Comprehensive phenotypic and genomic analyses of life-cycle variation in *Arabidopsis thaliana*

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

Der Mathematisch-Naturwissenschaftlichen Fakultät

Der Universität zu Köln

Vorgelegt von

Marilyne Debieu
(Aus Versailles; France)

Köln, 2011

Diese Arbeit wurde am Max-Planck-Institut für Pflanzenzüchtungsforschung in Köln in der Abteilung Pflanzenzüchtung und Genetik durchgeführt.
AG Juliette de Meaux

Prüfungsvorsitzender: Prof. Dr. Joachim Krug

Berichterstatter: Prof. Dr. Maarten Koornneef
Prof. Dr. Eric von Elert

Beisitzer: Dr. Wim J.J. Soppe

Tag der mündlichen Prüfung: 15/04/2010

Table of contents

Introduction

I. General Introduction

1. Evolution and life strategies

Life evolution and time

"Nothing in Biology Makes Sense Except in the Light of Evolution**",** (citation of Theodosius Dobzhansky, *The American Biology Teacher*, March 1973).

Plants and other organisms that we observe today are the results of short and long-term evolutionary processes. During short-time evolutionary processes, organisms which can reproduce before they die, are favored by natural selection. Long-term evolutionary processes favor the ability to create continuously variation allowing adaptation to environmental changes throughout time. Time at all scales and levels, from the timing of the expression of a specific gene, through the age at first reproduction, to the transmission of genetic information through several millions of generations, appears to be a crucial notion in the evolution of life.

Life strategies in plants

Plants are constrained to grow at the site where they germinate. Thus, in contrast with animals they face climatic variation along the seasons. Plants respond to this variation with diverse strategies. Perennial plant species grow and reproduce over several consecutive years, optimizing their fitness across multiple reproductive events. Instead, annuals plant species complete their life-cycle in a single year and their fitness is determined by a single reproductive event. Therefore, selection on the synchronization of germination, growth and reproduction with the timing and length of the optimal season is believed to be particularly strong in annual plant species. (Stearns 1992; Roff 2002).

First I describe developmental phases of annual plants which have been well characterized.

2. Physiological description of plant development

Germination and seed dormancy

The emergence of the radicle through the seed coat constitutes the first crucial developmental transition and is defined as germination (Bewley and Black 1994). Germination requires water, oxygen and an adequate temperature and can occur only after dormancy is released.

Seed dormancy, is defined as the ability of a viable seed *not* to germinate in conditions favourable to germination (Hilhorst 1995; Bewley 1997). Several classes of seed dormancy have been identified: physiological, morphological, morpho-physiological, physical and combinational.The most common form is the non-deep physiological dormancy which includes the model species *A. thaliana*. An embryo excised from non-deep physiological dormant seeds can produce a normal seedling. This dormancy can be broken by various chemical treatments (e.g. gibberellin, nitrate), after-ripening in dry conditions and cold or warm stratification (Baskin & Baskin 2004; Finch-Savage & Leubner-Metzger 2006). The plant hormones abscisic acid and gibberellins play an important role in germination and dormancy and have opposite effects: whereas abscisic acid promotes dormancy, gibberellins release dormancy and promote germination (Bewley and Black 1994).

Physiological dormancy release can be reversible; seeds can become fully dormant again after dormancy had been released. Two types of physiological dormancies can therefore be defined: primary dormancy, which is established on the mother plant during seed development and secondary dormancy, which can be established in imbibed seeds in response to environmental conditions that are not favourable for germination. Several secondary dormancy cycles can occur until conditions are favourable for germination (Vleeshouwers 1995; Baskin & Baskin 1998; Cadman et al 2006). Dormancy cycles are illustrated in figure 1.

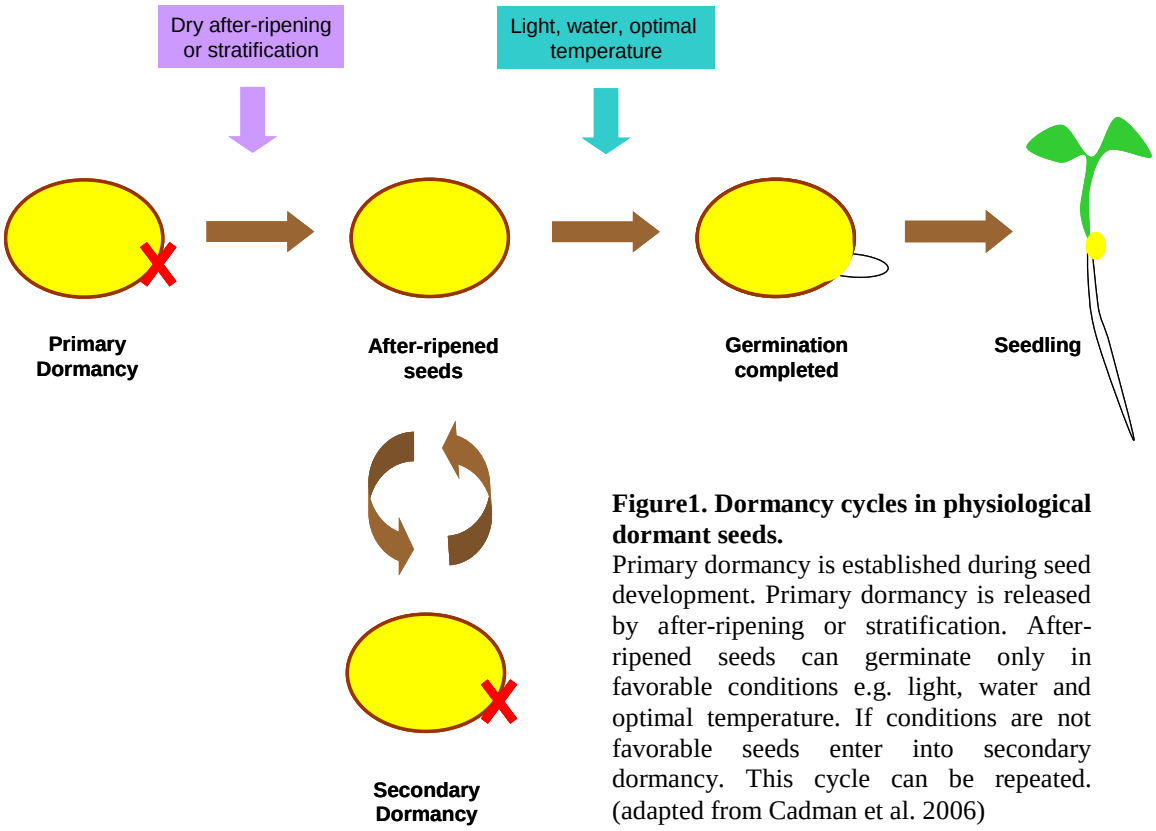


Figure1. Dormancy cycles in physiological dormant seeds.
Primary dormancy is established during seed development. Primary dormancy is released by after-ripening or stratification. After-ripened seeds can germinate only in favorable conditions e.g. light, water and optimal temperature. If conditions are not favorable seeds enter into secondary dormancy. This cycle can be repeated. (adapted from Cadman et al. 2006)

Vegetative, reproductive developments

Plant growth includes three major mechanisms cell division in meristems, cell elongation and differentiation. In this way, plants create organs during their life cycle. During vegetative growth meristems produce leaves, photosynthetic organs and roots, and during reproductive development meristems produce notably flowers (see reviews Gonzalez et al., 2009; Krizek, 2009). Plant growth is modulated by environment, especially affecting available resources (e.g. soil with minerals) and climatic conditions (water supply, temperature, daylength etc). Plant hormones and environmental factors play an important role in plant growth. If the conditions are not favourable plant development can be temporarily suspended.

3. Developmental traits in an evolutionary perspective

Germination, flowering timings and growth rate: major components of fitness

Fitness includes two major aspects: survival and fecundity (figure 2). Instead of estimating fitness which can be difficult, I focus on developmental traits influencing it. First I report studies which demonstrate the involvement in fitness of the individual developmental traits (fitness components).

Germination timing-seed dormancy and survival. Germination timing imposes conditions in which plant grows and influences seedling survival (Biere 1991; Gross and Smith 1991) and the plant's ability to complete its life cycle. Thus, germination timing influences strongly plant survival (Miller 1987; Kalisz 1989; Masuda and Washitani 1992). Seed dormancy, which controls germination timing, very likely plays a major role in plant survival, because seed dormancy prevents germination during unfavorable conditions and thus postpones germination to a suitable time of the year. Moreover, if we consider germination timing as the length between seed dispersal and germination, then we can assume that measurement of the level of seed dormancy estimates germination timing.

Flowering timing and fecundity. The timing of the switch between vegetative and reproductive developments determines maternal (mother plant size) and environmental conditions in which reproduction and seed maturation occur. Consequently, floral transition timing influences strongly fecundity rate. (Bernier, 1988; Geber, 1990; Stratton, 1998; Simpson and Dean, 2002).

Growth rate, survival and fecundity. During plant growth, some resources are stored to optimize reproductive performance and some resources can be used to perform survival mechanisms such as optimization of water use efficiency or defence. Thus, energetic activities related to plant growth, potentially can influence both components of fitness, survival and fecundity (Leister et al., 1999; McGraw and Garbutt, 1990).

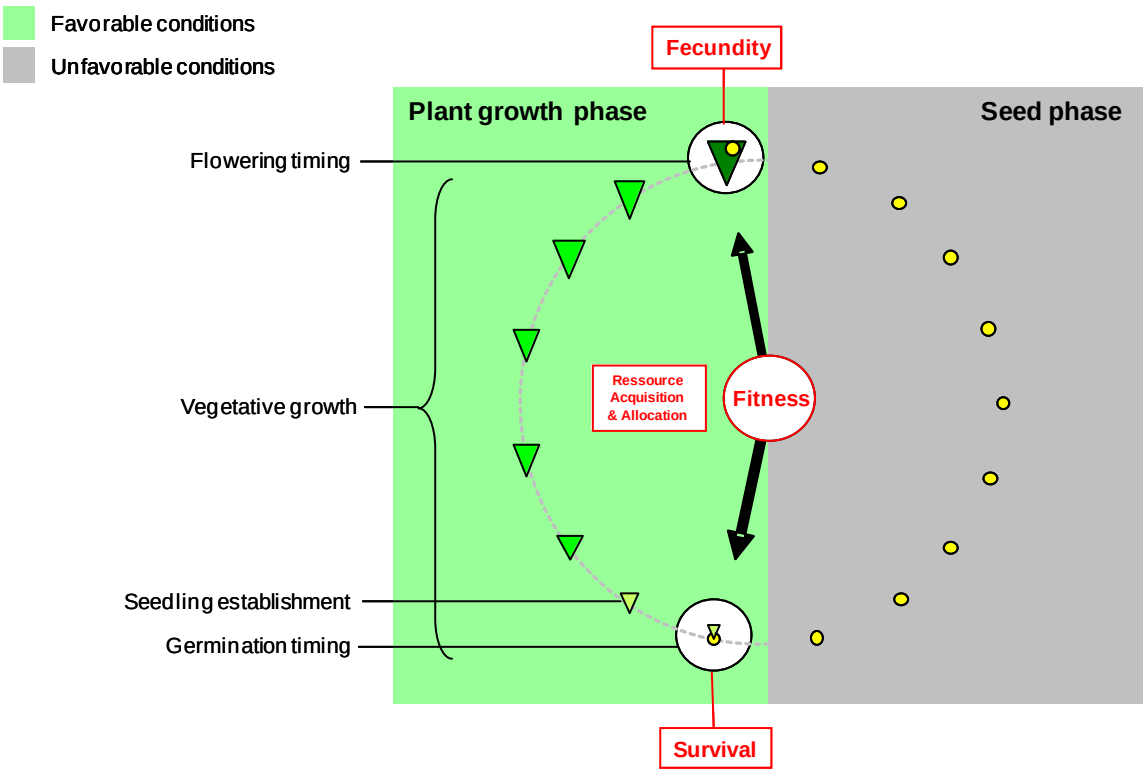


Figure 2. Life cycle and fitness components. Fitness includes two major aspects: survival and fecundity. Plant growth phase occurs in favorable conditions (green area) whereas the seed phase is associated with a period of unfavorable conditions for growth (grey area). Seed phase includes primary dormancy and potential secondary dormancy cycles as well as non dormant phases in non permissive conditions for germination. Germination timing determines environmental conditions in which plant (especially seedling) grows and thus influences strongly plant survival. Flowering timing determines maternal and environmental conditions in which reproduction and seed maturation occur and thus influences strongly plant fecundity. During vegetative growth, some resources are stored to optimize reproductive performance and some resources can be used to perform survival mechanisms (water use efficiency, defence…). Thus, energetic activities related to vegetative growth potentially influences both Survival and Fecundity.

Germination & flowering timings and growth rate influence each other

Whereas in the previous paragraph I reported the involvement in fitness of individual developmental traits, here I focus on how they influence each other. Because germination timing imposes environmental conditions in which plant grows, it influences expression of post-germination traits until flowering and the strength of selection they undergo (Evans and Cabin 1995; Donohue 2002). In return, flowering timing imposes environmental conditions in which seeds will mature and disperse. The conditions

experienced during seed maturation such as photoperiod and temperature (Gutterman and Evanari 1972; Pourrat and Jacques 1978) determine the conditions required to break seed dormancy after dispersal (Roach and Wulff 1987; Schmitt et al. 1992; reviewed in Baskin and Baskin 1998). Thus pre- and postdispersal conditions imposed by flowering timing determine germination timing at the next generation (Alexander and Wulff 1985; Munir et al. 2001). Moreover germination and flowering timings impose the length of vegetative growth and thus probably influence the strategy of resource allocation adopted by the plant during its vegetative development. Therefore each developmental phase depends on the timing of the previous phase and influences the next by imposing environmental conditions in which it will occur.

