Received Date : 24-May-2015
Revised Date : 01-Jul-2016
Accepted Date : 06-Jul-2016
Article type : Original Article

Extreme QTL mapping of germination speed in *Arabidopsis thaliana*

Wei Yuan[1], Jonathan M. Flowers,[1,2] Dustin J. Sahraie[1], Ian M. Ehrenreich[3] and Michael D. Purugganan[1*]

[1]Center for Genomics and Systems Biology, Department of Biology, 12 Waverly Place, New York University, New York, NY USA 10011

[2]Center for Genomics and Systems Biology, NYU Abu Dhabi Research Institute, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates

[3]Molecular and Computational Biology Section, Ray R. Irani Hall 201, University of Southern California, Los Angeles, CA 90089-2910, USA

Keywords: quantitative genomics, life history, bulk-segregant analysis, seed ecology

[*]Corresponding author. Tel. (212) 998 3800 Email: mp132@nyu.edu

Running Title: X-QTL mapping of rapid germination

ABSTRACT

Seed germination is a key life-history transition for annual plants, and partly determines lifetime performance and fitness. Germination speed, the elapsed time for a non-dormant seed to germinate, is a poorly understood trait important for plants' competitiveness and fitness in fluctuating environments. Germination speed varied by 30% among 18 *A. thaliana* populations measured, and exhibited weak negative correlation with flowering time and seed weight, with significant genotype effect ($p<0.005$). To dissect the genetic architecture of germination speed, we developed the Extreme QTL (X-QTL) mapping method in *Arabidopsis thaliana.* The method has been shown in yeast to increase QTL mapping power by integrating selective screening and bulk-segregant-analysis in a very large mapping population. By pooled genotyping of top 5% of rapid germinants from ~100,000 F_3 individuals, three X-QTLs regions were identified on chromosomes 1, 3, and 4. All regions were confirmed as QTL regions by sequencing 192 rapid germinants from an independent F_3 selection experiment. Positional overlaps were found between X-QTLs and previously-identified seed, life history, and fitness QTLs. Our method provides a rapid mapping platform in *A. thaliana* with potentially greater power. One can also relate identified X-QTLs to the *A. thaliana* physical map, facilitating candidate gene identification.

INTRODUCTION

Life history traits are a set of demographic traits associated with the schedule and duration of key developmental and reproductive events in an organism's life cycle (Vuorisalo and Mutikainen 1999, Braendle et al. 2011). The sporophyte life history of annual seed plants has three major stages: seed, vegetative growth, and reproduction. The timing of transitions between the stages (i.e. seed germination, seedling establishment and flowering) is vitally important as it

determines type and strength of selective pressure, and has been the focus of attention in studies of the ecology and life history evolution (Donohue et al. 2010, Burghardt et al. 2015, Vuorisalo and Mutikainen 1999). Germination as the initial step in a plant's life history can have a profound impact on fitness components during its life cycle (Donohue et al. 2010). The adaptive value of when a seed germinates depends upon the relative risk of mortality in seed versus germinant stage, seed dormancy, and environmental cueing (Rees et al. 1996, Donohue 2002, Donohue et al, 2005abc, Finch-Savage et al. 2006, Donohue et al. 2010, Huang et al. 2010).

The genetics and ecology of germination has been extensively studied in the model plant *Arabidopsis thaliana*, an annual, ruderal weedy species that colonizes disturbed habitats and the margins of agricultural fields (Koornneef et al. 2004, Bomblies et al. 2010). Seasonal cueing of germination is shown to have major consequences both for shaping life history strategies in this species and for local adaptation, and may partly determine species range (Donohue et al. 2005a). It was demonstrated, for example, that artificially shifting germination for <10 days during the fall season can determine whether plants will flower in the fall, or overwinter and flower in the spring (Wilczek et al. 2009). This provided links between the phenology of reproduction with the early stages in plant life histories.

Dormancy regulation was found to be a major factor of seasonal germination timing. Clinal variation for seed dormancy have been noted (Montesinos-Navarro et al. 2012). A latitudinal co-variation in seed dormancy and flowering time has also been observed (Debieu et al. 2013), indicating the potential role of dormancy underlying germination-flowering time relationship. QTLs controlling dormancy under a variety of conditions have been identified in

Arabidopsis and other species (Clerkx et al. 2004, Dias et al. 2011, Hayashi et al. 2008). For example, eleven *Delay of Germination (DOG)* loci have been mapped in *A. thaliana*, and together they explain 54% of natural variation in seed dormancy (Alonso-Blanco et al. 2003, Bentsink et al. 2010, Bentsink et al. 2006).

While dormancy has received much attention, natural variation in traits associated with the active phase of germination is not well understood. One of these key traits is germination speed, which we define as the elapsed time from the start of germination induction (e.g., supply of a favorable cue) through cotyledon expansion. In general, phenotypic variation in germination speed could arise as a bet-hedging strategy (Simons 2009), i.e. staggering the germinating seed cohorts in case of catastrophic event, or to exploit temporarily favorable environments. Very fast germination, for example, has been observed in ~20 species in the *Amaranthaceae,* and can occur in as little as 10 minutes after water imbibition (Parsons 2012). These species are generally found in high-stress environments (e.g., saline or arid), and the fast germination trait allows for exploitation of temporarily favorable conditions (Parsons 2012). It had also been suggested that early germination provides an advantage by outcompeting neighboring plants and increasing the growth period (Miller et al. 1994, Weinig 2000, Dubois and Cheptou 2012, Graebner et al. 2012). These results suggest that germination speed is associated with competition/colonization trade-off, pointing to a correlation between germination behavior and dispersal ability.

Genetic dissection of germination traits outside of dormancy is necessary if we aim to obtain insights in the genetic architecture of this key life history transition. Several QTL mapping studies have examined germination speed under laboratory or field conditions in a few species

(Clerkx et al. 2004, Foolad et al. 1999, Bettey et al. 2000). These studies provided evidence of germination speed being intrinsically variable in natural populations. The species range, life history, and extensive genetic resources of the model plant *A. thaliana* makes it an ideal model to perform a more comprehensive genetic dissection of germination speed natural variation.

In *A. thaliana*, quantitative genetic studies have largely relied on traditional QTL mapping approaches that use mapping populations of a limited number (n = 100-1000) of individuals (Mauricio 2001) and consequently limited in the number of recombination events that can be interrogated in the mapping process. Recently, a new QTL mapping method, Extreme QTL (X-QTL) mapping, was developed in *Saccharomyces cerevisiae,* relying on the ability to develop mapping populations with greater than 10^6 individuals in this unicellular eukaryote (Ehrenreich et al. 2010).

The assumption underlying X-QTL mapping is that chromosomal regions in which the imposed selection favors neither parental genotype will be found at equal allele frequencies ; whereas, for each locus affected by selection, there will be an overrepresentation of a single parent's genotype (Ehrenreich et al. 2010), which we henceforth call an X-QTL. X-QTL mapping promises to deliver more comprehensive mapping of selectable traits in highly fecund species in which large-scale selection screening and/or phenotyping is possible. The large size of plants imposes a logistical limit to X-QTL mapping, but seed and seedling traits (such as germination speed) can be readily screened with up to ~10^5 individuals, are therefore presumably amenable for analysis with this mapping approach.

Understanding the quantitative genetics of germination speed natural variation allow us to examine the genetic architecture of this key life history trait, and possibly enable us to further evaluate its importance in plant adaptation. In this paper, we develop an X-QTL mapping platform for seed traits in *A. thaliana*. We were able to use this to map QTLs for germination speed on three chromosomes and, furthermore, used whole genome re-sequencing of individual recombinants to confirm the X-QTL mapping results. Our study provides an estimate of the number and location of germination speed QTLs, and allows us to interrogate the genetic correlation of these with QTLs for other seed and life history traits.

MATERIALS AND METHODS

Development of Bs-2 × Col-0 F3 mapping population

All seeds were obtained from the Arabidopsis Biological Resource Center (Columbus, OH), and parental seed was selfed for at least 1 generation. F1 plants were obtained from a cross between the Bs-2 (**CS6628**) and Col-0 (**CS6673**) accessions of *A. thaliana*. F1 plants were selfed to produce an F2 population, and approximately 2,000 F2 plants were further selfed to generate $>10^6$ F3 individuals to be used as the mapping population.

Plant culture for seed bulking

All seeds were surface sterilized with 50% bleach (Clorox, Oakland, CA), triple-rinsed, and evenly plated onto a 50 cm^2 square Petri dishes containing solid culture media (0.25% Phytagel supplemented with Murashige/Skoog basal salt media and 0.5% sucrose). Stratification was completed by keeping the plates in darkness at 4°C for 48 hrs. Plates were then transferred to and randomly positioned in an environmental chamber set at 22°C, 65% humidity and long-day

(16 h light/8 h darkness) conditions. Seventy-two healthy seedlings from each accession were then transferred to soil (Metro-Mix 360). The position and orientation of plant trays were randomized and regularly shuffled in the chamber throughout their life cycle.

Germination speed scoring

For germination speed assays, we used 18 accessions (see Table 1). Seventy-two plants for each accession were grown in an environmental chamber set at 22°C and long day (16 hrs light/8 hrs darkness) conditions. Freshly harvested seeds from these accessions were after-ripened in airtight containers with silica gel beads for 8 weeks.

Triplicate scoring was carried out. Each replicate had 50 mg (approximately 2000) dry seeds, which were germinated in a 50 cm^2 square petri dish containing solid selection media (0.25% Phytagel supplemented with Murashige and Skoog basal salt media). 500-micron nylon mesh was applied to the interface of medium and seeds to prevent seed clumping. Stratification was completed by keeping the plates in darkness at 4°C for 48 hrs. Plates were allowed to equilibrate to 22°C in darkness for 4 hrs, then immediately transferred (designated 0 hour-after-stratification, HAS) to and randomly positioned into environmental chambers set at 22°C, 85% humidity, and continuous light conditions. Germination was scored at 6-hour intervals from 24 HAS to 120 HAS by visually identifying, counting and removing germinants that reached cotyledon expansion phase.

Germination curve analysis

Raw cumulative germination fraction of each replicate at each time point, *t*, was used as the fraction of germinants, G(*t*), in the 4-parameter Hill function (4PHF) model (El-Kassaby et al. 2008, Joosen et al. 2010)

$$G(t) = k + \frac{a \times t^b}{c^b + t^b}$$

Initial value of parameter *c*, germination speed, for each accession was estimated by solving the linear equation of the two data points spanning the 50% germination point. The initial value of *a* was set to 1 (the actual value of the final germinated fraction), and *k* and *b* were set at a small, arbitrary value. An automatic curve-fitting algorithm was executed iteratively. The parameters for each accession model were solved using the nls() function in R (R Core Team 2014) package chemCal (Ranke 2014), and with the least sum-of-squares method. Each round the algorithm calculates the residual sum of square, and exits when the residual sum-of-squares ceases to decrease. The maximum iteration was set to 100 rounds. Germination speed for each accession was estimated by averaging the final *c* values from all three replicates

Phenotyping of T-DNA insertion lines

We ordered T-DNA insertion lines for the QTL gene candidates (Supplementary Table 2) from TAIR (www.arabidopsis.org). All lines were documented to be heterozygous except for that of AZF2. The seeds were bulked for one generation using the protocol above. Rosette leaf tissues from individual plants were harvested and their DNA extracted with the CTAB method. Genotyping of the individual plants was performed by PCR using the primers and protocol provided by the SALK Institute Genomic Analysis Tool /T-DNA Express

(http://signal.salk.edu/tdnaprimers.2.html). Seeds were then harvested by single-seed descent and after-ripened for 8 weeks. Three homozygous T-DNA insertion lines for each gene candidate were randomly chosen, together with three randomly chosen wild-type (WT) lines segregated from the heterozygous seed lots. Germination speed of these chosen lines was scored and curve fitted with protocols above. Student's t-test was performed between each of the 4 T-DNA lines and the WT control.

