The Dissertation Committee for Ginnie Denise Morrison Certifies that this is the approved version of the following dissertation:

Germination Studies in *Arabidopsis thaliana* and *Sinapis arvensis*: Genetical and Ecological Perspectives

Committee:

C. Randal Linder, Supervisor

Thomas Juenger

Daniel Bolnick

Donald Levin

Stan Roux

Michael Singer

Germination Studies in *Arabidopsis thaliana* and *Sinapis arvensis*: Genetical and Ecological Perspectives

by

Ginnie Denise Morrison, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2013

Dedication

I dedicate this thesis to my parents, Karen and Larry Morrison, who I remember, back in undergrad, insisted that I did not need to go into science if I didn't want to. I wanted to. Without their continual support through some tough spots, I don't know if I would have mustered the courage to forge on.

Acknowledgements

I am afraid to do an acknowledgements page because I know I'll leave someone or something out. So I'll name a minimum of names. First, I think I was in just about the best Plant Bio/EEB/Integrative Biology (or whatever) cohort ever. You guys have been great. I just wish you all weren't so damn smart. I really need to thank Sandra Pelc and Teofil Nakov. Without them, this experience would have been much more trying, less invigorating, and a hell of a lot less entertaining. Also, I blame Teo for Chapter 3 existing; I think using RADseq started as a dare at the Turtle Pond.

Germination Studies in *Arabidopsis thaliana* and *Sinapis arvensis*: Genetical and Ecological Perspectives

Ginnie Denise Morrison, Ph.D.

The University of Texas at Austin, 2013

Supervisor: C. Randal Linder

The environment can exert strong selective pressures on an organism. When selective pressures on traits differ between environments local adaptation may occur. If there is gene flow between the environments, local adaptation may be slowed or prevented. In plants, particularly weedy ephemerals, germination is a life-history trait that can be a strong determinant on fitness. In this dissertation, I explore the germination traits of two weedy Brassicaceae species, *Arabidopsis thaliana* and *Sinapis arvensis*, having populations in different habitats to determine whether germination traits within and between populations vary based on environmental conditions and to assess the extent of local adaptation. In Chapter 1, I assessed which genomic regions of *A. thaliana* were associated with differences in germination traits due to genotype-by-environment interactions. I performed a genome-wide association study using 100 natural accessions of *A. thaliana* under four light and nutrient combinations. I found 20 single nucleotide polymorphisms significantly associated with different environments, but none associated

specifically with genotype-by-environment interactions. In Chapter 2, I assessed germination traits of *S. arvensis* collected from agricultural and non-agricultural habitats in the Bitterroot Valley of Montana. I discovered that the agricultural collection studied exhibited significantly different germination timing and amounts than the non-agricultural collections, which were statistically indistinguishable from each other. I also found evidence of a strong maternal effect on germination traits. In Chapter 3, I tested whether patterns of genetic variation between agricultural and non-agricultural collections of *S. arvensis* supported local adaptation to the two habitats even in the face of gene flow. While I expected to see some genetic differentiation between habitats, as seen in Chapter 2, no genetic differentiation was detected and markers putatively under selection were not associated with a particular habitat. I discuss why this might have occurred even though I have evidence for genetically-based phenotypic differentiation between agricultural and non-agricultural populations of *S. arvensis*.

Table of Contents

List of Tables

List of Figures

Introduction

The decision to germinate is one of the most important life events for a plant. When and where a seed germinates plays a major role in determining the environment that the adult plant will face (Donohue, 2005). This is especially true for annuals, which have one growing season to grow and reproduce (Rees & Long, 1992). In general, a temperate weedy ephemeral is expected to germinate as early as possible in order to establish itself before too much competition arises and to maximize growing time. To germinate at appropriate times and under optimal conditions, seeds may use multiple environmental cues including light, temperature, and nutrient conditions. Some seeds require special conditions, such as exposure to smoke (i.e., Keeley & Fotheringham, 1998), and many seeds require a prolonged period of cold prior to or during imbibition, called stratification, to be primed for germination once temperatures increase (Finch-Savage *et al*. 2006).

There are several ways for plants to confront environmental heterogeneity and germinate under the best possible circumstances given their habitat and life history. First, plants may use a bet-hedging strategy. Bet hedging can be on either an individual or population level. In the individual case, seeds produced by the same maternal plant will have different requirements for germination or different resources allocated to them by the maternal plant (Simons & Johnston, 2006). On a population level, different plants within a population produce seeds with different germination requirements (Simons, 2011). In either case, a fraction of the seeds may find the habitats and other conditions

they disperse into appropriate for germination in certain types of years or habitats. When the bet hedging is temporal, some seeds may remain dormant and postpone germinating until certain signals are received.

Another way that plants can react appropriately to environmental heterogeneity is adaptive phenotypic plasticity. Phenotypic plasticity is the expression of different phenotypes by the same genotype under different environmental conditions (Scheiner, 1993). In order to germinate under conditions most conducive to plant growth and fitness, seeds may exhibit some degree of phenotypic plasticity (Donohue, 2002). Because germination timing has a strong effect on future plant fitness, plasticity in germination traits could be adaptive (Donohue, 2002).

Finally, plants can also diverge genetically among subpopulations within a species such that they are genetically differentiated to produce different phenotypes. In this case, different genotypes are selected for in spatially separated environments, leading to a genotype-by-environment interaction (GxE) in germination traits, across habitats. GxE results in certain genotypes, and the phenotypes produced, having higher fitness in one environment than another. Local adaptation can occur even when the spatially separated environments are close enough to allow gene flow. However, for local adaptation to evolve despite gene flow, the selection-migration balance must be in favor of selection, i.e. selection must be a stronger force than gene flow (Keeley & Fotheringham 1998; Lenormand 2002; Donohue *et al*. 2005), but see (Edelaar *et al*. 2012).

Because germination in annual weedy ephemerals is so strongly affected by the environment and linked to fitness, germination is useful for studying local adaptation. In

addition, a number of weedy ephemerals are found both inside and outside of agricultural fields making them good subjects for studying the possibility of local adaptation in the presence of gene flow. Understanding local adaptation could be important for weed control. By understanding how weeds adapt to and invade cropland, we might be able to better control those invasions and retard or prevent adaptations that may arise.

Much of the underlying molecular basis of germination has been elucidated using mutant screens and knock-out experiments in *Arabidopsis thaliana*, or mutants identified in crops. We have learned that seeds are able to sense the different environmental factors using different sets of receptors. Changes in these molecular factors, induced by the environment, leads to the initiation of signal cascades, alteration of gene expression, and culminates in either the initiation of germination or not. It is not the embryo genotype alone that effects germination, however. Maternal effects, whether via physical means (i.e. testa pigmentation; Debeaujon & Koornneef 2000, Debeaujon *et al*. 2000), gene expression in the maternal tissues surrounding the seeds or maternal environmental effects (reviewed in Donohue, 2009) are also known to influence seed germination. What is unclear from these laboratory studies is if the genes identified also cause the phenotypic variation in germination observed in nature. Therefore, the genes identified by molecular methods are valuable for choosing *a priori* candidate genes for quantitative trait loci (QTL) studies and genome-wide association studies (GWAS) of germination traits. Both QTL and GWAS studies of germination traits have offered support for the importance of natural variation in some genes examined in molecular studies but have also identified novel genes (VanderSchaar *et al*. 1997; Alonso-Blanco *et al*. 2003,; Meng

et al. 2008; Laserna *et al*. 2008; Bentsink e*t al*. 2006; Atwell *et al*. 2010; DeRose-Wilson *et al*. 2011).

For my dissertation, I examined natural variation in germination and the potential for local adaptation in two species of weedy mustards (*Sinapis arvensis* and *A. thaliana*, both Brassicaceae). In the United States, both have been introduced from Eurasia. In my first chapter, I used a genome-wide association study in *A. thaliana* to examine natural variation for genotype-by-environment interactions affecting germination in 100 accessions. In chapter two, I tested whether habitat of origin (agricultural or non-agricultural) affected germination phenotypes of *S. arvensis* collected in the Bitterroot Valley of Montana. Finally, using samples collected from mature *S. arvensis* plants in agricultural and non-agricultural habitats in the Bitterroot Valley of Montana, I looked for genetic evidence of local adaptation to agricultural and non-agricultural habitats.

Chapter 1: Association Mapping of Germination Traits in *Arabidopsis thaliana* under Different Environmental Treatments.

INTRODUCTION

When a seed germinates often determines the environmental conditions a plant will face, thus affecting its lifetime fitness. Many species use environmental cues to either initiate germination or remain dormant. The reliability of these cues often provides a fitness advantage (Donohue 2005). These issues are especially important for annual plants (Rees & Long 1992; Donohue et al. 2005a), which have only one opportunity to reproduce after a single growing season and, therefore, less time for environmental conditions to change post-germination.

Environmental heterogeneity may select for different seed germination characteristics within a single species. Selection for different characteristics can result in at least two outcomes: adaptive plasticity (Via & Conner 1995) and local adaptation due to genotype-by-environment interactions (GxE). In the case of adaptive plasticity, selection operates such that a single genotype produces different, adaptive phenotypes depending upon the environment. For example, some species accommodate environmental heterogeneity via bet-hedging, i.e., different seeds with the same genotype, from the same maternal plant, have different requirements for germinating (Simons & Johnston 2006; Simons 2009). Local adaptation can be the result of GxE when a genotype expresses different phenotypes due to environmental conditions (Lynch & Walsh 1998), but in this case selection favors the phenotype produced in one environment, but not the other phenotype in an alternative environment (Ungerer et al.

2003). Populations with two or more genotypes may become locally adapted to multiple habitats by these means or by having genotypes with different, unvarying phenotypes. Since many annual plants are unable to change location post-germination, selection for local adaptation of germination via GxE may be common, as it would produce germination phenotypes appropriate in the home environment.

Light and nutrients affect germination timing and total germination in many weedy ephemeral species (Hilhorst & Karssen 1988; Adler et al. 1993; Weinig 2010). Light can strongly affect germination timing and cuing in some plant species by allowing a seed to sense whether it is buried or overtopped by neighbors (Rees & Brown 1991; Schmitt & Wulff 1993; Baskin & Baskin 2006). Although it is less well studied, nutrient availability in the environment may also influence germination because germinating under higher nutrient concentrations may confer a fitness advantage (Hilhorst & Karssen 1988). Because nutrient availability often varies seasonally, it could be a reliable cue for germination timing (Chapin 1980). The uptake and assimilation of nitrate in *Arabidopsis thaliana* seeds affects germination success and is influenced by at least four genes, as observed in several knock-out and mutant studies (Alboresi et al. 2005; Chopin et al. 2007; Finch-Savage et al. 2007).

Arabidopsis thaliana exhibits significant variation in germination timing and total amount of germination (Donohue et al. 2005b; a; Schmuths et al. 2006; Boyd et al. 2007; Huang et al. 2010). Quantitative trait loci (QTL) studies have revealed chromosomal locations important for germination responses under different environmental conditions (VanderSchaar et al. 1997; Alonso-Blanco et al. 2003; Schmuths et al. 2006; Laserna et

al. 2008; Meng et al. 2008; Bentsink et al. 2010; Huang et al. 2010; Silady et al. 2011). However, these studies have been limited to identifying broad chromosomal regions that contain tens to hundreds of genes and to regions where genetic variation existed between the mapping parents.

Genome-wide association studies (GWAS) can partially overcome the problems associated with QTL studies by using the natural variation present in a large number of geographically and genetically distinct individuals from many populations. Further, GWAS use much denser marker sets than traditional QTL studies, capturing more recombination events and, therefore, identifying smaller chromosomal ranges for loci influencing a trait under particular conditions. To date, few GWAS using *A. thaliana* have looked at germination traits but see (Atwell *et al*. 2010; DeRoseWilson *et al*. 2011), and no germination GWAS have examined GxE effects.

Here, I present the results of a GWAS on germination traits of 100 natural *A. thaliana* accessions and a publicly available set of 213,624 SNPs (Nordborg 250K dataset). I assessed these traits under fully factorial light and nutrient combinations. My goals were to identify genetic regions associated with and GxE effects on germination timing and total proportion of seeds germinated.

MATERIALS AND METHODS

Accession Selection

I used one hundred accessions from the Arabidopsis Biological Resource Center (ABRC, http://abrc.osu.edu/; Figure 1, Table S1). All accessions were part of an earlier

version of the *A. thaliana* RegMap panel (version 3.05, http://papya.usc.edu/2010/data/250k-data-version-3.05; Horton *et* al. 2012). I chose my accessions from an initial screen of 167 accessions. Out of this screen, I chose 100 accessions that flowered without a period of vernalization. Accessions in the initial screen were determined to not need vernalization if they bolted within one month of the earliest-bolting plants. This minimized possible genetic differences due to seasonal germination variation (Donohue 2005; Weinig 2005). The accessions originated from latitudes ranging from 37.79ºS to 61.36ºN, and longitudes from 123ºW to 141.35ºE (Table S1, Figure 1).

Seed Generation

To minimize environmental maternal effects and produce seed for the experiment, I grew the accessions under common garden conditions. I stratified seeds in 500 μL of ddH_2O at 4ºC for 7 days and then planted them in MetroMix 200 potting soil. Each accession had eight replicates. Plants were germinated in two growth chambers (Percival Scientific, Inc.) with identical settings (15 hr light, 22ºC and 9 hr dark, 18ºC). I randomized pots within flats and then rotated flats weekly in the growth chambers. Each growth chamber had an equal number of representatives of each accession. I removed pots with no germination or seedlings that died. Before bolting began, the number of pots was small enough to place the surviving plants in a single chamber. All plants flowered and produced seed in one chamber, minimizing chamber effects on the seeds.

Beginning two weeks after planting, I fertilized the plants every-other week with a half-strength solution of Peter's Professional 20-20-20. I cupped and sleeved plants as soon as they bolted (Aracon System) and stopped watering when the plants senesced. After the plants dried, I harvested seeds and stored them in coin envelopes in the dark at room temperature for at least 30d.

Germination Trial

I implemented a fully-factorial design of two nutrient treatments (High and Low) and two light treatments (Full-Light and Dark). The Low treatment used $1/16^{th}$ strength Peter's Professional 20-20-20 in ddH_2O and the High treatment used $1/8^{th}$ strength. Nutrient levels were selected based on previous trials using *A. thaliana* that showed these nutrient differences had germination effects (unpublished data). Complete darkness (Dark) and full-spectrum light (Full-Light) were chosen for the light treatments because germination can occur at different rates and in different proportions under these conditions (Adler *et al*. 1993; VanderSchaar *et al*. 1997; Meng *et al*. 2008), and prior experiments with *A. thaliana* had shown germination differences between full light and darkness but did not show a significant canopy-shade effect (unpublished data).

Due to time and space constraints, the experiment was temporally blocked. Each block included two replicates of each factorial combination for each accession, and the blocks were conducted two weeks apart. Overall, I had 1600 experimental units (100 accessions x 4 treatment combinations x 2 blocks x 2 replicates/block).

All seeds were surface sterilized for 8 min in a 3.5% (v/v) sodium hypochloride solution with Triton-X as a surfactant and then rinsed three times with filter-sterilized ddH_2O. The experimental unit was a 50 mm Petri dish (BD Falcon). Each dish held 25 seeds in a 5x5 array on doubled P7 filter paper (Fischer). I added 750μL of sterile Low or High nutrient solution to each prepared plate. All plates were placed on trays in two black acrylic boxes and stratified at 4ºC for 3-4 days. After stratification, trays were moved to a Percival Growth Cabinet (15 hr light, 22ºC and 9 hr dark, 18ºC); Full-Light plates were removed from the boxes under safe green light (Roscolux Moss Green Filter by Rosco) and placed in full light. Dark plates remained in the boxes in the growth chamber. Photographs of individual plates were taken under a safe green light every 12 hr for 5 days and then every 24 hr for the next 2 days using a Canon EOS digital camera with a macro lens. Plates were removed prior to the end of the experiment if all seeds had germinated or if mold was observed (16 of 1600 experimental units (1%) had mold; no treatment had a significantly higher rate of infection than any other; χ^2=4.45, df=3, p-value=0.22).

Phenotyping

Photographs were scored for germination. I considered a seed germinated when I observed its radicle protruding from the seed coat. Germination traits analyzed were final proportion germinated at the end of the two week period (Final), time of first germination (First), and two germination dynamics parameters of a germination curve fitted to each plate. The curve had the form of

$$g(t;\ k, tmax, gmax) = \ gmax * e^{\left(\frac{k*(t-\ tmax)}{k*(t-\ tmax)+1}\right)}$$

where *g* is the proportion of seeds germinated at time *t*. The model estimates the maximum number of seeds germinated (*gmax*), the maximum rate of germination (*k*), and the time of the maximum germination rate (*tmax*). Values for the three parameters were estimated using the nls() function in R. For analysis, I was only interested in *k* and *tmax* because I had a direct measure of the final proportion of seeds germinated (Final), and the two week germination period was sufficient to observe full germination of nondormant seeds for nearly all replicates. I only used maximum germination rate and time of maximum germination rate values with p-values < 0.05. The nls function calculates p-values based on profile likelihoods of the estimated model. Plates removed at any point due to mold were excluded from all analyses.

Linear Models for Germination Timing and Total Proportions

I explored the effects of genotype, environment, and GxE using generalized linear mixed models (GLMMs) for the Final and first phenotypes and linear models for the two parameters describing germination dynamics. The distributions for the Final and First phenotypes were non-normal (Shapiro-Wilkes test: Final, W=0.52, p<0.001; First, W=0.71, p<0.001). Neither an arcsine-square root transformation for the Final data nor a box-cox transformation (using bcpower in the car() package (Fox & Wiesberg 2011) of the First phenotype made the data more normal (Shapiro-Wilkes test: Final, W=0.62,

$p<0.001$; First, $W=0.85$, $p<0.001$). Therefore, for all analyses, the untransformed data were used. The GLMM used a binomial distribution for the Final phenotype (as it is a proportion), and the model for First used a Poisson distribution. In general the full models were

$$Y = L*N + L*A + N*A + B + Error,$$

where L (light quality: Full-Light or Dark) and N (nutrient level: High or Low) were fixed effects; and A (accession) and B (block) were random effects. I did not analyze the three-way interaction because the three-way models did not converge. To determine the simplest, most explanatory model, I sequentially reduced each model by removing the least significant term and then comparing the -2log likelihood score of the reduced model to the previous model. All models were run using glmer() in the R package lme4 (Bates *et al*.).

The distributions of significant maximum germination rates and times of maximum germination rate were also all non-normal (Shapiro-Wilkes test: time of maximum germination rate, $W=0.85$, $p<0.001$; maximum germination rate $W=0.97$, $p<0.001$). I found no transformations appropriate for the data, so linear mixed-models were run on the untransformed data. The initial models used and the methods to determine the best reduced models were the same as for the First and Final phenotypes.

Heritability

I calculated the broad-sense heritabilities (H^2) of the traits for each factorial combination using the SAS proc mixed procedure (SAS Institute Inc. 2008). The model Y = accession + block + error, with accession and block as a random effects, was used to obtain the variation due to genotype (accession) and total (accession + block + error) variance. The model fit with the accession term was compared to that without accession to test for a significant, genotypic effect. Although the data were not normal, I considered H^2 significant if the difference between the -2 loglikelihoods for the two models was greater than 3.84 (X^2 value significant at α=0.05) as most differences were much greater than 3.84.

Geographic Location and Germination

The genetic variation in *A. thaliana* shows genetic structure based on geographic location (Beck *et al*. 2008; Zhao *et al*. 2007). This variation is often correlated with longitude or latitude, and therefore with the kinship matrix used in association mapping (see below) Therefore, important phenotypic variation associated with genetic variation on latitudinal or longitudinal clines might masked by the kinship matrix (Zhao *et al*. 2007; Aranzana *et al*. 2005). To see if there was a correlation between the measured phenotypes and latitude or longitude, I ran two linear models testing each geographic feature separately. Generally, these models were

$$Y = L*N*P+Error$$

where Y is the phenotypic value; L and N are as before; and P represents either the latitude or longitude at an accession's reported collection site.

Candidate Genes

Prior to association mapping, I created a list of 132 *a priori* candidate genes known to be functionally important during germination or identified in previous studies of natural variation for germination traits (Table S2). Based upon gene function, I expected four genes might be subject to GxE under different nutrient conditions and 36 genes might be subject to GxE under different light conditions. I considered the remaining 92 genes to be 'general' germination genes, with no explicit hypotheses about which factors they would be most responsive to.

Association Mapping of Individual SNPs

To assess if I should use the MTMM or the simpler EMMA method, I calculated the phenotypic correlations between environments for phenotypes with a significant GxE interaction in R. I used Kendall's tau as the data were not bivariate normal and contained ties.

I used the Final phenotype and time of maximum germination rate for association mapping, as First and the maximum germination rate showed no significant GxE effects in the linear models. Using the SNP data from version 3.05 of the Nordborg dataset, I implemented the multi-trait mixed model method (MTMM) (Korte *et al*. 2012) to

perform the association analyses. The MTMM method uses a K-matrix, created with EMMA (Kang *et al*. 2008), to control for population structure. For the 2x2 factorial set up used in my study, MTMM first runs a model for each of the four factorial environments (environmental models), which are simply individual GWA analyses for each treatment combination (Full-Light/High, Full-Light/Low, Dark/High, and Dark/Low). MTMM also runs five complex models: (1) a model testing for significant effects of both genotype and environment against a null model (full model); (2) a model testing for significance of genotype alone against a null model (common model); (3) a model testing for GxE effects for light and nutrients together (GxE model); and (4) a model testing GxE for one factor (e.g, light model), and (5) a model testing GxE for the other factor (e.g., nutrient model) (Korte *et al*. 2012).

I removed all minor alleles (SNPs with frequencies ≤ 0.1) before the MTMM analyses because minor alleles are prone to spuriously low p-values (Atwell et al. 2010). I controlled experiment-wise Type I error by correcting the obtained p-values with a Benjamini-Hochberg false discovery rate (BH-FDR) correction (Benjamini & Hochberg 1995) in R using the p.adjust() function (The R Project for Statistical Computing).

To test for significant candidate genes, I used the MTMM model on two sets of SNPs: (1) only the SNPs present within 1000bp upstream of a candidate gene's start site and 500bp downstream of a candidate gene's stop site (a highly constrained view of the gene, ignoring linkage) and (2) all SNPs within 10kb up and downstream of a gene (the average size of haplotype blocks in *A. thaliana*) (Kim et al. 2007)

To do a genome-wide analysis of germination, I then used the MTMM model to look for associations between the full set of SNPs, minus minor alleles, and either the different final proportions of seeds germinated or time of maximum germination rate.

Genes Linked to Significant SNPs

I examined the gene(s) within 10kb up- and downstream of each significant SNP from the GWAS models, ignoring pseudogenes and transposable elements. I chose a ±10kb range as linkage disequilibrium generally decays within 10kb in *A. thaliana* (Kim et al. 2007). For genes linked to significant SNPs, I looked at TAIR10 gene descriptions and gene ontologies (GO's), when available, to identify gene function and assess whether the genes had any known effect on germination.

RESULTS

Germination Results

The average final proportion germination (Final) for each treatment ranged from 80.7% to 94.5% (Figure 2). Significantly fewer plates in the light had <50% total germination than plates in the darkness (45 of 793 in Full-Light and 120 of 791 in Dark; χ^2=30.8, df=1, $p < 0.0001$). Accessions that had plates with germination rates less than 80% likely had dormant seeds because the same accessions had rates of germination $\geq$ 81% when independent sets of seeds were treated with gibberillic acid (GA) to force germination (data not shown). It could be that these accessions preferred a warm/wet

stratification rather than a cold/wet one. The time of first germination did not vary greatly between the four factorial combinations (1.02-1.18 d Figure 2).

For the germination dynamics measures, I was able to fit curves for 1543 of 1584 experimental units. A total of 948 experimental units had significant estimated maximum germination rates, and 1525 experimental units had a significant estimated time of maximum germination rate. The maximum germination rate ranged from 0.06 to 1.75, measured as the proportion of seeds germinated per day (Figure 2). Numbers greater than 1 indicate that all seeds germinated in less than one day. The time of maximum germination rate ranged from 1.53 to 29.13 d (Figure 2).

Broad sense heritabilities ranged widely, depending on the germination trait. All Final and First heritabilities were significant. Within factorial combinations, broad-sense heritabilities for Final (0.44 to 0.55) were higher than those for First (0.16 to 0.26; Table 1). Heritabilities for maximum germination rate were also significant in all cases and fell between the values for Final and First (range: 0.25 to 0.37). Significant heritabilities of the time of maximum germination rate fell within a similar range (0.27 to 0.32) as those for the maximum germination rate and were significant in all cases except under the Full-Light/Low treatment ($H^2 = 0.10$).

Phenotype and GxE Effects

By sequentially removing nonsignificant terms, I determined which factors influenced each germination trait measured. Two GxE effects (light x accession and nutrient x accession) were significant for Final, as were the light and nutrient main effects

(Table 2). Full-Light and High nutrients increased the final proportion of seeds germinated. For First, the nutrient, light, and accession effects were all significant, but no GxE effects were significant. Full-light and High nutrients both caused the time of first germination to occur earlier. For the germination dynamics measures, there was a significant light-by-accession effect for time of maximum germination rate as well as significant light and nutrient effects. Full-Light and High caused the maximum germination rate to occur sooner than the Dark or Low treatments. The maximum germination rate itself depended upon the accession used, and light and nutrient conditions. Full-Light and High nutrients increased the maximum germination rate. Because there were only significant GxE effects for the Final and time of maximum germination rate phenotypes, I continued my analyses only on those two phenotypes.

Geographic Location and Final Phenotype

There was no association between the Final phenotype and either latitude ($F_{1,1580}$=0.99, p=0.32) or longitude ($F_{1,1580}$=1.14, p=0.29). The same held true for time of maximum germination rate (latitude: $F_{1,1520}$=1.01, p=0.32; longitude: $F_{1,1520}$=0.002, p=0.96). Therefore, I did not expect the kinship matrix to mask significant SNPs associated with either phenotype.

Significant Candidate Gene SNPs

Short Range

Three of the original 132 candidates did not have any SNPs within the short sequence length considered. A total of 762 SNPs in 129 genes were tested, and there were 1 to 33 SNPs/gene, averaging 5.91 SNPs/gene.

For the Final phenotype at an FDR level of 0.05, only one candidate gene SNP was significant. It was identified in the full, common, and Full-Light/High models (corrected p-values 0.006, 0.002, and 0.032, respectively). That SNP was within the gene TT12 (AT3G59030), a transparent testa gene affecting testa pigmentation and permeability. TT12 was classified as a general candidate as it might affect germination via light penetration (due to testa pigmentation), and nutrient availability (due to testa permeability).

For the time of maximum germination rate, two candidate-gene SNPs were significant. One was the same SNP in TT12 identified for the Final phenotype. It was significant in the common and the Full-Light/High model (corrected p-values 0.013 and <0.0001, respectively). The second SNP was identified in the Full-Light/High model and was within candidate gene FRS2, a light candidate gene, responsive to red/far-red light. As of writing, the function of FRS2 was unknown.

Long Range

When I enlarged the sequence range around a candidate gene to 10kb up- and downstream, all candidate genes had linked SNPs. A total of 4030 different SNPs were within the range examined, with each candidate gene containing between 3 and 87 SNPs (average: 31.7).

For the final phenotype at an FDR level of 0.05, I found three significant candidate gene SNPs associated with four candidate genes in in the full, common, light GxE, and Full-Light/Low models. One was the significant SNP in TT12 found under the shorter sequence conditions. Here it was also significant in the full and common models (adjusted p-values 0.03 and 0.01, respectively). That SNP was also within 10kb of candidate gene PIL6 (AT3G59060), a phytochrome-interacting gene that was a light candidate. One SNP was significant in the full model and light GxE model (adjusted p-values 0.03 and 0.02, respectively). This SNP was located in the range of candidate gene RAS1 (AT1G09950), a general gene with no described function. Finally, candidate gene NRT2.7 had a significant SNP in the Full-Light/Low model (adjusted p-value =0.01). NRT2.7 is a nutrient candidate gene and is involved in nitrogen transport in seeds.

Very similar results were seen for the time of maximum germination rate. I found two significant candidate gene SNPs associated with three candidate genes in the Full-Light/Low and the Full-Light/High models. In the Full-Light/High model, the same SNP in TT12 was identified again (corrected p-value = 0.025). As in the Final phenotype with the increased candidate gene range, PIL6 also had a significant SNP. And, just like the

Final phenotype analysis with the increased range, the same SNP in NRT2.7 was significant in the Full-Light/Low model (corrected p-value = 0.025)

Significant Genome-wide SNPs

I tested whether values for Final or time of maximum germination rate were correlated across environments to assess the appropriateness of using the MTMM method. I found that Final values measured within each light environment were highly correlated (Dark/Low v. Dark/High: Kendall's tau = 0.71, $p<0.001$; Full-Light/Low v. Full-Light/High: Kendall's tau = 0.76, $p<0.001$). Phenotypes within nutrient treatments had lower correlations but were still significant (Dark/Low v Full-Light/Low: Kendall's tau = 0.46, $p<0.001$; Dark/High v. Full-Light/High: Kendall's tau =0.48, $p<0.001$). Similarly, values within light environments and those within nutrient treatments were correlated for time of maximum germination rate but not as strongly as Final (Dark/Low v. Dark/High: Kendall's tau = 0.57, $p<0.001$; Full-Light/Low v. Full-Light/High: Kendall's tau = 0.36, $p<0.001$; Dark/Low v Full-Light/Low: Kendall's tau = 0.35, $p<0.001$; Dark/High v. Full-Light/High: Kendall's tau =0.47, $p<0.001$). The high correlations between these phenotypes indicated usage of models such as MTMM to disentangle GxE effects would be most appropriate (Korte *et al*. 2012).