Theoretical fitness trade-offs and strategies

Natural selection favors optimal survival and fecundity. Nevertheless the theoretical optimal fitness might never be achieved because of physiological constraints as well as resources and time limitations (Roff 2002). Thus natural selection favors the most suitable strategy of resource or time allocation.

Resource acquisition and allocation can be genetically variable (here I do not include variation due to environment; figure 3). We consider two components of fitness among several genotypes in the same environment. The genetic correlation between the components of fitness is positive when the genetic variance in acquisition is large and the genetic variance in allocation is small. On the contrary, the correlation is negative when the genetic variance in acquisition is small and the genetic variance in allocation is large. In this case an increase in fitness due to a change in one component is opposed by a decrease in fitness due to a concomitant change in the second component. These positive and negative correlations between components of fitness are defined as trade-offs.

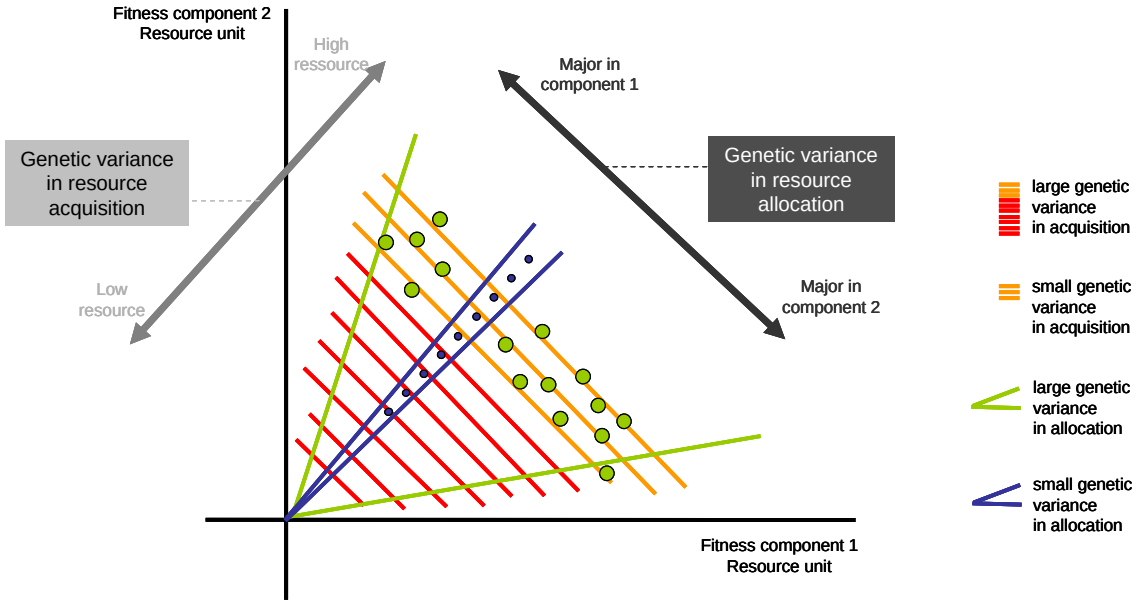


Figure 3. Theoretical fitness trade-offs.
To simplify we assume that for each individual (green and blue dots) there are two major fitness components 1 and 2 respectively axes X and Y, in which total acquired resources are allocated, in a constant environment. Orange and red lines represent different total amounts of acquired resource from low to high. Orange lines represent a small genetic variance in resource acquisition whereas all lines represent a large genetic variance in resource acquisition. The different strategies in resource allocation to component 1 and 2 are shown along each line. Between the green bars the genetic variance in resource allocation is large whereas between the blue bars the genetic variance in resource allocation is small. (adapted from Roff and Fairbairn 2007 by J. de Meaux and I)

In a large number of species a functional or statistical trade-off between survival and fecundity has been found (e.g. Reznick, 1985; Roff, 1992, 2002; Zuk, 1996). When the correlation is negative between survival and fecundity, a higher mortality rate before sexual maturity is compensated by a higher fecundity rate and a lower fecundity rate is compensated by a higher survival rate before sexual maturity. It defines different strategies in resource allocation.

Fitness trade-offs and strategies in annual plants

To complete their life cycle annual plants are strongly limited by time and have to synchronize developmental phases with seasons in the same year. Thus, time constitutes a major constraint. Plants can adopt different strategies concerning allocation of available time to the different developmental phases. There are two major developmental phases: the seed phase and the plant growth phase related respectively to survival and fecundity. Two major strategies can be identified figure 4. First, plants invest more time for plant growth to optimize their size at reproduction timing and thus fecundity, but have more risks to die. Second, plants invest more time in seed dormancy to escape unfavorable environmental conditions to optimize their chance to survive but plants are smaller at reproduction timing. Theoretically in a natural population we can expect a negative correlation between seed phase length and growth phase length.

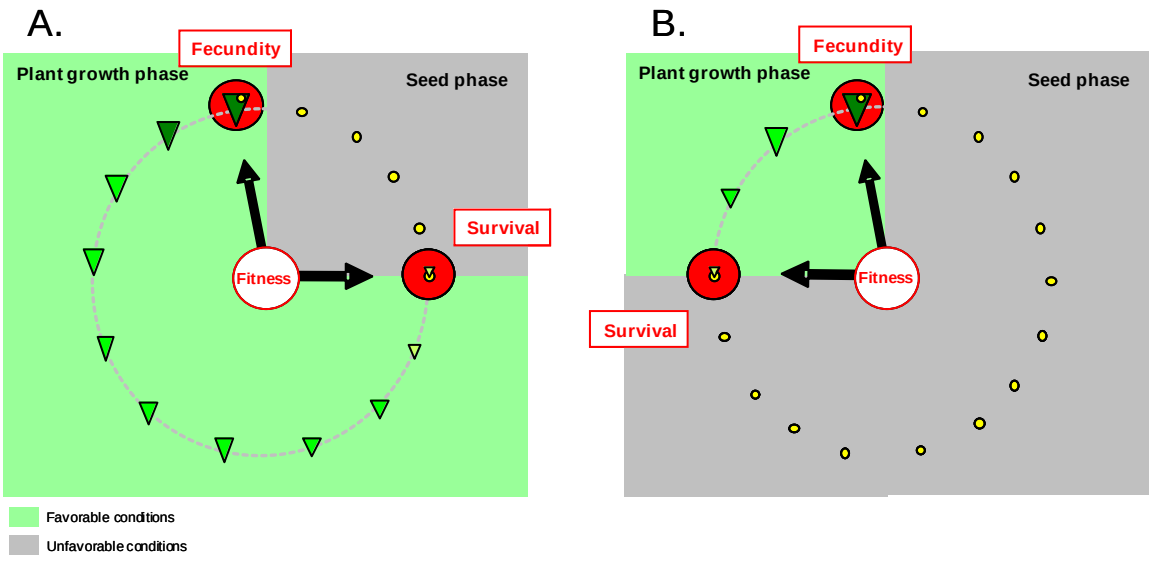


Figure 4. Acquisition is total life time which is allocated to seed phase and plant growth phase. In this case, we assume that resource acquisition is total life time, and that its genetic variance is small around one year. Time is divided in two major parts seed phase and plant growth phase. Seed phase includes primary dormancy and potential secondary dormancy cycles as well as non dormant phases in non permissive conditions for germination. Length of seed phase is a fitness component related to survival and length of plant growth phase is a fitness component related to fecundity. **A.** Length of seed phase is short and length of plant growth phase is long. The risk to die is higher but the plant produces more seeds. **B.** Length of seed phase is long and length of plant growth phase is short. The risk to die is lower but plant produces less seeds.

II. *Arabidopsis thaliana* and objectives

1. *Arabidopsis thaliana* : plant model, ideal for evolutionary investigations

Geographical distribution and climates

Arabidopsis thaliana (Brassicaceae) is an annual weedy plant, diploid and highly selfing (Abbott and Gomes 1989). Since the last glaciations, this model plant species has dramatically expanded its range through all continents and mostly in the northern hemisphere (Sharbel et al., 2000; Schmidt et al., 2005; Schmidt et al., 2006; Beck et al., 2008). The geographical distribution of *Arabidopsis thaliana* is shown figure 5. The range of *A. thaliana* largely exceeds that of its sister species suggesting an extensive adaptive evolution to survive in new environmental conditions (Hoffmann, 2002; Hoffmann 2005).

Thus *Arabidopsis thaliana* offers an ideal context to study how environmental variations influence the evolution of life-history traits.

Koornneef et al, 2004

Figure 5. Geographical distribution of *Arabidopsis thaliana.* Green area shows the geographical distribution of *Arabidopsis thaliana*; red dots are the available genotypes in 2004.

Genetic natural variation of life history traits

The first important question to answer in order to understand the evolutionary story of one trait is: is there natural genetic variation underlying it? Natural variation of germination, growth related traits and flowering time has been largely studied in *Arabidopsis thaliana* with field-collected genotypes or recombinant inbred lines. Primary dormancy variation which appears to be strong has only been studied on a few accessions (Schmutz et al. 2006) moreover natural variation of secondary dormancy remains still unknown in *Arabidopsis thaliana*. On the contrary, natural genetic variation of a large number of growth related traits (rosette diameter, rosette leaves number, relative growth rate, fresh or dry weight) has been reported in this species (Aarssen and Clauss 1992; Li at al. 1998; Perez-Perez et al. 2002 Koornneef et al. 2004). Finally, natural variation of flowering time has been intensively studied since the first researches on *Arabidopsis thaliana* (Laibach 1951, Shindo et al. 2005; Caicedo et al. 2004; Wilczek et al. 2009).

Life history traits and fitness studies

The second fundamental question concerning the evolutionary study of one trait is: does it influence fitness? In 2005, involvement in fitness of germination timing was extensively investigated (Donohue et al 2005). These investigations were performed in the field with a population of recombinant inbred lines created with a cross between two genotypes Cal and Tac respectively from England and the USA (Washington). Fitness was measured by the total number of seeds produced. The results proved clearly that germination timing influences fitness (Donohue et al. 2005). Field studies with a large number of genotypes showed that rosette diameter and flowering timing influence fitness as well (Griffith et al. 2004; Korves et al. 2007). In addition, these fitness studies revealed that influence of germination and growth related traits on fitness is variable.

2. Life history strategies in *A thaliana*

Life history strategies and plasticity

In *A. thaliana*, variation in the timing of germination and flowering at different seasons defines different life history strategies (Effmertova 1967; Lawrence 1976; Ratcliffe 1976; Nordborg and Bergelson 1999). Two major strategies have been described. On the one hand, "winter annuals" germinate in autumn, overwinter as rosette and flower in spring (Baskin and Baskin 1974, 1983). On the other hand, "rapid cyclers" germinate and flower during the same season (Thompson 1994; Griffith et al. 2004). It has been shown recently in field experiments that most of genotypes can express both strategies depending on germination timing, suggesting a strong plasticity in expression of life history strategies in this species. Only when germination occurs in autumn natural allelic variation of the flowering time gene *FRIGIDA* is expressed at the phenotypic level (Wilczek et al. 2009).

Genetic correlations and trade-off in *Arabidopsis thaliana*

Natural selection favors plants that flower early with a large size. This response to selection is opposed by a genetic trade-off between these two components of fitness due to physiological and time constraints (Mitchell-Olds et al. 1996). Flowering time is positively correlated as well with dehydration avoidance. This trade-off is genetically controlled by pleiotropic alleles at loci initially known to be responsible for natural variation in flowering time (*FRIGIDA* and *FLC*) (McKay et al. 2003). In addition, trade-offs were found between growth and plant defense (van Hulten et al. 2006). A fundamental question concerning variation in trade-off traits is what can maintain it. In contrasted environments different combinations of traits could be favored by selection (correlational selection). Thus, adaptation to specific/different environments could maintain variation in trade-off.

3. Adaptive and neutral evolution of traits

Genetic variation is modulated by neutral and or selective mechanisms. In *A. thaliana* population structure has been extensively studied (Nordborg et al. 2005; Schmid et al. 2006; Francois et al. 2008) allowing the estimation of the geographical distribution of neutral variation. In contrast latitudinal genetic variation provides an argument that a trait shows adaptive variation to an environmental gradient (Chapin 1974; Zhang and Lechowicz 1994). Nevertheless, adaptive and neutral evolutionary processes can influence each other especially adaptive evolution can promote itself population structure (Nosil et al. 2009).

Natural variation of life history traits along latitude has been extensively investigated to estimate the adaptive values of these traits. With a set of forty genotypes it has been shown that relative growth rate is negatively correlated with latitude (Li et al. 1998). Genetic variation behind responses to vernalization in leaf angle and leaf morphology follows a latitudinal cline (Hopkins et al. 2008).

Finally, natural variation of flowering time has been studied at the species scale and it revealed that the variation follows a latitudinal cline among genotypes carrying functional *FRIGIDA* allele (Stinchcombe et al. 2004; Caicedo et al. 2004). This Latitudinal cline has been confirmed controlling for population structure (Korves at al. 2007).

4. Identification of genes behind life history variation

Combining Linkage mapping and Genome wide Association mapping

To identify Quantitative Trait Loci (QTL) underlying natural variation of life history traits, two major methods of mapping are used: association mapping and linkage mapping illustrated in figure 6. Both methods are complementary and present different advantages and limits reviewed by Nordborg and Weigel (2008). A very important parameter which determines the success of QTL mapping is the sufficient number of recombination events which occurred in the population studied. Indeed a large number of recombination events allows QTL mapping with a high resolution with a dense marker coverage.

Genome Wide Association mapping (GWA) using a non-structured, old and large population allows QTL detection with a high resolution because the population carries historical recombination events. By contrast Linkage mapping is performed with an experimental population or recombinant inbred lines (RIL) generated by successive controlled crosses starting from two parental genotypes. The recombination events generated by these crosses allow a relatively low mapping resolution. Nevertheless it is possible to increase the resolution via fine mapping. This procedure requires both additional recombinant inbred lines and a higher number of markers. In addition to confirm the effect of a QTL minimizing complex epistatic interactions with other genes Nils (Near-isogenic line) and Hifs can be created.