Development of isothermal genotyping microarray

For markers, we used the SNP data from the 250K *Arabidopsis* SNP data (V 3.01, https://github.com/Gregor-Mendel-Institute/atpolydb/blob/master/250k_snp_data/call_method_32.tar.gz, Atwell et al. 2010). This version featured 214,533 SNPs from 1,179 accessions. SNPs for our array were chosen with the following criteria: (a) the SNP must be diallelic, (b) the SNP must not have another SNP within 17 bps on either side, (c) the SNP must be in a single-copy region of the genome, (d) minor-allele frequency must be above 0.35, and (e) the final SNPs featured on the array provides an even coverage across the genome. We chose moderate-frequency SNPs to allow the array to be maximally useful

We designed probes for an isothermal microarray (Gresham et al. 2010), where probes with virtually identical T_m values for hybridization are obtained by varying probe length according to local GC content. The maximum probe length was set at 27 bps, and the minimum at 21 bps, and the optimal T_m was set at 55°C. To maximize mismatch discrimination, the SNP was located in the middle region of the probe, and the hybridization temperature should be 5°C

higher than the mean value of T_m. Sequences 17 bps immediately upstream and downstream of each SNP were acquired from TAIR version 8. To further minimize binding efficiency bias introduced by T_m variation among probes, we rejected all probes whose T_m was lower than an arbitrary value of 53.7°C. The SNP array was manufactured by Agilent in Oligo-aCGH G3 8*60K format. The final array had 60,978 custom probes, which could be used to genotype 30,389 common SNP markers across the *A. thaliana* genome. We did not exclude SNPs in centromere regions, but this did not affect our mapping results since none of our X-QTL peaks were within centromere.

Microarray testing

DNA samples were labeled with Cy3 or Cy5 fluorescent dyes according to the Agilent Array-Based CGH Genomic DNA Analysis protocol. Differentially labeled samples were mixed in equal amounts before hybridizing to the array slide. Array scanning and feature extraction were performed with the Agilent hardware and software program, and with built-in rank-invariant normalization. Log_{10} ratio of fluorescent intensity for each probe reflects the relative hybridization to the probe by either DNA sample. For each SNP, the difference in log_{10} ratio of intensity (DLR) between the probe pair was calculated as a measurement of allele frequency bias. A DLR of 0 indicates no allele frequency bias between two samples.

There are 14,889 SNPs featured in the array that are polymorphic between Bs-2 and Col-0, the parental accessions of our mapping population. To test for the specificity of the array, DNA from both accessions was differentially labeled, and mixed in equal amount. Triplicate dye-swap experiment was performed by hybridizing the parental DNA mixture to the array. The

preferential binding of parental DNA to probes of their own allele genotype were tested with a two-way ANOVA model with genotype and dye as fixed effects. Multiple hypothesis testing was corrected by FDR=0.05. To test the sensitivity, a serial mixture of parental DNA representing a gradient of allele frequency (100% Bs-2, 100% Col-0, and Bs-2: Col-0 in the ratio of 9:1, 7:3, 5:5, 3:7, and 1:9) were hybridized to the array, together with equal amount of differentially-labeled Bs-2 × Col-0 F1 progeny DNA (having unbiased allele frequency). Linear regression was performed to model the DLR of each SNP in response to changing allele frequency. SNPs with slope value less than 1 standard deviation from the mean slope value, and SNPs whose r-squared values were less than 0.8 were dropped from further analyses. We found that 4,849 probe pairs/SNPs markers passed our specificity and sensitivity tests for our down-stream analysis.

X-QTL mapping in Bs-2 × Col-0 F3 mapping population

Phenotyping germination speed and selection of germination cohorts for genotyping were accomplished simultaneously. Two replicates of 100,000 F_3 seeds were phenotyped and sampled. For each replicate 50 vials of 50 mg F_3 dry seeds were weighed, surface sterilized, plated, and treated as described above (see germination speed scoring section). Germination was scored every 12 hrs for the first 36 HAS, and then every 8 hrs until 60 HAS. Individual germinants reaching cotyledon expansion stage were identified, counted, and removed from the plate. The last two time points, 60 HAS and 72 HAS, when the majority of the seeds reached cotyledon expansion, all germinants were collected from the plate by scraping and the number of germinants estimated. Approximately 100% germination was reached by 72 HAS. We designated the fast-germinating F_3 individuals for X-QTL mapping as the first 5000 germinants (5%) that were removed from the selection plate (spanning 36 HAS to 52 HAS). DNA was extracted from

each of the early, 60 HAS, and 72 HAS cohort pool using Qiagen DNeasy Plant Maxi Kit (Qiagen, Valencia, CA), and a final pool representing the total DNA of all 100,000 F_3 germinants was made by proportionally mixing together three DNA pools according to their DNA concentration and the number of individuals in each pool.

Dye-swap hybridization was carried out for each of the four pools with pool-extracted DNA from 2,000 random F2 individuals (equal in size to the founder population) as reference. Random F2s were used as reference to control for the possible segregation distortion in the founder population that might have biased our allele frequency estimates at specific SNP markers. Prior to the mapping hybridization, pooled 2,000 F2 DNA was hybridized against F1 DNA in a test run. No chromosomal region showed significant allele frequency deviation from the F1, indicating no segregation distortion in the F2 population as a whole.

Statistical analysis for microarray

All SNP coordinates featured on the array were updated from TAIR 8 to TAIR 10 version according to information provided by TAIR (ftp://ftp.arabidopsis.org/home/tair/Sequences/whole_chromosomes/TAIR9_assembly_updates_relative_to_TAIR8&TAIR9_assemblies.xls). DLR was calculated for each probe pair from the array hybridization. To detect genomic regions that respond to selection, the DLR of each germination cohort was tested against the DLR of all germinants. Taking linkage into consideration, we decided to only test the statistical significance of the mean allele frequency shift among consecutive SNPs. To accomplish this, we used a sliding window approach where the cohort DLR of all SNPs within a window were compared against the all-germinant DLR from

the corresponding window with a three-way ANOVA model*: DLR=Treatment (Replicate) + Dye +Treatment (Replicate) × Dye +Treatment × Replicate × Dye,* in which all three effects were fixed effects, with replicate nested within treatment. The F-statistic of treatment effect was recorded for each sliding window as the test statistics. The optimum sliding window size and significance threshold were determined by permutation. Permutation was performed with a series of sliding window sizes (from 25 SNPs to 150 SNPs, with 25-SNP increments). During each round of permutation, the association of the DLR values and SNP positions were randomly re-assigned, then the same ANOVA model tested on the permuted dataset, and the maximum F-statistics value (F_{max}) recorded. After 10,000 rounds of permutation, the most extreme 5% of F_{max} were used as significance threshold. The optimum window size was the window size at which the 5% significance threshold plateaued. Sliding window positions had a step size of 1 SNP, and the median position of the sliding window was used as its genomic coordinate.

The F-statistic was visualized via sliding window, and to avoid excessive data smoothing, a window containing 10 F-statistics was used regardless of the previous ANOVA window size. A secondary sliding window of 10 smoothed-F-statistics was used for peak calling. All smoothed F-statistics within a sliding window were fit to a linear model, and the slope value estimated. The genome was scanned for slope value sign change (i.e., positive to negative slope). A peak was called as the median position of all 10 smoothed F-statistics within the first negative-sloped window after the sign change. We applied this sliding-window-regression peak-calling algorithm both from the top and the bottom of each chromosome; corresponding peaks called from both directions were then paired and identified as one peak.

Simulations

Population genetic model and statistical analysis were established and performed following the protocol of Ehrenreich *et al.* 2010 with adjustments. To simulate a simplified *A. thaliana* genome, we assumed: (a) random-mating and independent assortment of the chromosomes in the mapping population, (b) that each of the five chromosomes had a genetic map length of 100 cM**,** (c) that one of the mapping parents (P_0) was homozygous with genotype 0, and the other parent was homozygous with genotype 1 for all polymorphic positions, and therefore the F1 hybrids were heterozygous with genotype 1/0 at all SNP positions, (d) uniform global recombination rate, and (e) that the number of recombination events per chromosome per generation followed a Poisson distribution of λ equals chromosomal map distance in Morgans (i.e. λ=1 in our case). Two-thousand F2 individuals were selfed to give rise to $>10^6$ F3s, and the fecundity of each F2 individual followed a normal distribution of μ=10,000 and σ^2=1500.

QTL positions were determined by randomly drawing chromosomal positions according to a genome-wide uniform distribution. To simulate the effect of linkage in QTL detection, we arbitrarily determined that each chromosome has 1, 2, or 5 QTLs (5,10, or 25 in total), and that any adjacent QTLs must be at least 25 cM apart for 5 QTLs in total, 10 cM for 10 QTLs, and 1 cM for 25 QTLs. We assumed all QTLs of equal, additive effects. In the first set of simulation, the parent carrying genotype 1 had the alleles at QTLs that are favored by selection, and in the second set, the favorable alleles are randomly assigned between parents.

Phenotypes of individuals were quantified according to Ehrenreich *et al.* 2010 with adjustment to ploidy and homo/heterozygosity. Phenotypic values of the individuals in the segregating population were then ranked. The individuals scoring the highest 5% were designated as "survivors".

The number of QTLs detected was calculated by taking all survivors and determining their pooled allele counts at each QTL position. χ^2 tests were performed to evaluate whether at each QTL the allele frequency significantly deviates from 1:1. A QTL was deemed detected if its χ^2 statistic surpasses a critical χ^2 value (Bonferroni-corrected α=0.05).

The precision of the X-QTL mapping was estimated by calculating the distance between the predetermined QTL positions and the "marker peak" within 10 cM on each side of the actual QTL position. The "marker peaks" were acquired by scanning the survivors' entire genome with markers spaced 0.5 cM apart. The allele counts at each marker site were also χ^2 tested and corrected for multiple hypotheses testing. We then used a 1 cM-sized sliding window to scan the genome, and performed linear regression within each window. A marker peak was defined as the inflection points where the slope of the linear model changes sign. Mapping precision was derived by assigning the marker peak to its closest QTL position (if within 10 cM), and calculating the genetic distance between the marker peak and its corresponding QTL.

Whole genome re-sequencing of accessions and F3 recombinants

An independent selection experiment was repeated with 100,000 F3 seeds derived from a separate set of 3,000 F2 founders. Plating, treatments and phenotyping scheme were performed as described above. The first 500 germinants were identified at 30 HAS, which were then carefully transferred onto the surface of solid culture media and placed in environmental chambers with long-day condition. Seedlings that had the first two true leaves developed were then transferred into pots and allowed to complete their life cycle. Four-hundred-and-thirty-four plants survived transplantation, and completed their life cycle. Their rosette leaves were individually sampled before bolting. One-hundred-and-ninety-two individuals were then randomly chosen to have their DNA extracted using Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, CA). Meanwhile, DNA was isolated from the Bs-2 and Col-0 accessions from pooled leaves of 72 individuals.

Libraries were constructed using the Illumina True-seq V2 DNA Library kit (the parental accessions) or Nextera library kits (the F3 individuals) with the Nextera DNA sample preparation index kit (96 indices, Illumina, San Diego, CA). Libraries were sequenced as 100-bp paired-end reads using Illumina HiSeq 2000 (the parental accession) and HiSeq 2500 (F_3 recombinants). Sequencing reads were mapped to the TAIR version 10 *A. thaliana* reference genome ("TAIR 10" hereafter, ftp://ftp.arabidopsis.org/home/tair/Sequences/whole_chromosomes) using Burrows-Wheeler Alignment (BWA, Li and Durbin 2009) and Sequence Alignment/Map (SAMtools, Li et al. 2009). Global realignment was performed for all samples after individual mapping (Genome Analysis Tool Kit, GATK, DePristo et al. 2011). GATK was also used for insertion/deletion-realignment and SNP/variant calling. SNPs were filtered according to the

following criteria: (a) aberrant transition/transversion ratios within upper and lower 10th percentile of variant confidence and depth, strand bias, allele count, and mapping quality, and (b) mis-typed or non-polymorphic between parental accessions. A total of 560,403 filtered SNPs from the individual F3 recombinants were obtained for further analysis. SNPs in the F3s were then converted to their corresponding parental genotype (Bs-2, Col-0, or Heterozygote)

A Hidden Markov Model (HMM, Baum and Petrie 1966) was executed to improve the genotype quality and enable haplotype inference, similar to Taylor and Ehrenreich 2014. For each recombinant, SNP positions with missing genotypes were excluded. An HMM was initiated independently for each chromosome. Both internal and observational states had three possible calls: Bs-2, Col-0, and heterozygote. The initial value of the transition probability matrix was determined based on the average recombination rate of *A. thaliana* chromosomes. The initial value of the emission probability matrix was determined by estimating the genotyping error rate from the raw SNP genotypes. The most likely path of HMM was solved by the Viterbi algorithm (Viterbi 1967). The programming was performed in the R statistical programming environment (R Core Team, 2014) and with add-on package HMM (http://CRAN.R-project.org/package=HMM).