Final Phenotype

For the MTMM models of the Final phenotype, I used the final proportion of seeds germinated for each accession in each environment (calculated using the GLMM)

for the response variable. The average number of seeds germinated for an accession ranged from 0.82 $\pm$ 0.27 to 0.95 $\pm$ 0.11 in each environment.

Overall, I identified 14 significant SNPs associated with final germination proportion. I refer to these SNPs as F-SNP1 through F-SNP14 (Table 3). Some of these SNPs were likely linked as they were within 10kb of each other (Table 3). In some cases, linked significant SNPs had intervening non-significant SNPs (Figures 3 and 4).

Twelve of the 14 significant SNPs were identified only in the environmental models, and ten were significant in only one factorial combination (Full-Light/Low: 7; Full-Light/High: 2; Dark/Low 1; Figure 3, Figure 4, Table 3). The remaining two SNPs found only in environmental models were significant in both Full-Light/Low and Full-Light/High treatments.

The two F-SNPs significant in models in addition to the environmental models were F-SNP5 and F-SNP2 (Table 3). F-SNP5 was identified in the full, common, Full-Light/Low, and Full-Light/High models. F-SNP2 was significant for the full, Full-Light/Low, and Full-Light/High models. These two SNPs are the ones most likely to be associated with GxE effects because of their significance in the full model. The full model assesses the significance of the SNPs by comparing tests that include and exclude both the genotype and environment terms. No significant SNPs were detected in the models explicitly testing for GxE overall, or GxE related to light, or nutrient effects alone.

Since GWAS is often treated as a hypothesis-generating tool, I also examined the SNPs that were significant at an FDR of 0.1. Eleven more SNPs were significant and a

twelfth SNP, previously significant at the 0.05 level, was significant in an additional model (Full-Light/High). The eleven new SNPs were named F-SNP15 through F-SNP25. One of these F-SNPs was significant in the Full-Light/High model. The other ten were significant in the Full-Light/Low model. Again, no SNPs were associated with any GxE effects and no new SNPs were identified in the full or common models.

Time to Maximum Germination

For the MTMM models of the time of maximum germination, I calculated the mean time for each accession by light by nutrient combination with a significant estimated time of maximum germination. At an FDR-corrected p-value level of 0.05, there were a total of six significant SNPs. I refer to them as tmax-SNP1 through tmax-SNP6 (Table 4). They were significant in the Full, Dark/Low, Full-Light/Low, and Full-Light/High models and no SNPs were significant in any GxE models (Figure 5 and Figure 6). The Dark/Low model had one significant SNP, tmax-SNP6, that was not significant in any other model. The Full-Light/Low, Full-Light/High, and Full models all shared one significant SNP, tmax-SNP1. The Full-Light/Low model identified tmax-SNP2 and tmax-SNP3. Finally, the Full-Light/High model identified the remaining two SNPs, tmax-SNP4 and 5 (Figure 5).

As with the Final phenotype, I examined SNPs that were significant at an FDR of 0.1. I identified an additional 31 significant SNPs at this level. I named these SNPs tmax-SNP7-37 (Table 4). The additional SNPs were identified in the Dark/Low, Full-Light/Low, and Full-Light/High models. One previously identified SNP, tmax-SNP2, was

significant at the 0.1 level in the Full-Light/High model. The three models shared one significant SNP, tmax-SNP15. Full-Light/Low and Full-Light/High shared one SNP, tmax-SNP20. Otherwise, each of the three models had unique significant SNPs (Dark/Low: 2; Full-Light/Low: 12; Full-Light/High: 16). Again, there were no SNPs identified in any GxE model.

At the 0.05 level, three SNPs were significant for both Final and time of maximum germination rate (F-SNP2, 8, and 11; tmax-SNP1, 3, 6). At the 0.1 level, an additional 13 were significant for both traits (Table 3 and Table 4).

Genes Linked to Genome-wide SNPs

Final Phenotype

Sixty-three genes resided within 10kb up- and downstream of the 14 significant F-SNPs at an FDR level of 0.05. Each F-SNP was linked to one to nine genes (Table A3). Only one of the 63 linked genes (AT5G14570, NRT2.7) was an *a priori* candidate. Six genes had a significant SNP located within an exon, 3'UTR, or intron (Table 5). In the remaining cases, F-SNPs were located outside the coding region, on average 4647 bp (94-9816bp) from a gene's transcription start or stop codon (Table 5).

I wanted to see if the set of genes linked to significant SNPs were enriched for any particular molecular functions or biological processes related to germination. When compared against the whole *A. thaliana* genome, the set of 63 genes was enriched for sinapate 1-glucosyltransferase activity (GO:0050284; p-value=0.024) (Carbon *et al.* 2009), which has not been connected to any germination process. Two genes,

AT4G15480 and AT4G15490, have this activity and both are linked to SNP9 in the Full-Light/High model.

Three linked genes had relatively direct ties to germination: candidate gene NRT2.7 (F-SNP10), IDD1 (AT5G66730, F-SNP14), and PIN1 (AT1G73590, F-SNP5). F-SNP10 and F-SNP14 were both significant in the Full-Light/Low model, and F-SNP 5 was significant in the Full-Light/Low, Full-Light/High, full, and common model. NRT2.7 has biological process GO's of nitrate transport, transmembrane transport and is known for its importance in nitrate uptake in seeds (Chopin *et al*. 2007). IDD1 has relevant biological process GO's that include regulation of GA-mediated signaling pathway, regulation of seed germination, and seed maturation. Relevant to germination and the environmental conditions under study, biological process GO's for PIN1 include embryo development, polarity specification of adaxial/abaxial axis, response to auxin stimulus, auxin polar transport, photomorphogenesis, and response to blue light.

Of the other 60 genes, 35 are expressed in the seed or embryo, but either had no known biological process GO or had a biological process GO that had no clear association with a seed or embryo phenotype.

Forty-one additional genes were linked to the 11 additional SNPs significant at the weaker FDR of 0.1. Four F-SNPs (15, 16, 18, 21) were linked only to genes that SNPs significant at the 0.05 level were linked to. The remaining seven F-SNPs were linked to between four and eleven genes and were the sole significant SNP linked to those genes. None of the linked genes were candidates and no new genes had obvious functions that would associate them with the environments examined or germination.

Time of Maximum Germination Rate

Twenty-nine genes were linked to tmax-SNPs significant at the 0.05 level. Each tmax-SNP was linked to one to seven genes (Table A4). Seventeen of the twenty-nine genes were expressed in the seed or embryo. Two linked genes were *a priori* candidates: TT12 and PIL6, also identified in the candidate gene screen for time of maximum germination rate. Both genes were linked to tmax-SNP4, significant in the Full-Light/High model. None of the remaining linked genes had a description that appeared to be directly associated with germination or light or nutrient responses.

When the FDR was loosened to 0.1, an additional 148 genes were linked to tmax-SNPs. Four tmax-SNPs at the 0.1 level were linked to eight genes linked to tmax-SNPs at the 0.05 level. One *a priori* candidate, NRT2.7, was identified; it was linked to tmax-SNP31 in the Full-Light/Low model. Two other genes involved in nitrate response and transport were linked to tmax-SNP22 and tmax-SNP32. These genes were ASFT (AT5G41040), and AT3G16460. Both SNPs were identified in the Full-Light/Low model, like tmax-SNP31. Also of interest were two genes responsive to GA (SHI, AT5G66350, tmax-SNP35; ERD, AT4G15430, tmax-SNP27). Both genes were identified in the Full-Light/Low model and are expressed in the seed or embryo. Another two genes, TOR (AT1G50030) and MEE36 (AT3G16440) are involved in embryo development and were both linked to tmax-SNPs from the Full-Light/Low model (tmax-SNP12 and tmax-SNP22). Additionally, tmax-SNP17, identified in the Full-Light/High model, was linked to NPY2 (AT2G14820), a gene responsive to light stimulus. Finally,

PIN1, linked to F-SNP5, was also identified. For time of maximum germination rate, PIN1 was linked to tmax-SNP15 in the Dark/Low treatment.

DISCUSSION

Determining the genetic causes of variation in germination across heterogeneous environments is important for our understanding of plant evolution and could have applications to agriculture. Considerable evidence shows that selection has acted on seeds for using the information in light and, to a lesser extent, nutrients as cues (Schmitt *et al*. 1992; Weinig 2010; Adler *et al*. 1993; Hilhorst & Karssen 1988).

Although the mixed model indicated significant interactions between genotype and light and nutrient conditions, no significant SNPs were identified in the models testing the light or nutrient environments by genotype interactions alone. Using the same method, Korte *et al*. (2012) found only one significant SNP in one GxE model, even though their study included approximately four times as many accessions (~400 vs. 100). Korte *et al*. (2012) suggest that their inability to find GxE SNPs could have been a result of the complexity of the model. Loci contributing to GxE seem to be difficult to identify and are often of small effect size (Smith & Kruglyak 2008).

The most pertinent results with respect to differing germination characteristics in differing environments are SNP5 and SNP2, identified in the full model, because they are potentially associated with GxE effects; the full model compares a model with genotypic and environmental effects against one with neither effect. These SNPs may be linked to a gene or genes that are responsive the environment and initiate responses that are

environment-dependent. While the models for the time of maximum germination rate did identify more SNPs and genes, only one SNP was significant in the Full model, and none of the linked genes had any obvious role in germination or light or nutrient responses.

Germination Dynamics

I had expected germination dynamics to show GxE effects because although germinating at high proportions could result in higher fitness on average, different environments, such as highly competitive environments or those with shorter growing seasons, could select for different optimal times to germinate after receiving a particular cue. Surprisingly, of the two traits I measured, only time of maximum germination rate showed any GxE effect. However, the maximum germination rate did show genotypic environmental effects. Environmental factors that may affect maximum germination rate could be different from the light quality and nutrient level environments I tested.

Candidate gene SNPs

I expected my candidate genes might have a large number of significant SNPs because of their previous associations with germination under other conditions. Even though I tested for significant SNPs within different ranges of the candidate genes, I identified very few candidate SNPs (one to three). The candidate list I made does not seem to have contained many genes underlying differences in the final proportion of seeds germinated or time of maximum germination rate between accessions or treatments. The candidates I selected from studies in knock-out or laboratory-created

mutants, though they represent genes important for germination, might not have any natural variants. Candidates I selected that had been identified by other association-mapping studies might not have detectible influence on the traits I measured under the treatments used here.

Genes linked to significant SNPs

I identified 104 new genes in this study that may affect germination totals in the four environments studied and, therefore, could be candidates for further study. Three linked genes, NRT2.7, IDD1, and PIN1, are of particular interest. Both NRT2.7 and IDD1 were linked to significant SNPs in the Full-Light/Low model. NRT2.7, the only *a priori* candidate gene identified in this study and also identified in the candidate gene analysis, is a nitrate transporter that operates in siliques and seeds (Chopin *et al*. 2007). It may be that different alleles of NRT2.7 affect germination under low nutrient conditions as opposed to under high nutrient conditions. Increasing the nitrate content of the maternal plant during fruit formation or directly in germination media can increase total germination proportions (Alboresi *et al*. 2005; Nambara & Marion-Poll 2003).

IDD1 is expressed during seed maturation and promotes germination through interactions with the phytochrome pathway. Its significance in a full light environment makes sense due to this interaction with the light pathway. Perhaps different alleles of IDD1 more positively affect light germination under low nutrient levels.

Several other members of the PIN family, but not PIN1, were *a priori* candidate genes because of their linkage with significant SNPs in other studies. The PIN family

consists of auxin efflux transporters. PIN1 is important for numerous auxin-related responses, including the development of the apical-basal axis in *A. thaliana* embryos (Friml *et al*. 2003). PIN1 is also up-regulated during germination and is eventually important for root establishment (Holdsworth *et al*. 2008). PIN1 also functions in photomorphogenesis, but has no described role in light sensing and germination. Therefore, PIN1 is clearly related to germination and seems like a good candidate for germination control. At this point, it is unclear how PIN1 is linked to germination differences between the four environments I used. Its role in root establishment and its significance in the full model may indicate some relation to nutrient sensing.

Furthermore, I identified 177 genes, 60 of which were also linked to significant F-SNPs, that may affect the time of maximum germination rate. When looking at all tmax-SNPs significant at the 0.1 level or lower, I found three *a priori* candidate genes linked tmax-SNPs and eight additional genes of interest linked to tmax-SNPs. One linked gene was PIN1, an *a posteriori* candidate also linked to an F-SNP. In this case, PIN1 was linked to a SNP significant in the Dark/Low, Full-Light/Low, and Full-Light/High treatments. Overall, sixty of the same genes were identified in both the Final and time of maximum germination rate analyses. This is not surprising as one would expect that genes effect one aspect of germination also effect another, as the two traits are likely to be at least somewhat dependent upon one another.

Although there was no nutrient-by-accession effect for the time of maximum germination rate mixed model (Table 2), three genes linked to tmax-SNPs (*a priori* candidate NRT2.7, ASFT, and AT3G16460) were of interest due to their roles in nitrate

transport and response. All three genes were linked to SNPs significant in the Full-Light/Low model.

NRT2.7 was linked to a significant SNP for the same model of the Final phenotype. As postulated above, genes affecting nutrient assimilation may be more important under low nutrient conditions. Also, the ability to assimilate nutrients might affect the dynamics of germination.

Other than PIN1, three other genes responsive to light stimulus, TT12, PIL6, and NPY2, were identified. All three were identified in the Full-Light/High model and TT12 and PIL6 were both *a priori* candidates. Like the other light-responsive gene, IDD1, identified in the Final analysis, these genes were all identified under Full-Light conditions.

Comparison with prior studies

QTL studies

Two QTL studies, using the Bay-0 x Sha RIL set (Laserna *et al*. 2008; Meng *et al*. 2008), examined different environmental effects on *A. thaliana* germination. Six QTL were associated with total germination in the dark at 6ºC (Meng *et al*. 2008), and three were associated with total germination after a red-light pulse (Laserna *et al*. 2008). Between the two studies, two QTL collocated. Because the physical locations of the Bay-0 x Sha markers are known (Loudet *et al*. 2002), I was able to estimate that two significant SNPs (SNP6 and SNP12) fell within these peaks. I also estimated that SNP9 collocates with another QTL identified by (Meng *et al*. 2008. Both Laserna *et al*. (2008)

and Meng *et al*. (2008) identified a QTL that collocated with both PHYB (AT2G18790) and PIL5 (AT2G20180), which were considered potential candidate genes in both studies. F-SNP6 collocated with the same QTL; however, the intervals under these QTL peaks are very large and PIL5 lies over 1500 kb from F-SNP6, with PHYB even more distant. Although PHYB and PIL5 are both involved in the light-regulated phytochrome pathways and affect *A. thaliana* germination (Oh *et al*. 2004), F-SNP6 is likely linked to a different gene due to its distance from the both genes. No candidate genes were identified in either study for the QTL that collocated with F-SNP9 or the one that collocated with F-SNP12.

GWA studies

Atwell *et al*. (2010) measured six phenotypes directly related to germination and dormancy: time to 50% germination after two different storage conditions and percent of (non-dormant) seeds germinated after a week under four conditions (in the dark at 4ºC and under 16 hr days at 10, 16, and 22ºC). They reported ten genes that were plausible germination- or dormancy-related genes within 20kb of the most significant SNPs. I did not identify any of the same SNPs or genes that they did and no SNP I identified was within 20kb of their SNPs. At best, four SNPs I identified were within 500kb of their SNPs.

There are at least three possible reasons why I found different significant SNPs from Atwell *et al*. (2010). First, the 100 accessions used in my study overlapped with at most 51 of their 199 accessions (not all of their accessions were used in each of their

treatments), so the responses of the accessions I did not have in common could have produced different results. Second, their experimental conditions and measured phenotypes were different from those I measured. Finally, our analyses were different. While I used the MTMM method and was interested in potential GxE interactions, Atwell *et al*. (2010) used EMMAX and were not expressly asking GxE questions. My finding of fewer and different SNPs using MTMM is not unique. Data for an earlier GWAS study using EMMAX found 92 significant SNPs across environments (Li *et al*. 2010). When that data was re-analyzed using the MTMM method, only 41 SNPs were found in all nine models tested, and, to the best of my knowledge, only nine SNPs were shared between the two studies (Li *et al*. 2010; Korte *et al*. 2012).

Conclusions

Natural variation in total germination in *A. thaliana* is partially dependent upon light and nutrients. Although I identified several genomic regions that influence differences in this trait, I likely identified a subset of the chromosomal regions that produce GxE interactions. To fully understand the role of natural variation in germination timing and cuing, we must study germination under a large variety of ecologically relevant conditions and begin testing the effects the genetic variants discovered by GWAS have on germination under natural and controlled conditions.

Table 1. Broad-sense heritability for each factorial combination and phenotype. Bold heritabilities are significant at $p < 0.05$. H2, broad-sense heritability

Phenotype	Treatment	H^2
Final	Dark/Low	**0.55**
	Dark/High	**0.47**
	Full-Light/High	**0.47**
	Full-Light/Low	**0.44**
First	Dark/Low	**0.26**
	Dark/High	**0.16**
	Full-Light/High	**0.21**
	Full-Light/Low	**0.22**
Max Germination Rate	Dark/Low	**0.37**
	Dark/High	**0.33**
	Full-Light/High	**0.32**
	Full-Light/Low	**0.25**
Time of Max Germination Rate	Dark/Low	**0.31**
	Dark/High	**0.27**
	Full-Light/High	**0.32**
	Full-Light/Low	0.10

Table 2. GLMM models run and their AIC and log likelihood scores.

Model[a,b]	AIC	log likelihood	Significant fixed terms
First			
Y~L+N+L*N+A+L*A+N*A+B	693.8	-334.9	
Y~L+N+L*N+A+L*A+N*A	724	-351	
Y~L+N+L*N+A+L*A+B	687.8	-334.9	
Y~L+N+L*N+A+B	681.8	-334.9	
Y~L+N+L*N+B	727.9	-359	
Y~L+N+A+B	**680.3**	**-335.1**	**Earlier germination in Full-Light and High**
Max Germination Rate			
Y~L+N+L*N+A+L*A+N*A+B	892.27	-433.13	
Y~L+N+A+L*A+N*A+B	890.58	-433.29	
Y~L+N+A+L*A+B	885.86	-433.93	
Y~L+N+A+B	**884.60**	**-436.30**	**Lower max rate under Dark, and under Low**
Time of Max Germination Rate			
Y~L+N+A+L*A+N*A+B	7845.8	-3910.9	
Y~L+N+L*N+A+L*A+B	**7839.8**	**-3910.9**	
Final			
Y~L+N+L*N+A+L*A+N*A+B	**6222**	**-3103**	**Higher proportions in Full-Light/High**
Y~L+N+L*N+A+L*A+B	6508	-3247	
Y~L+N+L*N+N*A+A+B	7980	-3983	

Table 3. The 25 SNPs significant for Final in at least one of the nine MTMM models. Raw p-values are shown for each SNP in each model. p-values in bold are significant at the 0.1 level after a B-H FDR correction; those with an asterisk are significant at the 0.05 level after a BH-FDR correction. SNPs in bold are shared with tmax-SNPs significant at the 0.1 level.

SNP	Chr	Position (bp)	Dark/Low	Dark/High	Full-Light/Low	Full-Light/High	Full	Common	GxE	Light	Nutrient
SNP1	1	2,757,164	3.11E-03	0.024	**1.82E-07***	**1.05E-06***	2.64E-06	1.85E-05	5.62E-03	5.33E-03	0.0514
SNP15	1	2,759,471	0.091	0.319	**4.35E-06**	2.19E-04	8.74E-05	1.61E-03	3.27E-03	3.70E-03	0.0372
SNP16	1	2,763,016	0.261	0.369	**8.66E-06**	4.54E-05	9.41E-05	1.50E-03	3.77E-03	9.09E-04	0.434
SNP2	1	2,765,047	0.029	0.049	**1.16E-09***	**9.63E-09***	**2.14E-7***	4.88E-06	1.50E-03	3.22E-04	0.521
SNP3	1	2,770,350	0.012	0.012	**2.86E-06***	1.42E-05	1.34E-04	3.89E-05	0.164	0.0628	0.876
SNP17	1	10,419,017	5.09E-04	1.83E-03	**9.30E-06**	7.41E-05	2.80E-04	5.83E-05	0.243	0.126	0.379
SNP4	1	21,916,027	7.42E-03	0.014	**3.18E-05**	**1.65E-08***	6.60E-05	1.44E-05	0.201	0.0862	0.764
SNP5	1	27,668,561	1.87E-05	3.60E-04	**1.31E-09***	**1.79E-07***	**3.49E-07***	**1.15E-7***	0.0907	0.109	0.0932
SNP6	2	10,297,188	0.064	0.17	**5.42E-07***	1.65E-04	1.21E-04	1.02E-03	6.94E-03	3.51E-03	0.122
SNP18	2	10,297,285	0.076	0.19	**7.50E-06**	1.64E-03	8.22E-04	2.94E-03	0.0198	0.0154	0.0892
SNP19	2	17,620,611	0.188	0.228	**3.63E-06**	1.80E-04	2.33E-04	2.52E-03	6.03E-03	1.49E-03	0.46
SNP20	3	5,837,328	0.025	0.096	**5.82E-06**	1.38E-03	2.21E-04	1.16E-03	0.0116	0.023	0.0276
SNP21	3	12,162,344	0.177	0.248	**7.91E-06**	1.21E-05	2.81E-04	6.42E-04	0.0258	7.37E-03	0.494
SNP7	3	12,162,371	0.082	0.116	**4.25E-07***	**1.23E-06***	2.49E-05	2.07E-04	5.91E-03	1.38E-03	0.595
SNP8	3	12,163,116	0.01	0.024	**2.23E-07***	**2.15E-06**	6.59E-05	3.09E-05	0.0976	0.0355	0.464
SNP22	4	5,556,326	0.02	0.078	**1.06E-05**	1.48E-04	3.30E-04	3.24E-04	0.0578	0.0471	0.119
SNP9	4	8,843,014	0.014	0.053	2.12E-05	**1.64E-06***	1.05E-04	4.65E-05	0.107	0.0624	0.226
SNP23	4	9,533,814	6.03E-04	4.32E-03	1.42E-05	**4.17E-06**	4.76E-05	8.61E-06	0.234	0.171	0.24
SNP24	4	13,491,707	0.049	0.138	**5.36E-06**	1.59E-04	2.19E-04	6.61E-04	0.0193	0.0135	0.101
SNP10	5	4,690,632	0.015	0.034	**2.76E-06***	2.38E-05	1.86E-04	3.38E-04	0.0305	0.0112	0.299

Table 3. Continued

SNP11	5	10,723,903	**2.07E-08***	6.21E-07	8.24E-04	0.027	9.93E-05	3.84E-05	0.122	0.117	0.261
SNP12	5	15,976,193	0.032	0.108	**7.93E-07***	6.21E-05	4.62E-05	8.58E-04	2.99E-03	1.99E-03	0.0702
SNP25	5	17,435,459	0.029	0.068	**4.36E-06**	1.02E-04	4.18E-04	6.30E-04	0.0399	0.0159	0.281
SNP13	5	26,628,368	0.034	0.053	**1.89E-06***	5.66E-05	1.94E-04	2.57E-04	0.0401	0.0116	0.668
SNP14	5	26,647,798	3.12E-03	6.63E-03	**1.15E-06***	3.15E-05	3.18E-04	1.83E-04	0.0949	0.0316	0.585

Table 4. The 37 SNPs significant for time of maximum germination rate in at least one of the nine MTMM models. Raw p-values are shown for each SNP in each model. p-values in bold are significant at the 0.1 level after a B-H FDR correction; those with an asterisk are significant at the 0.05 level after a BH-FDR correction. SNPs in bold are shared with F-SNPs significant at the 0.1 level.

SNP	Chr	Position (bp)	Dark/Low	Dark/High	Full-Light/Low	Full-Light/High	Full	Common	GxE	Light	Nutrient
SNP7	1	2,757,164	0.010	0.052	4.66E-05	**3.35E-06**	2.51E-05	2.86E-04	0.004	0.011	0.069
SNP8	1	2,763,016	0.191	0.112	8.86E-05	**7.26E-06**	4.43E-05	4.40E-04	0.005	0.002	0.607
SNP1	1	2,765,047	0.021	0.010	**4.17E-08***	**6.62E-09***	**1.76E-07***	2.38E-06	0.002	0.001	0.443
SNP9	1	2,770,350	0.009	0.003	**5.56E-06**	1.03E-04	1.76E-04	5.04E-05	0.171	0.274	0.161
SNP10	1	4,058,155	0.013	0.009	**4.68E-06**	0.013	0.001	0.003	0.024	0.395	0.012
SNP11	1	10,419,017	3.63E-05	0.001	1.17E-04	**3.63E-06**	1.50E-04	2.84E-05	0.249	0.344	0.206
SNP12	1	18,526,664	**1.28E-06**	4.89E-05	0.002	2.38E-04	1.50E-04	2.70E-05	0.262	0.247	0.202
SNP2	1	21,713,582	0.025	0.017	**1.53E-07***	**1.03E-05**	1.63E-05	9.06E-05	0.008	0.011	0.133
SNP13	1	21,916,027	2.13E-04	0.001	3.60E-04	**3.10E-06**	2.85E-04	2.00E-05	0.680	0.415	0.817
SNP14	1	23,494,937	**1.77E-06**	5.49E-05	0.002	7.55E-04	1.13E-04	6.22E-05	0.088	0.247	0.047
SNP15	1	27,668,561	**1.43E-06**	2.19E-05	**1.55E-06**	**1.86E-06**	5.32E-06	6.94E-07	0.267	0.962	0.108
SNP16	1	29,368,367	0.009	0.005	0.002	**9.67E-06**	2.74E-04	4.01E-05	0.337	0.278	0.267
SNP17	2	6,351,897	0.017	0.032	1.40E-04	**3.84E-06**	2.93E-04	3.99E-04	0.042	0.012	0.779
SNP18	2	8,960,447	2.37E-05	3.81E-05	7.38E-05	**4.44E-06**	3.24E-05	1.73E-06	0.724	0.943	0.431
SNP19	2	10,297,188	0.028	0.019	**1.27E-06**	7.48E-04	6.29E-04	0.001	0.030	0.054	0.104
SNP20	2	17,620,611	0.149	0.102	**5.33E-06**	**7.99E-06**	6.74E-05	5.09E-04	0.007	0.002	0.980
SNP21	3	4,786,505	0.084	0.024	**5.79E-06**	2.05E-04	8.96E-04	7.93E-04	0.072	0.024	0.859
SNP22	3	5,586,649	0.016	0.011	**9.14E-06**	3.90E-04	0.002	0.002	0.097	0.046	0.552
SNP23	3	12,162,371	0.017	0.013	**5.28E-06**	2.27E-04	7.42E-04	5.80E-04	0.079	0.071	0.249
SNP3	3	12,163,116	0.003	0.002	**2.36E-07***	2.14E-05	1.05E-04	3.11E-05	0.157	0.150	0.263

Table 4. Continued

SNP24	3	13,530,762	0.364	0.211	0.004	**1.02E-05**	8.30E-05	0.001	0.004	0.003	0.063
SNP25	3	17,718,905	0.025	0.020	0.001	**3.54E-06**	2.54E-04	2.34E-04	0.059	0.028	0.258
SNP4	3	21,818,882	0.001	0.007	0.002	**4.30E-08***	7.48E-05	1.75E-05	0.190	0.088	0.414
SNP26	4	7,287,800	0.596	0.379	1.45E-05	**7.79E-06**	8.51E-06	0.002	2.21E-04	4.20E-05	0.523
SNP5	4	7,657,583	0.066	0.187	7.61E-05	**2.77E-07***	1.97E-05	4.04E-04	0.002	5.92E-04	0.956
SNP27	4	8,841,131	0.076	0.016	**2.38E-06**	3.91E-05	1.23E-04	3.84E-04	0.018	0.005	0.890
SNP28	4	8,843,014	0.006	0.005	1.24E-05	**1.01E-05**	1.05E-04	3.85E-05	0.128	0.098	0.326
SNP29	4	8,843,150	0.198	0.026	**5.17E-06**	1.92E-04	3.15E-04	3.51E-04	0.051	0.018	0.720
SNP30	4	12,776,709	0.009	0.048	0.001	**2.45E-06**	3.22E-04	1.75E-04	0.100	0.072	0.316
SNP31	5	4,690,632	0.011	0.002	**6.20E-06**	0.001	0.003	0.001	0.303	0.207	0.453
SNP6	5	10,723,903	**1.07E-07***	7.86E-06	0.006	0.008	5.83E-05	1.90E-04	0.016	0.022	0.045
SNP32	5	16,425,024	0.008	0.001	**3.13E-06**	0.001	0.001	3.54E-04	0.161	0.220	0.186
SNP33	5	22,442,725	0.001	0.005	7.81E-05	**7.61E-06**	2.94E-04	7.22E-05	0.209	0.197	0.283
SNP34	5	26,393,336	0.002	0.004	9.82E-05	**3.14E-06**	2.21E-04	5.27E-05	0.208	0.088	0.771
SNP35	5	26,499,148	0.056	0.047	**2.52E-06**	0.002	0.001	0.003	0.018	0.058	0.056
SNP36	5	26,627,873	0.005	0.008	9.43E-05	**3.71E-06**	3.05E-04	7.91E-05	0.200	0.083	0.767
SNP37	5	26,628,368	0.051	0.012	**3.05E-06**	2.20E-05	1.36E-04	7.86E-05	0.086	0.033	0.725

Table 5. Distances, in base pairs upstream (u) and downstream (d), of each significant SNP from the start or stop site of the genes within +- 10 kb of the SNP (TAIR9). When the number of bp is 0, the SNP was within the transcript of a gene. When multiple splice variants with different start/stop sites exist, the main splice variant was used. Genes in bold are linked to more than one SNP.