In natural populations causal markers can be linked to population structure, which generate a high rate of false positives. In this case Linkage mapping is required to break the linkage disequilibrium between background markers and causal loci. Alternatively, to solve the problem of population structure statistical methods have been elaborated. Yu et al. (2006) have developed a unified mixed model approach to simultaneously account for two levels of relatedness: gross level population structure (Q) and familial relatedness from more recent coancestry (K: kinship). Few months later the same year, Price et al. (2006) proposed an alternative simple method using Principal components analysis (P) to control for population structure. Zhao et al. (2007) tested the different statistical methods for controlling for population structure in *A. thaliana* and found that the mixed model approach including Q and K matrix generally perform best. Moreover Stich et al. (2008) recommend to replace the K matrix by a K_T matrix which is based on a restricted maximum-likelihood (REML) estimate of the probability of two alleles at the same locus being identical in state but not identical by descent. Despite the statistical strategies developed, the problem of confounding between candidate gene and background markers still remains and the control for population structure introduces false negatives as well. Therefore it is necessary to complement the results from association mapping by the results from linkage mapping.

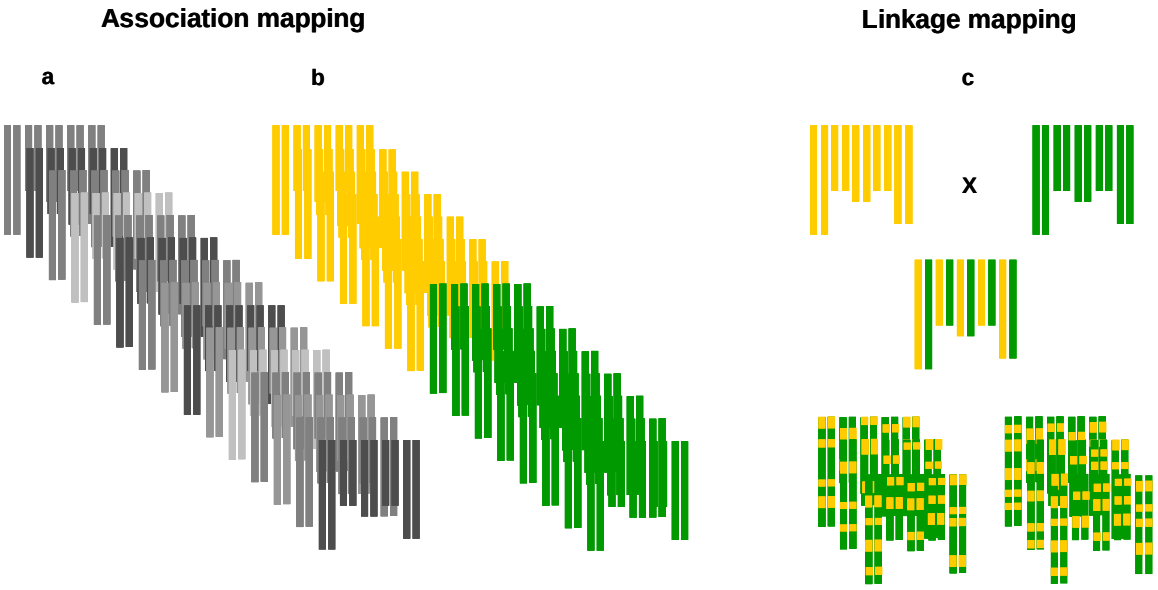


Figure 6. Linkage and Association mapping methods.
To detect QTL two major methods are used Association mapping and Linkage mapping. Here we illustrate the two methods in the highly selfing species *Arabidopsis thaliana* showing the 5 chromosomes of diploid inbred lines. Association mapping is effective with a non structured natural population carrying historical recombination events which allows QTL detection with a high resolution (**a**). Association mapping is not effective if there is a strong genetic differentiation between subpopulations because causal genes are linked with background markers and cannot be detected (**b**). Linkage mapping allows breaking linkage disequilibrium between causal genes and background markers by controlled crosses between two parents from the two subpopulations. It creates an experimental population or recombinant inbred lines (RIL) (**c**).

Candidate genes: *DOG1*, *FRIGIDA* & *FLC*

Dormancy variation is controlled by multiple loci. In *Arabidopsis thaliana*, *DELAY OF GERMINATION 1* (*DOG1*) gene was identified as a QTL involved in the control of seed dormancy using linkage mapping (Alonso-Blanco et al. 2003; Clerkx et al., 2004; Bentsink et al. 2010; Huang et al. 2010). *DOG1* encodes a protein of unknown function and is specifically expressed in the seed. A mutant approach indicated that this gene is absolutely required for seed dormancy induction (Bentsink et al. 2006). Functional and non functional alleles at *FRIGIDA* locus are major determinants of flowering time variation. It is known that functional *FRIGIDA* allele makes plants vernalization dependent (Johanson et al. 2000) but the function of this gene is not known. In this study, the influence of *FRIGIDA* alleles on evolutionary trajectories based on co-variation

between life history traits was investigated. Moreover, to control flowering time, *FRIGIDA* gene interacts with *FLOWERING LOCUS C* (*FLC*) gene. It has been shown recently that *FLC* gene regulates as well seed germination (Chiang et al. 2009). Seven other genes contributing to natural variation for flowering time have been isolated in *Arabidopsis thaliana* (reviewed in Alonso-Blanco et al. 2009) moreover a large number of germination QTL have been detected including eleven *DOG* QTL (Bentsink et al. 2010). In this study, two association mapping methods were used to confirm known QTL and to detect new QTL controlling natural variation of life history traits related to seed dormancy, growth and flowering time. Moreover the genetic architecture behind co-variation between traits was examined. Especially pleiotropic effects and linkage disequilibrium between genes which can be close physically or not were analyzed. The association mapping methods used in this study are: the unified mixed model approach (Yu et al. 2006; Stich et al. 2008) and the Genome Wide Association mapping method (Atwell et al. 2010) respectively in collaboration with B. Stich and M. Nordborg and his team. Finally, all the objectives are summarized below.

5. Objectives

First, the following questions related to phenotypic variation and co-variation as well as association of it with environmental parameters were addressed:

i) what are the characteristics of natural variation for primary and secondary seed dormancy, relative growth rate and flowering time after vernalization in *A. thaliana*,

ii) do life-history strategies vary along environmental gradients,

iii) is there co-variation among life-history traits,

iv) does co-variation in life-cycle traits depend on climatic variation

Second, questions related to genetic variation associated to phenotypic variation and co-variation were addressed:

i) do *FRIGIDA* alleles impact co-variation in life-cycle traits?

ii) does analysis of *DOG1* nucleotide diversity reveal selective marks?

iii) Does allelic variation of known candidate genes associate with the expected phenotype using the Genome Wide Association (GWA) mapping method?

iv) Is GWA method efficient to detect new candidates ?

v) What is the genetic variation which controls co-variation among life-history traits, in particular are they genes with pleiotropic effects or genes in linkage desequilibrium (physically close or not)?

6. Glossary

Fitness: represents the ability of an organism to survive until reproduction timing and to produce viable offspring. It includes two major aspects: survival and fertility. Plant fitness can be estimated by its probability to survive until reproduction timing multiplied by the number of seeds produced.

Component of fitness: trait related to survival or fecundity (or both).

Natural selection: favors individuals with the highest fitness. In a given environment if the phenotypic variation is genetically determined natural selection will favor alleles associated with the highest fitness.

Drift / Selection: random fluctuation of alleles frequencies through generations. This phenomenon is amplified in smaller population and can lead rapidly to the fixation or the disappearing of alleles. On the contrary, selection proceeds more efficiently in larger than in smaller populations. Drift, selection, mutations and migration are the major processes driving evolution.

Adaptation: constitutes successive selective events through several generations which tend to optimize individuals fitness in a specific environment.

Plasticity: is the variation of a phenotype measured in different environments associated with the same genotype.

Heritability: is the proportion of phenotypic variation in a population that is attributable to genetic variation among individuals. Phenotypic variation among individuals may be due to genetic and/or environmental factors.

Life history traits: traits which influence fitness. For example germination timing is a life history trait. The decision to germinate or not can be potentially subjected to selection.

Trade-off: a trade-off occurs when an increase of fitness due to a change in one trait is opposed by a decrease in fitness due to a concomitant change in the second trait.

Correlational selection: is a selection which favors combinations of traits. It can generate genetic correlations between traits. Genetic correlations can arise from linkage disequilibrium or pleiotropy (Sinervo & Svensson, 2002).

linkage disequilibrium: non random association between genes which can be physically close or distant.

Constraint: It means that there are evolutionary trajectories that are unavailable to selection, termed "evolutionarily forbidden trajectories" by Kirkpatrick & Lofsvold (1992) and "absolute evolutionary constraints" by Mezey & Houle (2005).

Materials and Methods

I. Phenotypes and climatic data

1. Plant Material

A total of 161 worldwide collected *A. thaliana* genotypes obtained from the ABRC and NASC stock centers were used for the study (Appendix 1).

2. Phenotyping

Measurements of vegetative plant growth rate:

Each accession was grown in four replicates and growth rate, floral traits and offspring germination were assessed. After a stratification at 4°C for two days (water imbibed seeds on filter paper in Petri Dishes), seeds were sown in potting soil, grown for one week in a greenhouse under long day (with supplementary light maintaining long day conditions), vernalized for 4 weeks (4°C, 16h light, 50% relative humidity) and finally placed back into the greenhouse until silique maturation. To estimate relative growth rate, we used a nondestructive method based on digital imaging described in (Leister et al. 1999 and El-Lithy et al. 2004). Plants were photographed from above with a charge coupled device (CCD) camera (Sony DSC-F828) at the first and the last day of the vernalization treatment, as well as after one week of growth in the greenhouse after vernalization. Leaf area at these time points was calculated using Image Pro Analyzer 6.0 (MediaCybernetics) in cm^2. Relative seedling growth rate was given by leaf area measured just before the cold treatment and divided by the number of days of growth since sowing (Seedling Growth in cm^2/day). Relative vegetative growth rates during and after vernalization (Vern Growth, After Vern Growth) were estimated as the increment of cm^2 leaf area per day respectively between the first and last day of vernalization and between the last day of cold treatment and after one week of growth in greenhouse.

Measurements of floral traits

Flowering time was scored as number of days from sowing the seeds in soil to the day that the petals of first flower were visible. In addition, the number (Nb) of leaves in rosette (Rosette Leaves Nb) and stem (Stem Leaves Nb) were counted at flowering. At the end of this experiment, only siliques of approximately similar age were bagged per plant for seed dormancy measurements. After harvest, seeds were stored in the same laboratory conditions in paper bags for all genotypes.

Measurements of germination-related traits:

All measurements of primary and secondary dormancy, for each genotype, were conducted with four independent seed batches (each batch contains seeds from one plant) collected from the experiment described above. Primary dormancy was measured as the progressive increase of germination rate measured after 7, 28, 56, 91, 133, 182 and 209-269 days of dry storage in laboratory conditions as described in (Alonso-Blanco et al. 2003). The germination percentage at each time point was measured after one week of imbibition in growth chamber (25°C 12h day/ 20°C 12h night). After approximately 9 months, the experiment was stopped and the viability of non germinating remaining seeds was confirmed by provoking germination with a mixture of 100-**µ**M gibberellin ($GA_{4/7}$ (ICI Ltd, Bracknel, UK) and 38-**µ**M fluridone (Dow Chemical Co., Hitchen, UK). Both fluridone and $GA_{4/7}$ were initially dissolved in ethanol and then diluted. The final concentration of ethanol was less than 0.03% (Cadman et al. 2006).

Dormancy was quantified as the Duration of Seed Dry Storage required for reaching 50% of seed germination ($DSDS_{50}$) as defined in Alonso-Blanco et al. (2003). Dormancy release followed different dynamics over time that could not be described by a single model. $DSDS_{50}$ was therefore simply estimated in number of days, by extrapolating from a straight line between the two time points at which less and more than 50% of the seeds had germinated.

Primary dormancy measurement was performed in three other independent experiments. The first experiment was conducted by Sigi Effgen with the same genotypes and 3 replicates that were grown in a different greenhouse experiment but tested for dormancy release in the same way as described above. A second experiment was conducted by Chunlao Tang with a subset of 96 genotypes. Plants were grown in environmentally controlled growth chambers with constantly maintained relative humidity (75%) and temperature (MTPS72 Conviron, Canada). After harvest, seeds were stored in cellulose paper bags and kept in a dark incubator at 20°C with 40% relative humidity. Dormancy release was measured as well by $DSDS_{50}$. Finally for a subset of 29 genotypes, germination rates of seeds harvested from a common garden experiment in the field (in Valencia, Spain) were also tested for germination after 2 months after-ripening by Emily Josephs.

We measured secondary dormancy as the reduction in germination rate of fully after-ripened seeds after a 6-week vs. 1-week long exposure to 4°C in darkness. For this, seeds were imbibed in laboratory conditions (21°C, light) in 500 µl of sterile water, on sterile filter paper, in sterile petri dishes and sown in a laminar flow hood. The experiment was conducted in sterile conditions to minimize fungal contaminations, which can complicate counting of germinating seeds. Seeds themselves were not sterilized to avoid effects of the sterilization treatment on the germination behaviour. Three batches were prepared simultaneously for the three different treatments. One batch of petri dishes was placed immediately in favorable conditions for germination in a 25°C 12h day / 20°C 12h night chamber for one week. The germination data of this batch corresponded to the last measurement of primary dormancy (see above). Not all genotypes reached 100 % of germination, therefore to disentangle residual primary dormancy from cold-induced secondary dormancy, only genotypes (124) that displayed more than 85% of germination (fully after-ripened) were used for secondary dormancy measurement. Two batches of petri dishes were closed with parafilm, wrapped in aluminum foil to ensure total darkness and placed in a chamber at 4°C for one or six weeks. Subsequently, cold-treated seeds were placed for germination in the 25°C 12h day/20 °C 12h night chamber for one week, whereafter germination was scored (Cadman et al. 2006). Secondary dormancy was

quantitified by the absolute value of the slope (percent of germinants per day) between the germination percentages of the fully after-ripened seeds and after six weeks of cold treatment; the lower the absolute value of the slope, the lower the proportion of seeds entering secondary dormancy. Viability of non germinating seeds after cold treatment was confirmed as described in Cadman et al., 2006.