The missing genotype of SNPs that were excluded prior to HMM was then substituted by the first non-missing genotype immediately upstream. This leads to a systematic back-shift in the estimation of recombination breakpoints. We did attempt several alternative approaches, which includes relying exclusively on HMM to impute all the missing genotypes. The current approach returns the data with the least noise due to potentially spurious genotypes.

A sliding window of 500 SNPs was employed to test the allele frequency shift in 191 sequenced rapid-germinating F3s. The average allele counts of all SNPs in each window were χ^2 tested and Bonferroni-corrected. The recombination breakpoints in all 191 individuals on chromosome 1, 3, and 4 were determined by plotting the counts of the favorable allele (Col-0 for chromosome 1 and 4, and Bs-2 for chromosome 3) at each SNP position along the chromosomes. The whole region whose F-statistic were above significance threshold from the array mapping study was further investigated. Peak boundaries were defined within these regions as the block carrying the highest allele frequency skew that are not due to spurious SNP genotypes. Spurious genotypes were identified as regions showing large fluctuation of allele counts between adjacent SNPs, and with less than 20 consecutive SNPs showing abrupt increase or decrease in their allele counts.

Identifying locations of other life history QTL

We identified 20 papers that reported germination, seed trait, life history and fitness QTLs in *A. thaliana* (see Supplementary Table 3 for references). We focused on reported QTLs found on chromosomes 1, 3, and 4. There is no standard way for locating QTLs to a physical map, especially since research groups have developed their own mapping lines from different accessions, and with different genetic markers. We therefore employed some of the following approaches to locate QTLs to the physical map of reference TAIR10 genome: (a) author supplied physical coordinate, (b) searching for the marker name on TAIR (https://www.arabidopsis.org), or (c) identifying the primer sequence, if available, of their closest-linked markers to the TAIR 10 genome through BLAST (Altschul et al. 1990, Morgulis et al. 2008). It should be noted,

however, that we were unable to delimit the physical boundaries of the confidence intervals for these previously published QTLs.

RESULTS

Quantitative variation of germination speed in natural population

We measured germination speed of a worldwide collection of 18 *A. thaliana* natural accessions (see Table 1). To estimate germination speed of each accession, the germination data was plotted as cumulative germination curve and fitted to a four-parameter Hill function (El-Kassaby et al. 2008):

$$G(t) = k + \frac{a \times t^b}{c^b + t^b} \ ,$$

where *G(t)* is the cumulative germination fraction at time t, *k* is the intercept on the y axis, *a* is the maximum germination proportion, *b* is a parameter controlling the shape and steepness of the curve, and *c* is the germination speed, which in an isogenic population equals the time it takes to reach 50% germination (Figure 1b-c).

Germination speed is highly variable among natural accessions (Figure 1a-b, Table 1). It ranges from 59.1 hours (Tsu-1, CS1640) to 77.4 hours (Ler-0, CS20). The >18-hour span of germination speed is approximately 30% of the trait value. Significant ($p < 10^{-12}$) genotype (accession) effect was found via one-way ANOVA, indicating genetic control of natural variation in germination speed.

Two accessions, Bs-2 and Col-0 were phenotyped separately. Germination speed differed significantly ($p < 0.05$, Student's t-test) between Bs-2 ($c = 63.65 \pm 0.28$ hrs) and Col-0 ($c = 60.30 \pm 1.19$ hrs). Both accessions reached full germination at comparable times (95 hours-after-stratification (HAS), $p = 0.238$), but from 30 to 65 HAS the germination dynamics differed between Bs-2 and Col-0 (Figure 1c, Student's t-test of germinated fraction at a series of given time points, $0.0001<$p-values < 0.041). Interestingly, we observed a density effect in germination speed, in that the difference in germination speed between Bs-2 and Col-0 was found at high density (~40 seeds/cm^2), but not under low (~0.5 seeds/cm^2) and medium (~4 seeds/cm^2) densities (data not shown).

We chose these two accessions to develop our X-QTL mapping population, despite the small (but significant) difference in their observed germination speed. The choice of these accessions was made, in part, to provide an unbiased sample of segregating genetic variation in germination speed, rather than focus on large effect alleles. We expected that transgressive segregation in our mapping population would uncover multiple loci for this trait. We thus measured the germination curve of the F3 X-QTL mapping population descended from the Bs-2 × Col-0 cross. We used F3 populations given the difficulty in obtaining large quantities of seed for F2 populations. Two replicates of 100,000 F3 bulks were phenotyped for germination speed by counting the number of germinants that reached the cotyledon expansion stage at 8- or 12-hour intervals from 0 HAS to 72 HAS. The F3 germination curve was fitted to the 4-parameter Hill function, and the distribution of germination speed in the F3 population is shown in Figure 1d. The range of germination speed in F3 (36 HAS – 90 HAS) exceeds that observed for the

original Col-0 and Bs-2 parental accessions (60.3±1.41 HAS, 63.65±0.08 HAS), indicating transgressive segregation for germination speed in the mapping population.

The X-QTL mapping platform in Arabidopsis thaliana

To map genetic loci for germination speed and also test its efficacy as a QTL mapping tool, we developed an X-QTL mapping platform in *A. thaliana*. Figure 2 illustrates the flow of the X-QTL mapping experiment. The central focus is a genotyping platform that is sensitive to local skew in allele frequencies. For that purpose, a custom isothermal SNP genotyping microarray (Gresham et al. 2010) was designed based on the polymorphism data from the 250K Arabidopsis SNP data (V 3.01, https://github.com/Gregor-Mendel-Institute/atpolydb/blob/master/250k_snp_data/call_method_32.tar.gz, Atwell et al. 2010). Our array features 30,389 diallelic common SNPs (minor allele frequency > 0.35), uniformly spaced across the entire genome with a mean inter-SNP distance of ~4 kb. Since the array features only common SNPs, it can be used to genotype any *A. thaliana* accession.

There were 14,889 SNPs in our array that were polymorphic between Bs-2 and Col-0. We tested the specificity and sensitivity of these SNP probes by using a serial mixture of Bs-2 and Col-0 DNA samples, and we kept only probe pairs with high fidelity to allele type, and those that responded linearly to allele frequency changes. After these assays, 4,849 probe pairs passed our criteria, and were included in our downstream analysis.

The difference in log-ratio (DLR) of fluorescence intensities of Cy3/Cy5-labeled DNA between probe pairs provides a measure of allele frequency bias at each SNP position. DLR was calculated for each individual SNP probe pair during mapping. For our *A. thaliana* X-QTL mapping platform, we generated and screened two replicates of ~100,000 F3 seeds from an initial Bs-2 × Col-0 cross, which was followed by two generations of selfing.

To examine the possible power and precision of X-QTL mapping in *A. thaliana*, we undertook simulations to model different quantitative genetic architectures. An average of 1, 2, and 5 uniformly distributed QTLs per chromosome of equal effect size (i.e. 5, 10, and 25 total QTLs) were simulated (see Figure 3, and Supplementary Figure 1). The selection strength for all simulations was set to 5%, and the population size equal to 100,000 individuals.

In all simulations, the increase in heritability of the trait also increased statistical power, to detect X-QTLs through allele frequency bias, as determined by an increase in the χ^2 statistic. The X-QTL mapping scheme provided the best precision when the underlying genetic architecture was simple (i.e, 1 QTL per chromosome). In this case, the heritability values did not affect the identification and alignment of the observed peaks to the pre-set true QTL position. Often the distances between marker peaks and QTLs were less than 0.1 cM.

However, as the number of QTLs per chromosome increased, peaks for individual QTLs on each chromosome merged, and linkage between QTLs strongly confounded the ability to resolve peaks, especially when the heritability was low. With multiple QTLs per chromosome,

the statistical significance of the peaks was determined largely by the linkage of QTLs. The more QTLs in linkage, the higher the significance level of the peak clusters (e.g., chromosome 5 was simulated to have 7 QTLs under 25-QTL scenario, and had the highest peak profile; see Figure 3c). It should be noted that the varying effect size could not be estimated via X-QTL mapping, and therefore was not simulated.

Considering the relatively low heritability of life history traits, the profile of X-QTL mapping should indicate the relative genetic complexity: sharp, well-separated peaks would suggest a single locus under each peak, while broad, elevated regions might arise from multiple loci in close linkage.

Genetic architecture of germination speed

We measured germination speed in two replicates of ~100,000 F3 seeds. Within each replicate, the first ~5,000 rapid germinants (5%) were pooled at three intervals at 52 HAS, while the rest of the germinants were harvested and pooled at two time points - 60 and 72 HAS. DNA was isolated from the pooled rapid germinant cohort, and separately for the 60- and 72-hour cohorts. A DNA sample of the total population was made by proportionally mixing three DNA pools according to their DNA concentration and the number of individuals in each pool. The rapid, 60-hour and 72-hour cohorts as well as the total population were then genotyped by hybridizing their DNA to the custom isothermal SNP array, with DNA from ~2,000 F2 pooled individuals as the DNA reference.

The DLR of each SNP position was calculated for each sample. To identify significant allele frequency bias, the DLR for each cohort pools (rapid, 60- and 72-hour cohorts) and the total population were compared in a sliding-window 3-way nested ANOVA, with treatment, replicate, and dye as fixed effects. The significance thresholds were determined via permutation.

Germinants that represent the later germinating cohorts (60- and 72-hour cohorts) did not show any significant allele frequency bias in genome-wide X-QTL scans (see Figure 4b). In contrast, for the rapid germination cohort, we identified three distinct regions on chromosomes 1, 3, and 4 that showed significant allele frequency bias (Figure 4a). In each case, the peaks were not sharp, but showed a broad area of elevated significance. The region on chromosome 1 has three small peaks, and the left arm of chromosome 3 has a single prominent peak with a smaller peak towards its tip. The right arm of chromosome 4 has the highest deviation in allele frequency, and shows a single peak towards the centromeric region and an elevated, unresolved peak cluster located more to the middle of the right arm. The peaks reported in X-QTL mapping are positions whose allele frequency skewed the most in response to selection. It should be noted that those regions that were distant from these peaks showed approximately equal frequencies of parental alleles.

Col-0 allele is favored for rapid germination in the chromosome 1 and 4 loci, while the Bs-2 allele is favored on chromosome 3. Both parents contributed genetically to rapid germination phenotype, which likely explains the transgressive segregation of germination speed in the F3 generation.

Germination speed X-QTLs confirmed by sequencing individual F3 rapid germinants

To confirm the X-QTL mapping results, we repeated the selection experiment in a new F3 mapping population of ~100,000 seeds derived from a separate set of 3,000 F2 progenitors. We identified the first 500 germinants (top 0.5%) in this population, and 192 individuals from this set of 500 were randomly selected and their whole genome re-sequenced to an average of about 3-4X sequencing coverage for each individual. The genomes of Bs-2 and Col-0 were independently re-sequenced to >100X coverage.

We were able to successfully map the sequencing reads from 191 F3 recombinants to the TAIR 10 reference genome (Col-0 accession). Sequencing reads for both the 191 recombinants and the parents were combined, and SNPs were identified. Altogether 560,403 high-quality SNPs homozygous in the Col-0 and Bs-2 parental accessions segregated in the 191 F3 recombinants. The genotypes of each recombinant at these SNP positions were classified based on their parental origins (Bs-2, Col-0, or heterozygous). To improve the quality of called genotypes and enable haplotype inference, we imputed the SNP genotype with Hidden Markov Modeling (HMM, Baum and Petrie 1966).

Allele counts for each SNP position having less than 10% missing data were determined. To detect regions of significant allele frequency skew, we again employed an overlapping sliding window of 500 SNPs in size. Bs-2 and Col-0 allele counts were averaged among the 500 SNPs, and a Pearson's χ^2 test was performed on the averaged allele counts (Figure 5). The permutation threshold was so low that nearly the entire genome was above the significance level, we therefore employed the much more stringent Bonferroni correction.

The allele frequency shift in the population of F3 recombinants we sequenced resembles the X-QTL mapping results using the genotyping isothermal microarray. Regions with strong bias in allele frequencies on chromosome 1 and 4 were well above the Bonferroni-corrected significance threshold, and the highest points in these regions coincide with the position of the highest peaks in the array-based X-QTL map. The region on chromosome 3 also showed a skew in allele frequency, although the p-value did not exceed the Bonferroni threshold.