SNP	Gene	bp	Gene	bp	Gene	bp	Gene	bp
SNP1	AT1G08640	5886 u	AT1G08660	0[a]	AT1G08650	3461 d		
	AT1G08680	5426 u	AT1G08670	3182 d				
SNP2	**AT1G08660**	5253 u	**AT1G08680**	0[b]	**AT1G08700**	4768 d	**AT1G08720**	8983 d
	AT1G08670	3474 u	AT1G08695	4009 d	**AT1G08710**	6586 d		
SNP3	AT1G08730	9613 u	**AT1G08720**	3680 u	**AT1G08695**	867 u	**AT1G08680**	1383 d
	AT1G08670	8777 u	**AT1G08710**	1283 u	**AT1G08700**	0[a]		
SNP4	AT1G59660	8666 u	AT1G59630	8072 u	AT1G59640	4826 u	AT1G59650	3718 u
SNP5	AT1G73610	9816 u	AT1G73600	2065 u	AT1G73603	2065 u		
	AT1G73607	5727 u	AT1G73602	2065 u	AT1G73590	5383 d		
SNP6	AT2G24200	7633 u	AT2G24210	0[a]	AT2G24205	5426 d		
	AT2G24220	2928 u	AT2G24230	4629 d				
SNP7	**AT3G30580**	5396 d						
SNP8	**AT3G30580**	4651 d						
SNP9	AT4G15475	2861 u	AT4G15470	553 u	AT4G15440	4338 d	AT4G15490	9684 d
	AT4G15460	1148 u	AT4G15450	1738 d	AT4G15480	5787 d		
SNP10	AT5G14520	6228 u	AT5G14560	3332 u	AT5G14550	374 d		
	AT5G14565	3994 u	AT5G14540	958 u	AT5G14570	4452 d		
	AT5G14530	3723 u	AT5G14545	390 u	AT5G14580	6608 d		
SNP11	AT5G28680	1890 u	AT5G28690	0[a]				
SNP12	AT5G39865	9020 u	AT5G39890	0[c]	AT5G39880	3956 d		
	AT5G39870	6147 u	AT5G39900	354 d	AT5G39920	5980 d		

Table 5. Continued

	AT5G39895	94 u	AT5G39910	3219 d	AT5G39930	8832 d		
SNP13	**AT5G66710**	8241 u	AT5G66680	7701 u	AT5G66700	5952 u	AT5G66690	1575 d
SNP14	AT5G66720	6982 u	AT5G66730	3396 u	AT5G66740	0[a]	**AT5G66710**	9004 d
	AT5G66760	5860 u	AT5G66750	1153 u	AT5G66755	5472 d		

[a] SNP lies within an exon
[b] SNP lies within an intron
[c] SNP lies in the 3'UTR

Figure 1. Locations of the 100 natural accessions used in this study.

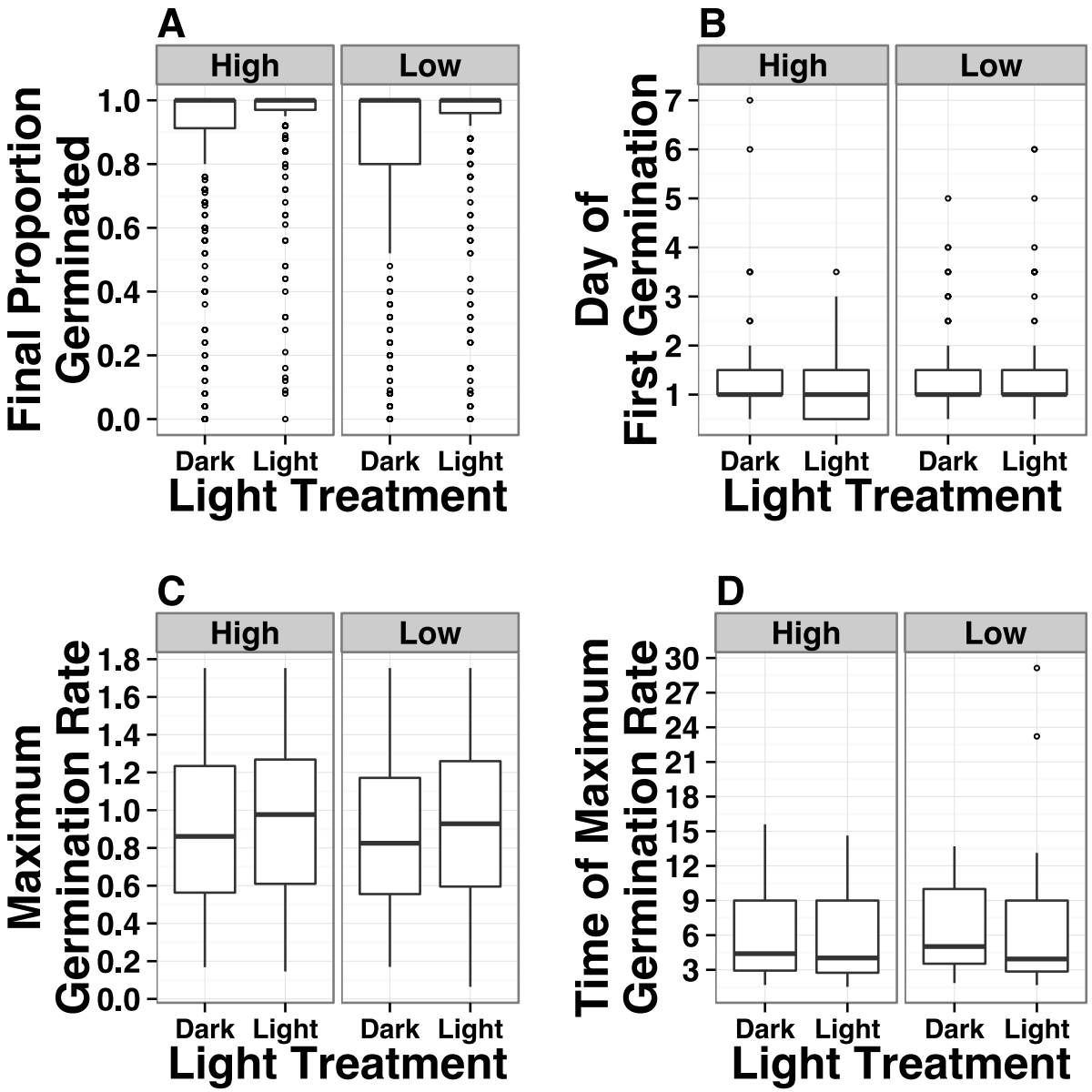

Figure 2. Box-and-whisker plots of **A.** Final, **B**. First, **C.** maximum germination rate, and **D.** time of maximum germination rate under each of the four factorial environments. In each figure, the left panel is low nutrients and either the Full-light or Dark treatment, and the right panel shows values under high nutrients and either the Full-light or Dark treatment.

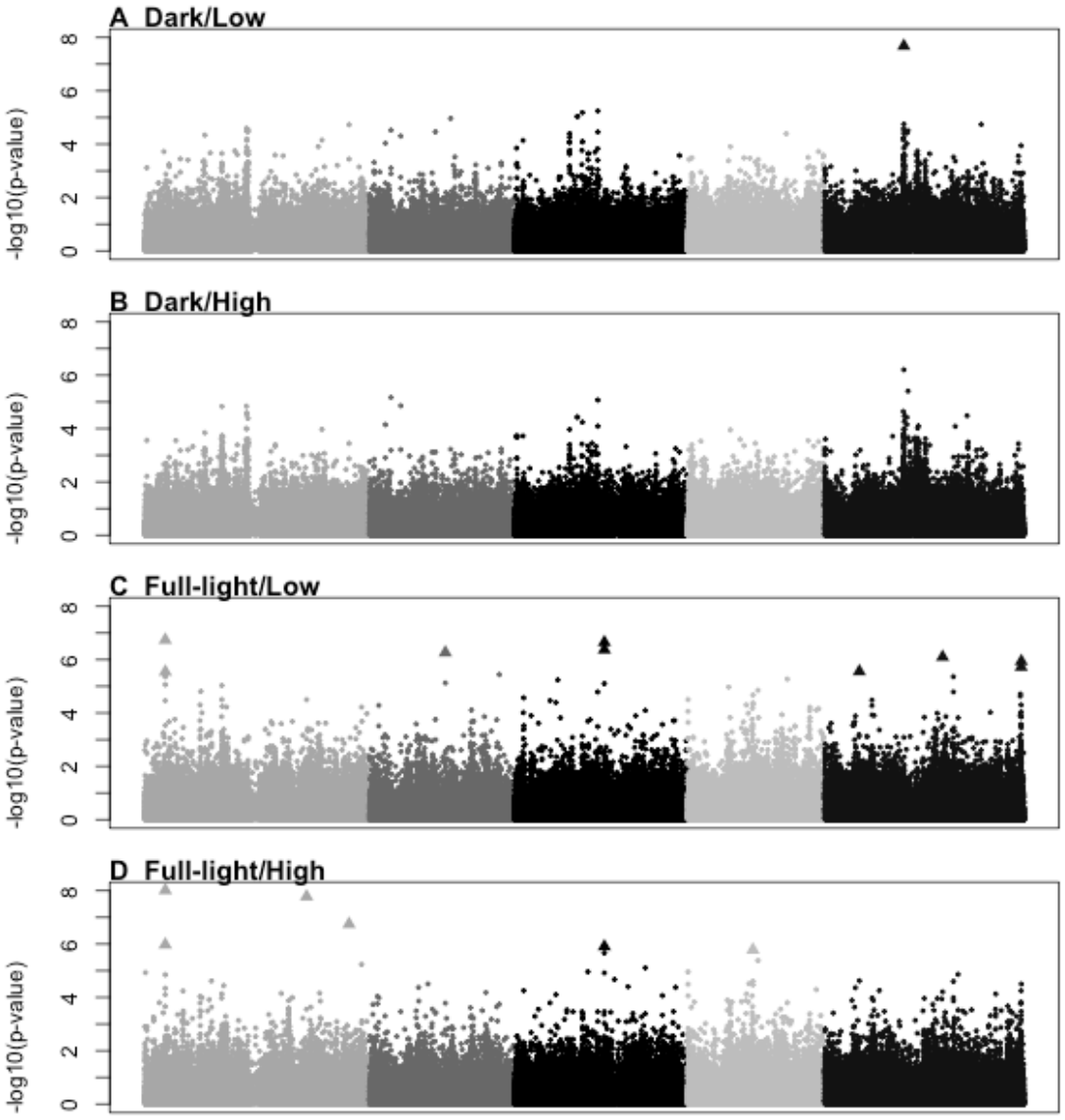

Figure 3. Manhattan plots of uncorrected p-values for the four environment models for Final. Triangles represent SNPs significant at the 0.05 level post-Benjamini-Hochberg correction. **A.** Dark/Low, **B.** Dark/High, **C.** Full-light/Low, and **D.** Full-light/High.

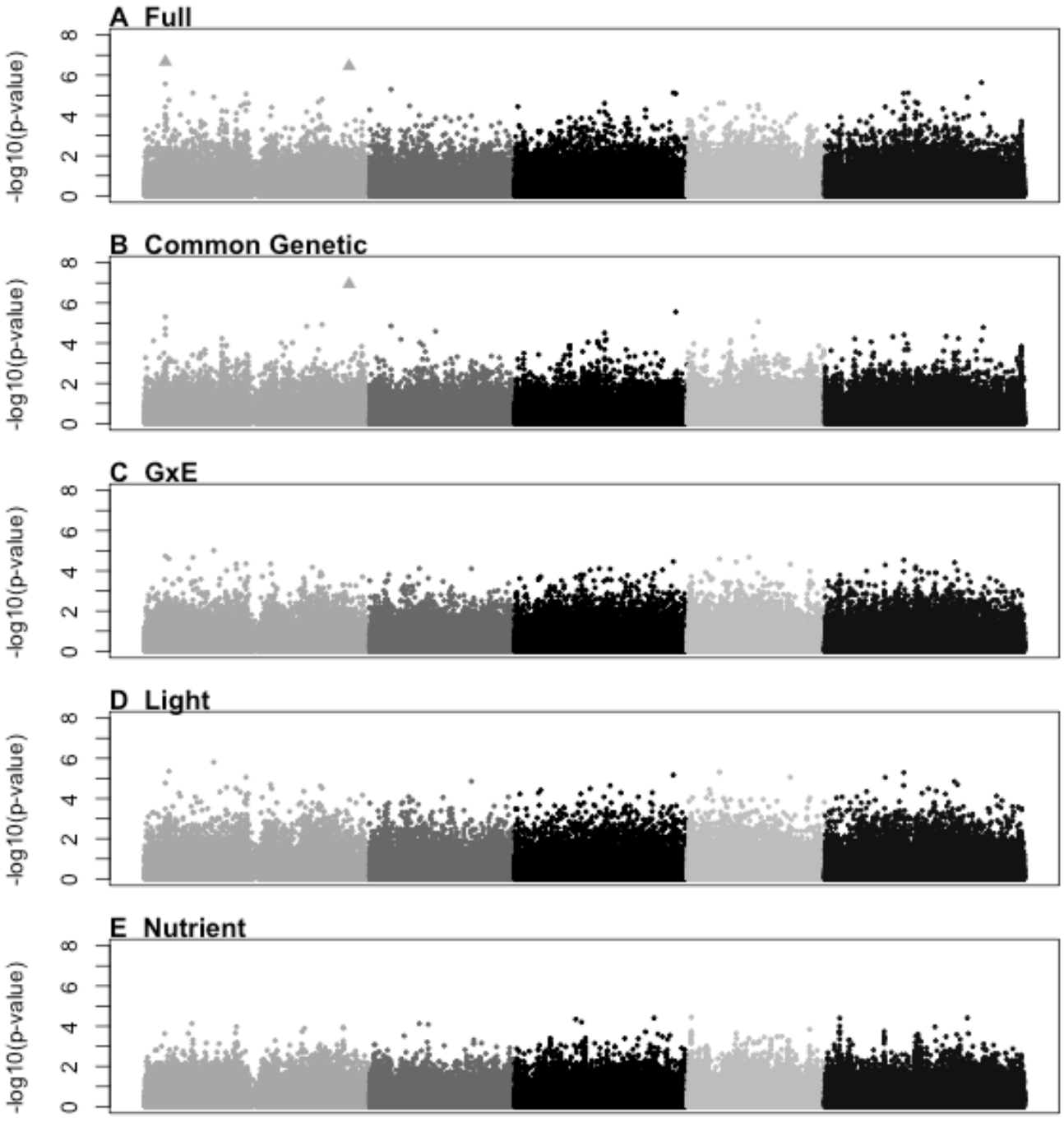

Figure 4. Manhattan plots of uncorrected p-values for Final in the **A.** Full, **B.** Common, **C.** GxE, and **D.** Light and **E.** Nutrient interaction models. Triangles represent SNPs significant at the 0.05 level post-Benjamini-Hochberg correction.

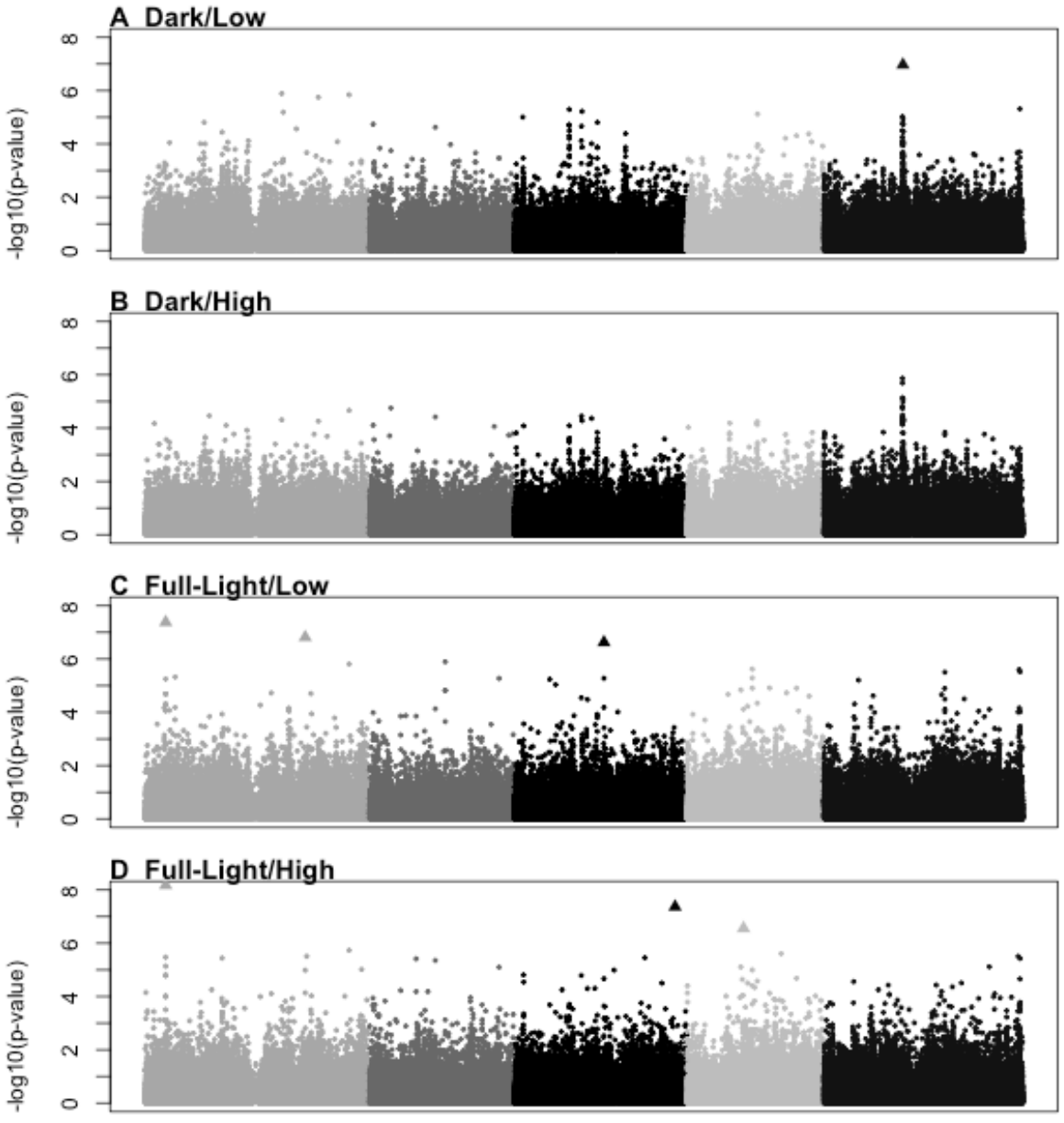

Figure 5. Manhattan plots of uncorrected p-values for the four environment models for time of maximum germination rate. Triangles represent SNPs significant at the 0.05 level post-Benjamini-Hochberg correction. **A.** Dark/Low, **B.** Dark/High, **C.** Full-light/Low, and **D.** Full-light/High.

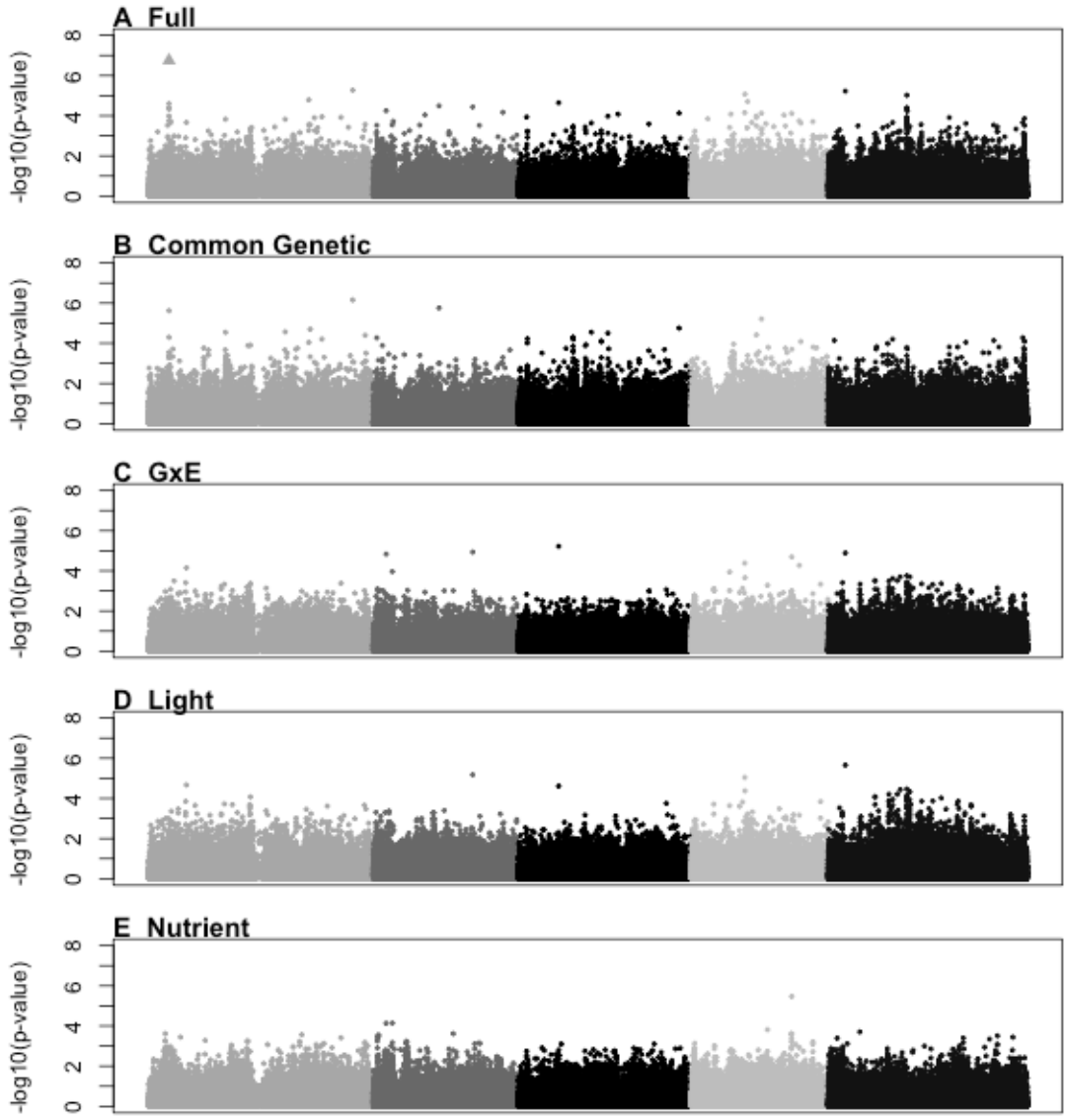

Figure 6. Manhattan plots of uncorrected p-values for the four environment models for time of maximum germination rate. Triangles represent SNPs significant at the 0.05 level post-Benjamini-Hochberg correction. **A.** Dark/Low, **B.** Dark/High, **C.** Full-light/Low, and **D.** Full-light/High.

Chapter 2: Germination Traits in *Sinapis arvensis* (Brassicaceae) from Agricultural and Non-agricultural Habitats in the Bitterroot Valley Show Signs of Local Adaptation.

INTRODUCTION

When environmental conditions differ enough between habitats to affect fitness, local adaptation and population differentiation can occur. Populations with highly restricted gene flow—generally ones at sufficient distances—can undergo local adaptation, but so can populations that are able to readily interbreed (reviewed in Kawecki & Ebert 2004). In the latter case, different phenotypes may be maintained despite gene flow between populations when selective forces act strongly between different habitats.

Agricultural fields and nearby non-agricultural habitats are prime places to study the phenotypic results of selection on plants in adjacent environments. Agricultural fields are frequently monocultures, experience disturbance (plowing and tilling) at regular intervals, and often have shortened growing seasons for annual crops (last date of plowing to harvest). Frequently, these fields also receive nutritional supplements and pest control. Surrounding non-agricultural lands (ditches, roadsides, field margins, etc.) often receive irregular and unpredictable disturbances, have longer growing seasons (date of last frost to date of first frost) and do not receive regular nutritional supplements. Therefore, species of weedy ephemerals occupy agricultural and non-agricultural habitats often experience very different selective regimens. Inside and outside fields, weedy ephemerals have been shown to differ in flowering time, germination timing,

morphology, mating systems, and other traits (Baker 1974; Clements *et al*. 2004), and, in most cases, these traits have a genetic basis (Clements *et al*. 2004; Weinig 2005).

Germination is a major life-history event for weedy ephemerals and usually responsive to external cues (Donohue 2005). The fitness of weedy annuals is strongly influenced by when germination occurs (Steber *et al*. 1998), making germination traits strong targets for selection and therefore phenotypic differentiation. Seeds from different habitats can have different sensitivities or responses to environmental conditions, including neighboring plants and photoperiod, which are often genetically based (Russell *et al*. 2000). As expected, differences in germination traits in agricultural and non-agricultural collections of weedy ephemerals have been observed in several species (Weinig 2005; Adler *et al*. 1993; Baker 1974). In general, seeds from agricultural habitats germinate later than those from non-agricultural and may be more strongly influenced by light quality and nutrient levels than seeds from non-agricultural habitats (Clements *et al*. 2004).

Genetic maternal effects can also have strong effects on germination timing and cueing (Donohue 2009). These maternal effects could affect gene flow via seed dispersal if seeds with a maternal parent from one habitat enter a non-maternal habitat and respond in a maladaptive manner. At the same time, these effects could promote introgression via pollination if the maternal tissues have a stronger effect on germination than the embryonic tissues. In this case, pollen from a different habitat could fertilize a plant and the F_1 offspring could be successful and produce offspring themselves.

Sinapis arvensis (Brassicaceae) is a weedy annual, an obligate out-crosser, and grows at many agricultural and non-agricultural sites in the Bitterroot Valley of Western Montana. Native to Eurasia, *S. arvensis* most likely arrived in the valley as a grain-seed contaminant a little more than a century ago (Fiege 2005). There were most likely multiple introduction events, and the initial *S. arvensis* seeds probably came from agricultural habitats. Plants in agricultural and non-agricultural habitats have overlapping flowering times, and bees have been observed pollinating between habitats (personal observation). These observations suggest that there is a high probability of gene flow via pollen. Gene flow by seeds may also occur, given the close proximity of many agricultural and non-agricultural sites and the possible anthropogenic movement of seeds on machinery and clothing.

A previous study on another weedy mustard introduced to the Bitterroot Valley, *Brassica rapa*, examined the effects of light quality and nutrient levels on the final germination proportions of seeds from agricultural and non-agricultural habitats Adler *et al*. 1993). They found that germination was more strongly suppressed by far-red light for agricultural than non-agricultural seeds, and that a lack of nutrients led to lower germination in agricultural than non-agricultural seeds. It is an open question whether *S. arvensis*, a close relative of *B. rapa* and found in the same habitats in the Bitterroot, has evolved similar differences in the germination responses of seeds from agricultural and non-agricultural sites.

Like Adler *et al*. (1993) and others (Weinig 2005; Weinig 2010), I was interested in how total germination may differ between seeds of *S. arvensis* from agricultural and

non-agricultural habitats and what effects light quality and nutrient availability have on the seed types. In addition, I was interested in assessing whether germination dynamics (maximum germination rate and time of maximum germination rate) also differed under these conditions. I tested five hypotheses: (1) seeds from non-agricultural sites do not vary significantly in their timing of germination and total proportion of seeds when germinated under similar conditions; (2) non-agricultural seeds have reduced total germination and later germination under dark and far-red light conditions and low nutrient conditions relative to full-light and high nutrient conditions, respectively; (3) agricultural seeds germinate later than non-agricultural seeds under similar light and nutrient conditions; (4) agricultural seeds' total germination and germination timing are lower and slower, respectively, than non-agricultural ones under low nutrient levels and far-red light; (5) seeds from between-habitat crosses will respond more like the maternal parent. I also discuss how my results for *S. arvensis* differ from those for *B. rapa*.

MATERIALS AND METHODS

Sinapis arvensis seed from four sites in the Bitterroot Valley, MT, was collected in August 2011, as part of a larger collection effort. Collections were made at sites designated as agricultural (Ag) or non-agricultural (Non-ag). An Ag site was defined as a cultivated field in which a crop was currently planted. A Non-ag site was defined as anything other than an Ag site, and consisted of roadsides, irrigation ditches, and areas immediately adjacent to Ag sites. Although seven Ag sites were identified in the Bitterroot Valley, seed was collected successfully from only one Ag site due to mowing

before seed set or fruits not reaching maturity in time for collection. Collection names were designated by a number followed by an N for seed collected from a Non-ag site or an A for seed collected from an Ag site. Seeds from a total of five collections (four Non-ag: 5N, 8N, 9N, and 22N; one Ag: 8A) were used in this study (Table 6, Figure 7). Collection 8N was located along the edge of the field in which collection 8A was growing. I refer to 8A and 8N as "paired" collections because of their close proximity. All field-collected seed was from fully mature fruits. When fruits were fully dry, the seeds were removed and stored in paper envelopes in the dark at room temperature for six months.

Seed production for the experiments

Field-collected seeds were dry-stratified at 4ºC in the dark for one month. After stratification, ten 4x4 inch pots were planted with four seeds per pot for each collection. Pots were filled with moistened Sunshine soil mix (SunGro Horticulture) and fertilized with Osmocote (Scott's) at a ratio of 40:1. Pots were randomized and grown in a glass house from March through May 2012 at the University of Texas at Austin without supplemental light. Once seedlings were established, 5 pots per collection were chosen for controlled crosses. *Sinapis arvensis* is an obligate out-crosser (Moodie *et al*. 1997), so flowers were not emasculated before crossing. Flowers from pairs of plants were reciprocally hand-crossed by brushing anthers removed from flowers of one plant against the stigma of a mature flower on the other plant. Reciprocal crosses were performed to account for maternal effects. I made five categories of crosses: (1) within-collection

crosses for all five Non-ag collections (N_i x N_i), (2) all possible combinations of between-collections crosses for the Non-ag collections (N_i x N_j), (3) within agricultural collection crosses (8Ax8A), (4) reciprocal crosses of collections 8A and 8N (8Ax8N, 8Nx8A), and (5) crosses with 8A using the Non-ag collections other than 8N as the female parents, designated Nx8A. 8AxN crosses were performed but produced too little seed for the germination experiment. Most of the 8AxN crosses were performed later in the plants' lives, and date of pollination affected the number of aborted or poorly filled seeds.