Some non-dormant seeds had already germinated after exposure to cold in the dark, indicating that some genotypes do not have an absolute light requirement (Meng et al. 2008). We measured this ability to germinate in the dark at 4°C (Germination at 4°C) as the percentage of non dormant seeds that can germinate within 1-week of cold exposure, in the absence of light.

3. Climatic data

The climatic profile of each genotype was defined by annual mean and coefficient of annual variation (CV = σ/μ) of monthly temperature (respectively average temperature and temperature contrasts) and precipitation (respectively average precipitation and precipitation contrasts). Mean of temperature and precipitation of each month were collected in the data base of the Natural Variation of *Arabidopsis thaliana* (http://dbsgap.versailles.inra.fr/vnat/), from the software **DIVA-GIS 4.1** (http://www.diva-gis.org/climate.htm) and from different meteorological stations (from J. Schmitt lab). The location of origin of each genotype ranges from 16 to 63.3°N in latitude. Based on that, we defined two climatic groups: "North" (latitude > 49.2°) and "South" (latitude < 49.2°), which displayed comparable annual temperature contrasts. This allowed the analysis of life strategies respectively in climates with relatively in average low temperature and low precipitation vs. climates with high temperature and high precipitation. Moreover we observe an annual more homogeneous precipitation in the north whereas annual precipitation in the south is heterogeneous.

4. Statistical analyses of phenotypes

Description of phenotypes

For each trait, the heritability (h^2 = MSgenotype / (MSgenotype + MSerror)) which is the proportion of phenotypic variance due to genetic variance, was obtained through an analysis of variance (ANOVA) with genotypes as fixed factor (SYSTAT 11).
In addition, for each genotype the adjusted entry mean was calculated for all traits using ASReml (Gilmour et al. 2006).

Correlations

Pearson genetic correlations were calculated between all life history traits and statistical significance of the correlations were calculated with Bonferroni correction (SYSTAT). These correlations were performed for the whole sample as well as for each latitudinal (North and South) and genetic (*FRI-F* and *FRI-N*) subsamples (see paragraph II.2).
Furthermore, path analysis was conducted to quantify the relative contribution of latitude and population structure (output by STRUCTURE; Pritchard et al. 2000; see paragraph II.3) to life-history variation. The model considered that latitude was controlled by four climatic variables (average temperature, temperature contrasts, average precipitation and precipitation contrasts). We considered 7 life-history traits, primary, secondary dormancy, germination at 4°C, seedling growth, growth rates during and after vernalization and flowering time. Estimates of standardized path coefficients and their significance were obtained with the structural equation modeling package *sem* in R (http://www.r-project.org, http://socserv.socsci.mcmater.ca/jfox/).

General Linear Model for multiple regressions

To detect statistical associations between life history traits and latitude or climatic parameters, the following General Linear Model (GLM) model was tested: $Y_j = \alpha + \sum \beta_j X_j + \varepsilon_j$ (Y_j: dependent variable, X_j: independent variables, α: intercept (value of Y when X=0), β_j: partial regression coefficient, ε_j: residual). All life-history traits were included as independent variables in the initial model. A threshold tolerance of 0.5, was chosen to exclude correlated variables. The multiple regressions were performed with six different dependent variables: latitude, average temperature, temperature contrasts, average precipitation, Precipitation contrasts and population structure (output by STRUCTURE see paragraph II.3). The distribution of the residuals for each multiple regression was analyzed.

With the same dependent and independent variables mentioned above, multiple regression analyses were performed with different samples: with the whole sample and each geographical (North and South) as well as with genetic (*FRI-F* and *FRI-N*) subsamples (see paragraph II.2).

II. Genotypes

1. *DOG1* genotyping and analysis

DOG1 sequencing and genotyping

DNA extraction: The total DNA of the 161 genotypes was isolated from leaves using the BioSprint 96 workstation and the BioSprint 96 DNA Plant Kit (Qiagen).

Amplification by PCR: First, for a subsample of 73-86 genotypes a fragment of 5Kb including promoter and coding regions of *DOG1* was amplified. The amplification was performed via amplification of three overlapped fragments of 1,5 – 2 Kb by PCR (PCR1, PCR2 and PCR3; figure 7). Second, for the 161 genotypes, exon 1 (~550 bp) of *DOG1* was amplified by PCR4 for genotyping (figure 7).

Sequencing: PCR products were purified with ExoSAP-IT or the Qiaquick PCR purification kit (Qiagen) and sequencing was performed with the Sanger method. The same mix (except Primers) and program were used for all the PCRs. PCR mix and program are shown in Table 1. Names and sequences of primers used for amplification and sequencing of *DOG1* are shown in appendix 2.

PCR mix	Concentration	Volume
DNA	100 ng/µl	1 µl
Primer 1	10 pmol/µl	1 µl
Primer 2	10 pmol/µl	1 µl
Buffer + $MgCl_2$	10X	5 µl
dNTP	25 mM	0.2 µl
Taq (Roche)	5U/µl	0.5 µl
H_2O		41.3 µl

	PCR Program	Temperature	Lenght
	Initial Denaturation	94°C	
35 cycles	Denaturation	94°C	
	Hybridization	55°C	
	Elongation	72°C	
1 cycle	Denaturation	94°C	
	Hybridization	55°C	
	Final Elongation	72°C	
			For ever

Table 1. PCR mix and program to amplify *DOG1* gene.

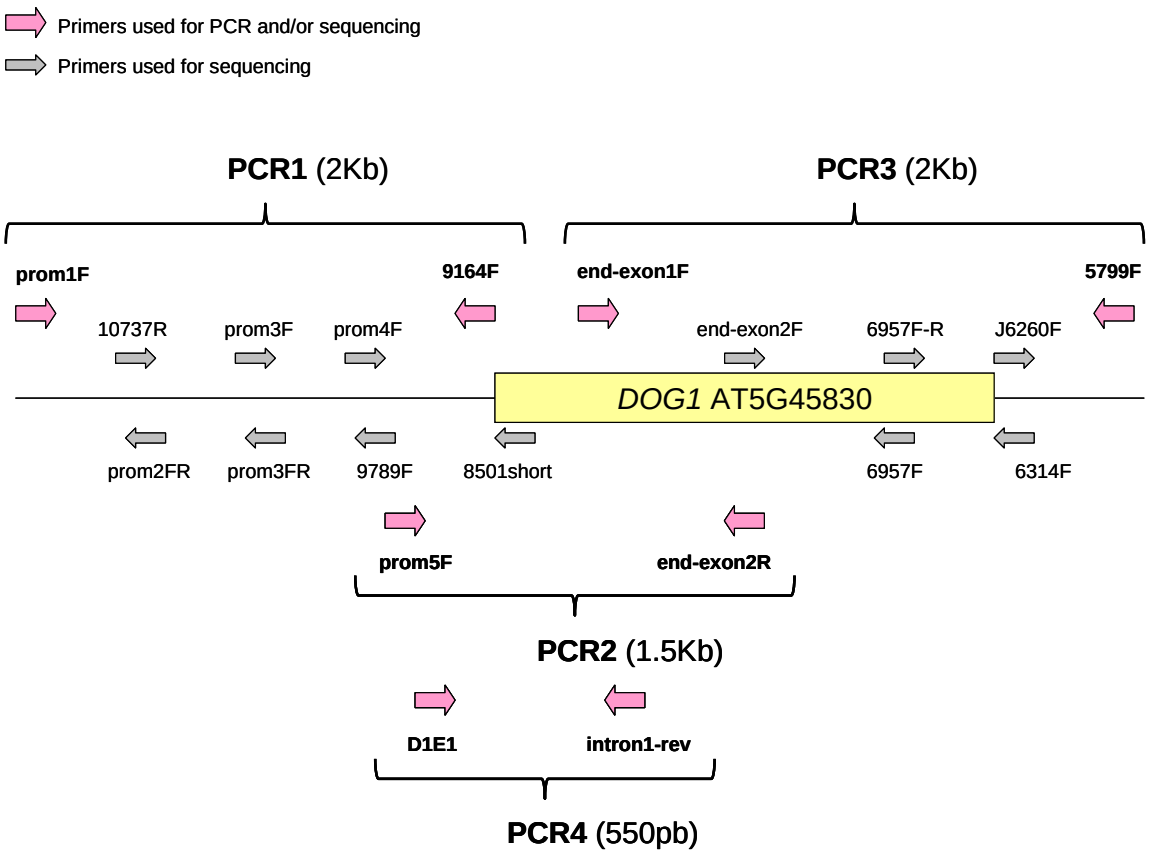


Figure 7. Scheme showing primers used for amplification and sequencing of *DOG1* gene.
Promoter and coding region of *DOG1* gene were sequenced for a subsample of 73-86 genotypes via amplification of three overlapped fragments (PCR1, PCR2 and PCR3). Exon1 was sequenced for the whole sample (161 genotypes) via PCR4. Primers used for PCR 1, 2, 3 and 4 and for sequencing are indicated with arrows.

Nucleotidic diversity & neutrality tests

To analyze patterns of nucleotide diversity of *DOG1*, the software DnaSP (DNA sequence polymorphism; Rozas and Rozas, 1999) was used. This software allows calculating diversity estimators such as nucleotide diversity θ_π (Nei and Li, 1979), θ_w (Watterson, 1975) and haplotype diversity. θ_π is the average number of nucleotide differences per site between two sequences. This estimate depends on allelic frequencies at polymorphic sites. θ_w depends on the SNP number, is normalized by sample size, but does not take allelic frequencies at polymorphic sites into account. The number of singletons, mutations appearing only once in the sample was provided as well. Concerning coding region, non-synonymous substitutions per non-synonymous site (KA)

and synonymous substitutions per synonymous site (KS) (Nei and Gojobori 1986) can be calculated. Moreover, DnaSP allows the estimation of minimum number of recombination events.

To detect selection on *DOG1*, we compared patterns of variation observed at this locus with the variation expected under standard neutral model, using different kinds of neutrality tests. Tajima' s *D* test (1989) reflects the difference between θ_π and θ_w. Under the neutrality hypothesis, $\theta_\pi = \theta_w$ and Tajima's $D = 0$. When $\theta_\pi > \theta_w$, Tajima's $D > 0$, there is an excess of polymorphisms in intermediate frequency and this can reflect a balancing selection. When $\theta_\pi < \theta_w$, Tajima's $D < 0$, there is an excess of rare polymorphisms (due to singletons) reflecting directional or purifying selection. Nevertheless this test is also sensitive to geographic isolation of populations or recent expansion, associated respectively with a positive and negative Tajima's *D*. Indeed such events as well as selection can affect nucleotide diversity similarly. *D* and *F* tests of Fu and Li (1993) are based on comparison between the number of singletons and θ_w and θ_π respectively. The interpretation of the tests of Fu and Li and Tajima are similar. While these tests are based on SNPs distribution, Fu's *Fs* test is based on haplotype frequencies. A negative value of *Fs* is evidence for an excess number of haplotypes, as would be expected from a recent population expansion or from directional selection whereas, a positive value of *Fs* is evidence for a deficiency of haplotypes, as would be expected from a recent population bottleneck or from balancing selection.

Selection can be detected as well using an outgroup sequence. *D* and *F* tests of Fu and Li can be calculated with an outgroup (*D'* and *F'*). Fay and Wu *H* neutrality test (Fay and Wu, 2000) is based on differences between the two diversity estimators: θ_π and θ_H. θ_H is based on the frequency of derived variants determined from an ancestral haplotype (outgroup). A sequence of *Arabidopsis lyrata* was used as orthologous outgroup sequence and derived polymorphisms were determined. A negative value of the Fay and Wu *H* test reflects an excess of derived variants at high frequency. Such pattern is produced by hitchhiking due to a recent positive selection. In contrast to Tajima's *D*, the Fay and Wu *H* is not sensitive to population expansion. However this test is sensitive to recent mixture of different populations. Thus, combining both tests allows distinguishing population expansion from purifying selection. Zeng et al. (2006) propose a normalized

version of Fay and Wu *H* test (*Hn*), using a new diversity estimator θ_L (the product of effective population size and neutral mutation rate), which is sensitive to the changes in high frequency variants. In addition, with DnaSP estimation of divergence as well as coalescence simulation can be performed. These simulations allow to build neutral expectations to establish the significance of neutrality tests.

There is a third series of tests that compare patterns of polymorphism and divergence across sites. The HKA test of Hudson, Kreitman and Aguadé (1987) is based on the "Neutral Theory of Molecular Evolution" (Kimura 1983), which predicts that regions of the genome that evolve at high rates, will also present high levels of polymorphism within species. The test requires data from one intra- and inter-specific polymorphism at two loci. Under neutrality, the ratio of silent polymorphism to divergence should be equal at both loci. Finally, the McDonald and Kreitman's test (1991) is based on a comparison of synonymous and non-synonymous sites at one locus, within and between species. Under neutrality, the ratio of replacement to synonymous fixed substitutions (differences) between species should be the same as the ratio of replacement to synonymous polymorphisms within species.