For chromosome 1 and 4, the regions corresponding to the X-QTL regions showed a Col-0 allele frequency of ~70% (a 20% allele frequency change) and ~76% (a 26% allele frequency change), respectively. The skew indicates that the extensive blocks of the Col-0 haplotype on chromosome 1 and 4 in the recombinants are strongly favored under selection pressure for rapid germination. Even the non-significant region on chromosome 3 displayed a 10% allele frequency shift towards the early germination Bs-2 allele.

Based on the X-QTL mapping results, we used an approach previously applied to yeast (Linder et al., 2016) and plotted the allele count for each SNP across the genomic regions associated with germination speed QTLs on chromosomes 1, 3 and 4 (Figure 6). This provides a higher resolution view of the skew in allele frequency across these genomic regions. In all three genomic regions, we can observe multiple peaks, although it is unclear to what extent this may represent sampling noise in either the recombinant individuals or the HMM imputation. Nevertheless, visual inspection each of these three regions do show at least 2-4 broad peaks of allele skew, suggesting that each of these genomic regions may contain more than one X-QTL peak. The highest peak for each genomic region ranges in size from ~56 kb to ~232 kb.

Germination speed of positional candidate genes

We acquired the T-DNA insertion lines of these 4 candidate genes from the Arabidopsis Information Resource (www.arabidopsis.org), and measured their germination speed using their original genetic background accession as control. These genes are *GTG1* (AT1G64990), *AZF2* (AT3G19580), *CPL5* (AT3G19600) and *FBA2* (AT4G38970). Two genes knockouts exhibited significant delay in germination. *AZF2* (p=0.0103, Student's t-test) and *FBA2* (p=0.0002, Student's t-test). The candidate genes and T-DNA line tested are listed in Supplementary Table 2

DISCUSSION

Germination speed is an intrinsic trait associated with early life history stages in plants, but little is known about this trait and its genetic basis. Using the model plant system *A. thaliana*, we surveyed natural variation in this trait, and demonstrated substantial diversity in germination speed among *A. thaliana* accessions (Figure 1).

The relationship of this trait to other life history and fitness traits is of interest, since it would suggest possible constraints and consequences of evolution in this trait. Integrating our data with previously-published work (Ehrenreich et al. 2009), we did find a negative correlation between germination speed and rosette leaf number among 13 accessions in both long-day and short-day conditions ($r = -0.35$ and $r = -0.44$ respectively, Spearman's correlation, see Supplementary Figure 2a-b), though these correlations were non-significant ($p < 0.25$).

We also found a negative correlation between germination speed and seed weight among 16 *A. thaliana* accessions (r = -0.23, Spearman's correlation, see Supplementary Figure 2c). As in the case of rosette leaf number, the correlation was not significant in our small sample size ($p < 0.35$). Nevertheless, the trend was consistent with previous studies that showed intraspecific selection for rapid germination can lead to an increase in seed size and evolution of invasiveness (Graebner et al. 2012, Tanveer et al. 2013, Torices et al. 2013). Given that ecological studies on germination speed have suggested a strong relationship with seed size and competitiveness (Dubois and Cheptou 2012, Lonnberg and Eriksson 2013, Afonso et al. 2014, Leverett and Jolls 2014), our finding may suggest that the competitiveness of larger seeds is partly mediated by rapid germination. We also found that several seed size and weight QTLs in *A. thaliana* (Gnan et al. 2014, Herridge et al. 2011, Alonso-Blanco et al. 1999, Moore et al. 2013, Guo et al. 2016) overlap with our germination speed X-QTLs. Interestingly, rapid germination in interspecific comparisons is associated with smaller (not larger) seeds with a high embryo-to-seed (E:S) ratio (Vandelook et al. 2012, Wang et al. 2012, Parsons et al. 2014). This is likely due to a trade-off between seed size and seed number, and stronger competition under high post-dispersal density. This conforms well with the characteristic of species in the *Brassicaceae* family, including *A. thaliana,* which have relatively small seeds and a very high E: S ratio compared to many other species (Forbis et al. 2002, Parsons et al. 2014).

Our study provided one of the few demonstrations of the genetic basis for germination speed variation (Foolad et al. 1999, Betty et al. 2000, Clerkx et al. 2004). By implementing the new technique of X-QTL mapping, we were able to identify three distinct genomic regions that controlled germination speed in this model plant species.

The role of pleiotropy in integrating or constraining life history evolution has long been an interest in evolutionary ecology (Stearns 1992, Braendle et al. 2011). For example, it has been demonstrated that an *Arabidopsis* flowering time regulatory gene, *FLC,* also regulates dormancy (Chiang et al. 2009), and therefore serves as a key element in the concerted expression of life history traits (Li et al. 2014). We examined the positional relationship between our germination speed X-QTLs and other previously-identified life history or seed trait QTLs. Determining whether our X-QTLs overlap with previously identified QTLs, however, is hampered by the low resolution of both our and previous QTL maps; thus any identified positional overlap should take this into account.

A summary of these positional overlaps can be found in Figure 7 and Table 2. We did find that our germination speed X-QTLs on chromosomes 1 and 4 overlap with a previously identified QTL for germination in water (Vallejo et al. 2010). Moreover, we found positional overlap between germination speed and maximum germination QTLs (Joosen et al. 2012), as well as a few minor dormancy QTLs (Clerkx et al. 2004, Bentsink et al. 2010).

We also find that flowering time (Brachi et al. 2010, Li et al. 2010, Grillo et al. 2013, Dittmar et al. 2014, Li, Cheng et al. 2014), morphological traits (Simon et al. 2008, Chardon et al. 2014, Oakley et al. 2014) and fitness QTLs (Ågren et al. 2013) overlap with our germination speed X-QTL regions. This suggests the possibility that germination speed could share a common genetic basis with these other traits. However, it should be stressed that the molecular identification and isolation of the relevant genes will be necessary to conclusively demonstrate the role of pleiotropy or close linkage in the co-localization of these X-QTLs.

Interestingly, several major dormancy QTLs (*DOG1* on chromosome 5, and *DOG6* on the right arm of chromosome 3) and flowering time loci (*FLC* on chromosome 5, and *FRI* on the left arm of chromosome 4) are absent in the germination speed X-QTL regions. This lack of overlap suggests germination speed is at least partially independent of other life history components, rather than subject to genetic co-variation with the other major life history traits that have been previously examined.

The availability of whole genome sequence for the parental Col-0 and Bs-2 accessions and individual rapid germinating F3s allowed us to identify recombination breakpoints within the X-QTL regions. This in turn facilitated our search for gene candidates for germination speed. We focused our attention on the highest peak for each genomic region based on the genome re-sequencing of 191 rapid germinating individuals selected from a population of ~100,000 F3 recombinants. While it is possible that there may be more than one X-QTL in each of these genomic regions (see Figure 6), we took a conservative approach and only considered the highest peak for each region for further scrutiny.

The highest peaks defined by the re-sequencing analysis had a total of 70 genes (chromosome 1), 19 genes (chromosome 3) and 22 genes (chromosome 4) located under each peak (see Figure 6 and Supplementary Table 1). We examined the TAIR 10 gene annotation and found 4 genes under these highest peak regions that were involved in ABA response or signaling. Two genes had homozygous T-DNA knockout lines with significant effects on germination speed. *AZF2* encodes a zinc-finger protein and has been described as a negative regulator of ABA signaling in seeds (Drechsel et al. 2010). The same homozygous T-DNA knockout line

showed hypersensitivity to ABA and delayed germination under ABA treatment. *FBA2* (AT4G38970) encodes an aldolase whose phosphorylation state is modulated in response to ABA in seeds (Ghelis et al. 2008), but no germination phenotype has been previously described.

While these candidate genes are promising, more detailed fine-mapping or functional studies are needed to precisely identify if any of these, or indeed other genes in these genomic regions, are responsible for the seed germination X-QTLs. Our results can also be extended in future work to examine other mapping populations, including those between parental accessions with large differences in germination speed; this would provide insights into possible large effect alleles, as compared to the unbiased approach we have taken in this study. It would also be interesting to examine germination speed under field conditions, and explore the genetics of this life history trait under natural environments.

Our X-QTL mapping study made use of a new mapping platform that allows the screening of $>10^5$ individuals for potentially greater power. Mapping complex traits with a combination of selection and bulk-segregant analysis (BSA) has gained attention over the past few years with the advent of new genomics technologies (e.g. Lai et al. 2007, Friedman et al. 2015, Guo et al. 2016, for review of BSA and next-generation sequencing, see Schlötterer et al. 2014). The pooling strategy inherent in X-QTL mapping, for example, enables the application of high-throughput genotyping, which promises less labor and higher marker densities. Although it has proven most powerful in yeast (Ehrenreich et al. 2010, Ehrenreich et al 2012), several studies have also established this strategy in multicellular eukaryotic systems. QTLs affecting life span in *Drosophila melanogaster*(Lai et al. 2007), annual/perennial variation in *Mimulus guttatus* (Friedman et al. 2015), and seedling vigor (Takagi et al. 2013) and cold tolerance in *O. sativa*

(Yang et al. 2013) have been identified by similar strategies. In *A. thaliana*, this approach has been modified as target-enriched extreme QTL (TEX-QTL) mapping to use sequencing of targeted markers to allow for greater sequencing depth and precision in allele frequency estimation (Guo et al. 2016).

Except for the *Drosophila* mapping study, all other studies that employed a bulk-segregant mapping design also resulted in broad peaks with low resolution. In general, mapping resolution should increase with population size, and our study has thus far been the largest study (~10^5 individuals) for a multicellular organism. Despite the very large mapping population size, our mapping resolution was still inferior to that observed in the yeast X-QTL mapping; defined peaks through re-sequencing of individual recombinants still contain ~20-70 genes that need to be further analyzed. Several factors may contribute to this, though the major factor is likely the disparity in recombination rate between *S. cerevisiae* and *A. thaliana.* The physical distance on the yeast genetic map is ~100 times shorter than that of *A. thaliana*, consistent with the greater recombination rates in the former. Although in principle this method may provide more power in detecting QTLs of smaller effect sizes, when it comes to *A. thaliana* it might be more suited for mapping traits with a lower number of unlinked causal genes if the goal is to obtain greater mapping resolution. Parenthetically, given our simulation results, our inability to resolve our X-QTLs might indicate that this life history trait is governed by multiple, linked loci.

Despite the limitations of X-QTL mapping, it is capable of rapidly providing information on QTLs in *A. thaliana* for early and selectable developmental traits, which has also been demonstrated in other similar approaches (Guo et al. 2016, Lai et al. 2007,Yang et al. 2013,

Takagi et al. 2013, Wolyn et al, 2004). The large population size used in this approach also provides greater power to detect QTLs of small effect. Our analysis of germination speed using this platform has resulted in the identification of key genomic regions in the *A. thaliana* genome that respond to strong selection for rapid germination, and allowed us to examine these QTLs in the context of other life history and morphological QTLs in the species. In principle, this method could also be employed to examine germination speed and other seed traits in a wide range of plant taxa. Such effort would allow for comparative QTL mapping for traits associated with seed traits and associated this important life history transition.

ACKNOWLEDGMENTS

We would like to thank Evan Baugh for the script to transform TAIR 8 genome coordinate to TAIR 10, Xu Wang for help in setting up Volocity image analysis software, Anne Plessis for the germination mutant gene list, and Sue Rhee for discussion. This work was supported in part by grants from the NSF Population and Evolutionary Processes and the Plant Genome Research Program.

REFERENCES

Afonso, A., S. l. Castro, J. o. Loureiro, L. Mota, J. Cerca de Oliveira and R. n. Torices (2014). The effects of achene type and germination time on plant performance in the heterocarpic *Anacyclus clavatus* (*Asteraceae*). *Amer. J. Bot.* **101**: 892-898.

Ågren, J., C. G. Oakley, J. K. McKay, J. T. Lovell and D. W. Schemske (2013). Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **110**: 21077-21082.

Alonso-Blanco, C., H. Blankestijn-de Vries, C. J. Hanhart and M. Koornneef (1999). Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **96**: 4710-4717.

Alonso-Blanco, C., Bentsink, L., Hanhart, C.J., Blankestijn-de Vries, H. and Koornneef, M. (2003). Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis. Genetics* **164**:711-729.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.