Germination experiment

Due to differing numbers of seeds available for the five cross types, two germination experiments were run simultaneously: one that used only seed from Non-ag x Non-ag crosses (the Non-ag experiment) and one that used pure Non-ag, pure Ag, and hybrid crosses (the Ag/Non-ag Experiment). None of the seeds used in the Non-ag experiment were used in the Ag/Non-ag experiment. Both experiments were fully factorial for light and nutrients but used slightly different light treatments. The Non-ag germination experiment had three light treatments: full-spectrum and far-red light and dark. The Ag/Non-ag experiment omitted the dark treatment due to a lack of seed. I chose to eliminate the dark treatment because others have observed a stronger suppressive effect of far-red light on germination than darkness (Adler *et al*. 1993). Nutrient treatments for both experiments were half-strength and quarter-strength Peters Professional 20-20-20 (Scott's; 1.2g/L or 0.6g/L).

Seeds were dry stratified at 4ºC in the dark for one month prior to trial. For each

replicate, ten seeds from a cross type were placed on sterilized, doubled filter paper in a 50 mm petri dish containing 1mL of filter-sterilized half- or quarter-strength fertilizer solution. All prep work was performed under safe-green light. I had four replicates of each treatment combination per cross type for a total of 120 dishes in the Non-ag experiment and 144 dishes in the Ag/Non-ag experiment. Before removal from the safe-green light, replicates of the dark treatment were placed in individual sealed foil wrappers and far-red replicates were placed in sealed green gel wrappers (Roscoluxe Moss green). Replicates were randomized and positioned in a Percival growth chamber set to 12hr light at 10ºC, 12hr dark at 6ºC. Temperatures and light cycles were set based on average photoperiod and high and low temperatures during February and March in the Bitterroot Valley.

Once the seeds were placed in the growth chamber, a photograph of each replicate was taken every 24hrs using a Canon EOS digital camera (10 MP). Once a replicate was observed to have completely germinated, it was removed from the chamber and no additional photos were taken. Seeds were scored for germination (visible emergence of the radicle). I recorded the total number of seeds germinated daily for each replicate as well as the total proportion of seeds germinated per replicate at the end of the experiment (after 14 days).

Seeds that had not germinated after two weeks had their nutrient solution replaced with a 0.1% w/v solution of gibberillic acid (GA_3; Sigma-Aldrich) and incubated at room temperature in the dark for three days to test for seed viability. Seeds that did not germinate under these conditions were considered inviable or in a state of deep

dormancy. I could not determine whether ungerminated seeds died prior to or during the study and made no distinctions. The seeds determined to be dead/highly dormant were removed from the analyses and not used in the calculations of proportions of seeds germinated.

Statistical Analyses

Separate statistical analyses were run for the Ag/Non-ag and Non-ag experiments. For each, I performed analyses to assess differences in the final proportion of seeds germinated for cross types under factorial light and nutrient conditions. I also performed analyses to examine the effects of the same factors on germination dynamics in each experiment. For germination dynamics, I fit a germination curve to each plate. The curve had the form of

$$g\,(t;\, k,\, tmax,\, gmax) = gmax * e^{\left(\frac{k*(t-t\,\max)}{k*(t-tmax)+1}\right)}$$

where g is the proportion of seeds germinated at time t. The model estimates the maximum number of seeds germinated ($gmax$), the maximum rate of germination (k), and the time of the maximum germination rate ($tmax$). Values for the three parameters were estimated using the nls() function in R. For analysis, I was only interested in k and $tmax$ because I was already directly measuring the final proportion of seeds germinated.

For the final proportion of seeds germinated, I used generalized linear models with binomial distributions. In the full models, cross, nutrient level, light quality, and their

interactions were fixed effects. The proportion of seeds germinated per plate after 14 days was the response variable. The analysis was implemented in SAS using proc GLIMMIX (SAS Institute, Inc. 2008). Factors that were not significant at the a = 0.05 level in the full model were excluded and reduced models were run. Once the best, reduced model was determined, I performed contrasts on cross, light, or nutrient if they were significant. For cross, I performed five contrasts designed to address my five hypotheses (Ag/Non-ag experiment). The contrasts compared: (1) the within and between collection crosses of the non-agricultural collections (N_ixN_i vs. the N_ixN_j), (2) the within agricultural collection crosses with all of the Non-ag x Non-ag collection crosses (8Ax8A vs. N_ixN_i and N_ixN_j), (3) the within Ag with the within Non-ag collection crosses (8Ax8A vs. N_ixN_i), (4) the hybrid crosses of the paired collections (8Nx8A vs. 8Ax8N), and (5) the Non-ag x Ag crosses of the paired collections with the all the other Non-ag x Ag collection crosses (8Nx8A vs. Nx8A). To compare cross effects in the Non-ag experiment, I performed a contrast to compare the within and between collection crosses of the non-agricultural collections (N_ixN_i vs. the N_ixN_j). For light, I performed either a contrast comparing the effects of full-light vs. far-red light (Ag/Non-ag experiment) or two contrasts comparing full-light vs. far-red and full-light vs. dark (Non-ag experiment). For nutrient, I performed a contrast of half-strength vs. quarter-strength nutrient levels (both experiments).

The same set of fixed effects (cross, light quality, nutrient levels, and their interactions) was used in models for both germination-dynamics traits. I only used maximum germination rate and time of maximum germination estimates for curves that

had a significantly good fit to the data based upon the p-value of the profile likelihood given by the nls function. I analyzed how the different fixed effects impacted these measures with proc GLM in SAS. As with the models analyzing the proportion of seeds germinated, I first ran full models followed by reduced models that excluded non-significant terms. On those reduced models, I performed the same contrasts used for the final proportion germinated.

RESULTS

End-of-Trial Seed Status

Overall, seeds with Non-ag mothers had very few dead/ deeply dormant seeds. In the Non-ag experiment one seed of 1,196 was dead/highly dormant. In the Ag/Non-ag experiment, two of 1,114 seeds with Non-ag mothers were dead/ deeply dormant. In contrast, 61 of 157 8Ax8N seeds, and 9 of 160 8Ax8A seeds had this condition. χ^2 tests confirmed that the rate of dead/highly dormant seeds with Non-ag mothers was the same between experiments (χ^2_1=0.004, p=0.953), and that the rate of death/deep dormancy in 8Ax8A and 8Ax8N seeds was significantly higher than that in seeds with Non-ag mothers (χ^2_1=44.21, p<0.0001 and χ^2_1=551.57 p<0.0001, respectively). The rate of death/deep dormancy was higher for 8Ax8N seeds than 8Ax8A seeds (χ^2_1=67.32 p<0.0001).

The numbers of seeds that germinated after application of GA followed the same pattern as the numbers of dead/deeply dormant seeds. In the Non-ag and Ag/Non-ag experiments, very few seeds with Non-ag mothers were released from dormancy by GA

(5 and 2 seeds, respectively). χ^2 tests confirmed that the rate of mild dormancy in the two experiments was indistinguishable (χ^2_1=0.45, p=0.50). A larger proportion of seeds germinated after GA addition for the 8Ax8A and 8Ax8N crosses when compared to the Non-ag maternal crosses in the Ag/Non-ag experiment (7 seeds, χ^2_1=20.51, p<0.0001; and 35 seeds, χ^2_1=346.74, p<0.0001, respectively). 8Ax8N seeds had a higher rate of mildly dormant seeds during the experiment compared to 8Ax8A (χ^2_1=39.89, p<0.0001), similar to the pattern of dead/deep dormancy seen for the same cross types. In *post-hoc* tests, the 8Ax8N hybrids had lower maximum germination rates and the maximum rate occurred later than that for the 8Ax8A cross type (k: $F_{1, 105}$=4.14, p=0.045; tmax: $F_{1, 102}$=89.71, p=<0.0001). In the case of the death/deep dormancy results, I cannot be certain whether the effect was due to inviability rather than dormancy. Further tests are required to determine the state of dead/deep dormant seeds. I could not detect any differences in the germination performance of seeds that had pure Non-ag parents and either 8Nx8A or Nx8A seeds.

Final Proportion Germinated

Non-ag experiment

Cross, light quality, and nutrient level did not influence the final proportion of seeds germinated in the Non-ag experiment. Nearly all the seeds germinated for all of the crosses (Figure 8, Table 7). Due to the lack of differences overall, no contrasts were performed for this experiment.

Ag/Non-ag experiment

Light quality and nutrient levels had no effect on the final proportion of seeds germinated; therefore, I ran a reduced model without them. In the reduced model, cross was significant ($F_{8, 135}$=9.71, p <0.0001). Of the five contrasts performed, only the contrast comparing 8Nx8A vs. 8Ax8N was significant (Table 8). A lower proportion of 8Ax8N seeds (mean: 0.67, SE=0.08) than 8Nx8A seeds (mean: 0.99, SE=0.006) germinated.

Germination Dynamics

Non-ag experiment

The average maximum germination rate ranged from 0.79 to 1.06 proportion germinated/day (Table 9). Analysis of the maximum germination rate revealed that light was the only factor with a significant effect (Table 10, Figure 9). Contrasts of maximum germination rate under full-light vs. far-red and full-light vs. dark were both significant (full-light vs. far-red, $F_{1,86}$=5.72, p=0.019; full-light vs. dark $F_{1,86}$=27.52, p<0.0001). The maximum germination rate was slower under full-light (mean: 0.73, SE=0.05) than under dark (mean: 1.17, se=0.06) or far-red conditions (mean: 0.94, SE=0.07)..

The situation for time of maximum germination rate was more complex. The average time of maximum germination rate ranged from 4.64 days to 5.18 days (Table 9). Cross, light, and the light-by-nutrient interaction were all significant (Table 11, Figure 9). Contrasts of time of maximum germination rate under full-light vs. far-red and full-light

vs. dark showed the same pattern as those for maximum germination rate. Seeds under full-light reached their maximum germination rate later than seeds under either far-red or dark conditions (full-light vs. far-red, $F_{1,102}=17.88$, $p<0.0001$; full-light vs. dark $F_{1,102}=43.13$, $p<0.0001$). The inter-collection cross vs. between-collection cross contrast detected no significant difference between these two groups ($F_{1,102}=3.47$, $p=0.07$). However, it appears that the 8Nx8N, 9Nx9N, and NxN crosses reached maximum germination rates around half a day earlier (mean: 4.64 to 4.80 days) than 22Nx22N and 5Nx5N (mean: 5.18 days for each; Table 9, Figure 9).

No contrasts were run for the light-by-nutrient interaction, as I had no expectation for how the different effects would be ranked. The significant light-by-nutrient interaction appeared to be due to an interaction in the dark treatment because half-strength nutrients caused the maximum germination rate to be reached later than under quarter-strength nutrients (Figure 9). This effect was not seen under far-red or full-light conditions, where half- and quarter-strength nutrients did not appear to have any effect on time of maximum germination rate (Figure 9).

Ag/Non-ag experiment

For maximum germination rate, cross was the only significant term and ranged, on average, from 0.28 to 1.01 proportion germinated/day (Table 9, Figure 10). Contrasts showed that, while all seeds with Non-ag mothers did not differ significantly in maximum germination rates, seeds with an Ag mother germinated at slower rates than those with Non-ag mothers (Table 12). A *post-hoc* test of the maximum germination rate

for the 8Ax8A cross vs. the 8Ax8N cross showed that the 8A8N cross had a significantly lower maximum germination rate ($F_{1, 105}$=4.14, p=0.045).

For the time of maximum germination rate, the three-way interaction of cross, light, and nutrient was significant (Table 11, Figure 11). Contrasts of cross effects alone showed the same pattern as those for maximum germination rate. Overall, seeds with Non-ag mothers reached their maximum germination rate at the same time, but those with Ag mothers reached maximum germination rates later than those with Non-ag mothers (Table 12).

Examination of the by-factorial results showed a complex interaction. Except for the case of the 8A8N cross, crosses under full-light reached the maximum germination rate either later than or at nearly the same time as those under far-red light. In the case of the 8A8N cross under half-strength nutrients, replicates reached their maximum germination rate later under full-light than far-red light, but under quarter-strength nutrients, that observation was reversed. In addition, under half strength nutrients and far-red light, all crosses having Non-ag mothers had consistently earlier times to maximum germination rate than crosses with Ag mothers. Under the same nutrient conditions, some of the Non-ag-mother crosses showed slightly delayed time to maximum germination rate under full-light. Under quarter strength nutrients, crosses with Non-ag mothers still had earlier times to maximum germination rate than crosses with Ag mothers, but the times to maximum germination rate varied more among the Non-ag-mother crosses whether under full-light or far-red conditions. For all nutrient and light conditions, the 8Ax8A crosses reached their maximum germination rates later than crosses with Non-ag mothers, and

the 8Ax8N had even later times of maximum germination rate (Table 12). A *post-hoc* test of the 8Ax8A cross vs. the 8Ax8N cross showed that the 8A8N cross reached its maximum germination rate significantly later than the 8Ax8A cross (Table 9). It should be noted that the standard errors around the means for the 8A8N cross were much greater than for any of the other crosses.

DISCUSSION

In this study, I tested whether Ag and Non-ag collections of *S. arvensis* have differences in their germination dynamics and total germination despite likely gene flow between the two habitats. Based on others' work with weedy annuals (Weinig:2005; Weinig 2010; Schmitt & Wulff 1993; Verdu & Traveset 2005; Adler *et al*. 1993), I expected some differences would be maintained in the face of gene flow and that light and nutrients would have different effects on germination.

I found evidence that Ag and Non-ag seeds differed from one another in their overall germination proportions and germination dynamics. Crosses with Ag mothers, no matter what the light or nutrient conditions, reached their maximum germination rate later than crosses with Non-ag mothers, and had lower maximum rates of germination than those crosses. There was also inconsistent evidence that the time of maximum germination rate had differentiated among my Non-ag collections of *S. arvensis* from the Bitterroot Valley, but this was seen in only the Non-ag experiment so the result may have been spurious. These differences did not depend upon light or nutrient conditions.

The four Non-ag collections tested were from a geographic range of approximately 20 km—nearly the full length of the part of the valley used for agriculture—but showed no signs that their total germination was different under any light or nutrient conditions. However, light conditions did affect germination dynamics. Non-ag seeds under full-light had lower maximum germination rates and reached those rates later than seeds growing under dark or far-red light. It is possible that *S. arvensis* seeds under lower light and nutrient conditions have been selected to germinate at a faster rate but the ecological or evolutionary reasons are unclear. Full-light and high nutrients are likely good indicators of a low competition environment, whereas lower light quality and nutrient levels may indicate the presence of competitors. Higher competition environments might favor earlier establishment to maximize competitive advantage for resources. Nonetheless, these results are somewhat perplexing because *S. arvensis* is capable of maintaining a viable seed bank. Depending on the probability of a seed surviving an additional season in a dormant condition, it could be more evolutionarily advantageous to remain dormant than to germinate under conditions perceived to be very competitive. For light and nutrients, it appears this is the strategy that has evolved for *B. rapa*. It is not clear why *S. arvensis* has evolved a different strategy for the same conditions. Perhaps it has evolved different cues for detecting competition and thereby maintaining dormancy.

The lack of differentiation among the Non-ag collections for final proportion of seeds germinated and maximum rate of germination has at least three explanations. First, the conditions in the Non-ag habitats could be similar enough that selection is operating

in the same way at all Non-ag sites. Second, Non-ag sites may be essentially panmictic with relatively weak forces of selection among Non-ag sites being insufficient to bring about differentiation.

The lack of differentiation among Non-ag collections in final proportion of seeds germinated and maximum germination rate and the clear differentiation that has occurred between the Ag collection and the Non-ag collections for all three phenotypes measured provides some support for the hypothesis that differences between the Ag and Non-ag habitats present selective pressures strong enough to promote differentiation. The generality of this difference needs to be confirmed with additional Ag collections. However, additional evidence exists for differentiation between Ag and Non-ag plants in my and others' personal observations in the greenhouse and field. For instance, Ag plants are generally smaller in size, both in the field and greenhouse. Likewise, Ag plants tend to flower for shorter periods in the greenhouse than Non-ag plants, which is consistent with the shorter flowering times they exhibit in the field.

Maternal and Other Asymmetric Effects

Overall, maternal effects appear to have the strongest influence on differences in germination timing and final germination proportions of Ag and Non-ag seeds. All seeds with Ag mothers had higher rates of death/deep dormancy and mild dormancy than seeds with Non-ag mothers. In addition, seeds with Ag mothers germinated later and at lower proportions overall than those with Non-ag mothers.

Although I cannot say how general the Ag seed responses are because I only had one Ag collection for this study, my results show evidence for asymmetrical nuclear or cytoplasmic effects. The 8Ax8N hybrids had higher rates of death/deep dormancy and mild dormancy than the 8Ax8A cross type. This result suggests that asymmetrical reproductive isolation (ARI) could be occurring (Tiffin *et al*. 2001). Despite its likely ancestral Ag habitat, some *S. arvensis* individuals must have been able to escape into and thrive in Non-ag habitats. Though this escape and subsequent evolution of ARI would have to have occurred in under 150 years, ARI does not rely on long divergence times. Also, strong ARI can be caused by a small number of loci Turelli & Moyle 2006), so that neither a fast mutation rate nor period of isolation need not be invoked. I do not propose that full ARI has evolved between individuals in the two habitats—indeed, hybrids with an Ag mother germinated on average at 67%. However, we could be observing ARI, which may or may not reach complete ARI. If this is ARI, selection could be driving its development, suggesting a strong disadvantage for the maternal Ag by Non-ag cross.

In the Ag collection, deep dormancy could be part of a strategy to maintain a seed bank. The greater death/deep dormancy and later and lower germination for Ag x Non-ag seeds could be a result of selection favoring incompatibility between Ag and Non-ag plants. On the other hand, the additional death/deep dormancy and later germination could be maladaptive if most of the dead/deeply dormant seeds are dead or too deeply dormant to germinate at an appropriate time. Inappropriate germination timing of the deeply dormant seeds would put emerging hybrid plants at a competitive disadvantage

relative to appropriately emerging Ag x Ag plants. If this is the case, selection may be acting to decrease the influx of potentially maladapted Non-ag alleles into Ag sites.

It is intriguing that all cross types having Non-ag mothers behaved similarly, showing no effect of hybridization. It is not clear what the underlying causes, proximal or final, might be for this difference with hybrids having Ag mothers. It could be that selection does not operate strongly against gene flow from Ag to Non-ag sites.

Germination Cuing Signals

Light and nutrient conditions did not have any measurable impact on final germination proportion in both the Non-ag and Ag/Non-ag experiments. I had expected significant differences caused by light treatments—particularly between full and far-red light conditions—and their interactions with cross type (Ag or Non-ag), as Adler *et al*. (1993) observed. This is a striking difference between the two species because *B. rapa* is a close relative of *S. arvensis* with a similar life history, and both grow in the same habitats in the Bitterroot Valley, sometimes side-by-side. Even though both species are known to produce large, long-lived seed banks (Fogg 1950), it appears that *S. arvensis* and *B. rapa* have evolved different mechanisms for maintaining their seedbanks and cuing germination. Although differences were observed for light conditions for both maximum germination rate and time of maximum germination rate and for light-by-nutrient conditions for time of maximum germination rate, I have no prior studies to compare these dynamics. Also, these differences in germination dynamics do not cause differences in final germination proportions.

The lack of a response to the two different nutrient levels was less puzzling. Another study with *S. arvensis* found no difference in germination between seeds supplemented with nitrogen versus those without added nitrogen (Luzuriaga *et al*. 2006). Studies in other mustards have shown varying results of added nutrients: in some instances promoting (Adler *et al*. 1993), suppressing (Tungate *et al*. 2002; Susko & Cavers 2008), or having no effect on (Susko & Cavers 2008) germination.

Maintenance of a Seedbank

Given my light and nutrient results, how *S. arvensis* maintains seedbanks in the Bitterroot Valley is an open question, particularly for the Non-ag collections. It is possible that the experimental conditions I used were not appropriate to identify mild or deep dormancy for seeds with Non-ag mothers or the Ag x Ag cross.

Several mechanisms of sensing burial, other than light, are known to operate in some species, including detecting reduced diurnal temperature fluctuation (Ghersa *et al*. 1992), and lack of oxygen and other volatiles (Holm 1972) when buried deeply. Diurnal temperature fluctuation could be a germination cue for *S. arvensis*. Since deep burial and an established plant canopy reduce diurnal temperature fluctuations in the soil, a lack of diurnal fluctuations could signal poor germination conditions (i.e., wrong time of year, high competition) and prevent germination, promoting the maintenance of a seed bank. In other studies performed in the field, *S. arvensis* seeds buried at depths greater than 5 cm did not germinate but remained viable, while seeds buried 5 cm or less germinated at much higher rates (Hails *et al*. 1997; Mennan & Ngouajio 2006). Diurnal temperatures

fluctuated naturally in the experiments, and there was a definite trend for those seeds to germinate in late spring or autumn (Hails *et al*. 1997; Mennan & Ngouajio 2006).

In the case of seeds with an Ag mother, plants may have been selected for a risk adverse germination strategy by producing seeds in at least three categories of dormancy: no dormancy, mild dormancy and deep dormancy. There might be no selection for multiple dormancy levels and alternative cuing mechanisms in Non-ag plants if they are rarely subjected to multiple disturbances in the spring the way Ag plants are, or the conditions of my experiment could have represented ideal conditions for Non-ag germination, causing high rates of germination. Supporting this hypothesis, *S. arvensis* seeds originating from latitudes above 45ºN have been observed to germinate at temperatures approaching 4ºC both in the field and in the lab (Edwards 1980, G. Morrison, personal observation).

Concluding Remarks

Final germination proportion and germination timing differed between Non-agricultural collections and an agricultural collection of *S. arvensis*. It is clear that not all seeds of different weedy ephemeral species accomplish the same ends (persisting between disturbances and cuing germination) by the same means. We still need to investigate the mechanisms by which *S. arvensis* senses a suitable germination environment while maintaining a long-lived seed bank, but at this point, it does not appear to do so via the anticipated light and nutrient cues. The question of how differences in germination phenotypes are maintained between Ag and Non-ag habitats in

the face of gene flow also remains. Further work looking at fitness differences in reciprocally planted Ag and Non-ag environments using multiple agricultural collections, and looking at allelic frequencies will help reveal how selection might be operating and if gene flow is restricted between habitats.

Table 6. GPS-determined latitudes and longitudes of collection sites used in this study.

Population	Latitude	Longitude
5N	46.39ºN	-114.09º
8N	46.33ºN	-114.11º
8A	46.33ºN	-114.11º
9N	46.26ºN	-114.13º
22N	46.47ºN	-114.09º

Table 7. Average total proportion of seeds germinated (standard error) by cross type.

	Final Proportion Germinated	
Cross	**Ag/Non-ag**	**Non-ag-only**
$N_i x N_i$	1.00 (0.003)	0.99 (0.003)
$N_i x N_j$	0.99 (0.006)	1.00 (0.00)
Nx8A	1.00 (0.00)	N/A
8Nx8A	0.99 (0.006)	N/A
8Ax8A	0.94 (0.05)	N/A
8Ax8N	0.67 (0.08)	N/A

Table 8. *A priori* orthogonal contrasts for final proportion germination.

Contrast	$F_{(1, 135)}$ score	p-value
8Ax8A vs. (N_ixN_i and N_ixN_j)	0.00	0.97
8Ax8A vs N_ixN_i	0.00	0.97
8Nx8A vs 8Ax8N	17.54	<0.0001
8Nx8A vs Nx8A	0.00	0.98
N_ixN_i vs. N_ixN_j	0.00	0.98

Table 9. Average time of maximum germination rate (standard error) in days and maximum germination rate (standard error) for each cross type in the Ag/Non-ag and Non-ag experiments.

	Ag/Non-ag		Non-ag	
Cross	**Time of Maximum Germination Rate***	**Maximum Germination Rate***	**Time of Maximum Germination Rate***	**Maximum Germination Rate**
5Nx5N	5.51 (0.17)	0.71 (0.07)	5.18 (0.13)	0.79 (0.06)
8Nx8N	5.25 (0.13)	0.92 (0.10)	5.18 (0.12)	0.91 (0.09)
9Nx9N	5.25 (0.22)	0.81 (0.12)	4.64 (0.15)	1.06 (0.09)
22Nx22N	4.97 (0.16)	0.97 (0.12)	4.80 (0.13)	0.97 (0.11)
$N_i x N_j$	4.91 (0.13)	0.95 (0.10)	4.70 (0.18)	1.03 (0.10)
Nx8A	5.02 (0.07)	1.27 (0.11)	N/A	N/A
8Nx8A	5.13 (0.11)	1.01 (0.05)	N/A	N/A
8Ax8A	6.96 (0.28)	0.54 (0.05)	N/A	N/A
8Ax8N	10.67 (0.81)	0.28 (0.04)	N/A	N/A

* The cross main effect was significant for these phenotypes.

Table 10. Significant effects in the reduced linear models on maximum germination rate in the Ag/Non-ag and Non-ag experiments.

Experiment	Effect	F-score	df	p-value
Ag/Non-ag	Cross	9.53	8, 105	<0.0001
Non-ag	Light	13.76	2, 86	<0.0001

Table 11. Significant effects in the reduced linear models on time to maximum germination rate in the Ag/Non-ag and Non-ag experiments.

Experiment	Effect	F-score	df	p-value
Ag/Non-ag	Cross	40.76	8,102	<0.0001
	Cross*Light*Nutrient	2.70	8,102	0.010
Non-ag	Cross	3.96	4, 102	0.005
	Light	22.23	2, 102	<0.0001
	Light*Nutrient	3.72	2, 102	0.028

Table 12. *A priori* contrasts for each germination timing phenotype.

	Maximum Germination Rate		Time of Maximum Germination Rate	
Contrast	**$F_{(1, 105)}$ score**	**p-value**	**$F_{(1, 102)}$ score**	**p-value**
8Ax8A vs. (N_ixN_i and N_ixN_j)	12.94	0.0005**	27.62	<0.0001**
8Ax8A vs N_ixN_i	10.97	0.0013**	24.38	<0.0001**
8Nx8A vs 8Ax8N	34.07	<0.0001**	193.66	<0.0001**
8Nx8A vs Nx8A	3.89	0.06	0.08	0.77
N_ixN_i vs. N_ixN_j	0.99	0.32	1.24	0.27

* = $p < 0.01$, ** = $p < 0.001$

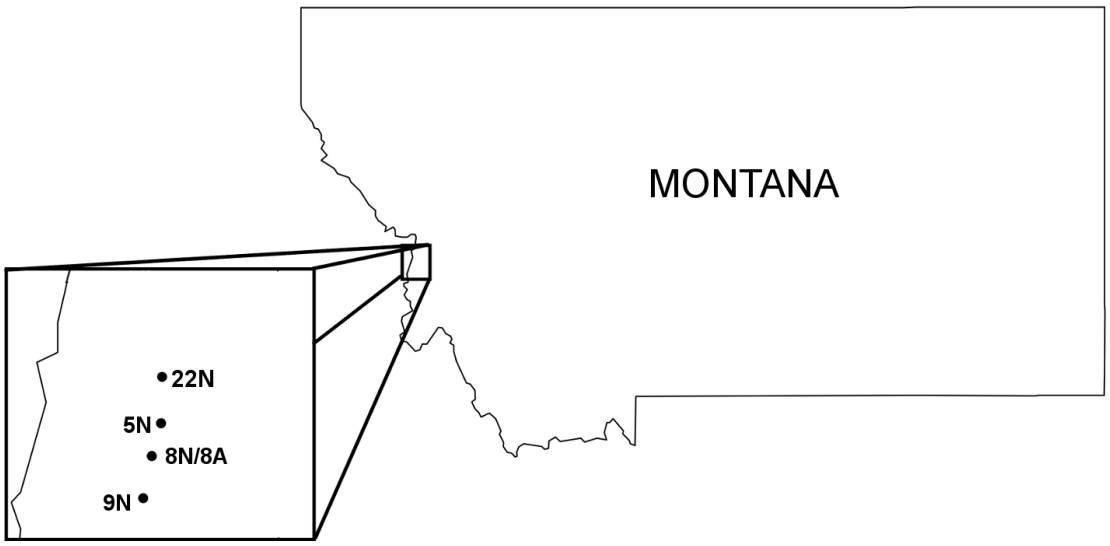

Figure 7. Locations of Collections Used in this Study. Inset shows the Bitterroot Valley of Montana, USA. Points represent the GPS-determined positions of the three unpaired Non-ag collections and one set of paired Ag and Non-ag collections.