These various tests make use of various sources of information (polymorphism, divergence, synonymous vs. non synonymous, derived vs. ancestral variants) and together help reveal putative footprints of selection. Moreover the pattern of diversity of the *DOG1* gene was compared with empirical distribution of neutrality tests derived from randomly sequenced genomic loci (Black et al. 2001; Luikart et al. 2003).

2. *FRIGIDA and FLC* genotyping

FRIGIDA

To assay the *FRI* genotype in our material (119 among 161 genotypes) we collected genotyping data from the literature (Caicedo et al. 2004; Aranzana et al. 2005; Lempe et al. 2005; Werner et al. 2005; Korves et al. 2007). Based on allelic variation at the *FRI* locus we defined two genetic subsamples: 77 genotypes carrying functional *FRI* allele ("*FRI-F"* group) and 42 genotypes carrying non-functional *FRI* alleles ("*FRI-N"* group).

FLC

Two major haplotypes were identified at *FLC* locus (Caicedo et al. 2004). One SNP was genotyped using Derived Cleaved Amplified Polymorphic Sequences (dCAPS) to distinguish these two alleles in our sample.

3. 149 SNPs genotyping and population structure (for phenotypic analyses)

149 SNPs genotyping

The plants were genotyped for a set of 149 single nucleotide polymorphism (SNP) markers described in (Warthmann et al. 2007) by Sequenom, inc. (San Diego, CA). Because of an excess of missing data Mt-0 was not included in the analysis. Out of the 149 SNP markers, 139 were polymorphic in the whole sample and showed a proportion of missing data < 0.04.

Population structure

We used the Software STRUCTURE 2.2 (Pritchard et al. 2000) to assess the existence of population structure in our sample. The random 139 snps were used as information basis together with a haploid setting and the "linkage model" with "correlated allele frequencies". The algorithm was run with a burn-in length of 200,000 Markov chain Monte Carlo (MCMC) iterations and then 100,000 iterations for estimating the parameters. This was repeated five times for each *K* (ranging from one to 8). The number of clusters (K) in our sample was detected as described in Evanno et al. 2005 via calculation of L(K) and ΔK.

4. 250K SNPs genotyping and population structure (for GWA)

250K SNPs genotyping

Among our sample of 161 genotypes, 111 were genotyped in collaboration with M. Nordborg and his team (Atwell et al. 2010) using the 250K SNP – tiling Affymetrix

array: tSNPtile1, containing probe sets for 248,584 SNPs (Kim et al. 2007). Each SNP has four probes (two alleles on sense and anti-sense strand, respectively). Genomic DNA was extracted using Wizard Magnetic 96 DNA Plant System (Promega). For each sample, approximately 250 ng genomic DNA (in 48 uL distilled water) was labeled using BioPrime DNA labeling system (Invitrogen), at conditions modified as previously described (Wolyn et al. 2004). 16 µg of labeled product was hybridized to each array using standard Affymetrix array washing and staining protocols. Samples were processed at the microarray core facilities at the University of Chicago and the Children's Hospital of Los Angeles (USC).

SNPs were called using a modified version of the Oligo package (Carvalho et al. 2007). The resulting calls were then filtered in several steps to remove substandard arrays and SNPs. Following these procedures, missing SNPs were imputed using NPUTE (Roberts et al. 2007).

Kinship matrix estimation

With the full 250k data it would take too much time to estimate population structure with STRUCTURE software. Thus, to assess population structure a kinship matrix, which reflects familial relatedness between genotypes, was estimated using identity-in-state (Atwell et al. in press). This method was successfully applied in previous studies (Zhao et al. 2007; Kang et al. 2008).

III. Genome Wide Association mapping

Genome wide Association mapping method was investigated in collaboration with M. Nordborg and his team (Atwell et al. 2010).

Phenotypes and Log transformation

Nine Phenotypes were used for the Genome Wide Association mapping. Three of them are related to growth: Seedling Growth, Vern Growth, After Vern Growth and six of

them are related to germination: Primary Dormancy, Secondary Dormancy, Germination at 4°C and germination percentage after 7, 28 and 56 days of dry Storage (respectively, Storage 7 days, Storage 28 days, Storage 56 days). These phenotypes are described in paragraph I.2. Except germination percentage after 56 days of dry storage and germination at 4°C, all germination traits were analyzed after log-transformation for Genome Wide Association mapping (GWA).

EMMA and Wilcoxon test

Two types of association mapping methods were used, a non-parametric method Wilcoxon rank-sum test and a parametric mixed model EMMA to handle confounding by population structure.

EMMA (Efficient Mixed-Model Association) (Kang et al. 2008) is an implementation of a linear mixed model which accounts for population structure by adding a genetic random effect with a fixed covariance structure. Specifically,

$$Y = \beta X + u + \in,$$

where Y denotes the vector of phenotypes, X the vector of genotypes at the locus being tested, β the fixed phenotypic effects attributable to this locus, $u \sim N_n(0, \sigma_g^2 K)$ and $\in \sim N_n(0, \sigma_e^2 I_n)$ are random effects meant to capture the variance due to background genetic factors and the environment, respectively.

The kinship matrix K accounts for genome-wide patterns of relatedness between the individuals and is estimated only once (see paragraph II.4). The parameters of the model β, σ_g^2 and σ_e^2 are estimated using REML (restricted maximum likelihood) for each marker (SNP). A general t-statistic testing the null hypothesis $\beta = 0$ is calculated to get the p-value. The percentage of phenotypic variation explained by each SNP is calculated as

$$\frac{\sum_i (x_i \hat{\beta} - \bar{x}_i \hat{\beta})^2}{\sum_i (y_i - \bar{y})^2}$$

Quantifying Population structure

To quantify the confounding effects of population structure, observed and expected genome wide p-values distribution were compared resulting from applying Wilcoxon's test and EMMA (quantile-quantile plot of p-values).

Enrichment for *a priori* candidates

The flowering gene list was initially constructed to enable the candidate enrichment ratio analysis. In addition gene lists were made for each phenotype to enable faster assignation of genes to suitable candidates that have been previously identified as connected to the phenotype. Gene lists were constructed by searching The Arabidopsis Information Resource (http://www.arabidopsis.org/), for genes with suitable annotation and literature searches, however, literature searches were not extensive. All gene lists were constructed without looking at the association plots. Whilst a proportion of these genes have not yet directly been shown to be functionally responsible for the traits described, they would be selected for further analysis in any mapping experiment. Some phenotypes were checked with several appropriate gene lists, for example all early development phenotypes (germination, dormancy) were checked against the dormancy and germination gene lists. Any genes that were selected post gene list construction are marked with an asterisk in the gene tables. Candidate gene lists can be found on the website: http://arabidopsis.usc.edu.

Results

I. Natural variation of life history traits

1. Population structure - climatic and *FRIGIDA* alleles partitioning

Using the software STRUCTURE, as well as the method described in Evanno et al. (2005) to identify the most likely number of clusters, we found two major genetic clusters in our sample of 160 genotypes with 139 SNPs as information basis, (Appendix 3). These two clusters "western" and "eastern" form a clear west-east genetic gradient, shown on the map (figures 8 and 9). Our results are consistent with the west-east genetic structure found in previous studies (Beck et al. 2008; Nordborg et al. 2005; Ostrowski et al. 2006; Schmid et 2006; Sharbel et al. 2000). The western and eastern clusters contained 118 and 42 genotypes respectively. Each genotype has a probability upper than 0.5 to belong to its genetic group (see more details in figure 8). Genotypes of western cluster are from Northern America and Europe whereas the eastern cluster regroups genotypes from Eastern Europe and Asia. This divergence between the European and Asian metapopulations is probably the result of a geographical isolation due to the last glaciations or older fragmentation events (Beck et al. 2008; Sharbel et al. 2000). Moreover, the admixed genetic diversity found in Europe (figure 8) is consistent as well with results of previous studies suggesting that admixed populations in Europe are the result of post-glacial recolonization from different refugia including Central Asia. An alternative interpretation of the west-east genetic distribution found in *Arabidopsis thaliana* is that recolonization of Eurasia started from a single refugium in Iberian Peninsula and during migration up to Asia a bottleneck occurred (Ostrowski et al. 2006). In addition, a north-south genetic differentiation was found as well in this species (Nordborg et al. 2005; Schmid et al 2006) however such north-south genetic gradient was not detected in our sample. Indeed, five genotypes from northern Sweden which belongs to an other cluster as shown in Nordborg et al. (2005) were not included in our sample. Similarly in southern Europe, with our sample no cluster was found in Iberian Peninsula,

which is a putative glacial refugia (Comes and Kadereit 1998; Pico et al. 2008), because of the lack of genotypes from this area. Genotypes from northern America were particularly close to genotypes from western Europe (figure 8) which is consistent with previous results suggesting a recent introduction of *Arabidopsis thaliana* in North America from western Europe due to human activities (Kawabe & Miyashita 1999; Hoffman et al. 2003). Interestingly, Japanese genotypes did not form a cluster but were assigned either to western or eastern group, probably due to recent migration, similar results were found in Ostrowski et al. (2006).

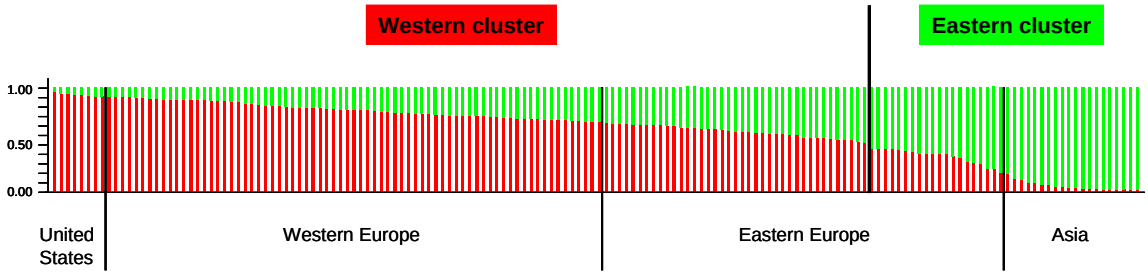


Figure 8. Population structure inferred with the STRUCTURE program with two genetic clusters (K=2). Each genotype is shown with its proportional assignation to the two clusters. We observe a clear west-east genetic gradient and we identify two major cluters: western cluster (red) and eastern cluster (green). Each cluster includes genotypes assigned to it with a probability upper than 0.5. We note the presence of admixed genotypes in eastern Europe, consistent with post-glacial recolonization events.

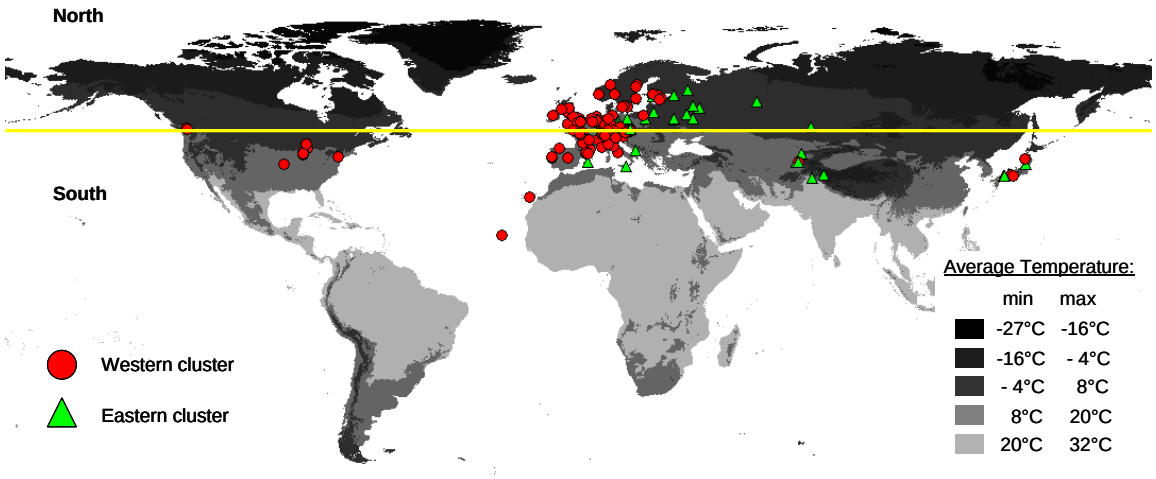


Figure 9. Population structure and climatic partionning. Our sample (160 genotypes) is devided in two genetically differenciated clusters: western cluster including genotypes from North America and Europe (in red circle), and eastern cluster including genotypes from eastern europe and Asia (in green triangle). This west-est genetic strucutre is consistent with post-glacial recolonization events. Japanese genotypes belong etheir to the western or eastern cluster due to recent migration. The yellow line divides the sample in two climatic groups: North and South. The annual average temperature (minimum and maximum) shown in legende, follows a latitudinal cline.

Temperature contrasts between seasons are similar in North and South but average temperature, average precipitation and precipitation contrasts along the year are significantly lower in the Northern subsample (Mann-Whitney test, minimum p=<0.001 (Appendix 4). The north-south climatic partitioning is independent from west-east genetic structure. Thus, these two subsets are ideal to analyze life history strategies and trait co-evolution in contrasted environments. Moreover, association between variances of climatic parameters and variances of life history traits was analyzed with the whole sample as well as in southern and northern subsamples.

FRIGIDA gene, an important determinant of flowering time carries two major alleles based on its functionality, *FRI-F* functional allele and *FRI-N* non functional allele. In this study we confirm that *FRIGIDA* gene is significantly associated with natural variation of flowering time (after four weeks of vernalization), nevertheless there is still an important variance in flowering timing in each *FRIGIDA* group, *FRI-F* and *FRI-N*. In this study we want to know if these two alleles co-evolve with other genes and if they are associated with different evolutionary trajectories concerning life history strategies. We therefore investigated life-cycle characteristics in each of these two *FRIGIDA* groups.

