G
35	Bay-0	T
37	Bay-0	T
38	Shahdara	G
40	Shahdara	G
41	Bay-0	T
42	Shahdara	G
46	Bay-0	T
47	Shahdara	G
48	Bay-0	T
49	Bay-0	T
50	Shahdara	G
51	Bay-0	T
52	Shahdara	G
55	Shahdara	G
56	Shahdara	G
57	Bay-0	T
58	Shahdara	G
59	Shahdara	G
61	Bay-0	T

Table B.5 continued

Line number	Genotype at HYPO5	Allele
62	Bay-0	T
65	Bay-0	T
70	Bay-0	T
72	Bay-0	T
74	Bay-0	T
76	Bay-0	T
82	Bay-0	T
83	Bay-0	T

Table B.6 Coarse sand beach lines used in burial assay to test effect of HYPO5**.**

Line name	Line ID	Allele
Angso 59	1313	G
Angso 74	1317	G
Angso 80	1318	T
Sim 1	9442	G
Sim 5	9446	T
TV-10	6258	G
TV-22	6268	T
TV-30	6276	T
TV-4	6252	T
TV-7	6255	T
Var2-1	7516	G
Var2-6	7517	G
VarA 1	9476	G

Table B.7 Plants removed from drought mapping experiment due to early flowering. Lines where over 50% of germinants flowered before vernalization were removed totally from the experiment. Early flowering plants of other lines were also removed, but later flowering individuals were retained.

Line Name	Line ID	Flowered/24	Removed
App1-12	5830	1	Individuals
Ba4-1	8258	21	Whole line
Ba5-1	8259	11	Whole line
Bil 3	5835	20	Whole line
Bot 1	9339	1	Individuals
Brosarp 21-140	1063	12	Whole line
Brosarp-51-157	1072	1	Individuals
Brosarp-53-159	1073	1	Individuals
Eden-5	6010	1	Individuals
EdJ 2	9363	1	Individuals
Had 1	9390	22	Whole line
Had 2	9391	10	Individuals
Had 3	9392	16	Whole line
Ham-27-256	1374	1	Individuals
Hen-16-268	1585	1	Individuals
Hovdala 2	6039	21	Whole line
Kor 1	9410	1	Individuals
Kor 3	9412	15	Whole line
Kor 4	9413	10	Individuals
Kru 3	9416	21	Whole line
Kulturen-1	8240	1	Individuals
Liarum	8241	3	Individuals
Lis 3	6041	17	Whole line
Rod 17-319	1435	19	Whole line
T610	6118	1	Individuals
TAD 03	6171	1	Individuals
TAL 07	6180	24	Whole line
Tdr 1	6188	22	Whole line
Tdr 11	6197	17	Whole line
Tdr 13	6198	11	Individuals
Tdr 2	6189	23	Whole line
Tdr 3	6190	22	Whole line
Tdr 4	6191	17	Whole line
Tdr 7	6193	19	Whole line
Tdr 9	6195	9	Individuals
TDr-16	6201	1	Individuals
TEDEN 02	6209	1	Individuals
Tottarp 2	6243	15	Whole line

Table B.7 continued

Line Name	Line ID	Flowered/24	Removed
TRA 01	6244	8	Individuals
Ull 1-1	8426	8	Whole line
Ull 2-13	8427	19	Whole line
Ull 2-3	6973	15	Whole line
UllA 1	9471	1	Individuals
Vinslov	9057	7	Individuals