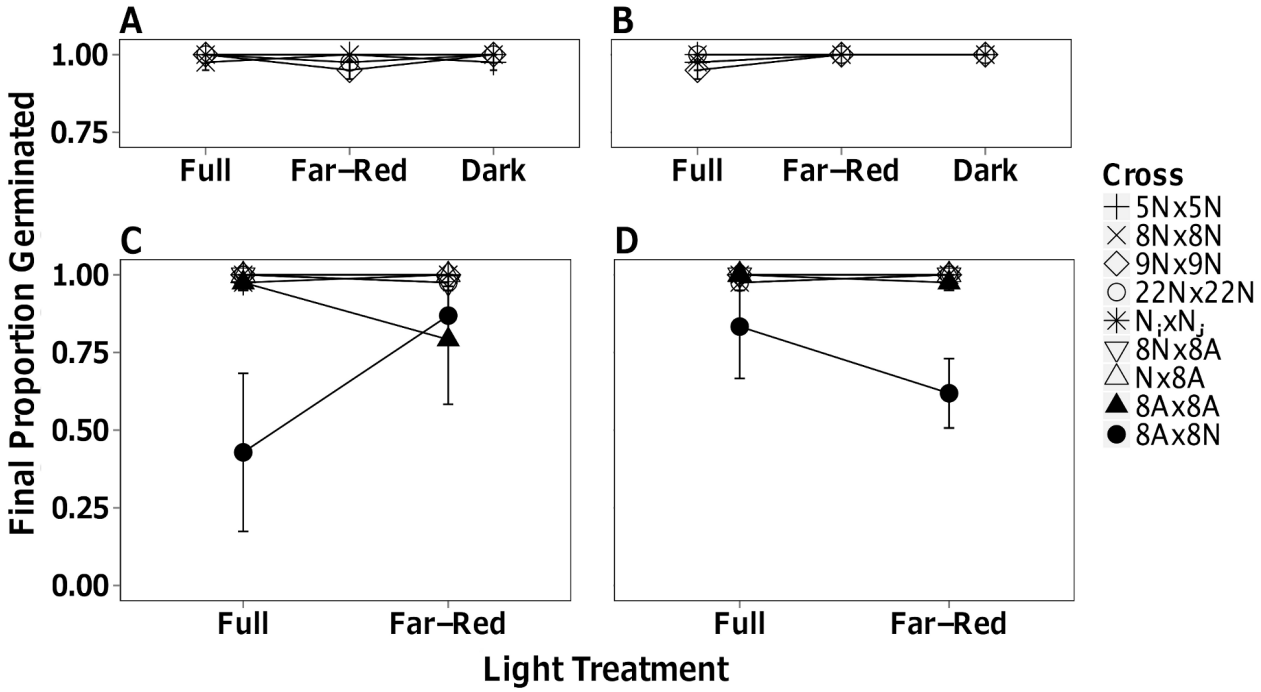

Figure 8. Final Proportion of Seeds Germinated in Both Experiments. The final proportion of seeds germinated for the Non-ag experiment under the half-strength nutrient treatment (**A**) and the quarter-strength nutrient treatment (**B**), and for the Ag/Non-ag experiment under the half-strength nutrient treatment (**C**) and under the quarter-strength nutrient treatment (**D**). Error bars represent standard error.

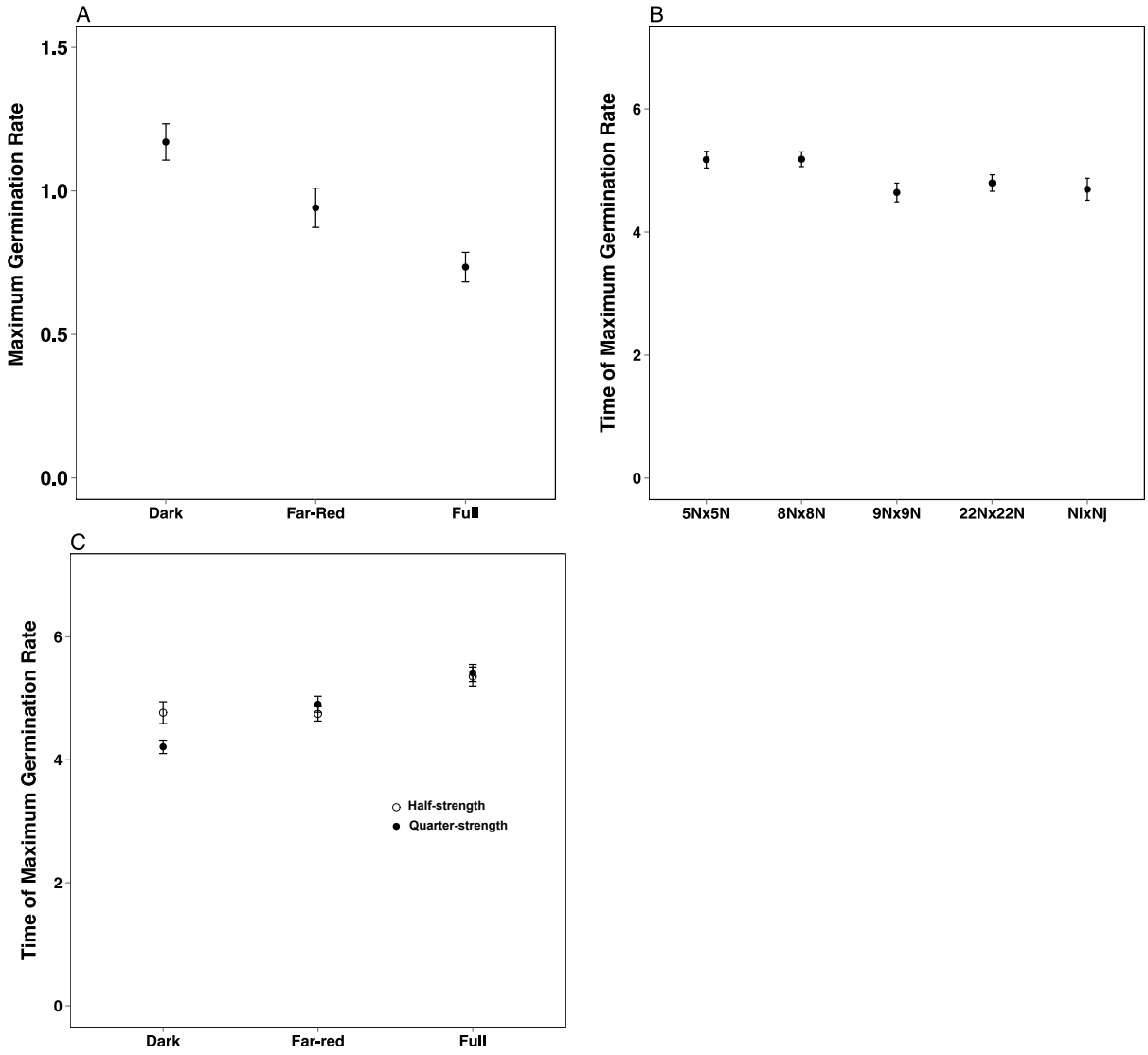

Figure 9. Significant Terms for Maximum Germination Rate and Time of Maximum Germination Rate in the Non-ag experiment. A. Germination rate and light conditions. **B.** Overall time to maximum germination rate for crosses. **C.** Time to maximum germination rate for factorial combinations of light and nutrients. Filled circles correspond to half-strength nutrients and empty circles to quarter-strength. Error bars represent standard error.

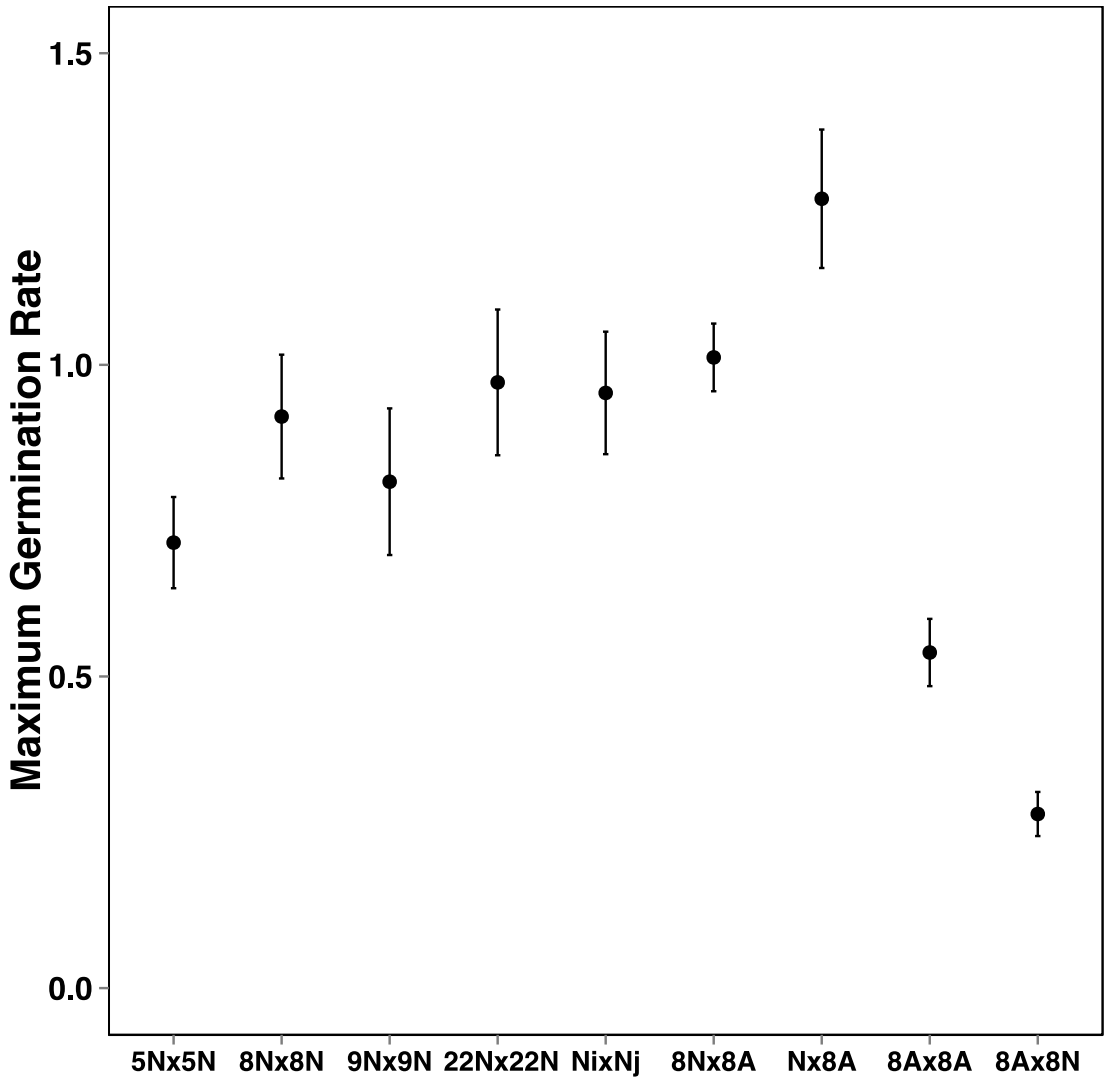

Figure 10. Maximum Germination Rates for Crosses in the Ag/Non-ag experiment. Effect of cross on maximum germination rate. Error bars represent standard error.

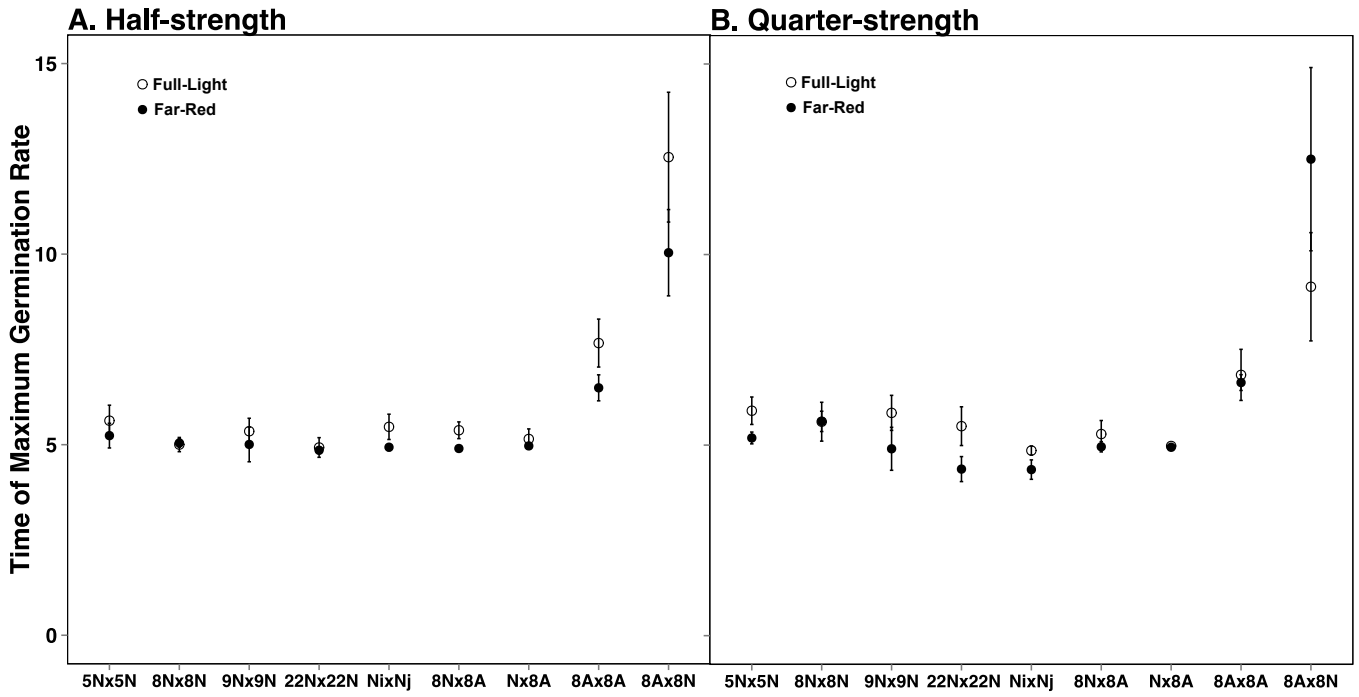

Figure 11. Interactions of Crosses, Light and Nutrients for Time to Maximum Germination Rate in the Ag/Non-ag experiment. A. Half-strength nutrients results for crosses and light conditions. **B.** Quarter-strength nutrient conditions for crosses and light conditions. Empty circles correspond to full-light and filled circles to far-red light. Error bars represent standard error.

Chapter 3: Genetic Analysis of Genome-wide Differences between Agricultural and Non-agricultural Populations of Sinapis arvensis (Brassicaceae) from the Bitterroot Valley, Montana

INTRODUCTION

Differing selective pressures between habitats can lead to local adaptation within a species. If the selection-migration balance favors selection, local adaptation can occur between nearby populations of the same species even when gene flow is occurring (Kawecki & Ebert 2004). While a full assessment of local adaptation should include fitness measurements in a population's home environment and other, non-home environments, genetic analyses of individuals from differing habitats can provide supporting evidence for local adaptation. A large body of studies has found associations between environmental factors and genetic differences between populations (reviewed in Schoville *et al*. 2012; Manel *et al*. 2010; see also Parisod & Chirstin 2008) supporting the possibility of local adaptation. The most common ways to discover loci that may underlie local adaptation are to look for correlations between genetic data and environmental and geographic distance data and to identify markers that show significantly more differentiation between populations (i.e. higher F_{st} than expected). These two methods are often combined to find a correlation between outlier markers and the environment (Freedman *et al*. 2010; Coop *et al*. 2010)

Agricultural landscapes, a mosaic of managed and marginal land, are a ready-made system in which to study local adaptation. Two major habitat types, cultivated agricultural fields (ag) and non-agricultural, disturbed land (non-ag; i.e., ditches, field

margins, and roadsides), exist near one another. Between these two habitats, there is the potential for strong selective differences. Ag fields are highly managed and often feature regular disturbances, nutrient additions, a shortened growing season, and herbicide and pesticide applications. Conversely, non-ag habitats generally have irregular disturbances and nutrient pulses, and a growing season restricted only by the weather in a given year.

Many ecologically weedy species have populations in both agricultural and non-agricultural habitats (Clements *et al*. 2004). Even though species that reproduce sexually have the potential for gene flow between populations in the two habitats (Clements *et al*. 2004), weedy populations often express phenotypes specific to the habitat in which they reside (Baker 1974; Begg *et al*. 2011; Weinig 2005; Clements *et al*. 2004). While some of these phenotypic differences are the result of plasticity (Baker 1974; Geng *et al*. 2006), there is increasing evidence that many traits that are distinct among habitat types have a genetic basis (Clements *et al*. 2004; Bommarco *et al*. 2010; Begg *et al*. 2011).

Besides being an evolutionarily interesting system, an understanding of gene flow and selection in agricultural landscapes is economically important. Weeds are responsible for billions of dollars in crop losses each year and decrease the productivity of cropland (Pimentel *et al*. 2000). Many studies have examined the population dynamics of weeds within agricultural fields. However, the surrounding, non-agricultural populations could play an important role in the spread of adaptive alleles, such as herbicide resistance. Thus, understanding patterns of gene flow and population differentiation between the agricultural and non-agricultural habitats could give insight into weed management.

Europeans began extensively farming the Bitterroot Valley of Montana in the late 1880's when an irrigation system was built (Fiege 2005), which is still in use today. A portion of the valley, stretching about 20 km from Hamilton in the south to Stevensville in the north, is a patchwork of grain fields, pastures, alfalfa, homesteads, ditches, and small towns. Several Eurasian weeds, now present in the valley, most likely arrived as seed contaminants in commercial seed at the end of the 1800's. Several of these weedy species live inside and outside of agricultural fields in the valley.

One of these species is the weedy mustard *Sinapis arvensis*. In this species, various phenological traits including germination characteristics, time to seed-set, and overall plant size and architecture differ between agricultural and non-agricultural collections (G. Morrison, *personal observation*). Another study (Chapter 2) has shown that there are significant, genetically-based differences in germination between agricultural and non-agricultural collections of *S. arvensis* even though *S. arvensis* is an obligate out-crosser (Lefol *et al*. 1996; Warwick *et al*. 2003), and individuals in agricultural and non-agricultural collections overlap in flowering time and pollination (G. Morrison, *personal observation*).

The genetic studies performed to date by others using *S. arvensis* fall into three categories: ones that have examined population structure (Lefol *et al*. 1996; Moodie *et al*. 1997), ones that examined the potential for gene flow from transgenic mustards into *S. arvensis* (Warwick *et al*. 2003; Moyes *et al*. 2002; Bing *et al*. 1996), and ones that examined the spread of herbicide resistance alleles between populations of *S. arvensis*

(Meikle *et al*. 1999; Warwick *et al*. 2005). None of these studies looked for local adaptation in *S. arvensis*.

To assess the possibility of local adaptation in *S. arvensis*, I looked for evidence of genetic differentiation between agricultural and non-agricultural collections. Because of the likelihood of increased gene flow between geographically closer collections, I also wanted to see if there was evidence that adjacent ag and non-ag collections were more similar to each other than more distant ag and non-ag collections or whether all collections simply showed increasing genetic distance with geographical distance. I used restriction-site associated DNA (RAD) sequencing to generate a large number of genetic markers from multiple agricultural and non-agricultural collections to test three hypotheses: (1) ag and non-ag collections show signs of genetic differentiation from each other, (2) isolation-by-distance is occurring, and (3) ag collections are more closely related to neighboring non-ag collections than more distant non-ag collections.

MATERIALS AND METHODS

Collections

I collected leaf tissue from the Bitterroot Valley, Montana, during July and August, 2011, at sites located along an approximately 20km long stretch of the East Side Highway between Hamilton and Stevensville (Figure 12). This area is where the majority of *S. arvensis* collections occurred. The site of each collection was designated as either agricultural (within a planted, maintained field) or non-agricultural (any other location). I obtained GPS coordinates for each collection with a Garmin eTrex Legend H GPS unit

from a collection's approximate center. I collected one or more leaves from up to 5 plants in each collection (Table 13). A total of 124 individuals from 30 collections were collected and used in the genetic analyses. Of the 30 collection sites, seven were ag and 23 were non-ag. Four of the seven ag collections were 'paired' with a non-ag collection. I called collections paired when a non-ag collection was located adjacent to an ag collection. In the absence of strong habitat selection, I expected these non-ag collections to be more closely related to their paired ag collection than another, more distant ag collection. Because *S. arvensis* is not clonal, each plant represented a genetically distinct individual. I placed each individual's tissue into separate paper envelopes, which were then placed in a sealed plastic bag with silica gel. Saturated silica gel was replaced with new silica gel as needed.

Genetic Methods

DNA extraction

I extracted DNA from dried plant tissue using a CTAB protocol (Porebski *et al.* 1997) with the following modifications. I replaced 24:1 chloroform:octanol with 12:1 chloroform:isoamyl alcohol, repeated the cleaning step twice, precipitated the DNA in isopropanol at -20ºC overnight, followed by 10min incubation at room temperature with 800μL of 76% Ethanol/0.01 M ammonia acetate, and did a final elution, after samples were fully dried, into TE buffer. I measured the DNA concentration with a DQ 300 Fluorometer (Hoefer, Inc.) and re-precipitated and dissolved the DNA to a concentration of ~250 μg/μL.

2-b RAD

I digested 4μL of the concentrated samples with BcgI [NEB; 0.6μL 10x Buffer 3, 0.4μL 32mM SAM, 1.0μL BcgI] for 1 hr at 37ºC and inactivated the enzyme for 20 min at 65ºC. I ligated 1/16 reduction adaptors (Table 14) onto the digested fragments using T4 DNA ligase in a total volume of 25μL [NEB; 2.5μL 3' adaptor, 2.5μL 5' adaptor, 11.5μL nuclease-free water, 0.5μL 10mM ATP, 1.0μL 10x T4 buffer, 1.0μL T4 DNA ligase]. Adaptors were created by annealing the 5' or 3' oligo to the anti-oligo (Table 14) in equal proportions [95ºC for 5min, decreasing by 2.5ºC every 30 sec to 25ºC]. Nuclease-free water was then added for a concentration of 4μM. To determine the minimum number of PCR cycles, I did test runs using 4μL of the ligation product, and assessed amplification after 16, 18, 20 and 22 cycles. I determined the correct number of cycles to be either 18 or 20 cycles (applied as appropriate to individual samples), I amplified 20μL of the ligation product using Phusion Taq [NEB; 32.5μL NFW, 12.5μL 10mM dNTPs, 2.0 10μM forward primer, 2.0μL 10μM reverse primer, 5.0μL 2μM P5 adaptor, 5.0μL 2μM barcoded P7 adaptor, 10x HF buffer, 1μL Phusion Taq; NEB] (Table 14). To eliminate adaptor dimers, I ran the PCR product on a 2% agarose gel in 1x TB for 1hr at 70V, stained the gel with ethidium bromide, and selected the band at ~154bp [using NEB Low Mass ladder]. The correct-sized product was then eluted overnight in TE buffer at 4º. I used Illumina HiSeq single end sequencing to sequence the samples.

To prepare the sequences for analysis, I removed any remaining adaptor sequence from the sequenced reads and filtered for high-quality sequences (fastx toolkit v.0.0.13.2; Phred quality $\geq$ 24). I then used a regular expression search to extract only the enzyme

cut site ($NGN_{10}CGAN_6TGCN_{10}CN$). The number of filtered sequence reads per individual used for locus calls ranged from 4667 to 5,871,283 (mean: 814,581.5). I ran the initial filtering and all following analyses on the lonestar server at the Texas Advanced Computing Center (TACC; http://www.tacc.utexas.edu).

Genotyping

To assign reads to markers, I used Stacks (v. 0.99994; Catchen *et al*. 2011) with (1) a minimum stack depth of five reads for a marker, (2) a maximum distance of one between stacks, and (3) the number of mismatches allowed between sample tags set to two. As the *S. arvensis* genome has not been sequenced, distinguishing between error and paralogs, and correctly assigning alleles to the correct markers in the case of the later, is not possible. Therefore, I chose to exclude any markers that were polyallelic (more than two alleles for a marker in an individual) from further analyses. In addition, I used only markers present in at least two individuals and that were not monomorphic.

Population Genetics Measurements

Hardy-Weinburg Equilibrium

I tested if markers were in Hardy-Weinberg Equilibrium, to assess whether they were likely to be neutral, using the HWE.test() function in the R genetics package [ref]. Because 34,321 independent tests were run, all p-values were corrected using a Benjamini-Hochberg FDR correction (BH-FDR; (Benjamini & Hochberg 1995);

p.adjust() function in R). I considered a marker to be in Hardy-Weinberg Equilibrium if its BH-FDR-corrected p-value was ≤ 0.05.

Pair-wise F_{st}

I calculated the genome-wide average pair-wise F_{st} values between each collection pair using the population module in Stacks. Only markers present in at least 20 of the 30 collections (30,523) were used in the calculation. These F_{st} values were used to make a genetic distance matrix (see below).

Distance and Habitat Influences

I looked for correlations between habitat and geographic distance and genotype using two similar methods. First, I used a distance-based redundancy analysis (db-RDA) to look for correlations between the genetic distances between collections and habitat type and/or geographic distance (Legendre & Anderson1999; Borcard & Legendre 2002; Balkenhol *et al*. 2009; Legendre & Fortin 2010; Borcard *et al*. 2011). Geographic distances were calculated from collection GPS coordinates using a haversine function in Python. First, the homogeneity of the dispersion of genetic distance and habitat or latitude was tested using the betadisper() function in R to ensure that the variances were homogenous. db-RDAs were run with the vegan package in R (Oksanen *et al*. 2013). Using the capscale() function in vegan, I tested for correlations between genetic distance and three different classes (habitat, latitude, and habitat-by-latitude interaction). I also used principle of coordinates of neighbor matrices analysis (PCNM; Legendre *et al*.

2012) to find the significant axes of dissimilarity based on geographic distance (as described in Borcard *et al*. 2011).

Similarly, I used a redundancy analysis (RDA; rda() function; Oksanen *et al*. 2013) to examine the counts of each allele in a collection at a single locus for an association with habitat and/or distance. Using the PCNM() function, I first checked for significant positive distance axes (Borcard *et al*. 2011). If the positive PCNMs had a significant effect, I used the forward.sel() function to identify the significant PCNM vectors. The significant vectors were then used as factors in an RDA that used allele counts per collection as the independent variable and habitat type and significant PCNM vectors as independent variables. If the PCNMs had no significant effect, allele counts per collection was the independent variable and habitat type was the dependent variable and the geographic distance p-value was considered 1.0. Again, all p-values were corrected using a BH-FDR correction.

Loci Under Selection

I used BayeScan (v.2.1; Foll & Gaggiotti 2008) to estimate per-marker F_{st} and identify markers with larger or smaller than expected F_{st}-values. I ran BayeScan with the default settings with different prior odds to make sure the prior did not overly influence the results (Foll & Gaggiotti 2008). I designated markers as outliers at a q-value level of 0.05, using the R programs packaged with BayeScan 2.1. Outlier markers fell into two categories: those under diversifying selection (indicated by a positive alpha) and those under purifying or balancing selection (indicated by a negative alpha). I ran BayeScan

comparing all 30 collections and comparing the aggregated ag and aggregated non-ag collections to each other.

RESULTS

Markers

Stacks identified a total of 201,240 markers from the filtered sequence data. Of these markers, 9,614 were polyallelic in at least one individual. 34,321 markers were (1) not monomorphic, (2) present in at least two individuals, and (3) not potential paralogs. The average stack depth of these 34,321 markers was 48.48 reads/marker (range 1-1937; s.d.=103.95). Most markers were in Hardy-Weinberg Equilibrium (32,780, 95.51%) indicating that the vast majority were likely neutral.

F_{st}-based Analyses of Distance and Habitat

Genome-wide pair-wise F_{st} values were calculated using the 30,523 markers present in a least one individual in at least 20 of the 30 collections genotyped. The genome-wide pair-wise F_{st} between collections ranged from 0.035 to 0.28 (mean=0.11, Table 15). A multivariate test of homogeneity of group variances showed that group variances were homogeneous for habitat and latitude (Table 16). Permutation tests of a db-RDA showed that neither habitat nor latitude significantly explained the variation in genetic distances (Table 17).

To see if geographic distance, rather than latitude, had an effect on pair-wise F_{st}, I used a PCNM test to break down the geographic distances between collections into

eigenvectors. The full model comparing genetic distances with all the positive PCNM vectors was not significant ($F_{7, 22}$=1.10, p=0.31), so I could not reject the null hypothesis that the genetic distances of collections are unrelated to their geographic distances.

Per Marker Analyses of Distance and Habitat

Although no overall collection differentiation was seen based on habitat or distance, it is possible that on a marker-by-marker basis some markers had significant associations with those variables. Therefore, I looked for marker-based associations by analyzing the association between the number of each allele for a marker present in a collection and geographic distance or habitat type. With no FDR correction, 2,425 markers had a significant association with habitat and 2,635 markers were significantly associated with geographic distance (via PCNM). Uncorrected, 135 of these markers were significant for both habitat and geographic distance. After a BH-FDR correction, however, no markers were significant for either habitat or geographic distance. Because geographic distance was not significant, I was unable to reject the null hypothesis that ag/non-ag paired collections are no more similar to each other than to more distant collections.

Markers Under Selection

BayeScan identified different numbers of outlier markers and markers putatively under directional selection depending upon the prior odds used (Table 18, Figure 13). However, the markers identified under increasingly larger priors were always a subset of

those previously identified under a lower prior and a relatively stable number of markers were identified when the priors were between 1000 and 7000. I was particularly interested in markers under putative directional selection as they could be important for differentiation between ag and non-ag habitats. At a prior odds of 1000, three markers (M30694, M31594, and M82172) were significant, positive outliers (Table 18, Figure 13). M30694 and M82172 remained significant prior odds through 7000. M82172 was the only marker that was a significant, positive outlier at a prior odds of 10000. The authors of BayeScan suggest a prior odds of 3000 should be appropriate a sample of ~35,000 markers. Therefore, I considered M30694 and M82172 to have strong evidence as markers under selection among the 30 collections.

When I compared the aggregated ag and non-ag collections, no loci showed significant signs of selection, even with a prior odds of 100.

DISCUSSION

The results showed no genetic evidence of selection based on habitat type or of genetic correlation with habitat. While two markers showed evidence for positive selection among all 30 collections, they were not under selection based on habitat. There were also no signs of genetic relatedness varying with distance.

There are many potential reasons why I did not detect selection between habitats First, I might have collected material in the field before selection had had a chance to occur. This is unlikely since I collected plant tissue after plants had already flowered and were setting seed (July 21-August 11, 2011).