Atwell, S., Huang, Y., Vilhjamsson, B., Willems, G. et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465: 627–631.

Baum, L. E. and T. Petrie (1966). Statistical inference for probabilistic functions of finite state Markov Chains. *Ann. Math. Statist.* 37: 1554-1563.

Bentsink, L., Jowett, J., Hanhart, C.J. and Koornneef, M. (2006). Cloning of *DOG1*, a quantitative trait locus controlling seed dormancy in *Arabidopsis. Proc. Natl. Acad. Sci. USA* **103**: 17042-17047.

Bentsink, L., Hanson, J., Hanhart, C., Blankestijn-de Vries, H. *et al*. (2010). Natural variation for seed dormancy in *Arabidopsis* is regulated by additive genetic and molecular pathways. *Proc. Natl. Acad. Sci. USA* **107**: 4264-4269.

Bettey, M., W. E. Finch-Savage, G. J. King and J. R. Lynn (2000). Quantitative genetic analysis of seed vigour and pre-emergence seedling growth traits in *Brassica oleracea*. *New Phytologist* **14**8: 277-286.

Bomblies, K., L. Yant, R. A. Laitinen *et al.* (2010). Local-Scale Patterns of Genetic Variability, Outcrossing, and Spatial Structure in Natural Stands of Arabidopsis thaliana. *PLoS Genet*. **6**: e1000890.

Brachi, B., N. Faure, M. Horton, E. Flahauw, A. Vazquez, M. Nordborg, J. Bergelson, J. Cuguen and F. Roux (2010). Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *Plos Genetics* **6**: e1000940

Braendle, C., A. Heyland and T. Flatt (2011). Integrating mechanistic and evolutionary analysis of life history variation. In *Mechanisms of Life History Evolution: The Genetics and Physiology*

of Life History Traits and Trade-Offs. (eds., T. Flatt and A. Heyland) Oxford University Press, Oxford.

Burghardt, L.T., Metcalf, C.J.E., Wilczek, A.M. Schmitt, J., and Kathleen Donohue (2015). Modeling the influence of genetic and environmental variation on the expression of plant life cycles across landscapes. *Amer. Nat.* **185**: 212-227.

Chardon, F., S. Jasinski, M. Durandet, A. Lecureuil, F. Soulay, M. Bedu, P. Guerche and C. Masclaux-Daubresse (2014). QTL meta-analysis in *Arabidopsis* reveals an interaction between leaf senescence and resource allocation to seeds. *J. Exp. Bot.* **65**: 3949-3962.

Chiang, G. C., D. Barua, E. M. Kramer, R. M. Amasino and K. Donohue (2009). Major flowering time gene, *FLC*, regulates seed germination in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **106**: 11661-11666.

Clerkx, E.J.M., El-Lithy, M.E., Vierling, E., Ruys, G.J., Blankestijn-De Vries, H., Groot, S.P., Vreugdenhil, D., Koornneef, M. (2004). Analysis of natural allelic variation of *Arabidopsis* seed germination and seed longevity traits between the accessions Ler and Shakdara, using a new recombinant inbred line population. *Plant Physiol*. **135**: 432-443.

Debieu, M., Tang, C., Stich, B., Sikosek, T., Effgen, S., Josephs, E., Schmitt, J., Nordborg, M., Koornneef, M., De Meaux, J. (2013). Co-variation between seed dormancy, growth rate and flowering time changes with latitude in *Arabidopsis thaliana. PLoS One* **8**: e61075

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler and M. J. Daly (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**: 491-498.

Dias, P.M., Brunel-Muguet, S., Dürr, C., Huguet, T., Demilly, D., Wagner, M.H., Teulat-Merah, B. (2011). QTL analysis of seed germination and pre-emergence growth at extreme temperatures in *Medicago truncatula. Theor. Appl. Genetics* **122**: 429-444.

Dittmar, E. L., C. G. Oakley, J. Agren and D. W. Schemske (2014). Flowering time QTL in natural populations of Arabidopsis thaliana and implications for their adaptive value. *Mol. Ecol.* **23**: 4291-4303.

Donohue, K. (2002). Germination timing influences natural selection on life-history characters in *Arabidopsis thaliana. Ecology* **83**: 1006-1016.

Donohue, K., Dorn, L., Griffith, E., Kim, S. *et al*. (2005a). Niche construction through germination cueing: Life-history responses to timing of germination in *Arabidopsis thaliana. Evolution* **59**: 771-785.

Donohue, K., Dorn, L., Griffith, C., Kim, S., Aguilera, C., Polisetty, C., Schmitt, J. (2005b). The evolutionary ecology of seed germination of Arabidopsis thaliana: Variable natural selection on germination timing. *Evolution* **59**: 758-770.

Donohue, K., Dorn, L., Griffith, C., Kim, S., Aguilera, C., Polisetty, C., Schmitt, J. (2005c). Environmental and genetic influences on the germination of *Arabidopsis thaliana* in the field. *Evolution* **59**: 740 - 757.

Donohue, K., de Casas, R., Burghardt, L., Kovach, K., Willis, C.G. (2010). Germination, postgermination adaptation, and species ecological ranges. *Ann. Rev. Ecol. Evol. Syst.* **41**: 293 - 319.

Drechsel, G., S. Raab and S. Hoth (2010). Arabidopsis zinc-finger protein 2 is a negative regulator of ABA signaling during seed germination. *J. Plant Physiol.* **167**: 1418-1421.

Dubois, J. and P. O. Cheptou (2012). Competition/colonization syndrome mediated by early germination in non-dispersing achenes in the heteromorphic species *Crepis sancta*. *Ann. Bot.* **110**: 1245-1251.

Ehrenreich, I. M., Y. Hanzawa, L. Chou, J. L. Roe, P. X. Kover and M. D. Purugganan (2009). Candidate gene association mapping of *Arabidopsis* flowering time. *Genetics* **183**: 325-335.

Ehrenreich, I. M., N. Torabi, Y. Jia, J. Kent, S. Martis, J. A. Shapiro, D. Gresham, A. A. Caudy and L. Kruglyak (2010). Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**: 1039-1042.

El-Kassaby, Y. A., I. Moss, D. Kolotelo and M. Stoehr (2008). Seed germination: Mathematical representation and parameters extraction. *Forest Science* **54**: 220-227.

Finch-Savage, W.E. and G. Leubner-Metzgar (2006). Seed dormancy and the control of germination. *New Phytol.* **171**: 501-523.

Foolad, M. R., G. Y. Lin and F. Q. Chen (1999). Comparison of QTLs for seed germination under non-stress, cold stress and salt stress in tomato. *Plant Breeding* **118**: 167-173.

Forbis, T. A., S. K. Floyd and A. d. Queiroz (2002). The evolution of embryo size in angiosperms and other seed plants: implications for the evolution of seed dormancy. *Evolution* **56**: 2112-2125.

Friedman, J., A. D. Twyford, J. H. Willis and B. K. Blackman (2015). The extent and genetic basis of phenotypic divergence in life history traits in *Mimulus guttatus. Mol. Ecol.* **24**: 111-122.

Ghelis, T., G. Bolbach, G. Clodic, Y. Habricot, E. Miginiac, B. Sotta and E. Jeannette (2008). Protein tyrosine kinases and protein tyrosine phosphatases are involved in abscisic acid-dependent processes in *Arabidopsis* seeds and suspension sells. *Plant Physiol.* **148**: 1668-1680.

Gnan, S., A. Priest and P. X. Kover (2014). The genetic basis of natural variation in seed size and seed number and their trade-off using *Arabidopsis thaliana* MAGIC Lines. *Genetics* **198**: 1751.

Graebner, R. C., R. M. Callaway and D. Montesinos (2012). Invasive species grows faster, competes better, and shows greater evolution toward increased seed size and growth than exotic non-invasive congeners. *Plant Ecol.* **213**: 545-553.

Gresham, D., B. Curry, A. Ward, D. B. Gordon, L. Brizuela, L. Kruglyak and D. Botstein (2010). Optimized detection of sequence variation in heterozygous genomes using DNA microarrays with isothermal-melting probes. *Proc. Natl. Acad. Sci. USA.* **107**: 1482-1487.

Grillo, M. A., C. B. Li, M. Hammond, L. J. Wang and D. W. Schemske (2013). Genetic architecture of flowering time differentiation between locally adapted populations of *Arabidopsis thaliana*. *New Phytol.* **197**: 1321-1331.

Guo, J., J. Fan, B. A. Hauser and S. Y. Rhee (2016). Target enrichment improves mapping of complex traits by deep sequencing. *G3: Genes | Genomes | Genetics*, **6** (1): 67-77

Hayashi, E., Aoyama, N. and Still, D.W. (2008). Quantitative trait loci associated with lettuce seed germination under different temperature and light environments. *Genome* **51**: 928-947.

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

Huang, X., Schmitt, J., Dorn, L.A., Griffith, C. *et al.* (2010). The earliest stages of adaptation in an experimental plant population: strong selection on QTLs for seed dormancy. *Mol. Ecol*. **19**: 1335-1351

Joosen, R. V., J. Kodde, L. A. Willems, W. Ligterink, L. H. van der Plas and H. W. Hilhorst (2010). GERMINATOR: a software package for high-throughput scoring and curve fitting of *Arabidopsis* seed germination. *Plant J.* **62**: 148-159.

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

Koornneef, M., C. Alonso-Blanco, D. Vreugdenhil (2004). Naturally occurring genetic variation in *Arabidopsis thaliana. Annu. Rev. Plant Biol.* **55**: 141-172.

Lai, C. Q., J. Leips, W. Zou, J. F. Roberts, K. R. Wollenberg, L. D. Parnell, Z. B. Zeng, J. M. Ordovas and T. F. C. Mackay (2007). Speed-mapping quantitative trait loci using microarrays. *Nature Methods* **4**: 839-841.

Leverett, L. D. and C. L. Jolls (2014). Cryptic seed heteromorphism in *Packera tomentosa* (Asteraceae): differences in mass and germination. *Plant Species Biol.* **29**: 169-180.

Li, H. and R. Durbin (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis and R. Durbin (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li, Y., Y. Huang, J. Bergelson, M. Nordborg and J. O. Borevitz (2010). Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA.* **107**: 21199-21204.

Li, Y., R. Y. Cheng, K. A. Spokas, A. A. Palmer and J. O. Borevitz (2014). Genetic variation for life history sensitivity to seasonal warming in *Arabidopsis thaliana*. *Genetics* **196**: 569-.

Linder, R., F. Seidl, K. Ha and I.M. Ehrenreich (2016). The complex genetic and molecular basis of a model quantitative trait. *Mol Biol Cell* **27**(1): 209-218.

Lonnberg, K. and O. Eriksson (2013). Rules of the seed size game: contests between large-seeded and small-seeded species. *Oikos* **122**: 1080-1084.

Mauricio R (2001). Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology. *Nat. Rev. Genetics* **2**: 370-381.

Miller, T.E., Winn, A., Schemske, D. (1994). The effects of density and spatial distribution on selection for emergence time in *Prunella vulgaris. Amer. J. Botany* **81**: 1-6.

Montesinos-Navarro, A., Pico, F.X., Tonsor, S.J. (2012) Clinal variation in seed traits influencing life cycle timing in *Arabidopsis thaliana.2 Evolution* **66**: 3417 – 3431.

Moore, C. R., D. S. Gronwall, N. D. Miller and E. P. Spalding (2013). Mapping quantitative trait loci affecting *Arabidopsis thaliana* seed morphology features extracted computationally from images. *G3* **3**: 109-118.

Morgulis, A., G. Coulouris, Y. Raytselis, T. L. Madden, R. Agarwala and A. A. Schaffer (2008). Database indexing for production MegaBLAST searches *Bioinformatics* **24**: 1757-1764.

Oakley, C. G., J. Agren, R. A. Atchison and D. W. Schemske (2014). QTL mapping of freezing tolerance: links to fitness and adaptive trade-offs. *Molecular Ecol.* **23**: 4304-4315.

Parsons, R.F. (2012) Incidence and ecology of very fast germination. *Seed Science Research* **22**: 161-167.

Parsons, R. F., F. Vandelook and S. B. Janssens (2014). Very fast germination: additional records and relationship to embryo size and phylogeny. *Seed Science Research* **24**: 159-163.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ranke, J. (2014). chemCal: Calibration functions for analytical chemistry. R package version 0.1-34.