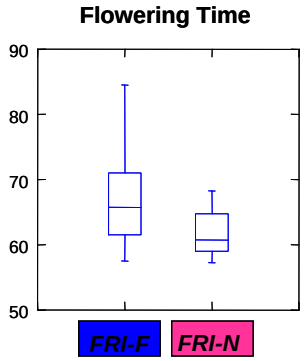


Figure 10. Significant association with our data between *FRIGIDA* alleles and natural variation of Flowering time with 77 genotypes carrying *FRI-F* and 42 carrying *FRI-N* tested with Mann-Whitney test (p<0.001). Association very well documented in the litterature.

The geographical repartition of *FRI-F* and *FRI-N* did not overlap with the north-south climatic groups examined above. Indeed, the frequencies of *FRIGIDA* alleles, *FRI-F* and *FRI-N* and missing data were equally distributed in North and South, they were respectively 50%, 25%, 25% and 46%, 27%, 27%. Moreover, despite genotypes carrying *FRI-N* allele are located mostly in intermediate latitude, climatic conditions are not significantly different between *FRI-F* and *FRI-N* subsamples (tested with Mann-Whitney test). Finally, using a non parametric test (Mann-Whitney test) we showed that the two

groups *FRI-F* and *FRI-N* were not associated with the genetic structure based on the 139 SNPs.

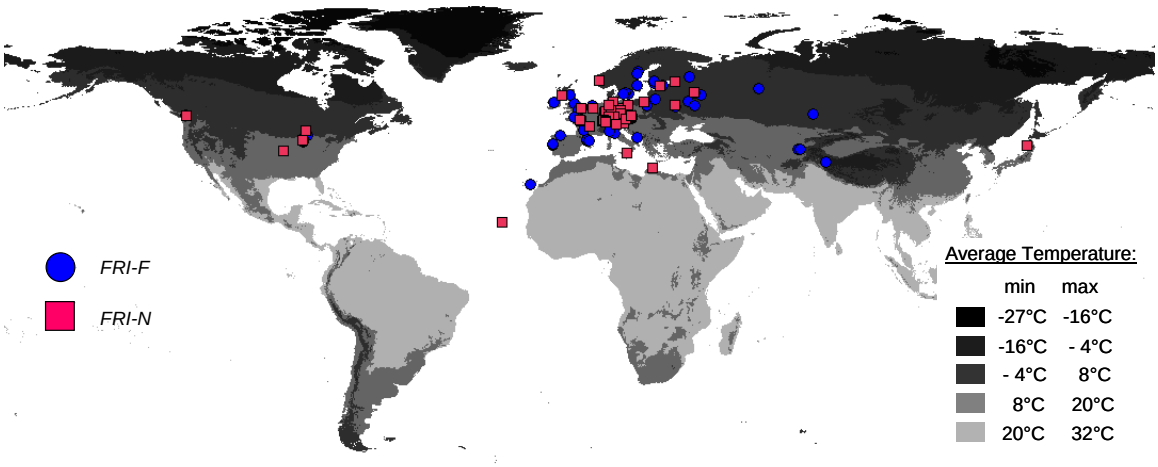


Figure 11. Geographical distribution of *FRIGIDA* alleles. *FRIGIDA* gene carries two major alleles: functionnal allele (*FRI-F*) and non fuctionnal allele (*FRI-N*). Genotypes carrying *FRI-F* are blue circles. Genotypes carrying *FRI-N* are pink square. The geographical distribution of *FRIGIDA* alleles is independent of north-south climatic groups and population structure (shown figures 8,9).

2. Natural variation in life-history

Primary dormancy is established during seed maturation on the mother plant. For most of the 161 genotypes tested, primary dormancy release was monotonous but occurred at variable speed (Figure 12A). DSDS50 ranged from 3.5 to 264 days, with an average of 74 days, a median value of 53 days and a standard deviation of 68.9 days (Figure 12A). Heritability of DSDS50 values was high: 0.95. This experiment was replicated independently with another batch of seeds harvested from greenhouse grown plants, with a DSDS50 ranging from 3.5 to 201 days, with an average of 23 days and a median value of 16 days. Although dormancy levels were generally lower in this independent experiment (data from Sigi Effgen), both experiments were correlated ($R= 0.49$, $p < 0.001$). For a subset of 96 genotypes, seed dormancy was measured independently using seeds grown in a growth chamber experiment under uniform light and temperature conditions, which again yielded correlated results ($R= 0.46$, $p < 0.001$). Moreover, for a subset of 29 genotypes, germination rates of seeds after ripened for approximately 2

months after harvest in the field (in Valencia, Spain) were scored and were significantly although less correlated with dormancy levels observed in the main experiment ($R = 0.25$ $p = 0.001$).

A 6-week-long exposure to cold, decreased germination in 103 genotypes among a total of 124 fully after-ripened genotypes. Secondary dormancy ranged from – 0.18 (no secondary dormancy) to 2.29 (percentage germination decrease/day) with an average of 0.58, a median value of 0.27 and standard deviation 0.68 (Figure 12A). Heritability of secondary dormancy was 0.91.

We observed that a number of genotypes displayed the ability to germinate at 4°C in darkness indicating that they do not require light for germination. We observed that the germination percentage in darkness at 4°C after 1 week is highly correlated with the percentage of germination in darkness after 6 weeks ($R=0.87$; $p<0.001$). Thus, germination speed is not influencing the proportion of seeds germinating in those conditions. The ability to germinate in darkness at 4°C was measured as the proportion of non-dormant seeds germinating after 1 week at 4°C. The non dormant seeds, which did not germinate in darkness at 4°C after 1 week, did so after one week at higher temperature with light, thus cold exposure did not impair their viability. Germination at 4°C in darkness ranged from 0 to 100% of germination with an average of 53%, a median value of 57.5% and a standard deviation of 35% (Figure 12A). Heritability of germination at 4°C was 0.91.

We also found natural variation for all growth rate related traits, which were measured at seedling establishment (Seedling Growth), during vernalization (Vern Growth) and after vernalization (After Vern Growth). The mean for these traits was respectively 0.00775, 0.00423 and 0.104 (cm2/day) and the heritability for all was relatively high respectively 0.78, 0.88 and 0.84. (Figure 12B)

Finally, we found important variation in flowering timing which reflects variation in response to vernalization (figure 12C). Thus, all genotypes responded completely to vernalization and consistently most of Scandinavian genotypes which are known to need

a much longer vernalization treatment flowered among the latest (after 70 days after sowing). Mean and the median of flowering time, were respectively 65 and 64 days. The heritability of this trait was very high (h^2 = 0.97). Mean of rosette and stem leaves number was respectively 18 and 5. Heritability of those traits was respectively 0.96 and 0.95. (figure 12C)

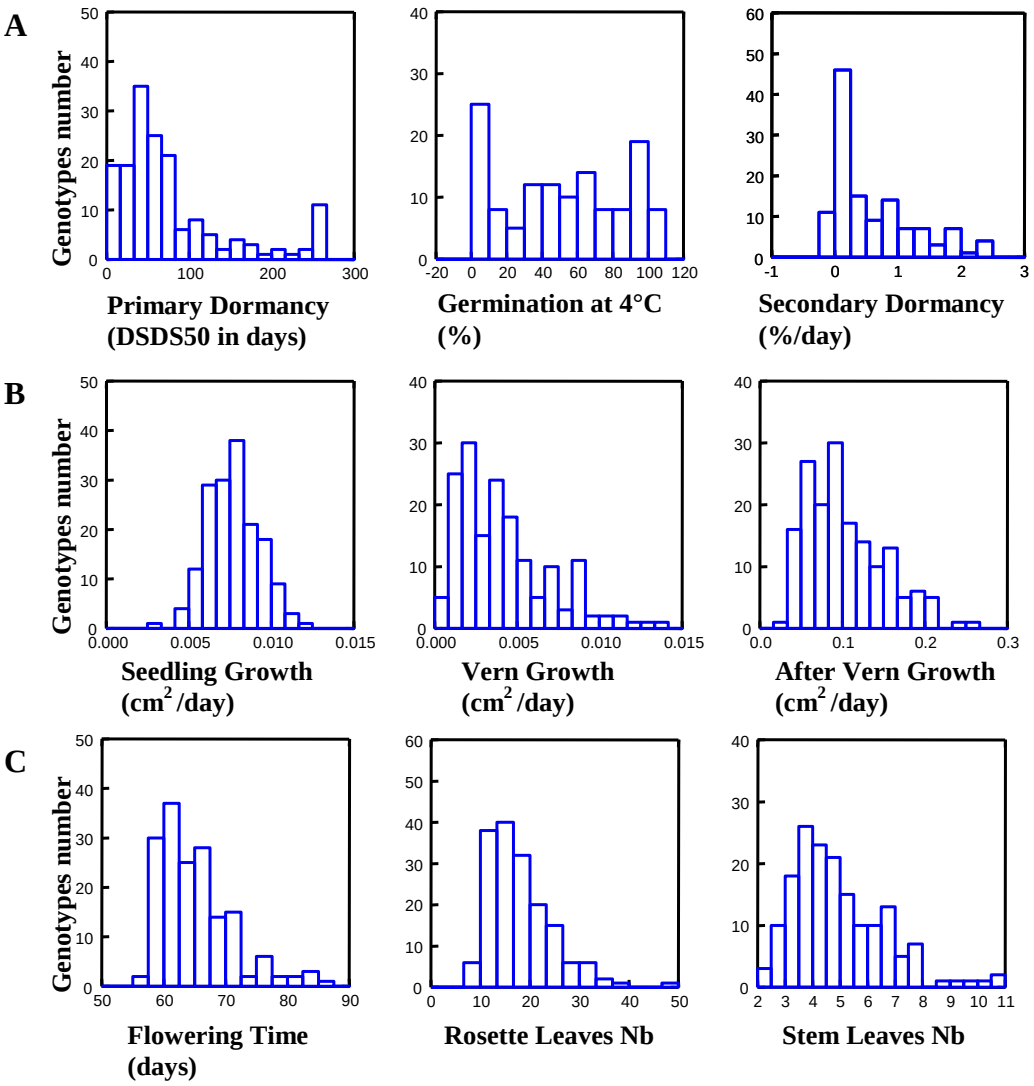


Figure.12. Histograms of phenotypes with adjusted means. **A. Germination related traits.** Primary Dormancy: measured by calculation of DSDS50 in days (number of days required to reach 50 % of germination). Germination at 4°C: percentage of germination among fully after ripened seeds at 4°C in the dark. Secondary Dormancy: measured by decrease rate of germination percentage (in % per day) of fully after-ripened seeds after treatment at 4°C in the dark. **B. Vegetative growth related traits.** Seedling Growth: growth rate of seedling in Cm2/day. Vern Growth: growth rate during vernalization in Cm2/day. After Vern Growth: growth rate after vernalization in greenhouse conditions in Cm2/day. **C. Flowering time related traits:** Flowering time: number of days at first flower after snowing. Rosette Leaves Nb: number of leaves at rosette at flowering timing. Stem Leaves Nb: number of leaves at stem at flowering timing.

	Heritabilities
Primary Dormancy (days)	0.95
Secondary Dormancy (%/day)	0.91
Germination at 4°C (%)	0.90
Seedling Growth (Cm^2 /day)	0.78
Vern Growth (Cm^2 /day)	0.88
After Vern Growth (Cm^2 /day)	0.84
Flowering Time (days)	0.97
Rosette Leaves Nb	0.96
Stem Leaves Nb	0.95

Table 2. Heritabilities of life history traits. For each trait, the heritability (h^2 = MSgenotype / (MSgenotype + MSerror)) which is the proportion of phenotypic variance due to genetic variance, was obtained through an analysis of variance (ANOVA) with genotypes as fixed factor (SYSTAT 11).

3. Correlations between life history traits

Correlation between life-history traits may reveal constraints on the evolution of life-history strategies. These constraints may vary both with the environment and with the genetic background. We examined correlations among traits over the whole sample as well as over subsamples partitioned by latitude (Northern vs. Southern) or by the genetic control of flowering time (*FRI-F* and *FRI-N*).

Most of the traits that describe the same life-cycle phases (e.g. germination, growth or flowering) are very strongly correlated (Table 3). Instead, no significant correlation was observed between traits describing different phases. For germination and growth rate, however, the pattern was more complex. Secondary dormancy is significantly negatively correlated with the ability to germinate at 4°C ($r=-0.73$, $p < 0.001$), as expected since seeds that have germinated in the cold cannot display any dormancy. But no significant correlation was found between primary dormancy and other germination-related traits. For vegetative growth, we found that growth rates during and after vernalization were highly correlated ($r=0.88$, $p<0.001$) but uncorrelated to seedling growth rate (minimum $p=0.607$). Floral traits were all highly correlated among each other (minimum $r = 0.57$, $p<0.001$). These patterns were not modified by the partition into geographical or allelic subsamples (Table 3).

Interestingly, the Northern subsample displayed a significant negative correlation between flowering time and growth rate after vermalisation ($r= -0.42$, $p=0.032$). Instead, in the Southern subsample, there was a positive correlation between rosette leaf number and growth rate after vernalization but not significant ($r= 0.38$, $p=0.074$). Because rosette

leaf number was measured at flowering and therefore is strongly correlated to flowering time (Table 3), this shows that the relationship between flowering time and growth is strongly contrasted between latitudinal groups, an effect that remained invisible over the whole sample.