Second, it could be that there has not been any genetic differentiation between agricultural and non-agricultural collections because gene flow is overriding selection. This explanation is also not likely because many genetically based phenotypic differences were observed when *S. arvensis* ag and non-ag collections were grown under common garden conditions. Height, branching, flowering time and length of flowering period differed consistently between collections from the two habitats (G. Morrison, *personal observation*). In addition, these common garden differences corresponded with differences observed in numerous collections in the wild (G. Morrison, *personal observation*). Finally, in Chapter 2, I recorded statistically significant differences in germination timing and total proportion of seeds germinated based on habitat for seeds that were generated under common garden conditions. Germination characteristics for reciprocal crosses between a single ag collection and several non-ag collections exhibited consistent differences for direction of cross; the ag collection was different from the non-ag collections; and the non-ag collections were statistically indistinguishable from one another. Alternatively, these seemingly genetically-based effects could be an epigenetic effect (Bergelson & Roux 2010).

A third possibility is that the method for assessing genetic differentiation (RAD markers and F_{st}) between the collections was inadequate. The *S. arvensis* genome is 367 Mbp (Arumuganathan & Earle 1991) and has nine chromosomes. I had average coverage of approximately 3,813 markers per chromosome (34,321 markers/9 chromosomes) with markers spaced ~10,700 bp apart on average. This is relatively fine scale coverage, but might not have been sufficient if the average haplotype block size for *S. arvensis* is

smaller than the distance between markers or if the genomic distribution of RAD markers was clumped rather than random. There is no estimate of the size of an average haplotype block in *S. arvensis*; however, average haplotype blocks could be relatively small as it is an obligate outcrosser, unlike *Arabidopsis thaliana*, which has haplotype blocks of approximately 10 Kb.

With respect to the distribution of RAD markers, when 2b-RAD was used in *Arabidopsis thaliana*, relatively even genome-wide coverage was obtained (Wang *et al*. 2012), but I used a different endonuclease than Wang et al., which may have led to a different distribution of markers throughout the genome. Currently, there is no *S. arvensis* genome published or under construction, so it is not possible to map reads to even a scaffold. A previous attempt to map reads to the *Brassica rapa* genome, a close relative, using bowtie (Langmead *et al*. 2009) was unsuccessful (G. Morrison, *unpublished data*).

With respect to F_{st} as a measure of among collection variation, Arnold *et al*. (2013) found that RADseq methods tend to estimate F_{st} fairly well even though they can underestimate other population genetics measures, such as pi and Tajima's D. If anything, F_{st} is overestimated, due to missing data.

So how might the observed genetically-based phenotypic differences between ag and non-ag collections be explained? Since *S. arvensis* has been in the Bitterroott Valley for less than 120 generations and there are no barriers to gene flow between ag and non-ag collections, lineages may not have had enough time to differentiate broadly. The genetic variation causing the phenotypic variation seen between ag and non-ag collections might be at just a small number of very specific loci. It is possible that

creating a linkage map and performing a QTL study would be fruitful in this situation, particularly if the trait differences that have been observed are due to a small number of large-effect loci.

Alternatively, the process by which *S. arvensis* entered the valley was probably multiple colonization events as a contaminant in different batches of crop seed over many years or even decades. Most likely these colonizers had been selected to do well in agricultural fields, and non-ag types were selected for when ag seeds escaped into non-ag habitat. If widely divergent *S. arvensis* ag genotypes entered the valley, I might not expect to see differentiation along habitat lines as the ag types might still harbor distantly related individuals in different fields. Likewise, the non-ag types, while phenotypically similar, might be genetically varied for those phenotypes due to multiple migrations into non-ag habitat followed by convergent evolution on favored non-ag traits.

While not considered a noxious weed in Montana, *S. arvensis* is noxious in Canada, where it is frequently found in canola and sunflower fields. Because of the clear evidence for gene flow occurring between the ag and non-ag habitats, genes conferring herbicide resistance may more easily spread between fields via non-ag populations. If the populations of *S. arvensis* in Canada are similar to those studied in the Bitterroot Valley, management of *S. arvensis* populations between fields might be a viable way to retard the spread of herbicide resistance genes.

Table 13. Populations, number of individuals collected, habitat type, and location.

Population[a]	No. of Plants	Habitat	Latitude[b]	Longitude
9N	5	Non-ag	46.262	-114.135
13N	5	Non-ag	46.309	-114.122
14N	4	Non-ag	46.324	-114.122
26A	5	Ag	46.327	-114.122
8A	5	Ag	46.335	-114.114
8N	4	Non-ag	46.335	-114.114
6A	5	Ag	46.339	-114.114
7A	1	Ag	46.341	-114.109
7N	4	Non-ag	46.341	-114.109
10N	5	Non-ag	46.346	-114.103
3N	5	Non-ag	46.367	-114.104
4N	5	Non-ag	46.374	-114.104
5A	5	Ag	46.391	-114.09
5N	5	Non-ag	46.391	-114.09
10-92-N1	5	Non-ag	46.396	-114.094
10-92-N2	5	Non-ag	46.4	-114.094
10-92-N3	5	Non-ag	46.4	-114.094
10-92-N4	5	Non-ag	46.4	-114.094
16N	5	Non-ag	46.4	-114.094
16N1	5	Non-ag	46.4	-114.094
12N	5	Non-ag	46.404	-114.094
24A	5	Ag	46.411	-114.097
24N	5	Non-ag	46.411	-114.097
1N	5	Non-ag	46.425	-114.09
1N1	5	Non-ag	46.425	-114.09
2A	5	Ag	46.428	-114.09
20N	5	Non-ag	46.437	-114.094
21N	5	Non-ag	46.458	-114.093
22N	5	Non-ag	46.471	-114.088
19N	4	Non-ag	46.501	-114.072

[a]Populations with the same number and habitat were directly across a road (1,16) or railroad ditch (10-92). Populations with the same number and different habitat were paired—located adjacently.

[b]Populations are ordered by increasing latitude from South to North.

Table 14. Oligonucleotides used for 2b-RAD library preparation.

Adaptors	**Sequence**
3'	CAGACGTGTGCTCTTCCGATCTNG[1]
5'	CTACACGACGCTCTTCCGATCTNG[1]
anti[3]	AGATCGGAAGAGCT[2]

Amplification Primers	**Sequence**
forward	AATGATACGGCGACCACCGA
reverse	CAAGCAGAAGACGGCATACGA

Illumina Adaptors	**Sequence**
P5 (forward)	AATGATACGGCGACCACCGAAAAATACACTCTTT CCCTACACGACGCTCTTCCGATCT
P7 (reverse)	CAAGCAGAAGACGGCATACGAGATXXXXXXGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATC[4]

[1] N is any nucleotide (ATCG)
[2] The 3' T is an inverted T
[3] Annealed to both the 5' and 3' oligonucleotide
[4] XXXXXX is the 6bp barcode used

Table 15. Genome-wide, average, pair-wise F_{st} values as calculated by Stacks.

	1N	1N1	10N	10-92-N1	10-92-N2	10-92-N3	10-92-N4	12N	14N	16N	16N1	19N	2A	20N
1N	0.000													
1N1	0.065	0.000												
10N	0.149	0.146	0.000											
10-92-N1	0.191	0.187	0.257	0.000										
10-92-N2	0.054	0.132	0.074	0.097	0.000									
10-92-N3	0.096	0.150	0.104	0.131	0.076	0.000								
10-92-N4	0.144	0.135	0.064	0.142	0.070	0.121	0.000							
12N	0.065	0.136	0.052	0.129	0.101	0.067	0.059	0.000						
14N	0.117	0.170	0.118	0.115	0.132	0.123	0.103	0.149	0.000					
16N	0.142	0.091	0.122	0.075	0.069	0.205	0.126	0.165	0.106	0.000				
16N1	0.063	0.129	0.082	0.101	0.081	0.074	0.089	0.145	0.080	0.106	0.000			
19N	0.058	0.164	0.079	0.111	0.067	0.065	0.078	0.210	0.101	0.129	0.150	0.000		
2A	0.076	0.085	0.080	0.191	0.096	0.063	0.065	0.139	0.077	0.120	0.055	0.147	0.000	
20N	0.135	0.101	0.119	0.279	0.078	0.055	0.082	0.175	0.081	0.104	0.106	0.119	0.112	0.000
21N	0.055	0.072	0.180	0.119	0.148	0.085	0.086	0.145	0.116	0.159	0.079	0.137	0.135	0.102
22N	0.056	0.072	0.095	0.156	0.132	0.142	0.136	0.134	0.093	0.103	0.061	0.188	0.141	0.063
24A	0.047	0.132	0.091	0.130	0.083	0.113	0.069	0.062	0.098	0.062	0.092	0.114	0.136	0.115
24N	0.053	0.166	0.067	0.126	0.175	0.087	0.099	0.106	0.159	0.065	0.070	0.127	0.132	0.185
26A	0.114	0.073	0.074	0.151	0.108	0.099	0.075	0.153	0.101	0.127	0.086	0.115	0.229	0.079
3N	0.048	0.101	0.130	0.052	0.126	0.118	0.115	0.084	0.109	0.061	0.102	0.100	0.169	0.090
4N	0.057	0.166	0.128	0.202	0.110	0.104	0.144	0.082	0.114	0.076	0.116	0.168	0.130	0.080
5A	0.035	0.083	0.085	0.139	0.097	0.123	0.236	0.067	0.086	0.040	0.139	0.103	0.284	0.085
5N	0.072	0.080	0.095	0.152	0.131	0.081	0.097	0.072	0.109	0.048	0.106	0.113	0.127	0.109
6A	0.130	0.063	0.140	0.178	0.089	0.089	0.116	0.086	0.163	0.089	0.090	0.091	0.088	0.116
7A	0.185	0.062	0.087	0.112	0.133	0.075	0.101	0.126	0.130	0.120	0.096	0.108	0.207	0.120
7N	0.086	0.095	0.096	0.073	0.117	0.095	0.087	0.117	0.124	0.062	0.096	0.227	0.103	0.105
8A	0.085	0.156	0.189	0.084	0.082	0.160	0.141	0.095	0.086	0.055	0.115	0.252	0.199	0.133
8N	0.094	0.075	0.099	0.146	0.105	0.077	0.224	0.079	0.103	0.054	0.153	0.114	0.140	0.121
9N	0.058	0.071	0.134	0.075	0.108	0.088	0.143	0.095	0.140	0.075	0.100	0.099	0.128	0.103
13N	0.147	0.147	0.098	0.093	0.040	0.061	0.126	0.092	0.219	0.092	0.117	0.096	0.089	0.152

Table 15. Continued.

	21N	22N	24A	24N	26A	3N	4N	5A	5N	6A	7A	7N	8A	8N	9N	13N
1N																
1N1																
10N																
10-92-N1																
10-92-N2																
10-92-N3																
10-92-N4																
12N																
14N																
16N																
16N1																
19N																
2A																
20N																
21N	0.000															
22N	0.107	0.000														
24A	0.117	0.089	0.000													
24N	0.105	0.073	0.160	0.000												
26A	0.085	0.123	0.098	0.056	0.000											
3N	0.142	0.172	0.103	0.105	0.068	0.000										
4N	0.078	0.101	0.069	0.114	0.136	0.138	0.000									
5A	0.113	0.094	0.135	0.085	0.081	0.091	0.095	0.000								
5N	0.070	0.121	0.202	0.080	0.080	0.095	0.132	0.113	0.000							
6A	0.107	0.113	0.096	0.057	0.048	0.097	0.192	0.071	0.081	0.000						
7A	0.165	0.115	0.086	0.076	0.111	0.083	0.083	0.104	0.133	0.106	0.000					
7N	0.088	0.185	0.085	0.079	0.128	0.118	0.108	0.070	0.070	0.098	0.093	0.000				
8A	0.146	0.115	0.152	0.154	0.103	0.139	0.097	0.102	0.073	0.083	0.066	0.094	0.000			
8N	0.210	0.137	0.079	0.080	0.077	0.095	0.083	0.125	0.055	0.067	0.086	0.038	0.095	0.000		
9N	0.094	0.121	0.090	0.098	0.095	0.109	0.146	0.114	0.045	0.060	0.058	0.092	0.087	0.080	0.000	
13N	0.086	0.095	0.050	0.089	0.081	0.127	0.091	0.044	0.065	0.073	0.056	0.053	0.056	0.066	0.065	0.000

Table 16. Tests for homogeneity-of-groups dispersion (variances).

Genetic Distance by:	df	F-value	p-value
Habitat	1	0.43	0.52
Latitude	20	2.43	0.09

Table 17. Results of db-RDA testing using 999 permutations.

Model[a]	df	F-score	p-value	% variance explained	adjusted R^2
gd = habitat + latitude + habitat*latitude	3	0.90	0.70	9.4	0.122
gd = habitat + latitude	2	0.86	0.69	6.0	0.071
gd = habitat	1	0.50	0.96	1.8	0.006
gd= latitude	1	1.15	0.32	3.9	0.057

[a]gd=genetic distance, genome-wide average pair-wise F_{st} as calculated by Stacks (Finch-Savage & Leubner-Metzger 2006; Catchen *et al*. 2011).

Table 18. Total number of outliers and outliers with positive values found at different prior odds using BayeScan v2.1 at a q-value $\leq$ 0.05 (in bold). F_{st} Marker columns indicate the F_{st} of the three markers with positive outlier F_{st} values at a prior odds of 1000 or higher (q-value).

Prior Odds	**Total Outliers**	**Positive Outliers**	**F_{st} Marker 30694**	**F_{st} Marker 31594**	**F_{st} Marker 82172**
100	486	33			
500	245	7			
1000	176	3	**0.79 (0.002)**	**0.76 (0.013)**	**0.74 (0.001)**
3000	111	2	**0.77 (0.009)**	0.66 (0.058)	**0.72 (0.007)**
5000	93	2	**0.75 (0.014)**	0.58 (0.087)	**0.70 (0.017)**
7000	83	2	**0.71 (0.031)**	0.52 (0.121)	**0.68 (0.026)**
10000	75	1	0.65 (0.052)	0.47 (0.167)	**0.65 (0.041)**

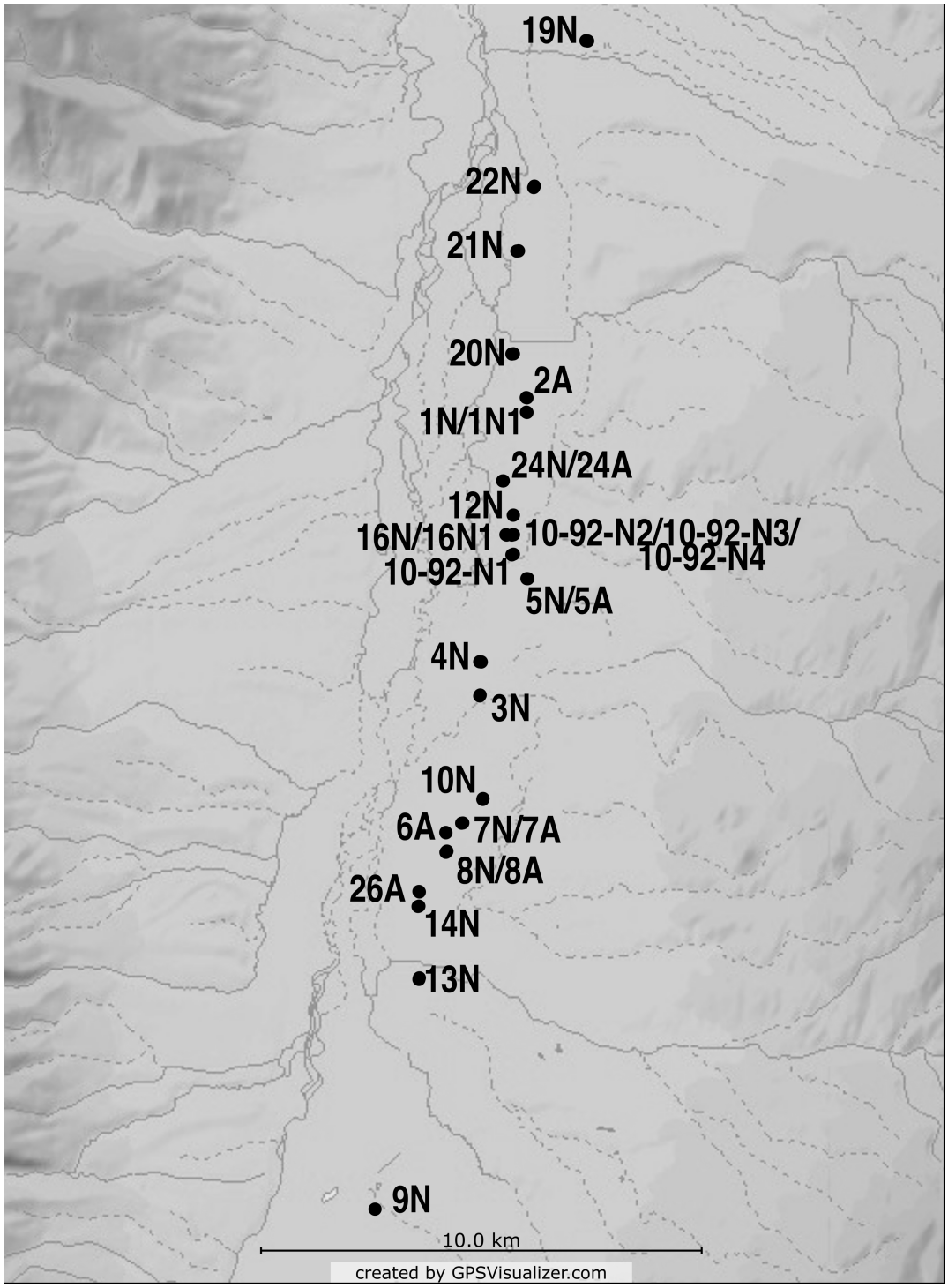

Figure 12. Map of the area sampled in the Bitterroot Valley. Dots represent collection sites with population name(s). North is at the top of the map.

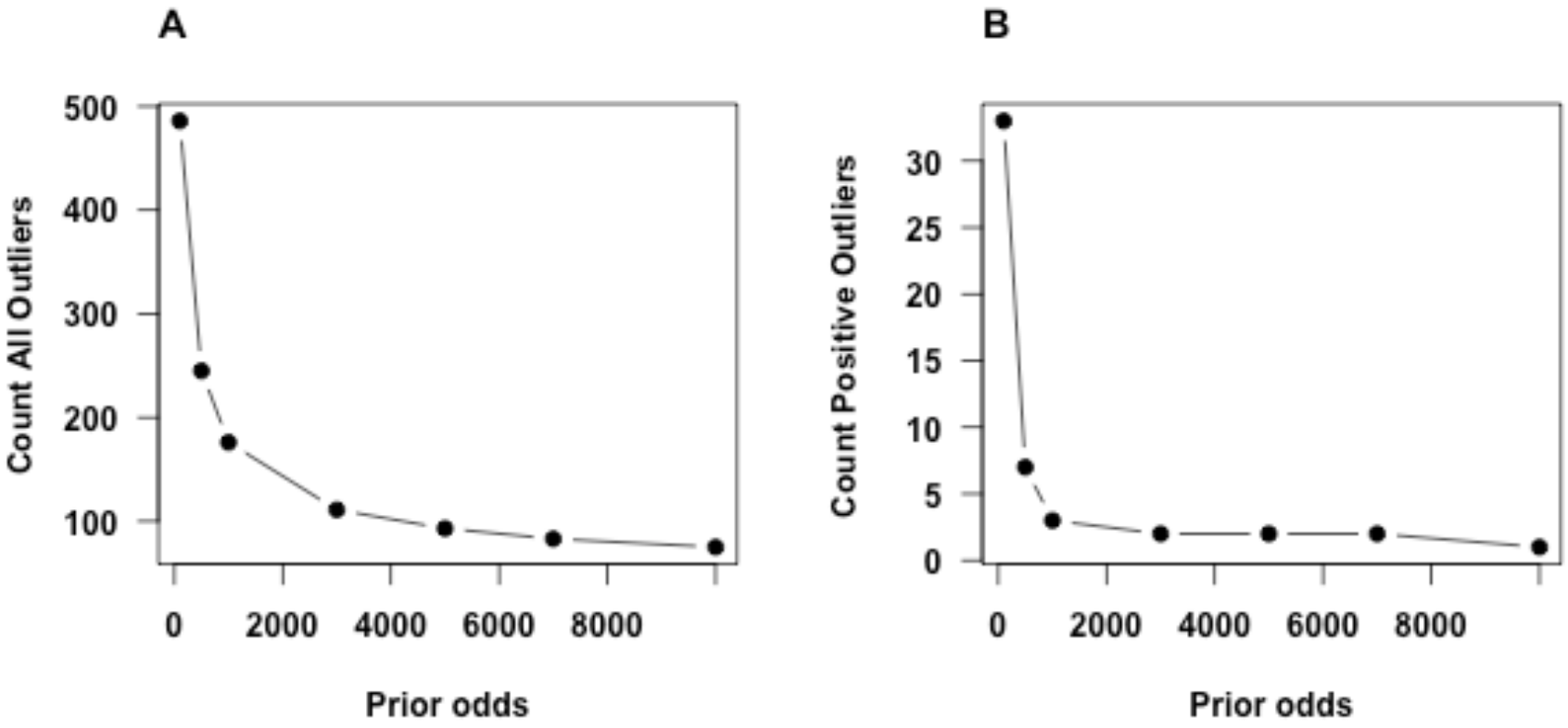

Figure 13. Plots of the (A) total number of outliers or (B) number of positive outliers given a particular prior odds.

Appendix A

Table A1. The 100 spring-germinating accessions used in this study

Accession[1]	Germplasm[1]	Latitude[2]	Longitude[2]
Aa-0	CS6600	50.9167 N	9.57073 E
Ag-0	CS22630	45 N	1.3 E
Ak-1	CS6602	48.0683 N	7.62551 E
Alc-0	CS1656	40.31 N	3.22 W
ALL1-2	CS76089	45.2667 N	1.48333 E
Alst-1	CS22550	54.8 N	2.4333 W
Amel-1	CS22526	53.448 N	5.73 E
An-1	CS22626	51.2167 N	4.4 E
Ang-0	CS6605	50.3 N	5.3 E
Baa-1	CS22529	51.3333 N	6.1 E
Bch-1	CS6609	49.5166 N	9.3166 E
Bd-0	CS6612	52.4584 N	13.287 E
Boot-1	CS22551	54.4 N	3.2667 W
Bor-1	CS22590	49.4013 N	16.2326 E
Bor-4	CS22591	49.4013 N	16.2326 E
Bsch-0	CS6630	50.0167 N	8.6667 E
Bsch-2	CS6631	50.0167 N	8.6667 E
Chat-1	CS22521	48.0717 N	1.33867 E
CIBC-17	CS22603	51.4083 N	0.6383 W
CIBC-5	CS22602	51.4083 N	0.6383 W
Cvi-0	CS22614	15.1111 N	23.6167 W
Db-0	CS6677	50.3055 N	8.324 E
Dr-0	CS6684	51.051 N	13.7336 E
Dra-2	CS6687	49.4167 N	16.2667E
Ei-2	CS22616	50.3 N	6.3 E
El-0	CS6694	51.5105 N	9.68253 E
Ema-1	CS6923	51.3 N	0.5 E
Fei-0	CS22645	40.5 N	8.32 W
Ga-0	CS22634	50.3 N	8 E
Gel-1	CS22533	51.0167 N	5.86667 E
Gie-0	CS6720	50.584 N	8.67825 E
Gu-0	CS22617	50.3 N	8 E
Gy-0	CS22631	49 N	2 E
H55	CS923	49 N	15 E
Hey-1	CS22534	51.25 N	5.9 E
Hl-3	CS6904	52.1444 N	9.37827 E
HR-10	CS22597	51.4083 N	0.6383 W
HR-5	CS22596	51.4083 N	0.6383 W
Is-1	CS6906	50.5 N	7.5 E
Kin-0	CS22654	44.46 N	85.37 W
Kz-1	CS22606	49.5 N	73.1 E

Table A1. Continued

Kz-9	CS22607	49.5 N	73.1 E
Li-2:1	CS6772	50.3833 N	8.0666 E
LI-OF-095	CS76165	40.7777 N	72.9069 W
LL-0	CS22650	41.59 N	2.49 E
Lm-2	CS1345	48 N	0.5 E
Lp2-2	CS22594	49.38 N	16.81 E
Lp2-6	CS22595	49.38 N	16.81 E
Lz-0	CS22615	46 N	3.3 E
Me-0	CS1364	51.9183 N	10.1138 E
Mh-0	CS6792	50.95 N	7.5 E
MIB-15	CS76181	47.3833 N	5.31667 E
MIB-22	CS76182	47.3833 N	5.31667 E
MIB-28	CS76183	47.3833 N	5.31667 E
MIB-84	CS76184	47.3833 N	5.31667 E

[1]From TAIR (www.arabidopsis.org/index.jsp).
[2]Nordborg dataset (http://papya.usc.edu/2010/data/250k-data-version-3.05).

Table A2. *A priori* candidate genes and the environmental factor to which they are known to respond.

Gene Name[1]	Abbreviated Name[1]	Source	Environmental Factor[2]
AT1G03120	ATRAB28	(Atwell *et al*. 2010)	General
AT1G05010	EFE	(Atwell *et al*. 2010)	General
AT1G09950	RAS1	(Ren *et al*. 2010)	General
AT1G12610	DDF1	(Atwell *et al*. 2010)	General
AT1G14920	GAI	(Lee *et al*. 2002)	General
AT1G15550	GA3ox1	(Kucera *et al*. 2005)	General
AT1G24260	SEPALLATA3	(Atwell *et al*. 2010)	General
AT1G28560	SRD2	(Atwell *et al*. 2010)	General
AT1G30040	GA2ox2	(Cadman et al. 2006)	General
AT1G34790	TT1	(Debeaujon *et al*. 2000)	General
AT1G43620	TT15	(Debeaujon *et al*. 2000)	General
AT1G49040	SCD1	(Atwell *et al*. 2010)	General
AT1G49480	RTV1	(Atwell *et al*. 2010)	General
AT1G52340	ABA2	(Koornneef *et al*. 1982)	General
AT1G66350	RGL1	(Lee *et al*. 2002)	General
AT1G72560	PSD	(Atwell *et al*. 2010)	General
AT1G72770	HAB1	(Atwell *et al*. 2010)	General
AT1G72830	HAP2C	(Atwell *et al*. 2010)	General
AT1G78240	TSD2	(Atwell *et al*. 2010)	General
AT1G78390	NCED9	(Cadman et al. 2006)	General
AT1G80340	GA3ox2	(Kucera *et al*.2006)	General
AT2G04240	XERICO	(Zentella *et al*. 2007)	General
AT2G06210	ELF8	(Atwell *et al*. 2010)	General
AT2G19560	EER5	(Atwell *et al*. 2010)	General
AT2G20000	HBT	(Atwell *et al*. 2010)	General
AT2G25170	PKL	(Atwell *et al*. 2010)	General
AT2G27380	ATEPR1	(Atwell *et al*. 2010)	General
AT2G33830		(Atwell *et al*. 2010)	General
AT2G34900	IMB1	(Duque & Chua 2003)	General
AT2G36270	ABI5	(Finkelstein 1994)	General
AT2G39810	HOS1	(Atwell *et al*. 2010)	General
AT2G40220	ABI4	(Finkelstein 1994)	General
AT2G42830	SHP2	(Atwell *et al*. 2010)	General
AT2G44950	RDO4/HUB1	(Liu *et al*. 2007)	General
AT2G45660	AGL20	(Atwell *et al*. 2010)	General

Table A2. Continued

AT3G03450	RGL2	(Lee *et al*. 2002)	General
AT3G05120	GID1	(Griffiths *et al*. 2006)	General
AT3G05890	RCI2B	(Atwell *et al*. 2010)	General
AT3G11440	ATMYB65	(Atwell *et al*. 2010)	General
AT3G11540	SPY	(Jacobsen & Olszewski 1993)	General
AT3G20780	ATTOP6B	(Atwell *et al*. 2010)	General
AT3G24220	NCED6	(Cadman *et al*. 2006)	General
AT3G24440	VRN5	(Atwell *et al*. 2010)	General
AT3G24650	ABI3	(Koornneef *et al*. 1989)	General
AT3G26120	TEL1	(Atwell *et al*. 2010)	General
AT3G54810	BME3	(Atwell *et al*. 2010)	General
AT3G54990	SMZ	(Atwell *et al*. 2010)	General
AT3G55120	TT5	(Debeaujon *et al*. 2000)	General
AT3G59030	TT12	(Debeaujon *et al*. 2000)	General
AT3G63010	ATGID1B	(Atwell *et al*. 2010)	General
AT4G02020	EZA1	(Atwell *et al*. 2010)	General
AT4G02570	ATCUL1	(Atwell *et al*. 2010)	General
AT4G02780	GA1	(Raz *et al*. 2001)	General
AT4G16280	FCA	(Atwell *et al*. 2010)	General
AT4G18660	sim DOG1	(Atwell *et al*. 2010)	General
AT4G24210	SLY1	(Steber *et al*. 1998)	General
AT4G24540	AGL24	(Atwell *et al*. 2010)	General
AT4G24620	PGI1	(Atwell *et al*. 2010)	General
AT4G25140	OLEO1	(Atwell *et al*. 2010)	General
AT4G25530	FWA	(Atwell *et al*. 2010)	General
AT4G26080	ABI1	(Atwell *et al*. 2010)	General
AT4G33280	sim VRN1	(Atwell *et al*. 2010)	General
AT4G39850	COMATOSE	(Russell *et al*.2000)	General
AT5G01560		(Atwell *et al*. 2010)	General
AT5G02310	PRT6	(Holman *et al*. 2009)	General
AT5G04040	SDP1	(Atwell *et al*. 2010)	General
AT5G07190	ATS3	(Atwell *et al*. 2010)	General
AT5G07280	EMS1	(Atwell *et al*. 2010)	General
AT5G09810	ACT7	(Atwell *et al*. 2010)	General
AT5G09820	TT8	(Debeaujon *et al*. 2000)	General
AT5G10140	FLC	(Atwell *et al*. 2010)	General
AT5G13790	AGL15	(Atwell *et al*. 2010)	General
AT5G13930	TT4	(Debeaujon *et a*l. 2000)	General

Table A2. Continued

AT5G14750	WER1	(Atwell *et al*. 2010)	General
AT5G15100	PIN8	(Atwell *et al*. 2010)	General
AT5G16320	FRL1	(Atwell *et al*. 2010)	General
AT5G23150	HUA2	(Atwell *et al*. 2010)	General
AT5G24520	TTG1	(Debeaujon *et al*. 2000)	General
AT5G24630	BIN4	(Atwell *et al*. 2010)	General
AT5G27320	GID1	(Griffiths *et al*. 2006)	General
AT5G35550	TT2	(Debeaujon *et al*. 2000)	General
AT5G42800	TT3	(Debeaujon *et al*. 2000)	General
AT5G45830	DOG1	(Bentsink *et al*. 2006)	General
AT5G47010	LBA1	(Atwell *et al*. 2010)	General
AT5G48100	TT10	(Debeaujon *et al*. 2000)	General
AT5G57380	VIN3	(Atwell *et al*. 2010)	General
AT5G59710	VIP2	(Atwell *et al*. 2010)	General
AT5G61850	LFY	(Atwell *et al*. 2010)	General
AT5G62000	ARF2	(Atwell *et al*. 2010)	General
AT5G64210	AOX2	(Atwell *et al*. 2010)	General
AT5G65420	CYCD4;1	(Atwell *et al*. 2010)	General
AT5G67030	ABA1	(Koornneef *et al*.1989)	General
AT1G01060	LHY	(Atwell *et al*. 2010)	Light
AT1G03790	SOM	(Kim *et al*. 2008)	Light
AT1G09530	PIF3	(Martinez-Garcia *et al*. 2000)	Light
AT1G09570	PHYA	(Shinomura *et al*.1994)	Light
AT1G14280	PKS2	(Atwell *et al*. 2010)	Light
AT1G52830	IAA6	(Atwell *et al*. 2010)	Light
AT1G53090	SPA4	(Atwell *et al*. 2010)	Light
AT1G65480	FT	(Atwell *et al*. 2010)	Light
AT1G70940	PIN3	(Atwell *et al*. 2010)	Light
AT1G80730	ZFP1	(Atwell *et al*. 2010)	Light
AT2G01570	RGA	(Dill & Sun 2001)	Light
AT2G18790	PHYB	(Shinomura *et al*. 1994)	Light
AT2G20180	PIL5	(Oh *et al*. 2004)	Light
AT2G32250	FRS2	(Atwell *et al*. 2010)	Light
AT2G37678	FHY1	(Atwell *et al*. 2010)	Light
AT2G40080	ELF4	(Atwell *et al*. 2010)	Light
AT2G42260	UVI4	(Atwell *et al*. 2010)	Light
AT3G07650	COL9	(Atwell *et al*. 2010)	Light
AT3G09150	HY2	(Atwell *et al*. 2010)	Light

Table A2. Continued

AT3G19820	DWF1	(Atwell *et al*. 2010)	Light
AT3G22380	TIC	(Atwell *et al*. 2010)	Light
AT3G59060	PIL6	(Atwell *et al*. 2010)	Light
AT4G02560	LD	(Atwell *et al*. 2010)	Light
AT4G03400	DFL2	(Atwell *et al*. 2010)	Light
AT4G11110	SPA2	(Atwell *et al*. 2010)	Light
AT4G16250	PHYD	(Aukerman *et al*. 1997)	Light
AT4G18130	PHYE	(Hennig *et al*. 2002)	Light
AT4G19990	FRS1	(Atwell *et al*. 2010)	Light
AT4G36930	SPT	(Penfield *et al* .2005)	Light
AT4G37580	HLS1	(Atwell *et al*. 2010)	Light
AT5G25220	KNAT3	(Atwell *et al*. 2010)	Light
AT5G54510	DFL1	(Atwell *et al*. 2010)	Light
AT5G58960	GIL1	(Atwell *et al*. 2010)	Light
AT5G61380	TOC1	(Atwell *et al*. 2010)	Light
AT5G62640	ELF5	(Atwell *et al*. 2010)	Light
AT5G64330	NPH3	(Atwell *et al*. 2010)	Light
AT1G12110	NRT1.1	(Alboresi *et al*. 2005)	Nutrient
AT1G37130	NIA2	(Finch-Savage *et al*. 2007)	Nutrient
AT1G77760	NIA1	(Finch-Savage *et al*. 2007)	Nutrient
AT5G14570	NRT2.7	(Chopin *et al*. 2007)	Nutrient

[1]Gene names are from TAIR 9 (www.arabidopsis.org/index.jsp).