Rees, M. (1996). Evolutionary ecology of seed dormancy and seed size. *Phil. Trans. Royal Soc. London B.* **351**: 1299-1308.

Schlötterer, C., R. Tobler, R. Kofler and V. Nolte (2014). Sequencing pools of individuals [mdash] mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet*. **15**: 749-763.

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

Simons, A.M. (2009) Fluctuating natural selection accounts for the evolution of diversification bet hedging. *Proc. Royal Soc. B* **276**: 1987-1992.

Stearns, S. C. (2000). Life history evolution: successes, limitations, and prospects. *Naturwiss.* **87**: 476-486.

Takagi, H., A. Abe, K. Yoshida, S. Kosugi, S. Natsume, C. Mitsuoka, A. Uemura, H. Utsushi, M. Tamiru, S. Takuno, H. Innan, L. M. Cano, S. Kamoun and R. Terauchi (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* **74**: 174-183.

Tanveer, A., M. Tasneem, A. Khaliq, M. M. Javaid and M. N. Chaudhry (2013). Influence of seed size and ecological factors on the germination and emergence of field bindweed (*Convolvulus arvensis*). *Planta Daninha* **31**: 39-51.

Taylor, M. B. and I. M. Ehrenreich (2014). Genetic interactions involving five or more genes contribute to a complex trait in yeast. *PLoS Genet.* **10**: e1004324.

Torices, R. n., A. Agudo and I. Álvarez (2013). Not only size matters: achene morphology affects time of seedling emergence in three heterocarpic species of *Anacyclus* (Anthemideae, Asteraceae). *Anales del Jardín Botánico de Madrid* **70**:48-55.

Vallejo, A. J., M. J. Yanovsky and J. F. Botto (2010). Germination variation in *Arabidopsis thaliana* accessions under moderate osmotic and salt stresses. *Ann. Bot.* **106**: 833-842.

Vandelook, F., M. Verdu and O. Honnay (2012). The role of seed traits in determining the phylogenetic structure of temperate plant communities. *Ann*. Bot. **110**: 629-636.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions* **13**: 260-269.

Vuorisalo, T.O. and Mutikainen, P.K. (1999). *Life history evolution in plants*. Dordrecht ; Boston, Kluwer Academic Publishers.

Wang, J. H., W. Chen, C. C. Baskin, J. M. Baskin, X. L. Cui, Y. Zhang, W. Y. Qiang and G. Z. Du (2012). Variation in seed germination of 86 subalpine forest species from the eastern Tibetan Plateau: phylogeny and life-history correlates. *Ecol. Res.* **27**: 453-465.

Weinig, C. (2000). Differing selection in alternative competitive environments: shade-avoidance responses and germination timing. *Evolution* **54**: 124-136.

Wilczek, A.M., Roe, J.L., Knapp, M.C., Cooper, M.D. *et al.* (2009) Effects of genetic perturbation on seasonal life history plasticity. *Science* **323**: 930-934.

Wolyn, D. J., J. O. Borevitz. O. Loudet, C. Schwartz, J. Maloof, J. R. Ecker, C. C. Berry and J. Chory (2004) Light-response quantitative trait loci identified with composite interval and eXtreme array mapping in *Arabidopsis thaliana. Genetics* **167**: 907- 917

Yang, Z., Huang, D., Tamg, W., Zheng, Y., Liang, K., Cutler, A., Wu, W. (2013). Mapping of quantitative trait loci underlying cold tolerance in rice seedlings via high-throughput sequencing of pooled extremes. *PLoS One* **8**: e68433

DATA ACCESSIBILITY

Data uploaded to Dryad doi:10.5061/dryad.kt5nq

- Germination scoring of 18 natural accessions
- Iso-thermal SNP array probe sequences
- Bulk-segregant microarray genotyping data

Data uploaded to Sequence Read Archive: PRJNA329531

- Whole genome sequence of parental accessions Bs-2 and Col-0
- Whole genome sequence of 191 rapid-germinating F_3 recombinants

AUTHOR CONTRIBUTIONS

W.Y. and M.D.P. conceived the study. I.E. helped develop resources for study. W.Y. and D.S. collected the data, W.Y. and J.F. analyzed the data. W.Y. and M.D.P. wrote the paper.

Table 1. *A. thaliana* Natural accessions phenotyped for germination speed

Accession	CS #	Name	Locality	Germination Speed/h	SD
Ak-1	6602	Achkarren	Germany	66.54	2.61
Bs-2*	6628	Basil	Switzerland	63.64	0.28
Bu-18	6649	Buchschlag	Germany	73.01	0.89
Col-0*	6673	Columbia	United States	60.3	1.19
Fi-0	6704	Frickhofen	Germany	74.18	2.36
Gr-1	6723	Graz	Austria	68.15	0.6
Hl-0	6737	Holtensen	Germany	72.09	4.05
HOG	6178	Hodja-Obi-Garm	Tadjikistan	65.29	1.23
Ler-0	20	Landsberg erecta	Germany	75.78	1.45
Ob-0	6816	Oberursel	Germany	62.58	0.52
Oy-0	6824	Oystese	Norway	67.84	0.43
Pa-1	6825	Palermo	Italy	74.63	1.54
Tsu-1	1640	Tsushima	Japan	58.93	1.53
Uk-1	6879	Umkirch	Germany	68.71	0.13
Ws	915	Wassilewskija	Russia	71.85	1.35
X-0	6898	X-0	N/A	65.28	1.08
XX-0	6899	XX-0	N/A	68.89	2.36
Zu-1	6903	Zurich	Switzerland	67.93	1.32

*Accessions whose germination speed was phenotyped separately from the rest, hence excluded from correlation analysis

Table 2. Seed size, inflorescence development, and fitness QTLs that overlaps with germination speed X-QTLs by genetic map

Author	Trait	Closest Marker(s)	Chrom o-some	cM	Description
Alonso-Blanco et al, 1999	Seed weight	GD.86L	1	24	Correlation of seed size loci with other life history traits in Ler × Cvi RILs
		GB.120C-Col/GAPC	3	3	
		CD.329C-Col	4	48	
		GB.490C	4	69	
	Seed length	GD.86L	1	24	
		EG.75L	3	8	
		FD.167L-Col	4	53	
		GB.750C	4	76	
	Ovule length	CD.329C-Col	4	50	
Moore et al, 2013	Seed area RIL1	c1.loc74	1	74	Seed area loci in two sets of Ler × Cvi RILs and NILs
		c3.loc4	3	4	
		c4.loc64	4	64	
	Seed area RIL2	c1.loc73	1	73	
		C3.loc1	3	1	
		DF.108L-Col	4	40.6	
		GB.750C	4	67.8	
	Seed area NIL	c1.loc72	1	72.2	
		c3.loc3	3	2.9	
Guo et al, 2016	Seed weight		1	107.9-114.1	Seed weight loci in Sha × Tsu F2 population via TEX-QTL mapping
			4	52.8-57.2	
			4	62.3-68.3	

Table 2(continue)

Author	Trait	Closest Marker(s)	Chrom o-some	cM	Description
Ungerer et al, 2002	Length of reproductive phase of main axis	mi185	1	113.2	Inflorescence development QTLs in Ler × Col RILs
	Rosette leaves at bolting	mi185	1	108.7	
	Plant height	CDs11	4	72.3	
	Length of reproductive phase of main axis	HH.375L	1	90.7	Inflorescence development QTLs in Ler × Cvi RILs
	Rosette diameter	DF.77C	3	0.0	
		GH.390L	3	21.5	
		g4539	4	43.4	
	Plant height	GD.160C	1	86.5	
		FD.111L	3	12.1	
	Axillary fruits	GB.120C	3	4.5	
	Total fruits	GB.120C	3	3.0	
Weinig et al, 2003	RI spring fecundity	mi208	1	78.8	Geographical heterogeneous selection of fitness components, Ler × Col RILs
		g4564a	4	46.0	
	NC fall over-winter survivorship	agp64	1	129.4	

FIGURE LEGENDS

Figure 1. Natural variation of germination speed in *Arabidopsis thaliana*. (a) Natural variation in germination speed in 16 natural accessions. Germination speed was measured as the time for a population of 2,000 seeds to reach 50% germination. Shown are mean germination speed for each accession, error bar indicates standard error. (b) Cumulative germination plot of the fastest (Tsu-1) and the slowest (Ler-0) germinating accessions. Gray dots show the percentage germinated at each sampling time point, with the center of the dot indicating mean germination percentage, and the error bar showing the standard error. The black lines show the fit using a 4-parameter Hill function, with 100 iterations using the least sum-of-square method. (c) Cumulative germination plot of the mapping parents Bs-2 and Col-0. Both germination speed and germination percentage from 30-65 HAS significantly differed between the accessions ($p<0.05$). (d) Histogram showing the distribution of germination speed in ~100,000 Bs-2 × Col-0 F3 population. Data for plotting the histogram was derived from fitting raw F3 germination data to the Hill function. Most individuals germinate between 60-65 HAS, with early germinants at <40 hours and late germinants >70 hours. The germination speed of Bs-2 and Col-0 are indicated by upward triangles.

Figure 2. The *A. thaliana* Extreme-QTL (X-QTL) mapping procedure. Two *A. thaliana* natural accessions were intercrossed, and 2000 F2s were used as founder to bulk ~10^6 F3 segregants for selection screening. For each replicate, ~100,000 F3 seed was subjected to selective screening, and the 5% population with extreme phenotypes were pooled and genotyped independently from the entire population (no selection) using a custom-designed iso-thermal SNP genotyping microarray. The genome-wide allele frequency of the extreme population was then

compared against that of the entire population. Regions with significant allele frequency skew within the extreme population are identified as the extreme-QTLs (X-QTLs)

Figure 3. X-QTL mapping power analysis with consideration of genetic architecture complexity. Simulations were performed to visualize the power and precision of X-QTL mapping in *A.thaliana*. Scenarios of an average of (a) one, (b) two, and (c) five QTLs per chromosome were simulated. Under each scenario, 100,000 F3 were derived from 2000 F2 progenitors, and individuals of the most extreme 5% phenotypic value were designated as "survivors". The survivors are then "genotyped" with 1000 equidistant markers along the entire genome. The figures show the χ^2 value (× 1000) of a Pearson's goodness-of-fit test for allele frequency bias in the survivors. The positions of QTLs are indicated as the dark tick marks within the gray chromosome bars on top of each panel. A series of trait heritabilities were simulated, denoted with different shades of lines. All scenarios were repeated 50 times.

Figure 4. The genetic architecture of germination speed in *A. thaliana*. (a) Allele frequency of rapid germinants, and (b) the rest of the population were fitted to a sliding-window three-way-nested ANOVA model. The F-statistic was plotted along the genome, with positive values indicating Col-0 allele and negative values Bs-2 bias. Significance threshold (shown as grey dashed lines) and optimal window size were established via permutation. The red shaded areas denote individual X-QTL regions above the significance threshold.

Figure 5. The genetic map of F3 rapid germinants derived by individual re-sequencing. A χ^2 test was performed to test local allele frequency skews in 191 rapid F3 germinants sequenced and

SNPs called individually. The log-transformed, Bonferroni-corrected p-value was plotted along the genome, with positive values indicating Col-0 allele, and negative being Bs-2. The red shaded areas denote the full region above significance threshold (α=0.05, grey dashed lines).

Figure 6. Recombination breakpoints of the haplotype with highest allele frequency skew is identified by re-sequencing. Genotypes of 191 F_3 individuals at the three QTL regions were determined, and the count of the selected allele along each chromosome is shown. The grey line indicates region outside of the peak boundaries, while the red line indicates the peak region. Areas likely harboring gene candidates for germination speed are indicated by the dark horizontal bars on top of each chromosome. The insets magnify and illustrate the regions in the vicinities of allele frequency peaks. The light-red shaded areas indicate the haplotype blocks carrying the highest allele frequency skew. Gene models with ABA-related functional annotation in these highest peaks are showcased as block arrows (red) with their gene names, together with genes that are directly upstream and downstream (grey)

Figure 7. Seed size, inflorescence development, and fitness QTLs that overlaps with germination speed X-QTLs by physical location. Physical location of QTL peaks from individual published studies were acquired and plotted in correspondence to our X-QTL map. Fifteen studies were classified into three general groups, seed traits (red), flowering time (teal), and fitness and related morphological traits (purple). Within each group, each line of tick marks indicates the QTLs identified on chromosome 1, 3, and 4. Red tinted areas indicate the full peak region, and therefore tick marks fallen within the tinted areas are considered overlapping with our X-QTLs.