We further asked whether the functionality of *FRI* influenced the patterns of correlation between life cycle traits. Hundred and nineteen genotypes of our sample were previously genotyped at the *FRI* locus. Correlation analyses were performed among all traits with different *FRI* allele. Interestingly the genotypes carrying *FRI-F* showed a significant negative correlation between primary and secondary Dormancy (r= -0.43, p= 0.020, Table 3). In a *FRI*-functional background, seeds which released primary dormancy slowly did not enter in secondary dormancy and generally germinated at 4°C in the dark. In addition, within this subsample, we observed a significant negative correlation between primary dormancy and flowering time (r= -0.40, p= 0.050, Table 3). The genotypes that flower the latest tended to display low levels of dormancy. Instead, we did not observe genotypes showing strong dormancy and late flowering time (except Mr-0 and Omo2-1 respectively from Italy and Sweden). On the other hand, in the subsample with a non-functional *FRI*-N allele, rosette leaf number and growth rate after vernalisation were positively correlated (r= 0.53, p= 0.035, Table 3). This correlation was slightly stronger than the one found across the Southern subsample of genotypes.

pair correlations	**All**	**North**	**South**	***FRI-F***	***FRI-N***
Germination at 4°C - Secondary Dormancy	-0.73***	-0.84***	-0.67***	-0.80***	-0.64***
Vern Growth - Vern After Growth	0.88***	0.85***	0.91***	0.87***	0.90***
Rosette Leaves Nb - Stem Leaves Nb	0.63***	0.59***	0.67***	0.60***	0.74***
Flowering Time - Rosette Leaves Nb	0.85***	0.88***	0.81***	0.87***	0.74***
Flowering Time - Stem Leaves Nb	0.57***	0.50***	0.64***	0.51***	0.73***
Primary Dormancy - Secondary Dormancy	ns	ns	ns	-0.43*	ns
Primary Dormancy - Flowering Time	ns	ns	ns	-0.40*	ns
After Vern Growth - Flowering Time	ns	-0.42*	ns	ns	ns
After Vern Growth - Rosette Leaves Nb	ns	ns	0.38ns	ns	0.53*

Table 3. Coefficients of correlation between life history traits in the whole sample as well as in climatic and *FRIGIDA* alleles subsamples. Pearson genetic correlations were calculated between all life history traits and statistical significance of the correlations were calculated with Bonferroni correction (SYSTAT), in the whole sample and in climatic and *FRIGIDA* alleles subsamples. In this table only pair of traits which show at least one significant correlation in one of the category are shown. The coefficient of correlation (r) and the p-values are indicated (p-values: *=< 0.05; **=<0.01; ***=<0.001).

4. Path Analysis and contribution of latitude and population structure on life-history variation

We used path analysis to quantify the respective contribution and associated significance of latitude and population structure (output by STRUCTURE K=2) on life-history variation (Figure 13). Variation of all climatic parameters: Averages temperature and precipitation as well as Temperature and Precipitations contrasts, were significantly correlated with latitude. Average temperature was the most correlated with latitude. When latitude and population structure were introduced either as co-varying or under the control of one or the other, the model did not converge. As these two variables are in fact not correlated (R^2=-0.07, p>0.1), we erased this link from the model. The results show that each of the 7 traits included in this analysis was controlled by one or both components: latitude and population structure. Flowering time and seedling growth variation were both controlled by latitudinal variation but not by variation in population structure. Instead, variation in germination at 4°C and growth rate during vernalization were correlated with population structure but not with latitude. Interestingly, both primary and secondary dormancy associated significantly with both the latitudinal gradient and population structure. In each of these cases, the path coefficients towards latitude were comparable to (and even larger than) those towards population structure (Figure 13). However, the model accounted relatively poorly for the genetic variance observed (Chi2= 378.5, ddf= 62, p<<1.e-06, NFI= 0.6 << 0.9). This analysis was restricted to the whole sample only and not applied to the subsamples because of convergence problems that hindered the estimation of parameters.

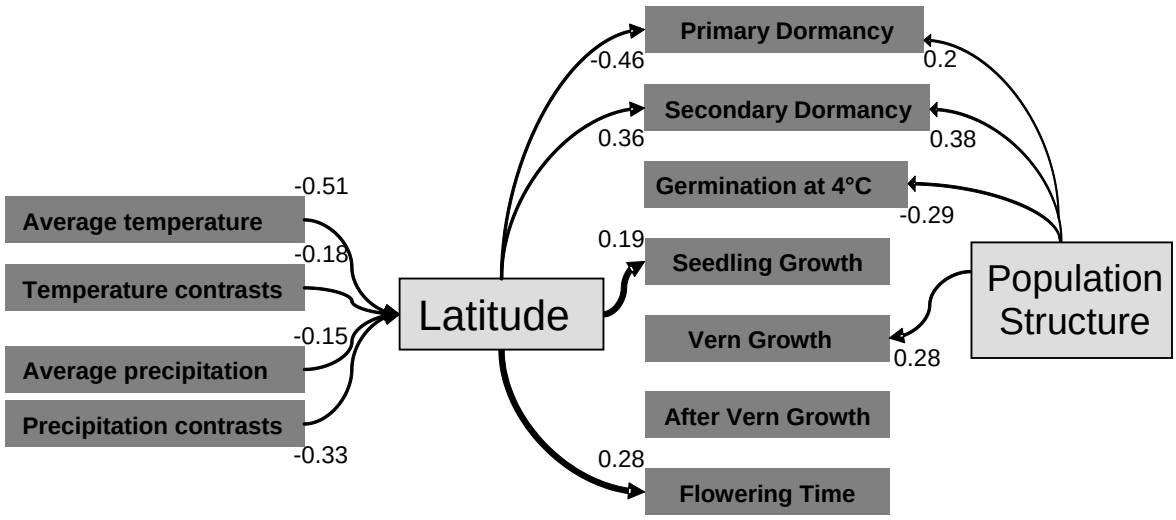


Figure 13. Path analysis showing the influence of latitude and population structure on life-history traits variation. Standardized path coefficients are given. Only the path that are significant (P<0.05) are represented.

5. Association of life-history strategies with environmental parameters

We used a multiple regression approach to identify life-cycle traits that best explain environmental variables such as latitude, average temperature, average precipitation, temperature contrasts, and precipitation contrasts (table 4). We did the analysis this way to integrate the life history traits as a whole. Moreover latitudinal variation does not include only climatic parameters and therefore temperature and precipitation were studied separately to determine whether some traits are influenced specifically by one or both of those climatic parameters. In addition, we identified the traits co-varying with population structure, in order to identify association related to the divergence between European and Asian genotypes. In this analysis, when traits display a correlation (as shown previously), only the most orthogonal to the other traits in the model was maintained. This analysis reveals that four life-cycle traits explain 20% of the latitudinal variance (Table 4). Primary dormancy (p= 0.007), secondary dormancy (P= 0.029), seedling growth rate (p= 0.031) and number of stem leaves (p = 0.038) significantly follow the latitudinal gradient without being correlated to each other. In this

way, these four traits define a four dimensional parameter space in which a "life-cycle strategy" at any given latitude can be represented. Secondary dormancy is also associated with population structure (R^2= 0.13, p< 0.001, Table 4). This is an indication that variation in secondary dormancy may be confounded by demographic history.

The same approach showed that in our sample, dormancy traits together explain 14% of the variance in temperature (p= 0.001 and p= 0.038, for primary and secondary dormancy, respectively), although only primary dormancy was independent from population structure. Average precipitation and precipitation contrasts instead were only weakly explained by primary dormancy (not significant) and stem leaf number respectively (R^2= 0.02, p= 0.080 and R^2= 0.04, p=0.013 respectively Table 4). Thus, the expected significant association between primary dormancy and precipitation was not found. Nevertheless, it appears that strong degree of primary dormancy was associated either with law Average precipitation either with high precipitation contrasts which explains why a significant association was not found.

We reasoned that adaptation of the life-cycle may require different coordination between traits along the climatic range. To test this, we repeated multiple regressions after partitioning our sample into a Northern and a Southern sub-sample.

Primary dormancy was significantly lower in the Northern subsample and the opposite was true for secondary dormancy and stem leaf number, which were lower in the Southern subsample (Mann-Whitney test, respectively p= 0.005, p= 0.014 and p=0.04). Overall, there was no significant difference in seedling growth rate across subsamples (Mann-Whitney test, p =0.63)

In the Northern subsample, flowering time explained on its own 21% of the latitudinal gradient (p <0.001, Table 4). Variation of other traits including seedling growth, instead, associated with population structure. On contrary, in the Southern subsample, only seedling growth explained up to 5% (p = 0.040) of the latitudinal gradient and variation of flowering time depended on population structure (Table 4). Therefore, flowering time and seedling growth rate appeared to be adaptive traits in only part of the climatic range. In North, variation of germination related traits (primary and secondary dormancies) whereas in South variation of a floral related trait (Stem leaves

number) associated with Temperature contrasts. Moreover only in the southern subsample some germination traits (Secondary dormancy and germination at 4°C) explained significantly the variation of a climatic parameter related to precipitation. Germination at 4°C is the only trait that explained significantly precipitation contrasts along the year, independently of population structure. Thus, the same climatic parameter could influence different life history traits in northern and southern environments and some climatic parameters influenced life cycle variation only in a restricted geographical area.

Secondary dormancy and rosette leaf number explained 23% of latitudinal variance among the genotypes with a functional *FRI* allele, although the former trait also displayed some association with population structure. Furthermore, in this group, secondary dormancy and flowering time explained a significant part of variation in average temperature. By contrast, among the genotypes with a null *FRI* allele, latitudinal variance (20%) was explained by secondary dormancy as well as seedling growth and variation in average temperature (22%) was explained by seedling growth. In addition, rosette and stem leaf numbers associated respectively with contrasts in temperature and average precipitation which was not observed among genotypes carrying *FRI-F*. Despite they are no significant climatic differences between *FRI-F* and *FRI-N* subsamples, interestingly environmental parameters influenced differently life history variation in these two groups. Thus, *FRI-F* and *FRI-N* alleles were associated with different evolutionary trajectories.

6. Comparison path analysis and multiple regressions

With the whole sample, both analyses path and multiple regression, showed that variation of primary dormancy, secondary dormancy, seedling growth and a floral trait (respectively flowering time and Stem leaves number) associated with latitudinal variation. Variation of secondary dormancy was associated with population structure in both analyses but nevertheless path analysis revealed that primary dormancy, germination at 4°C and Growth rate during vernalization depend on population structure as well.

	All						North						South						FRI-F						FRI-N					
	Latitude	mean-TP	coef.variation-TP	mean-PP	coef.variation-PP	Structure	Latitude	mean-TP	coef.variation-TP	mean-PP	coef.variation-PP	Structure	Latitude	mean-TP	coef.variation-TP	mean-PP	coef.variation-PP	Structure	Latitude	mean-TP	coef.variation-TP	mean-PP	coef.variation-PP	Structure	Latitude	mean-TP	coef.variation-TP	mean-PP	coef.variation-PP	Structure
Primary Dormancy	**	***		*					*			***																		*
Secondary Dormancy	*	*				***			*			*					*	***	**	**				***	*					*
Germination at 4°C																	*													
Seedling Growth	*											*	*												*	*				
Growth at 4°C																														
Growth after 4°C																														
Flowering Time							***											*		*										
Rosette Leaves Nb																			**								**			
Stem Leaves Nb	*				*										***		**	*								*		**		
R-square	0.20	0.14		0.02	0.04	0.13	0.21		0.20			0.28	0.05		0.16		0.17	0.28	0.23	0.21				0.23	0.20	0.22	0.16	0.18		0.30
Nb genotypes	124	123		156	157	123	80		62			62	81		78		61	61	62	62				62	36	42	42	42		35

Table 4. Latitudinal and climatic gradients. Using general linear multivariate model, associations between life history traits and latitude, climatic variables and population structure were performed. R-square, p-values (p-values: '=< 0.08; *=< 0.05; **=<0.01; ***=<0.001) and number of genotypes are indicated. The multiple regression was performed with the whole sample as well as geographical and *FRIGIDA* groups.

II. Analysis of nucleotide diversity of DOG1 gene

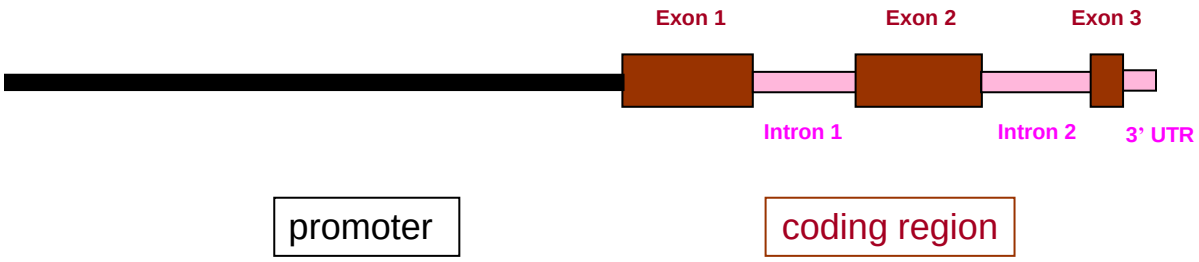


Figure 14. *DOG1* structure. *DOG1* coding region (1.9kb) contains three exons, two introns and a 3' UTR region. Extremity of Exon 2, Intron 2 and Exon 3 are involved in alternative splicing. The putative promoter is slightly longer with a size of 2.6 kb.

1. Nucleotide diversity pattern in *DOG1* gene

Recombination events in *DOG1* gene

In the *DOG1* promoter two recombination events were detected. Moreover, at least three recombination events occurred in the coding region. We note that they did not occur in exon1.

Comparison *DOG1* promoter/coding (exons, introns, UTR) regions

The coding region of *DOG1* contained more SNPs than the promoter region (82 against 28) although the coding region is slightly shorter. Moreover, based on the diversity estimators θ_π and θ_w, the nucleotide diversity in the coding region was higher than in the promoter region (θ_π = 0.0053; θ_w = 0.0091 and θ_π = 0.0042; θ_w = 0.0063, respectively) (Table 5). In both regions, Tajima´s *D* test, Fu and Li's *D* and *F* tests did not detect any significant departure from neutral expectations (Table 5). Nevertheless, the values of all those tests were negative which reflects most likely the recent post-glacial expansion of *Arabidopsis thaliana*. Similar diversity patterns were found in the whole genome (Nordborg et al. 2005; Schmid et al. 2005). The value of Fu's *Fs* test was negative as well and revealed an excess of number of alleles in both coding and promoter regions confirming a recent expansion. Fu's *Fs* test was the only significant negative neutrality test which is consistent because this test is more sensitive to population expansion (or purifying selection) than Tajima´s *D* test (Fu 1997). The ratio KA/KS including all *DOG1* exons was smaller than 1 as found in multiple random exons in the genome (Schmid et al. 2005) (Table 5).