[2]General = not known to respond specifically light or nutrient cues. Light = light responsive or in light signaling pathway. Nutrient = responsive to nutrient levels.

Table A3. Genes located within 10kb up- or downstream from a significant SNP, the SNPs they are linked to and model in which the significant SNP was found. Names, descriptions, expression, and GO information from TAIR.

Gene[a]	Name	SNP	Model(s)	Description	Expressed[b]	GO Biological Process
AT1G08640	CJD1	SNP1	L1N0, L1N1		y	fatty acid metabolism
AT1G08650	PPCK1	SNP1	L1N0, L1N1		n	protein phosphorylation
AT1G08660	MGP2	SNP1, SNP2	L1N0, L1N1, Full		y	metabolic process, microtubule nucleation
AT1G08670		SNP1,SNP2, SNP3	L1N0, L1N1, Full		n	iron ion transport, nitrate transport, response to nitrate
AT1G08680	ZIGA4	SNP1,SNP2, SNP3	L1N0, L1N1, Full		y	protein autophosphorylation, regulation of ARF GTPase activity
AT1G08695	SCRL3	SNP2, SNP3	L1N0, L1N1, Full		n	signal transduction
AT1G08700	PS1	SNP2, SNP3	L1N0, L1N1, Full		y	calcium-mediated signaling, intracellular signal transduction, metabolic process
AT1G08710		SNP2, SNP3	L1N0, L1N1, Full	F-box family	y	
AT1G08720	EDR1	SNP2, SNP3	L1N0, L1N1, Full		y	MAPK cascade, cell death, defense response to oomycetes, negative regulation of abscisic acid mediated signaling pathway, protein autophosphorylation, regulation of salicylic acid mediated signaling pathway, response to bacterium, response to ethylene stimulus, response to fungus, response to water deprivation

Table A3. Continued

AT1G08730	XIC	SNP3	L1N1	myosin complex, motor protein	y	Golgi localization, actin cytoskeleton organization, actin filament-based movement, mitochondrion localization, peroxisome localization
AT1G59630		SNP4	L1N0	F-box associated	n	
AT1G59640	BPE	SNP4	L1N0	transcription factor	y	fatty acid catabolic process, jasmonic acid metabolic process, ovule development, plant-type cell wall modification, regulation of transcription, DNA-dependent, seed dormancy process
AT1G59650	CW14	SNP4	L1N0		y	N-terminal protein myristoylation
AT1G59660		SNP4	L1N0		y	transport
AT1G73590	PIN1	SNP5	L1N0, L1N1, Full, Common		y	anthocyanin accumulation in tissues in response to UV light, auxin polar transport, cell wall macromolecule metabolic process, determination of bilateral symmetry, embryo development, gravitropism, meristem initiation, meristem maintenance, organ morphogenesis, photomorphogenesis, polarity specification of adaxial/abaxial axis, response to auxin stimulus, response to blue light, root development, shoot system development,
AT1G73600		SNP5	L1N0, L1N1, Full, Common		n	maltose metabolic process, metabolic process, starch biosynthetic process

Table A3. Continued

AT1G73602	CPUORF32	SNP5	L1N0, L1N1, Full, Common		n	
AT1G73603	LCR64	SNP5	L1N0, L1N1, Full, Common		n	
AT1G73607	LCR65	SNP5	L1N0, L1N1, Full, Common		n	
AT1G73610		SNP5	L1N0, L1N1, Full, Common		y	lipid metabolic process
AT2G24200	LAP1	SNP6	L1N0		y	Golgi organization, gluconeogenesis, glycolysis, hyperosmotic response, protein metabolic process, protein targeting to vacuole, proteolysis, response to cadmium ion, response to salt stress, response to temperature stimulus, water transport
AT2G24205		SNP6	L1N0	ECA1 gametogenesis related family protein	n	
AT2G24210	TPS10	SNP6	L1N0		y	meristem development, metabolic process, monoterpenoid biosynthetic process, response to jasmonic acid stimulus, response to wounding
AT2G24220	PUP5	SNP6	L1N0		y	nucleobase-containing compound transport protein
AT2G24230		SNP6	L1N0		y	phosphorylation, transmembrane receptor protein tyrosine kinase signaling pathway

Table A3. Continued

AT3G30580		SNP7, SNP8	L1N0	unknown	n	
AT4G15440	HPL1	SNP9	L1N1	fatty acid metabolism, auxin biosynthesis	y	fatty acid metabolic process, indoleacetic acid biosynthetic process, oxidation-reduction process, response to wounding, tryptophan catabolic process
AT4G15450		SNP9	L1N1	Senescence/ dehydration -associated protein-related	n	
AT4G15460		SNP9	L1N1	glycine-rich protein	n	
AT4G15470		SNP9	L1N1	Bax inhibitor-1 family protein	y	Golgi vesicle transport, cellular membrane fusion, endosomal transport, protein targeting to vacuole, vesicle-mediated transport
AT4G15475		SNP9	L1N1	F-box family	y	protein glycosylation, ubiquitin-dependent protein catabolic process
AT4G15480	UGT84A1	SNP9	L1N1		y	flavonoid biosynthetic process, metabolic process, response to UV-B, response to sucrose stimulus
AT4G15490	UGT84A3	SNP9	L1N1		y	metabolic process
AT5G14520		SNP10	L1N0		y	protein import into nucleus, protein maturation, ribosome biogenesis
AT5G14530		SNP10	L1N0	ubiquitin ligase complex	y	photoperiodism, flowering
AT5G14540		SNP10	L1N0	unknown	y	
AT5G14545	MIR398B	SNP10	L1N0	microRNA	n	cellular response to phosphate starvation, defense response to bacterium, response to copper ion, response to ozone, response to salt stress
AT5G14550		SNP10	L1N0		y	circadian rhythm
AT5G14560		SNP10	L1N0	unknown	n	

Table A3. Continued

AT5G14565	MIR398C	SNP10	L1N0	microRNA	n	cellular response to phosphate starvation, defense response to bacterium, response to copper ion, response to ozone, response to salt stress
AT5G14570	NRT2.7	SNP10	L1N0		y	nitrate transport, transmembrane transport
AT5G14580		SNP10	L1N0		y	mitochondrial RNA catabolic process, mitochondrial RNA processing,
AT5G28680	ANX2	SNP11	L0N0		y	transcription factor import into nucleus protein phosphorylation,
AT5G28690		SNP11	L0N0	unknown	n	
AT5G39865		SNP12	L1N0	Glutaredoxin family protein	n	N-terminal protein myristoylation
AT5G39870		SNP12	L1N0		n	plant-type cell wall modification
AT5G39880		SNP12	L1N0	unknown	y	
AT5G39890		SNP12	L1N0		y	cell wall macromolecule metabolic process, oxidation-reduction process, regulation of hydrogen peroxide metabolic process, response to hypoxia, salicylic acid mediated signaling pathway
AT5G39895	pre-tRNA	SNP12	L1N0	pre-Ala	n	(see also Parisod & Christin 2008; 2008; Manel et al. 2010; reviewed in Schoville et al. 2012; 2012)
AT5G39900		SNP12	L1N0	Small GTP-binding protein	y	
AT5G39910		SNP12	L1N0	Pectin lyase-like superfamily protein	n	carbohydrate metabolic process

Table A3. Continued

AT5G39920		SNP12	L1N0	Pre-mRNA cleavage complex	n	
AT5G39930	CLPS5	SNP12	L1N0		n	Mo-molybdopterin cofactor biosynthetic process
AT5G66680	DGL1	SNP13	L1N0		y	Golgi vesicle transport, RNA methylation, cell wall modification, cellulose biosynthetic process, plant-type cell wall organization, lignin metabolic process, metabolic process, regulation of anion channel activity, transition metal ion transport
AT5G66690	UGT72E2	SNP13	L1N0		y	
AT5G66700	HB53	SNP13	L1N0		y	regulation of transcription, DNA-dependent, response to auxin stimulus, root development
AT5G66710		SNP13, SNP14	L1N0		y	protein phosphorylation
AT5G66720		SNP14	L1N0	Protein phosphatase 2C family protein	n	
AT5G66730	IDD1	SNP14	L1N0		y	regulation of gibberellic acid mediated signaling pathway, regulation of seed germination, regulation of transcription, DNA-dependent, seed maturation
AT5G66740		SNP14	L1N0		n	plant-type cell wall modification
AT5G66750	CHR1	SNP14	L1N0		y	DNA mediated transformation, RNA interference, chromatin silencing by small RNA, histone H3-K9 methylation, histone phosphorylation, negative regulation of histone H4 acetylation, regulation of gene expression by genetic imprinting

Table A3. Continued

AT5G66755	pre-tRNA	SNP14	L1N0	pre-Glu	n	translational elongation
AT5G66760	SDH1-1	SNP14	L1N0		y	electron transport chain, mitochondrial electron transport, succinate to ubiquinone, oxidation-reduction process, tricarboxylic acid cycle

[a]TAIR gene identifier
[b] y = gene is expressed in the seed or embryo, n = not known to be expressed in embryo or seed.

Table A4. Genes located within 10kb up- or downstream from a significant time of maximum rate of germination SNP, the position of the SNP(s) they are linked to and model in which the significant SNP was found. Names, descriptions, expression, and GO information from TAIR.

Gene[a]	Name	SNP	Model(s)	Description	Expressed[b]	GO Biological Process
AT1G08640	CJD1	SNP7	Full-Light/High		y	fatty acid metabolism
AT1G08650	PPCK1	SNP7, SNP8	Full-Light/High		n	protein phosphorylation
AT1G08660	MGP2	SNP1, SNP7, SNP8	Full-Light/Low, Full-Light/High, Full		y	metabolic process, microtubule nucleation
AT1G08670		SNP1, SNP7, SNP8, SNP9	Full-Light/Low, Full-Light/High, Full		n	iron ion transport, nitrate transport, response to nitrate
AT1G08680	ZIGA4	SNP1, SNP7, SNP8, SNP9	Full-Light/Low, Full-Light/High, Full		y	protein autophosphorylation, regulation of ARF GTPase activity
AT1G08695	SCRL3	SNP1, SNP8, SNP9	Full-Light/Low, Full-Light/High, Full		n	signal transduction
AT1G08700	PS1	SNP1, SNP8, SNP9	Full-Light/Low, Full-Light/High, Full		y	calcium-mediated signaling, intracellular signal transduction, metabolic process
AT1G08710		SNP1, SNP8, SNP9	Full-Light/Low, Full-Light/High, Full	F-box family	n	

Table A4. Continued

AT1G08720	EDR1	SNP1, SNP9	Full-Light/Low, Full-Light/High, Full		y	MAPK cascade, cell death, defense response to oomycetes, negative regulation of abscisic acid mediated signaling pathway, protein autophosphorylation, regulation of salicylic acid mediated signaling pathway, response to bacterium, response to ethylene stimulus, response to fungus, response to water deprivation
AT1G08730	XIC	SNP9	Full-Light/Low		y	Golgi localization, actin cytoskeleton organization, actin filament-based movement, mitochondrion localization, peroxisome localization
AT1G11990		SNP10	Full-Light/Low	O-fucosyltransferase family protein	n	
AT1G12000		SNP10	Full-Light/Low	Phosphofructokinase family protein	y	acetyl-CoA biosynthetic process, acetyl-CoA metabolic process, chromatin silencing, glycolysis, histone H3-K9 methylation, photosynthesis, response to cadmium ion
AT1G12010		SNP10	Full-Light/Low		n	cellular response to fatty acid, ethylene biosynthetic process, oxidation-reduction process
AT1G12013	SNOR111	SNP10	Full-Light/Low		n	rRNA modification
AT1G12015		SNP10	Full-Light/Low		n	rRNA modification
AT1G12020		SNP10	Full-Light/Low	unknown	y	
AT1G12030		SNP10	Full-Light/Low	unknown	n	
AT1G29740		SNP11	Full-Light/High		n	protein phosphorylation
AT1G29750	RKF1	SNP11	Full-Light/High		y	oligopeptide transport, protein phosphorylation, transmembrane receptor protein tyrosine kinase signaling pathway

Table A4. Continued

AT1G29760		SNP11	Full-Light/High	Putative adipose-regulatory protein	y	
AT1G29770		SNP11	Full-Light/High	Haloacid dehalogenase-like hydrolase	n	
AT1G29780		SNP11	Full-Light/High	Haloacid dehalogenase-like hydrolase	n	
AT1G29785		SNP11	Full-Light/High	other RNA	n	
AT1G50000		SNP12	Dark/Low		y	rRNA processing, ubiquinone biosynthetic process
AT1G50010	TUA2	SNP12	Dark/Low		y	GTP catabolic process, microtubule-based movement, microtubule-based process, protein polymerization, response to salt stress
AT1G50020		SNP12	Dark/Low	unknown	y	

Table A4. Continued

AT1G50030	TOR	SNP12	Dark/Low		y	actin nucleation, cell adhesion, cell division, cell wall organization, cytokinesis by cell plate formation, embryo development, embryo development ending in seed dormancy, embryonic pattern specification, meiotic DNA double-strand break formation, meiotic chromosome segregation, negative regulation of autophagy, organ morphogenesis, positive regulation of cell growth, positive regulation of embryonic development, positive regulation of organelle organization, positive regulation of rRNA processing, positive regulation of transcription, DNA-dependent, post-embryonic development, primary shoot apical meristem specification, rRNA transcription, reciprocal meiotic recombination, regulation of cell differentiation, regulation of chromosome organization, root hair cell differentiation, seed development, seed maturation, sister chromatid cohesion, tissue development, toxin catabolic process, trichome morphogenesis, vegetative to reproductive phase transition of meristem
AT1G58410		SNP2	Full-Light/Low, Full-Light/High		y	defense response

Table A4. Continued

AT1G58420		SNP2	Full-Light/Low, Full-Light/High	unknown	y	
AT1G58430	RXF26	SNP2	Full-Light/Low, Full-Light/High	GDSL-like Lipase/Acylhydrolase superfamily protein	y	lipid metabolic process, ovule development
AT1G58440	DRY2	SNP2	Full-Light/Low, Full-Light/High		y	brassinosteroid biosynthetic process, metabolic process, oxidation-reduction process, pentacyclic triterpenoid biosynthetic process, polysaccharide biosynthetic process, response to water deprivation, sterol biosynthetic process, thiamine biosynthetic process
AT1G58450	TPR6	SNP2	Full-Light/Low, Full-Light/High	Tetratricopeptide repeat-like superfamily protein	n	
AT1G58460		SNP2	Full-Light/Low, Full-Light/High	unknown	n	
AT1G59630		SNP13	Full-Light/High	F-box associated	n	
AT1G59640	BPE	SNP13	Full-Light/High		y	fatty acid catabolic process, jasmonic acid metabolic process, ovule development, plant-type cell wall modification, regulation of transcription, DNA-dependent, seed dormancy process
AT1G59650	CW14	SNP13	Full-Light/High		y	N-terminal protein myristoylation
AT1G59660		SNP13	Full-Light/High		y	transport
AT1G63300		SNP14	Dark/Low	Myosin heavy chain-related protein	y	
AT1G63310		SNP14	Dark/Low	unknown	y	
AT1G63320		SNP14	Dark/Low	Pentatricopeptide repeat superfamily protein	n	

Table A4. Continued

AT1G63330		SNP14	Dark/Low	Pentatricopeptide repeat superfamily protein	n	
AT1G63340		SNP14	Dark/Low	FAD/NAD(P)-binding oxidoreductase family protein	n	oxidation-reduction process
AT1G63350		SNP14	Dark/Low		n	N-terminal protein myristoylation
AT1G63360		SNP14	Dark/Low		n	N-terminal protein myristoylation
AT1G63370		SNP14	Dark/Low	FAD/NAD(P)-binding oxidoreductase family protein	n	oxidation-reduction process
AT1G73590	PIN1	SNP15	Dark/Low, Full-Light/High, Full-Light/Low		y	anthocyanin accumulation in tissues in response to UV light, auxin polar transport, cell wall macromolecule metabolic process, determination of bilateral symmetry, embryo development, gravitropism, meristem initiation, meristem maintenance, organ morphogenesis, photomorphogenesis, polarity specification of adaxial/abaxial axis, response to auxin stimulus, response to blue light, root development, shoot system development
AT1G73600		SNP15	Dark/Low, Full-Light/High, Full-Light/Low		n	maltose metabolic process, metabolic process, starch biosynthetic process
AT1G73602	CPUORF32	SNP15	Dark/Low, Full-Light/High, Full-Light/Low		n	
AT1G73603		SNP15	Dark/Low, Full-Light/High, Full-Light/Low		n	

Table A4. Continued

AT1G73607		SNP15	Dark/Low, Full-Light/High, Full-Light/Low		n	
AT1G73610		SNP15	Dark/Low, Full-Light/High, Full-Light/Low		y	lipid metabolic process
AT1G78070		SNP16	Full-Light/High	Transducin/WD40 repeat-like superfamily protein	y	
AT1G78080	WIND1	SNP16	Full-Light/High		y	cellular response to salt stress, cytokinin mediated signaling pathway, ethylene mediated signaling pathway, red or far-red light signaling pathway, regulation of cell differentiation, regulation of transcription, DNA-dependent, response to cold, response to light stimulus, response to osmotic stress, response to salt stress, response to water deprivation, response to wounding
AT2G14800		SNP17	Full-Light/High		y	reciprocal meiotic recombination, synapsis
AT2G14810		SNP17	Full-Light/High	unknown	n	
AT2G14820	NPY2	SNP17	Full-Light/High		y	positive gravitropism, response to light stimulus
AT2G20790		SNP18	Full-Light/High		y	intracellular protein transport, vesicle-mediated transport
AT2G20800	NDB4	SNP18	Full-Light/High		n	oxidation-reduction process
AT2G20805		SNP18	Full-Light/High	unknown	n	
AT2G20810	GAUT10	SNP18	Full-Light/High		y	carbohydrate biosynthetic process
AT2G20815		SNP18	Full-Light/High	unknown	n	
AT2G20820		SNP18	Full-Light/High		y	photorespiration

Table A4. Continued

AT2G20825	ULT2	SNP18	Full-Light/High		y	
AT2G20830		SNP18	Full-Light/High		y	metabolic process
AT2G24200	LAP1	SNP19	Full-Light/Low		y	Golgi organization, gluconeogenesis, glycolysis, hyperosmotic response, protein metabolic process, protein targeting to vacuole, proteolysis, response to cadmium ion, response to salt stress, response to temperature stimulus, water transport
AT2G24205		SNP19	Full-Light/Low	ECA1 gametogenesis related family protein	n	
AT2G24210	TPS10	SNP19	Full-Light/Low		y	meristem development, metabolic process, monoterpenoid biosynthetic process, response to jasmonic acid stimulus, response to wounding
AT2G24220	PUP5	SNP19	Full-Light/Low		y	nucleobase-containing compound transport
AT2G24230		SNP19	Full-Light/Low		y	protein phosphorylation, transmembrane receptor protein tyrosine kinase signaling pathway
AT2G42270		SNP20	Full-Light/High, Full-Light/Low	U5 small nuclear ribonucleoprotein helicase	y	
AT2G42280		SNP20	Full-Light/High, Full-Light/Low		n	photoperiodism, flowering, regulation of transcription, DNA-dependent, response to arsenic-containing substance
AT2G42290		SNP20	Full-Light/High, Full-Light/Low		y	protein phosphorylation, transmembrane receptor protein tyrosine kinase signaling pathway
AT2G42300		SNP20	Full-Light/High, Full-Light/Low		y	regulation of transcription, DNA-dependent

Table A4. Continued

AT2G42310		SNP20	Full-Light/High, Full-Light/Low		y	photorespiration, proteasome core complex assembly, response to misfolded protein, ubiquitin-dependent protein catabolic process
AT2G42320		SNP20	Full-Light/High, Full-Light/Low		y	cysteine biosynthetic process
AT3G14320		SNP21	Full-Light/Low		n	zinc ion binding
AT3G14330		SNP21	Full-Light/Low	Tetratricopeptide repeat-like superfamily protein	y	mRNA modification
AT3G14340		SNP21	Full-Light/Low	unknown	y	
AT3G14350	SRF7	SNP21	Full-Light/Low		y	protein phosphorylation, transmembrane receptor protein tyrosine kinase signaling pathway
AT3G14360		SNP21	Full-Light/Low		y	lipid metabolic process
AT3G14362	RTFL10	SNP21	Full-Light/Low		n	shoot system development
AT3G16420	PBP1	SNP22	Full-Light/Low		y	protein folding, regulation of hydrolase activity
AT3G16430	PBP2	SNP22	Full-Light/Low		n	
AT3G16432		SNP22	Full-Light/Low	unknown	n	
AT3G16440	MEE36	SNP22	Full-Light/Low		n	embryo development ending in seed dormancy, transition metal ion transport
AT3G16450	JAL33	SNP22	Full-Light/Low		n	response to cold, response to zinc ion
AT3G16460		SNP22	Full-Light/Low		y	brassinosteroid biosynthetic process, nitrate transport, response to cold, response to nitrate, sterol biosynthetic process
AT3G16470	JAL35	SNP22	Full-Light/Low		y	Golgi organization, calcium ion transport, cysteine biosynthetic process, response to cold, response to jasmonic acid stimulus, response to salt stress, response to wounding

Table A4. Continued

AT3G30580		SNP3, SNP23	Full-Light/Low	unknown	n	
AT3G32980		SNP24	Full-Light/High		n	oxidation-reduction process, response to oxidative stress
AT3G47965		SNP25	Full-Light/High	unknown	n	
AT3G47980		SNP25	Full-Light/High	Integral membrane HPP family protein	n	
AT3G47990	SIS3	SNP25	Full-Light/High		y	glucuronoxylan metabolic process, protein ubiquitination, response to high light intensity, response to hydrogen peroxide, sugar mediated signaling pathway, xylan biosynthetic process
AT3G48000	ALDH2	SNP25	Full-Light/High		y	metabolic process, oxidation-reduction process, response to cadmium ion
AT3G48010	CNGC16	SNP25	Full-Light/High		n	ion transport, transmembrane transport
AT3G48020		SNP25	Full-Light/High	unknown	n	
AT3G48030		SNP25	Full-Light/High		n	response to hypoxia
AT3G59020		SNP4	Full-Light/High		y	intracellular protein transport, protein import into nucleus, docking
AT3G59030	TT12	SNP4	Full-Light/High		y	drug transmembrane transport, maintenance of seed dormancy, proanthocyanidin biosynthetic process, purine nucleobase transport, transmembrane transport
AT3G59040		SNP4	Full-Light/High	Tetratricopeptide repeat-like superfamily protein	y	chloroplast organization, pentose-phosphate shunt, rRNA processing, tRNA metabolic process
AT3G59050	PAO3	SNP4	Full-Light/High		n	cellular modified amino acid biosynthetic process, coumarin biosynthetic process, oxidation-reduction process, phenylpropanoid metabolic process, polyamine catabolic process, response to wounding

Table A4. Continued

AT3G59052	CPUORF18	SNP4	Full-Light/High		n	
AT3G59060	PIL6	SNP4	Full-Light/High		y	aromatic amino acid family biosynthetic process, aromatic amino acid family metabolic process, cell differentiation, cellular amino acid biosynthetic process, cellular cation homeostasis, chlorophyll biosynthetic process, circadian rhythm, coenzyme biosynthetic process, cysteine biosynthetic process, ethylene biosynthetic process, glucosinolate biosynthetic process, glycine catabolic process, gravitropism, indoleacetic acid biosynthetic process, isopentenyl diphosphate biosynthetic process, mevalonate-independent pathway, jasmonic acid biosynthetic process, leaf morphogenesis, lipoate metabolic process, nucleotide metabolic process, oxidoreduction coenzyme metabolic process, positive regulation of transcription, DNA-dependent, red or far-red light signaling pathway, red, far-red light phototransduction, regulation of auxin biosynthetic process, regulation of auxin mediated signaling pathway, secondary metabolic process, sulfur amino acid metabolic process, sulfur compound biosynthetic process, unsaturated fatty acid biosynthetic process, vitamin metabolic process

Table A4. Continued

AT4G12190		SNP26	Full-Light/High	RING/U-box superfamily protein	n	
AT4G12210		SNP26	Full-Light/High	RING/U-box superfamily protein	n	
AT4G12220		SNP26	Full-Light/High	unknown	n	
AT4G12230		SNP26	Full-Light/High		y	
AT4G12240		SNP26	Full-Light/High		y	regulation of transcription, DNA-dependent
AT4G12250	GAE5	SNP26	Full-Light/High		y	carbohydrate metabolic process, cellular metabolic process, cellular response to phosphate starvation, cellular response to water deprivation, galactolipid biosynthetic process, nucleotide-sugar metabolic process
AT4G13130		SNP5	Full-Light/High		n	oxidation-reduction process
AT4G13150		SNP5	Full-Light/High	unknown	n	
AT4G13160		SNP5	Full-Light/High		y	protein autophosphorylation
AT4G13170		SNP5	Full-Light/High		y	translation
AT4G13180		SNP5	Full-Light/High		y	metabolic process, response to arsenic-containing substance
AT4G13190		SNP5	Full-Light/High	Protein kinase superfamily protein	n	protein phosphorylation
AT4G13195	CLE44	SNP5	Full-Light/High		y	axillary shoot meristem initiation, phloem or xylem histogenesis, procambium histogenesis, xylem development