Supplementary Figure 1. X-QTL mapping power analysis with favorable alleles derived from both parents. Additional simulation was performed to visualize how parental origin of the favorable alleles affects the power and precision of X-QTL mapping in *A.thaliana*. Scenarios of an average of one (a), two (b), and five (c) QTLs per chromosome were simulated. The set up of the population and selection experiment are consistent with that of Figure 3. The positions of QTLs are indicated as the tick marks within the gray chromosome bars on top of each panel, light and dark tick marks discriminate different parental origins. A series of trait heritability was simulated, denoted with different shades of lines. All scenarios were repeated 50 times.

Supplementary Figure 2. Correlation between (a-b) germination speed and flowering time, or (c) seed weight. Scatter plot of germination speed vs. rosette leaf number in 13 natural accessions under (a) long day and (b) short day conditions, and (c) scatter plot of germination speed vs. seed weight. The grey line shows the linear correlation of the trait pairs. Each pair of traits showed non-significant negative correlation.

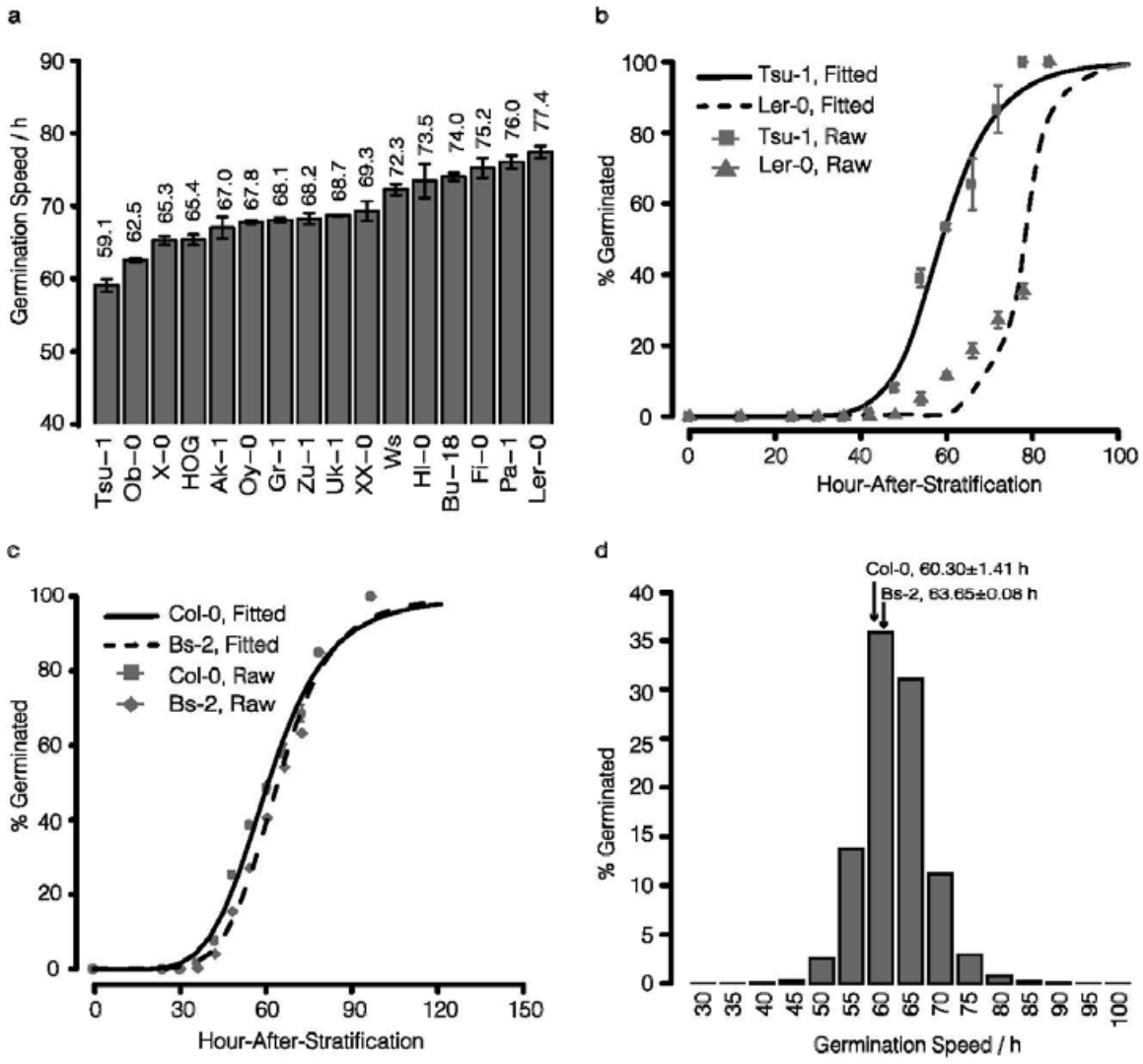

Figure 1. Natural variation of germination speed in *Arabidopsis thaliana*

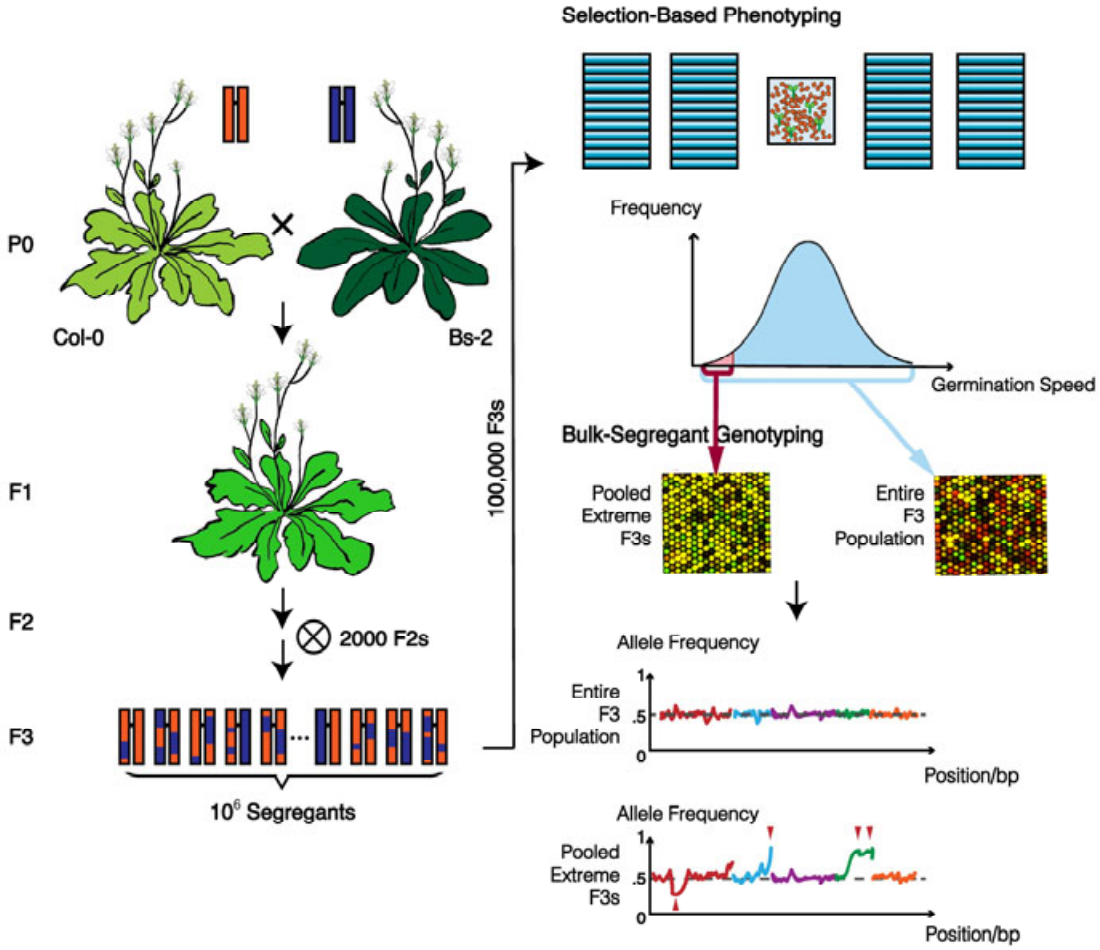

Figure 2. The *A. thaliana* Extreme-QTL (X-QTL) mapping procedure

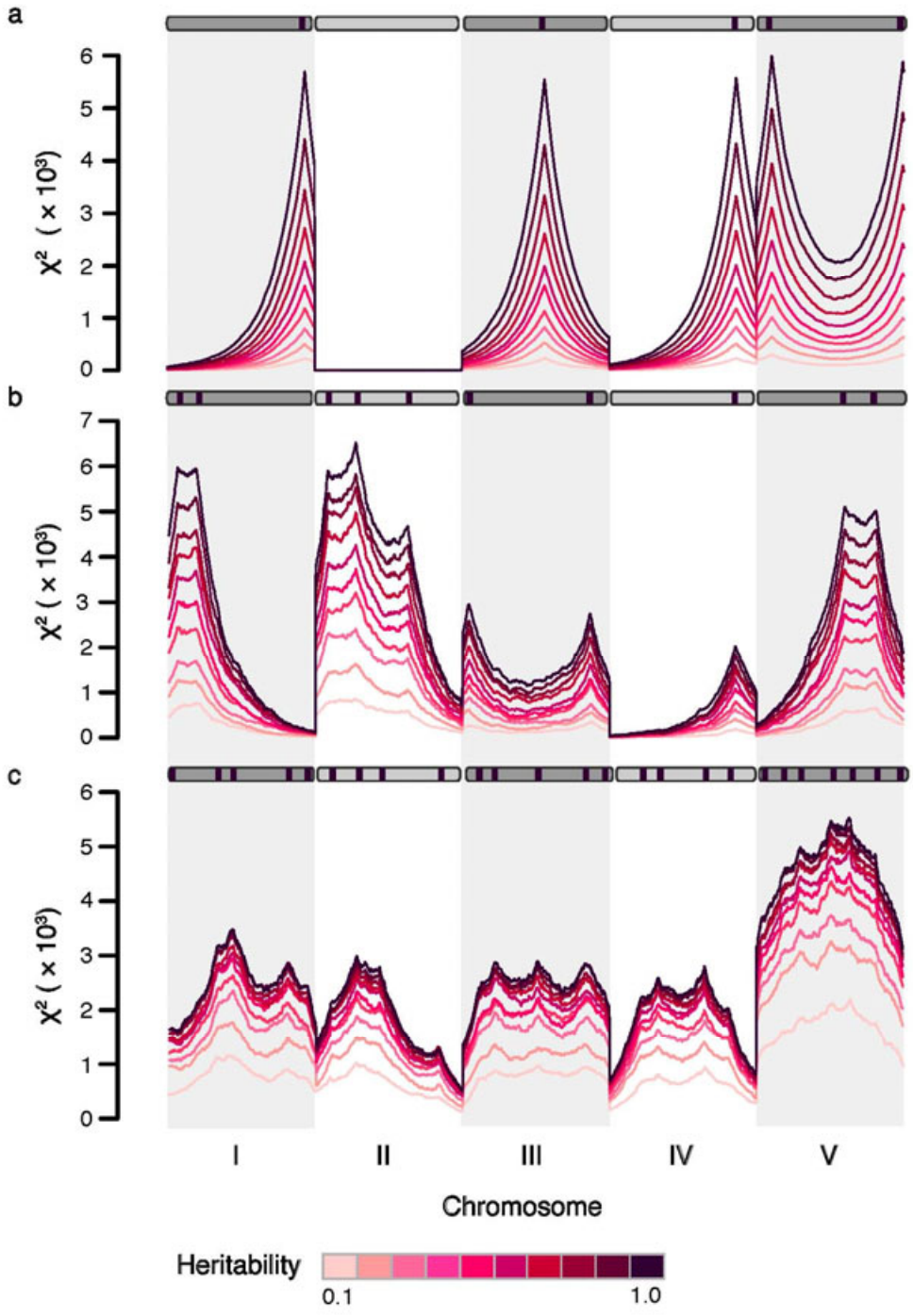

Figure 3. X-QTL mapping power analysis with consideration of genetic architecture complexity