APPENDIX C: SUMMARY OF BURIAL PILOT EXPERIMENTS

Experiment	Substrates used	lines	n	Problems/Improvements
HYPO II	soil, sand, clay, peb	28	784	Seeds were buried too deeply; emergence was very low
HYPO III	soil, sand, peb	36	756	Seeds were still buried too deeply; emergence was very low
HYPO IV	soil, sand, peb	34	714	Emergence very low; could not calculate accurate emergence frequency because multiple seeds were planted per cell
HYPO V	soil, sand, peb	34	1428	Planted one seed per cell; emergence still low
HYPO VI	felt>fine sand>coarse sand	18	294	Developed layered fine sand/coarse sand protocol to more accurately reflect beach environment; coarse sand lines outperformed inland and fine sand beach conspecifics, but no significant difference between coarse sand lines with T vs G allele at HYPO5
HYPO VII	felt>fine sand>coarse sand	14	392	Restricted lines tested to coarse sand beach lines; germination rates do not differ but T allele lines seem to have increased survival
HYPO VIII	felt>fine sand>coarse sand	14	392	Opposite effect to that expected (G allele has better survival), but emergence very low, especially for the coarse sand beach lines; difficulty in keeping moisture levels in healthy range
HYPO IX	cheesecloth>fine sand>coarse sand	21	588	Include unburied control to test for differences in germination; germination very low especially in T allele lines; difficulty in keeping moisture levels in healthy range
HYPO X	cheesecloth>fine sand>coarse sand	21	588	Improved watering protocol; replaced all seed supplies from stocks which improved germination; no significant difference between lines with alternate alleles
HYPO XI	cheesecloth>fine sand>coarse sand	21	1176	Tested range of burial depths; difference between G and T alleles evident when buried at intermediate depth

Table C.1 Summary of burial pilot experiments

APPENDIX D: HYPOCOTYL LENGTH OF DARK GROWN PLANTS

METHODS

I lined 38 petri dishes with filter paper saturated with water. I spread ~10 seeds in each plate, with two plates for each line in Table E.1 (three for TV7 and TV22 due to low germination rates in pilot experiments), sealed each plate with parafilm, placed in boxes, and stratified for seven days in the refrigerator. Plates were exposed to light in a growth cabinet for 22 hours to induce germination, then returned to boxes for eleven days, after which hypocotyls were measured from the root-shoot junction to the apical hook.

RESULTS

Lines with the G and T allele did not differ significantly in hypocotyl length when grown in darkness.

Name	Allele	Mean hypocotyl length (mm)
Angso 57	G	8.7
Angso 59	G	9.3
Angso 74	G	7.5
Angso 80	T	8.8
HolA2 2	G	10.8
Hov 2-1	G	9.7
Rev 1	G	8.3
Sim 1	G	9.8
Sim 5	T	8.3
TV10	G	8.2
TV22	T	12.4
TV30	T	11.8
TV4	T	9.5
TV7	T	8.3
Ull 2-3	G	10.0
Var A1	G	10.9
Var 2-1	G	7.7
Var2-6	G	11.8
Yst 1	G	8.6

Table D.1 Average hypocotyl length when grown in darkness for 19 lines.

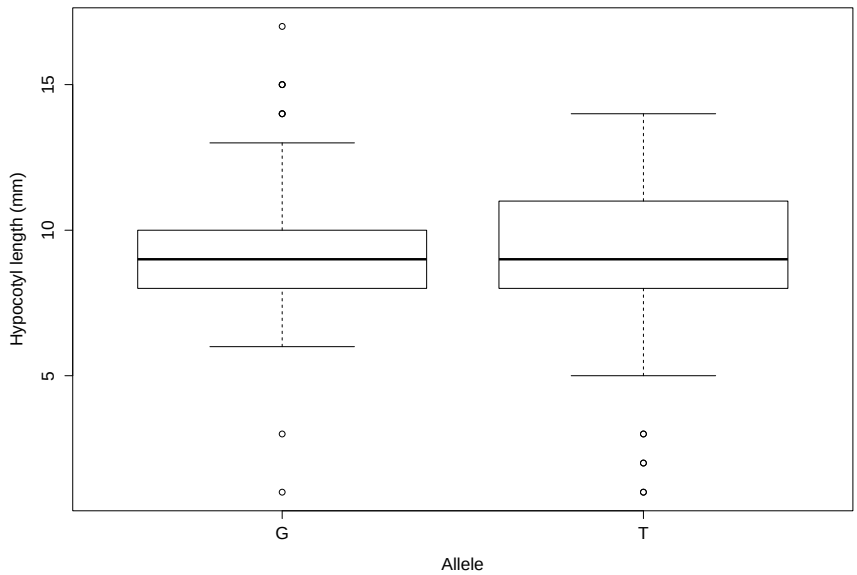

Figure D.1 Average hypocotyl length when grown in darkness of 19 lines with alternate alleles at HYPO5 (n = 332; mean G = 9.2 mm; mean T = 9.3 mm, Student's t-test p-value = 0.70)