Table A4. Continued

AT4G15430	ERD	SNP27	Full-Light/Low		y	circadian rhythm, long-day photoperiodism, flowering, regulation of transcription, DNA-dependent, response to abscisic acid stimulus, response to auxin stimulus, response to ethylene stimulus, response to gibberellin stimulus, response to jasmonic acid stimulus, response to salicylic acid stimulus, response to salt stress
AT4G15440	HPL1	SNP27, SNP28	Full-Light/High, Full-Light/Low	fatty acid metabolism, auxin biosynthesis	y	fatty acid metabolic process, indoleacetic acid biosynthetic process, oxidation-reduction process, response to wounding, tryptophan catabolic process
AT4G15450		SNP27, SNP28, SNP29	Full-Light/High, Full-Light/Low	Senescence/dehydration-associated protein-related	n	
AT4G15460		SNP27, SNP28, SNP29	Full-Light/High, Full-Light/Low	glycine-rich protein	n	
AT4G15470		SNP27, SNP28, SNP29	Full-Light/High, Full-Light/Low	Bax inhibitor-1 family protein	y	Golgi vesicle transport, cellular membrane fusion, endosomal transport, protein targeting to vacuole, vesicle-mediated transport
AT4G15475		SNP27, SNP28, SNP29	Full-Light/High, Full-Light/Low	F-box family	y	protein glycosylation, ubiquitin-dependent protein catabolic process
AT4G15480	UGT84A1	SNP27, SNP28, SNP29	Full-Light/High, Full-Light/Low		y	flavonoid biosynthetic process, metabolic process, response to UV-B, response to sucrose stimulus
AT4G15490	UGT84A3	SNP28, SNP29	Full-Light/High, Full-Light/Low		y	metabolic process

Table A4. Continued

AT4G24770	RBP31	SNP30	Full-Light/High		y	RNA modification, RNA processing, RNA stabilization, aromatic amino acid family biosynthetic process, base conversion or substitution editing, cell differentiation, cold acclimation, innate immune response, iron-sulfur cluster assembly, leaf morphogenesis, positive regulation of transcription, DNA-dependent, protein targeting to chloroplast
AT4G24780		SNP30	Full-Light/High	Pectin lyase-like superfamily protein	y	syncytium formation
AT4G24790		SNP30	Full-Light/High		y	DNA replication
AT4G24800	ECIP1	SNP30	Full-Light/High		y	response to ethylene stimulus, response to salt stress
AT4G24805		SNP30	Full-Light/High	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	n	metabolic process
AT4G24810		SNP30	Full-Light/High		y	mRNA modification, protein phosphorylation
AT5G14520		SNP31	Full-Light/Low		y	protein import into nucleus, protein maturation, ribosome biogenesis
AT5G14530		SNP31	Full-Light/Low	ubiquitin ligase complex	y	photoperiodism, flowering
AT5G14540		SNP31	Full-Light/Low	unknown	n	
AT5G14545	MIR398B	SNP31	Full-Light/Low	microRNA	n	cellular response to phosphate starvation, defense response to bacterium, response to copper ion, response to ozone, response to salt stress
AT5G14550		SNP31	Full-Light/Low		y	circadian rhythum
AT5G14560		SNP31	Full-Light/Low	unknown	n	

Table A4. Continued

AT5G14565	MIR398C	SNP31	Full-Light/Low	microRNA	n	cellular response to phosphate starvation, defense response to bacterium, response to copper ion, response to ozone, response to salt stress
AT5G14570	NRT2.7	SNP31	Full-Light/Low		y	nitrate transport, transmembrane transport
AT5G14580		SNP31	Full-Light/Low		y	mitochondrial RNA catabolic process, mitochondrial RNA processing, transcription factor import into nucleus
AT5G28680	ANX2	SNP6	Dark/Low		y	protein phosphorylation
AT5G28690		SNP6	Dark/Low	unknown	n	
AT5G40980		SNP32	Full-Light/Low	unknown	y	
AT5G40981		SNP32	Full-Light/Low	unknown	n	
AT5G40990	GLIP1	SNP32	Full-Light/Low		n	defense response to bacterium, defense response to fungus, defense response to fungus, incompatible interaction, induced systemic resistance, ethylene mediated signaling pathway, jasmonic acid and ethylene-dependent systemic resistance, ethylene mediated signaling pathway, lipid metabolic process, response to fungus, response to salicylic acid stimulus, systemic acquired resistance
AT5G41000	YSL4	SNP32	Full-Light/Low		y	oligopeptide transport, transmembrane transport
AT5G41010	NRPB12	SNP32	Full-Light/Low		y	RNA splicing, via endonucleolytic cleavage and ligation, transcription from RNA polymerase II promoter, transcription, DNA-dependent
AT5G41020		SNP32	Full-Light/Low	myb family transcription factor	n	
AT5G41030		SNP32	Full-Light/Low	TCP family transcription factor	n	regulation of transcription, DNA-dependent

Table A4. Continued

AT5G41040	ASFT	SNP32	Full-Light/Low		y	cell wall pectin biosynthetic process, nitrate transport, response to nitrate, suberin biosynthetic process
AT5G41050		SNP32	Full-Light/Low		y	
AT5G55310	TOP1	SNP33	Full-Light/High		y	DNA topological change, defense response to insect, fatty acid biosynthetic process, long-chain fatty acid metabolic process
AT5G55320		SNP33	Full-Light/High	membrane bound O-acyl transferase family protein	n	
AT5G55330		SNP33	Full-Light/High	membrane bound O-acyl transferase family protein	y	
AT5G55340		SNP33	Full-Light/High	membrane bound O-acyl transferase family protein	y	
AT5G55350		SNP33	Full-Light/High	membrane bound O-acyl transferase family protein	n	
AT5G55360		SNP33	Full-Light/High	membrane bound O-acyl transferase family protein	n	
AT5G55370		SNP33	Full-Light/High	membrane bound O-acyl transferase family protein	y	
AT5G55380		SNP33	Full-Light/High	membrane bound O-acyl transferase family protein	y	
AT5G55390	EDM2	SNP33	Full-Light/High		y	defense response to fungus, positive regulation of flower development, signal transduction, vegetative to reproductive phase transition of meristem
AT5G65950		SNP34	Full-Light/High		y	Golgi vesicle transport, biological_process, cellulose biosynthetic process, protein glycosylation
AT5G65960		SNP34	Full-Light/High		y	small GTPase mediated signal transduction
AT5G65970	MLO10	SNP34	Full-Light/High		n	cell death, defense response
AT5G65980		SNP34	Full-Light/High		n	auxin polar transport, transmembrane transport

Table A4. Continued

AT5G65990		SNP34	Full-Light/High		y	amino acid transport
AT5G66000		SNP34	Full-Light/High	unknown	y	
AT5G66005		SNP34	Full-Light/High		n	
AT5G66010		SNP34	Full-Light/High	RNA-binding family protein	y	
AT5G66020	SAC1B	SNP34	Full-Light/High		n	inositol phosphate dephosphorylation, inositol trisphosphate metabolic process, response to salt stress
AT5G66310		SNP35	Full-Light/Low		y	
AT5G66320	GATA5	SNP35	Full-Light/Low		y	microtubule-based movement
AT5G66330		SNP35	Full-Light/Low	Leucine-rich repeat family protein	y	positive regulation of transcription, DNA-dependent, regulation of transcription, DNA-dependent
AT5G66340		SNP35	Full-Light/Low	unknown	y	signal transduction
AT5G66350	SHI	SNP35	Full-Light/Low		y	
AT5G66680	DGL1	SNP36, SNP37	Full-Light/High, Full-Light/Low		y	gibberellic acid mediated signaling pathway, response to gibberellin stimulus
AT5G66690	UGT72E2	SNP36, SNP37	Full-Light/High, Full-Light/Low		y	Golgi vesicle transport, RNA methylation, cell wall modification, cellulose biosynthetic process, plant-type cell wall organization
AT5G66700	HB53	SNP36, SNP37	Full-Light/High, Full-Light/Low		y	lignin metabolic process, metabolic process, regulation of anion channel activity, transition metal ion transport
AT5G66710		SNP36, SNP37	Full-Light/High, Full-Light/Low		y	regulation of transcription, DNA-dependent, response to auxin stimulus, root development

[a]TAIR gene identifier

[b]y = gene is expressed in the seed or embryo, n = not known to be expressed in embryo or seed.

References

Adler L, Wikler K, Wyndham F, Linder C, Schmitt J (1993) Potential for Persistence of Genes Escaped from Canola: Germination Cues in Crop, Wild, and Crop-Wild Hybrid Brassica rapa. Functional Ecology, 7, 736–745.

Alboresi A, Gestin C, Leydecker M-T et al. (2005) Nitrate, a signal relieving seed dormancy in Arabidopsis. Plant Cell And Environment, 28, 500–512.

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

Aranzana M, Kim S, Zhao K et al. (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genetics, 1, 531–539.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Molecular Ecology, 22, 3179-3190.

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter, 9, 208–218.

Atwell S, Huang YS, Vilhjalmsson BJ et al. (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature, 465, 627–631.

Aukerman M, Hirschfeld M, Wester L et al. (1997) A deletion in the PHYD gene of the Arabidopsis Wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing. The Plant Cell, 9, 1317–1326.

Baker H (1974) The Evolution of Weeds. Annual Review of Ecology and Systematics, 5, 1–24.

Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography, 32, 818–830.

Baskin C, Baskin J (2006) The natural history of soil seed banks of arable land. Weed Science, 54, 549–557.

Bates D, Maechler M, Bolker B lme4. GPL–2 GPL–3.

Beck JB, Schmuths H, Schaal BA (2008) Native range genetic variation in Arabidopsis thaliana is strongly geographically structured and reflects Pleistocene glacial dynamics. Molecular Ecology, 17, 902–915.

Begg GS, Wishart J, Young MW, Squire GR, Iannetta PPM (2011) Genetic structure among arable populations of Capsella bursa-pastoris is linked to functional traits and in-field conditions. Ecography, 35, 446–457.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B, 57, 289-300.

Bentsink L, Jowett J, Hanhart CJ, Koornneef M (2006) Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. Proceedings of the National Academy of Sciences, 103, 17042–17047.

Bergelson J, Roux F (2010) Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. Nature Reviews Genetics, 11, 867–879.

Bing D, Downey R, Rakow G (1996) Assessment of transgene escape from Brassica rapa (B. campestris) into B. nigra or Sinapis arvensis. Plant Breeding, 115, 1–4.

Bommarco R, Lönn M, Danzer U, Pålsson KJ, Torstensson P (2010) Genetic and phenotypic differences between thistle populations in response to habitat and weed management practices. Biological Journal Of The Linnean Society, 99, 797–807.

Borcard, D, Gillet F, Legendre P (2011) Numerical Ecology with R (Use R).

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological Modelling, 153, 51–68.

Borthwick H, Hendricks S, Parker M, Toole E, Toole V (1952) A Reversible Photoreaction Controlling Seed Germination. Proceedings Of The National Academy Of Sciences Of The United States Of America, 38, 662–666.

Boyd E, Dorn L, Weinig C, Schmitt J (2007) Maternal effects and germination timing mediate the expression of winter and spring annual life histories in Arabidopsis thaliana. International Journal of Plant Sciences, 168, 205–214.

Cadman CSC, Toorop PE, Hilhorst HWM, Finch-Savage WE (2006) Gene expression profiles of Arabidopsis Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism. The Plant Journal, 46, 805–822.

Carbon S, Ireland A, Mungall CJ et al. (2009) AmiGO: online access to ontology and annotation data. Bioinformatics, 25, 288–289.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping Loci de novo from short-read sequences. G3, 1, 171–182.

Chapin FS (1980) The mineral nutrition of wild plants. Annual Review of Ecology and Systematics, 11, 233–260.

Childs DZ, Metcalf CJE, Rees M (2010) Evolutionary bet-hedging in the real world: empirical evidence and challenges revealed by plants. Proceedings Of The Royal Society B-Biological Sciences, 277, 3055–3064.

Chopin F, Orsel M, Dorbe M-F et al. (2007) The Arabidopsis ATNRT2.7 nitrate transporter controls nitrate content in seeds. The Plant Cell, 19, 1590–1602.

Clements D, DiTommaso A, Jordan N et al. (2004) Adaptability of plants invading north American cropland. Agriculture Ecosystems & Environment, 104, 379–398.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using Environmental Correlations to Identify Loci Underlying Local Adaptation. Genetics, 185, 1411–1423.

Debeaujon I, Koornneef M (2000) Gibberellin requirement for Arabidopsis seed germination is determined both by testa characteristics and embryonic abscisic acid. Plant Physiology, 122, 415–424.

Debeaujon I, Léon-Kloosterziel KM, Koornneef M (2000) Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. Plant Physiology, 122, 403–414.

DeRose-Wilson L, Gaut BS (2011) Mapping Salinity Tolerance during Arabidopsis thaliana Germination and Seedling Growth. PLoS ONE, 6, e22832.

Dill A, Sun T (2001) Synergistic Repression of gibberellin signaling by removing RGA and GAI function in Arabidopsis thaliana. Genetics, 159, 777–785.

Donohue K (2002) Germination Timing Influences Natural Selection on Life-History Characters in Arabidopsis thaliana. Ecology, 83, 1006–1016.

Donohue K (2009) Completing the cycle: maternal effects as the missing link in plant life histories. Philosophical Transactions Of The Royal Society B-Biological Sciences, 364, 1059–1074.

Donohue K (2005) Seeds and Seasons: interpreting germination timing in the field. Seed Science Research, 15, 175-187.

Donohue K, Dorn L, Griffith C, Kim E, Aguilera A, Polisetty C, Schmitt J (2005a) Environmental and Genetic Influences on the Germination of Arabidopsis thaliana in the Field. Evolution, 59, 740–757.

Donohue K, Dorn L, Griffith C, Kim E, Aguilera A, Polisetty C, Schmitt J (2005b) Niche Construction through Germination Cueing: Life-History Responses to Timing of Germination in Arabidopsis thaliana. Evolution, 59, 771–785.

Duque P, Chua N (2003) IMB1, a bromodomain protein induced during seed imbibition, regulates ABA- and phyA-mediated responses of germination in Arabidopsis. The Plant Journal, 35, 787–799.

Edelaar P, Bolnick DI (2012) Non-random gene flow: an underappreciated force in evolution and ecology. Trends in Ecology & Evolution, 27, 659–665.

Edwards M (1980) Aspects of the Population Ecology of Charlock. Journal of Applied Ecology, 17, 151-171.

Fiege M (2005) The weedy west: mobile nature, boundaries, and common space in the Montana landscape. The Western Historical Quarterly, 35, 22–47.

Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination. New Phytologist, 171, 501–523.

Finch-Savage WE, Cadman CSC, Toorop PE, Lynn JR, Hilhorst HWM (2007) Seed dormancy release in Arabidopsis Cvi by dry after-ripening, low temperature, nitrate and light shows common quantitative patterns of gene expression directed by environmentally specific sensing. The Plant Journal, 51, 60–78.

Finkelstein R (1994) Maternal Effects Govern Variable Dominance of Two Abscisic Acid Response Mutations in Arabidopsis thaliana. Plant Physiology, 105, 1203–1208.

Fogg G (1950) Sinapis Arvensis L. Journal Of Ecology, 38, 415–429.

Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics, 180, 977–993.

Fox J, Wiesberg S (2011) An {R} Companion to Applied Regression. Sage, Thousand

Oaks, CA.

Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. Molecular Ecology, 19, 3773–3788.

Friml J, Vieten A, Sauer M et al. (2003) Efflux-dependent auxin gradients establish the apical-basal axis of Arabidopsis. Nature, 426, 147–153.

Geng Y-P, Pan X-Y, Xu C-Y et al. (2006) Phenotypic plasticity rather than locally adapted ecotypes allows the invasive alligator weed to colonize a wide range of habitats. Biological Invasions, 9, 245–256.

Ghersa CM, Arnold RLB, Martinez-Ghersa MA (1992) The role of fluctuating temperatures in germination and establishment of Sorghum halepense. Regulation of germination at increasing depths. Functional Ecology, 460–468.

Griffiths J, Murase K, Rieu I et al. (2006) Genetic characterization and functional analysis of the GID1 gibberellin receptors in Arabidopsis. The Plant Cell, 18, 3399–3414.

Hails R, Rees M, Kohn D, Crawley M (1997) Burial and seed survival in Brassica napus subsp oleifera and Sinapis arvensis including a comparison of transgenic and non-transgenic lines of the crop. Philosophical Transactions Of The Royal Society B-Biological Sciences, 264, 1–7.

Hennig L, Stoddart W, Dieterle M, Whitelam G, Schafer E (2002) Phytochrome E controls light-induced germination of Arabidopsis. Plant Physiology, 128, 194–200.

Heschel M, Butler C, Barua D et al. (2008) New Roles of Phytochromes during Seed Germination. International Journal of Plant Sciences, 169, 531–540.

Hilhorst HWM (1990) Dose-response analysis of factors involved in germination and secondary dormancy of seeds of Sisymbrium officinale. Plant Physiology, 94, 1096–1102.

Hilhorst HWM, Karssen CM (1988) Dual effect of light on the gibberellin-and nitrate-stimulated seed germination of Sisymbrium officinale and Arabidopsis thaliana. Plant Physiology, 86, 591.

Holdsworth MJ, Finch-Savage WE, Grappin P, Job D (2008) Post-genomics dissection of seed dormancy and germination. Trends in Plant Science, 13, 7–13.

Holm R (1972) Volatile Metabolites Controlling Germination in Buried Weed Seeds. Plant Physiology, 50, 293–297.

Holman TJ, Jones PD, Russell L et al. (2009) The N-end rule pathway promotes seed germination and establishment through removal of ABA sensitivity in Arabidopsis. Proceedings Of The National Academy Of Sciences Of The United States Of America, 106, 4549–4554.

Horton MW, Hancock AM, Huang YS et al. (2012) Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nature Genetics, 44, 212–216.

Huang X, Schmitt J, Dorn L et al. (2010) The earliest stages of adaptation in an experimental plant population: strong selection on QTLs for seed dormancy. Molecular Ecology, 19, 1335–1351.

Jacobsen SE, Olszewski NE (1993) Mutations at the SPINDLY Locus of Arabidopsis Alter Gibberellin Signal-transduction. The Plant Cell, 5, 887–896.

Kang HM, Zaitlen NA, Wade CM et al. (2008) Efficient control of population structure in model organism association mapping. Genetics, 178, 1709–1723.

Kawecki T, Ebert D (2004) Conceptual issues in local adaptation. Ecology Letters, 7, 1225–1241.

Keeley JE, Fotheringham CJ (1998) Smoke-Induced Seed Germination in California Chaparral. Ecology, 79, 2320–2336.

Kim DH, Yamaguchi S, Lim S et al. (2008) SOMNUS, a CCCH-type zinc finger protein in Arabidopsis, negatively regulates light-dependent seed germination downstream of PIL5. The Plant Cell, 20, 1260–1277.

Kim S, Plagnol V, Hu TT et al. (2007) Recombination and linkage disequilibrium in Arabidopsis thaliana. Nature Genetics, 39, 1151–1155.

Koornneef M, Hanhart CJ, Hilhorst HWM, Karssen C (1989) In Vivo Inhibition of Seed Development and Reserve Protein Accumulation in Recombinants of Abscisic Acid Biosynthesis and Responsiveness Mutants in Arabidopsis thaliana. Plant Physiology, 90, 463–469.

Koornneef M, Jorna ML, Derswan D, Karssen C (1982) The Isolation of Abscisic-acid (ABA) Deficient Mutants by Selection of Induced Revertants in Non-germinating Gibberellin Sensitive Lines of Arabidopsis thaliana (L.) Heynh. Theoretical and

Applied Genetics, 61, 385–393.

Korte A, Vilhjalmsson BJ, Segura V et al. (2012) A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nature Genetics, 44, 1066–1071.

Kucera B, Cohn M, Leubner-Metzger G (2005) Plant hormone interactions during seed dormancy release and germination. Seed Science Research, 15, 281–307.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology, 10, R25.

Laserna MP, Sanchez RA, Botto JF (2008) Light-related loci controlling seed germination in Ler 3 Cvi and Bay-0 x Sha recombinant inbred-line populations of Arabidopsis thaliana. Annals Of Botany, 102, 631–642.

Lee S, Cheng H, King K et al. (2002) Gibberellin regulates Arabidopsis seed germination via RGL2, a GAI/RGA-like gene whose expression is up-regulated following imbibition. Genes & Development, 16, 646–658.

Lefol E, Danielou V, Fleury A, Darmency H (1996) Gene flow within a population of the outbreeding Sinapis arvensis: Isozyme analysis of half-sib families. Weed Research, 36, 189–195.

Legendre P, Anderson MJ (1999) Distance-Based Redundancy Analysis: Testing Multispecies Responses in Multifactorial Ecological Experiments. Ecological Monographs, 69, 1–24.

Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Molecular Ecology Resources, 10, 831–844.

Legendre P, Borcard D, Blanchet FG, Dray S (2012) PCNM: MEM spatial eigenfunction and principle coordinate analyses.

Lenormand T (2002) Gene flow and the limits to natural selection. TRENDS in Ecology & Evolution, 17, 183–189.

Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO (2010) Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. Proceedings of the National Academy of Sciences, 107, 21199–21204.

Liu Y, Koornneef M, Soppe WJJ (2007) The absence of histone H2B monoubiquitination

in the Arabidopsis hub1 (rdo4) mutant reveals a role for chromatin remodeling in seed dormancy. The Plant Cell, 19, 433–444.

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.

Luzuriaga A, Escudero A, Perez-Garcia F (2006) Environmental maternal effects on seed morphology and germination in Sinapis arvensis (Cruciferae). Weed Research, 46, 163–174.

Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates.

Manel S, Joost S, Epperson BK et al. (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Molecular Ecology, 19, 3760–3772.

Martinez-Garcia J, Huq E, Quail P (2000) Direct targeting of light signals to a promoter element-bound transcription factor. Science, 288, 859–863.

Meikle A, Finch R, McRoberts N, Marshall G (1999) A molecular genetic assessment of herbicide-resistant Sinapis arvensis. Weed Research, 39, 149–158.

Meng P, Macquet A, Loudet O, Marion-Poll A, North H (2008) Analysis of natural allelic variation controlling Arabidopsis thaliana seed germinability in response to cold and dark: identification of three major quantitative trait loci. Molecular Plant, 1, 145.

Mennan H, Ngouajio M (2006) Seasonal cycles in germination and seedling emergence of summer and winter populations of catchweed bedstraw (Galium aparine) and wild mustard (Brassica kaber). Weed Science, 54, 114–120.

Mercer KL, Alexander HM, Snow AA (2011) Selection on seedling emergence timing and size in an annual plant, Helianthus annuus (common sunflower, Asteraceae). American Journal Of Botany, 98, 975–985.

Moodie M, Finch R, Marshall G (1997) Analysis of genetic variation in wild mustard (Sinapis arvensis) using molecular markers. Weed Science, 45, 102–107.

Moyes C, Lilley J, Casais C et al. (2002) Barriers to gene flow from oilseed rape (Brassica napus) into populations of Sinapis arvensis. Molecular Ecology, 11, 103–

112.

Nambara E, Marion-Poll A (2003) ABA action and interactions in seeds. Trends In Plant Science, 8, 213–217.

Oh E, Kim J, Park E et al. (2004) PIL5, a phytochrome-interacting basic helix-loop-helix protein, is a key negative regulator of seed germination in Arabidopsis thaliana. The Plant Cell, 16, 3045–3058.

Oksanen J, Blanchet FG, Roeland K et al. (2013) vegan: Community Ecology Package.

Parisod C, Christin P-A (2008) Genome-wide association to fine-scale ecological heterogeneity within a continuous population of Biscutella laevigata (Brassicaceae). New Phytologist, 178, 436–447.

Penfield S, Josse E, Kannangara R et al. (2005) Cold and light control seed germination through the bHLH transcription factor SPATULA. Current Biology, 15, 1998–2006.

Pimentel D, Lach L, Zuniga R, Morrison D (2000) Environmental and economic costs of nonindigenous species in the United States. BioScience, 50, 53–65.

Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Molecular Biology Reporter, 15, 8–15.

Raz V, Bergervoet J, Koornneef M (2001) Sequential steps for developmental arrest in Arabidopsis seeds. Development, 128, 243–252.

Rees M, Long MJ (1992) Germnation Biology and the Ecology of Annual Plants. American Naturalist, 139, 484–508.

Rees M, Brown V, (null) (1991) The Effect of Established Plants on Recruitment in the Annual Forb Sinapis arvensis. Oecologia, 87, 58–62.

Ren Z, Zheng Z, Chinnusamy V et al. (2010) RAS1, a quantitative trait locus for salt tolerance and ABA sensitivity in Arabidopsis. Proceedings of the National Academy of Sciences, 107, 5669.

Russell L, Larner V, Kurup S, Bougourd S, Holdsworth M (2000) The Arabidopsis COMATOSE locus regulates germination potential. Development, 127, 3759–3767.

SAS Institute Inc. (2008) SAS/STAT® 9.2 User's Guide. SAS Institute Inc., Cary, NC.

Scheiner S (1993) Genetics and Evolution of Phenotypic Plasticity. Annual Review of Ecology and Systematics, 24, 35–68.

Schmitt J, Wulff RD (1993) Light spectral quality, phytochrome and plant competition. Trends in ecology & evolution (Personal edition), 8, 47–51.

Schmitt J, Niles J, Wulff R (1992) Norms of Reaction of Seed Traits to Maternal Environments in Plantago lanceolata. The American Naturalist, 139, 451–466.

Schmuths H, Bachmann K, Weber WE, Horres R, Hoffmann MH (2006) Effects of preconditioning and temperature during germination of 73 natural accessions of Arabidopsis thaliana. Annals Of Botany, 97, 623–634.

Schoville SD, Bonin A, François O et al. (2012) Adaptive Genetic Variation on the Landscape: Methods and Cases. Annual Review Of Ecology Evolution And Systematics, 43, 23–43.

Shinomura T, Nagatani A, Chory J, Furuya M (1994) The Induction of Seed Germination in Arabidopsis thaliana is Regulated Principally by Phytochrome B and Secondarily by Phytochrome A. Plant Physiology, 104, 363–371.

Simons AM (2009) Fluctuating natural selection accounts for the evolution of diversification bet hedging. Proceedings Of The Royal Society B-Biological Sciences, 276, 1987–1992.

Simons AM (2011) Modes of response to environmental change and the elusive empirical evidence for bet hedging. Proceedings Of The Royal Society B-Biological Sciences, 278, 1601–1609.

Simons AM, Johnston MO (2006) Environmental and genetic sources of diversification in the timing of seed germination: Implications for the evolution of bet hedging. Evolution, 60, 2280–2292.

Smith EN, Kruglyak L (2008) Gene–Environment Interaction in Yeast Gene Expression (T Mackay, Ed,). PLoS Biology, 6, e83.

Steber CM, Cooney SE, McCourt P (1998) Isolation of the GA-response mutant sly1 as a suppressor of ABI1-1 in Arabidopsis thaliana. Genetics, 149, 509–521.

Susko DJ, Cavers PB (2008) Seed size effects and competitive ability in Thlaspi arvense L. (Brassicaceae). Botany, 86, 259–267.

Tiffin P, Olson S, Moyle LC (2001) Asymmetrical crossing barriers in angiosperms.

Proceedings Of The Royal Society B-Biological Sciences, 268, 861–867.

Tungate KD, Susko DJ, Rufty TW (2002) Reproduction and offspring competitiveness of Senna obtusifolia are influenced by nutrient availability. New Phytologist, 154, 661–669.

Turelli M, Moyle LC (2006) Asymmetric Postmating Isolation: Darwin"s Corollary to Haldane"s Rule. Genetics, 176, 1059–1088.

Ungerer M, Rieseberg L (2003) Genetic architecture of a selection response in Arabidopsis thaliana. Evolution; international journal of organic evolution, 57, 2531–2539.

VanderSchaar W, AlonsoBlanco C, LeonKloosterziel K et al. (1997) QTL analysis of seed dormancy in Arabidopsis using recombinant inbred lines and MQM mapping. Heredity, 79, 190–200.

Verdu M, Traveset A (2005) Early emergence enhances plant fitness: A phylogenetically controlled meta-analysis. Ecology, 86, 1385–1394.

Via S, Conner J (1995) Evolution in heterogenous environments: genetic variability within and across different grains in Tribolium castaneum. Heredity, 74, 80-90.

Via S, Gomulkiewicz R, De Jong G et al. (2004) Adaptive phenotypic plasticity: consensus and controversy. Trends in Ecology & Evolution, 10, 212–217.

Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. Nature methods, 9, 808–810.

Warwick S, Sauder C, Beckie H (2005) Resistance in Canadian Biotypes of Wild Mustard (Sinapis arvensis) to Acetolactate Synthase Inhibiting Herbicides. Weed Science, 53, 631–639.

Warwick S, Simard M, Legere A et al. (2003) Hybridization between transgenic Brassica napus L. and its wild relatives: Brassica rapa L., Raphanus raphanistrum L., Sinapis arvensis L., and Erucastrum gallicum (Willd.) OE Schulz. Theoretical and Applied Genetics, 107, 528–539.

Weinig C (2005) Rapid evolutionary responses to selection in heterogeneous environments among agricultural and nonagricultural weeds. International Journal of Plant Sciences, 166, 641–647.

Weinig C (2010) Differing selection in alternative competitive environments: shade-

avoidance responses and germination timing. Evolution; international journal of organic evolution, 54, 124–136.

Zentella R, Zhang Z-L, Park M et al. (2007) Global analysis of DELLA direct targets in early gibberellin signaling in Arabidopsis. The Plant Cell, 19, 3037–3057.

Zhao K, Aranzana MJ, Kim S et al. (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genetics, 3,

Vita

Ginnie Denise Morrison grew up in Glen Ellyn, Illinois. She entered the Weinburg College of Arts and Sciences at Northwestern University, in Evanston, Illinois, in 2002. After graduating with a B.A. in 2006, she worked for a year in Kalamazoo, Michigan, in the necropsy department at MPI Research. In September, 2007, she entered the Plant Biology Program at the University of Texas at Austin in hopes of obtaining her Ph.D.

Permanent email: ginnie.gintastic@gmail.com

This dissertation was typed by the author.