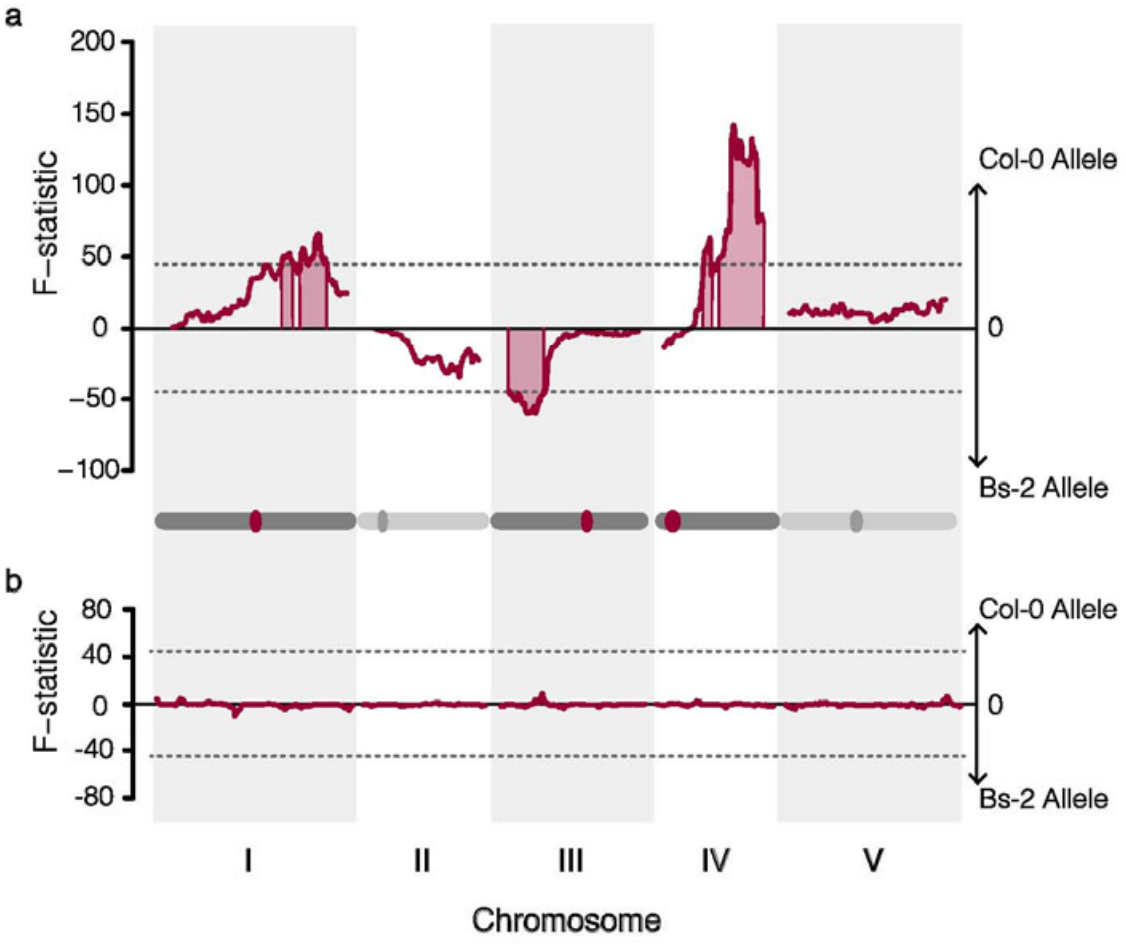

Figure 4. The genetic architecture of germination speed in *A.thaliana*

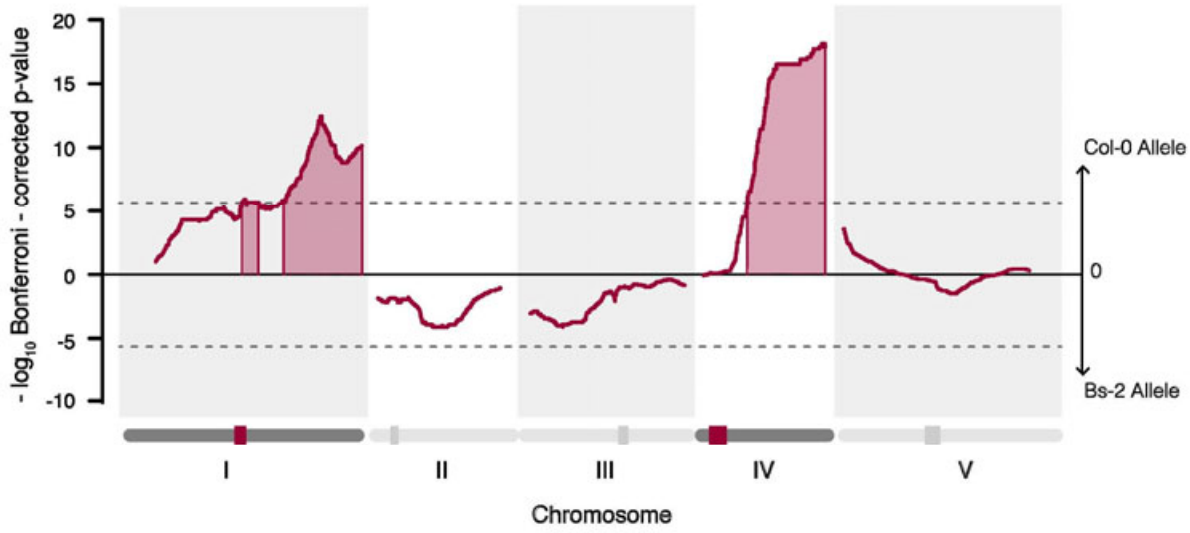

Figure 5. The genetic map of F_3 rapid germinants derived by individual re-sequencing.

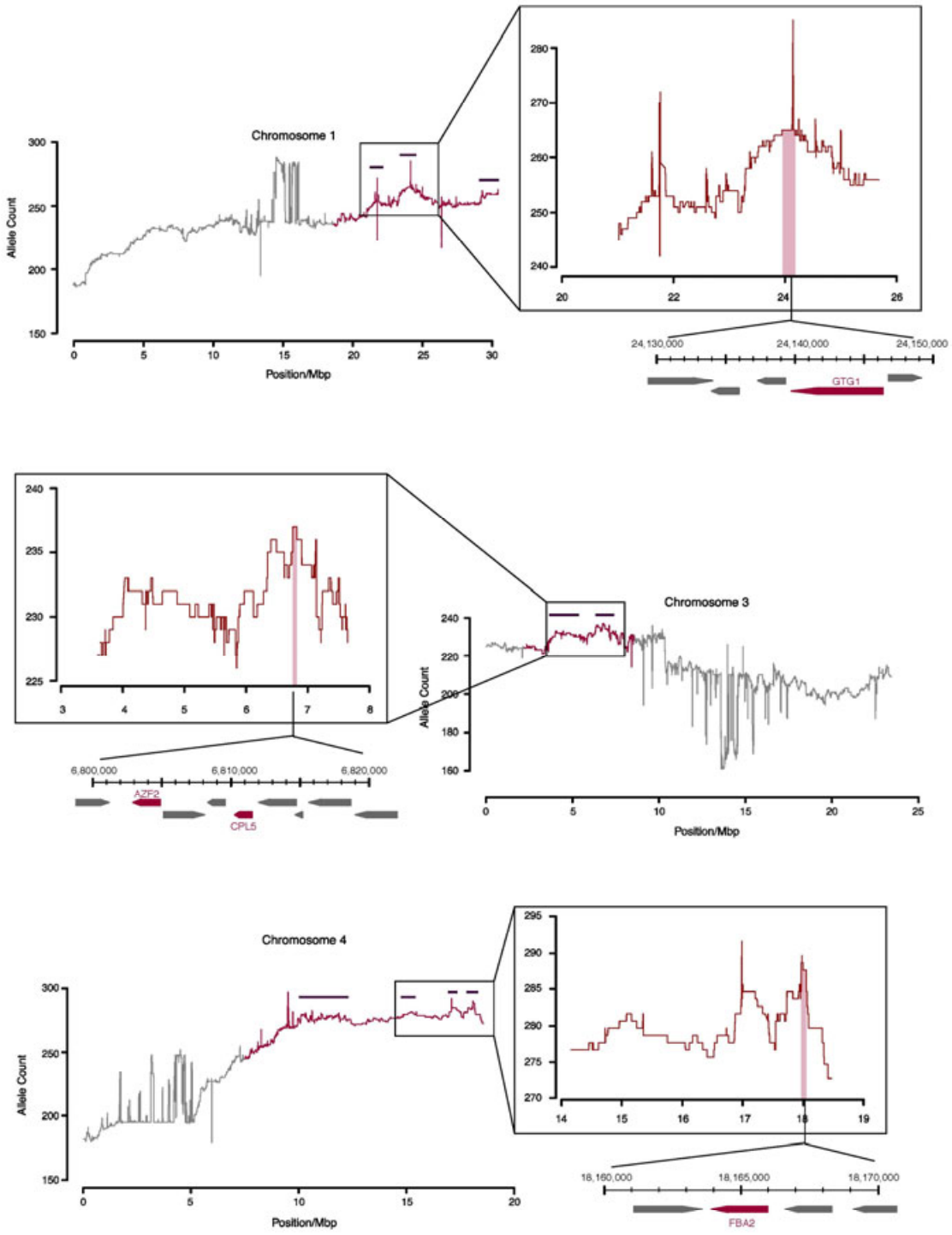

Figure 6. Recombination breakpoints of the haplotype with highest allele frequency skew was identified by resequencing

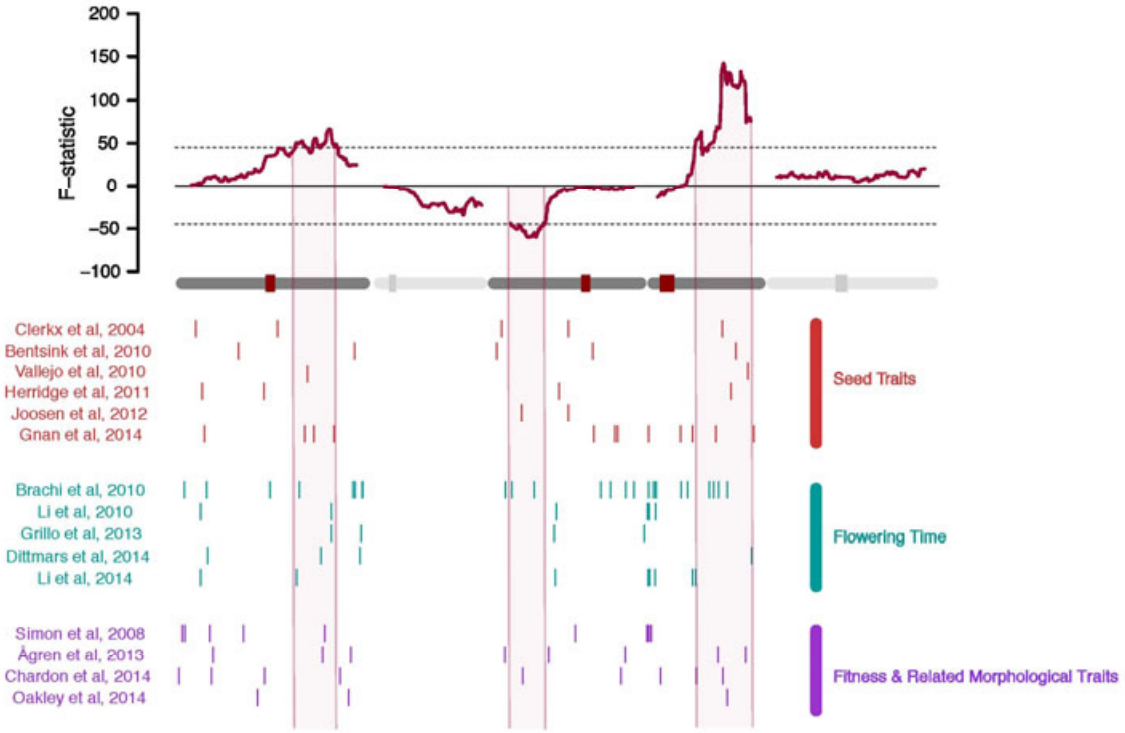

Figure 7. Seed size, inflorescence development, and fitness QTLs overlaps with germination speed X-QTLs