Thus, the diversity pattern of the promoter and the whole coding region of *DOG1*, reflected the recent demographic expansion of *Arabidopsis thaliana*, as found previously in multiple random genomic fragments (Nordborg et al. 2005; Schmid et al. 2005). To understand more in details the diversity in *DOG1* coding region, the different functional parts of this gene were analyzed and compared.

The reduction of the size of the studied sequences can decrease the statistical power. Nevertheless exons 1 and 2 as well as introns 1 and 2 contained more than 10

SNPs. Moreover, the size of the sequences of exon 1, exon 2, intron 1 and intron 2 are comparable. Thus, the context was ideal to compare these different parts of the gene.

Comparison between exon1/exon2

The diversity of the exon 1 (θ_π = 0,0075; θ_w = 0.011) was around twice higher than the diversity of the exon 2 (θ_π = 0.0046; θ_w = 0.0061) and much higher than exon diversity in average in *Arabidopsis thaliana* (Nordborg et al. 2005). Moreover the ratio KA/KS is higher than 1 (KA/KS = 1.64) for the exon 1, whereas this ratio is smaller than 1 for the exon 2 (KA/KS = 0.16) (Figure 15). To explain the high rate of non synonymous substitutions in exon 1, there are three major interpretations. First, selection is relaxed in this part of the gene. Second, the amino acid sequence of exon 1 is subjected to some evolutionary force maintaining various alleles within the population. Third, this high diversity in amino acid is due to a recent mixture of several populations. For exon1, the *H* value of the Fay and Wu test, which is not sensitive to population expansion, was very negative (*Hn* = -2.41). This test revealed an excess of derived variants at high frequency, which is the result of hitchhiking due to positive selection. But it could be as well the result of a recent mixture of populations. For exon 1, the values of tests *D* and *F* of Fu and Li, were not significant but were positive which is consistent with the excess of alleles in high frequency found with *H* test. Moreover, for exon 1, the Fu's *Fs* test was close to zero. Thus, this test confirmed that the diversity in this part of the gene which contained only one singleton did not reflect the recent population expansion and contrasted with observations made over the whole coding region or across exons (Table 5). Indeed, diversity of exon 2, which contained more singletons than exon 1 (4 against 1), reflected significantly population expansion based on Fu's *Fs* test (*Fs* = -3.12 ; $p < 0.05$). HKA and McDonald and Kreitman tests were not significant, nevertheless based on ratio KA/KS divergence between *DOG1* exons (exons1, exons2) and the orthologue sequences of *A. lyrata*, showed that exon 1 has evolved faster than exon 2, which is consistent with a possible selection detected in exon1 (table 5).

Thus, the pattern of diversity observed at exon 1 shows an unusually high level of amino-acid polymorphism. Nevertheless, the evolutionary scenario which can explain this diversity is not clear.

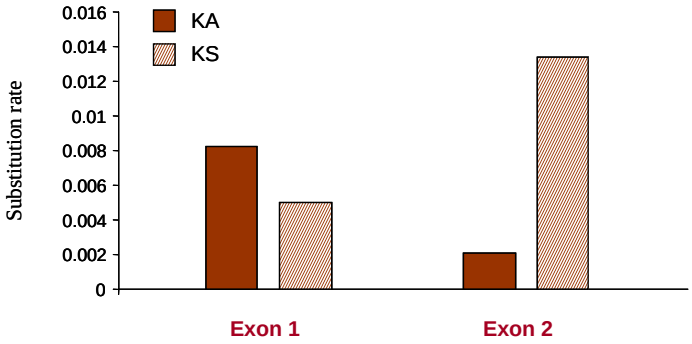


Figure 15. Rate of polymorphism at non-synonymous (KA) and synonymous (KS) position in exon1 and exon2 of *DOG1*. KA: non synonymous substitution rate. KS: synonymous substitution rate. In exon 1, ratio KA/KS was higher than 1 whereas in exon 2 ratio KA/KS was lower than 1.

Comparison between intron1/intron2

The diversity of the intron 2 (θ_π = 0,0019) is 3 times lower than the diversity of the intron 1 (θ_π =0,0068) (table 5) and much lower than intronic diversity in average in *Arabidopsis thaliana* (Nordborg et al. 2005). The very low diversity in intron 2, could be explained by selection on sites involved in alternative splicing. For intron 2, Fay and Wu *H* test did not detect any significant departure from neutral expectation (Table 5). However, both Tajima's *D* and Fu's *Fs* tests were significantly negative (-1.82 $p<0.05$ and -10.69 $p<0.001$, respectively). These tests indicate that diversity in intron 2 can reflect either population expansion or directional selection. By contrast, for intron 1 Fay and Wu *H* test was negative whereas Tajima's *D* and Fu's *Fs* tests were not significant (Table 5). Diversity of intron 1 seemed to be influenced by diversity pattern of exon 1 due to hitchhiking (Figure 16, 17), however the presence of 5 singletons in intron 1 suggests that new mutations have already occurred since the possible selection on exon 1.

Diversity distribution from promoter to last intron of *DOG1* gene

Based on θ_π and θ_w diversity estimators, a peak of diversity was observed in exon 1, and on both sides of this exon diversity decreased (Figure 16). Neutrality tests displayed the same pattern (Figure 17). Fu's *Fs* test was particularly negative and significant in *DOG1*

promoter and intron2 (*Fs* = -6.37, p<0.001; *Fs* = -10.69 p<0.001, respectively) whereas close to zero in exon1. By contrast Fay and Wu *H*/*Hn* test was negative in exon1 and close to zero in intron2. Moreover, Fu and Li's *D/D'* and *F/F'* tests were positive in exon1 and negative in all the other parts of the *DOG1* gene. Taken together, these results suggest that the evolution of exon1 departs from the other regions of DOG1 (Figure 16, 17). Interestingly, a signature of positive selection is apparent in intron 2, which is involved in alternative splicing. Alternative splicing was shown to be absolutely required for the function of DOG1 in dormancy, in *A. thaliana*.

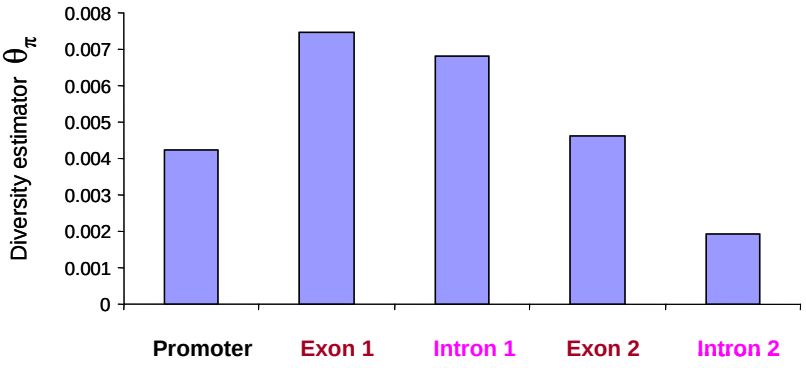


Figure 16. Distribution of nucleotidic diversity of *DOG1* gene. θ_π uses the average number of pairwise differences. There is a peak of diversity in exon1 and diversity decreases on both sides of this exon. There is particularly a low variation in intron2 involved in alternative splicing.

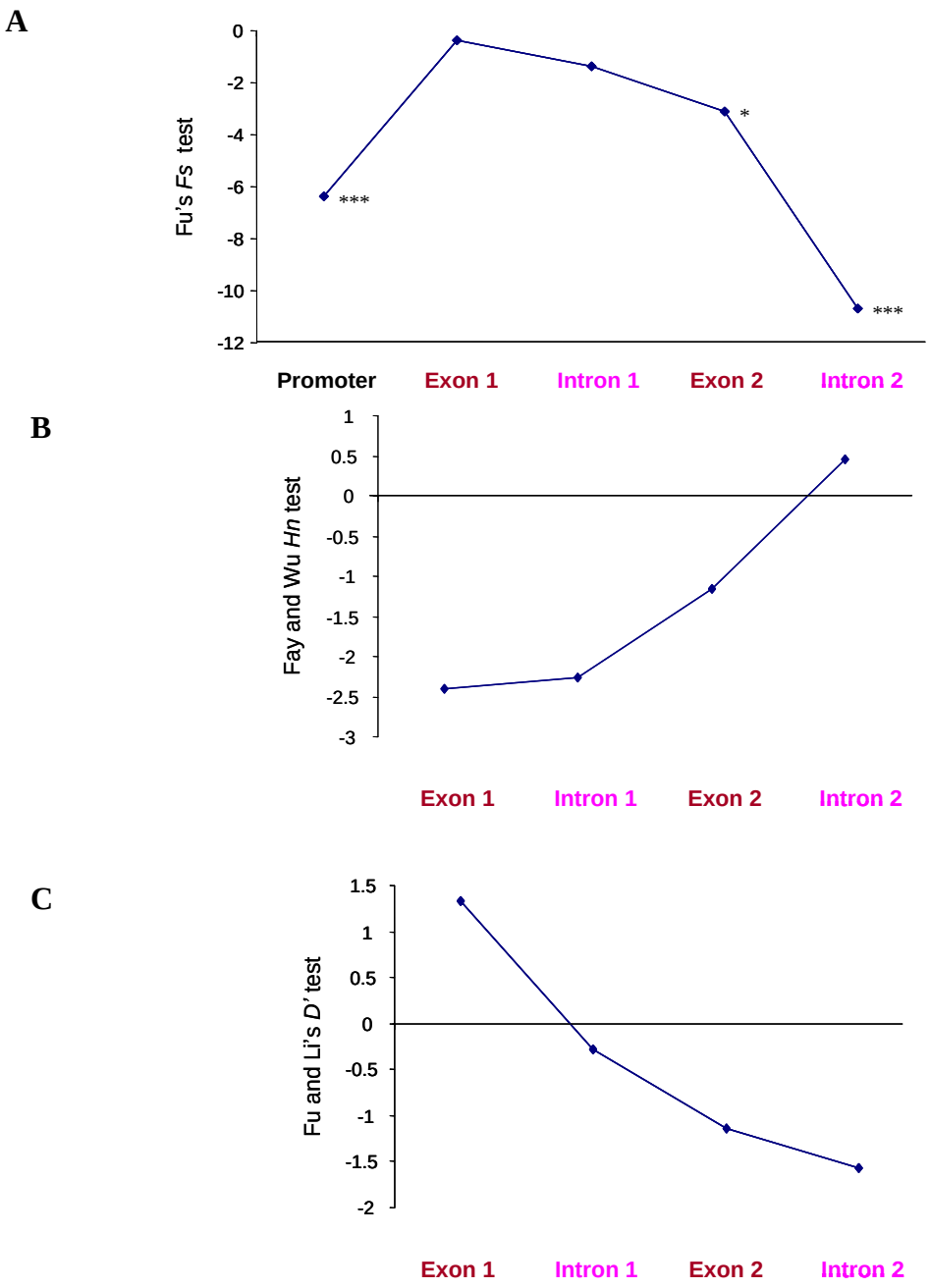


Figure 17. Distribution of neutrality tests in *DOG1* gene. **A.** *Fs* test is significantly negative for promoter, exon2 and intron2 of *DOG1* gene. *Fs* is negative but not significant in intron 1 and close to zero in exon1. * = p<0.05 and *** = p<0.001. **B.** Fay and Wu *Hn* test; *Hn* increases progressively until to be close to zero in intron 2. **C.** Fu and Li *D* test is positive in exon1 and negative in the other parts of the gene.

		Promoter	conding regoin	Exon1	Intron1	Exon2	Intron2
	Number of lines	86	73	73	73	73	73
	Number of sites	2639	1931	393	355	408	504
	Number of SNPs	28	82	19	12	12	12
	Number of singletons	6	17	1	5	4	5
Diversity estimators	θπ	0.00423	0.0053	0.00747	0.00679	0.00463	0.00191
	θw	0.00627	0.00915	0.011	0.00728	0.0061	0.00536
	KA	_	0.00499	0.00821	_	0.00209	_
	KS	_	0.00965	0.00501	_	0.0134	_
	KA/KS	_	0.52	1.64	_	0.16	_
	Number of haplotypes	22	32	11	11	12	14
	Haplotypic diversity	0.896	0.954	0.835	0.82	0.749	0.68
Neutrality tests	Tajima's *D* (+)	- 1.00 ns	- 1.423 ns	- 0.879 ns	- 0.188 ns	- 0.668 ns	- 1.816*
	Fu and Li's *D*	- 0.122 ns	- 0.0612 ns	1.307 ns	- 1.472 ns	- 0.882 ns	- 1.836 ns
	Fu and Li's *F*	- 0.546 ns	- 0.724 ns	0.611 ns	- 1.217 ns	- 0.958 ns	- 2.171 ns
	Fu's Fs (++)	- 6.375***	- 5.734**	- 0.365 ns	- 1.369 ns	- 3.125*	- 10.688***
(outgroup)	Fu and Li's *D'*	_	_	1.33ns	-0.28ns	-1.14ns	-1.57ns
(outgroup)	Fu and Li's *F'*	_	_	0.56ns	-0.24ns	-1.32ns	-1.95ns
(outgroup)	Fay and Wu's H/Hn (-)	_	_	-6.54/-2.41	-3.08/-2.26	-2.01/-1.16	0.85/0.46
(outgroup)	HKA	_	_	0.42ns	0.49ns	0.42ns	0.49ns
(outgroup)	McDonald & Kreitman	_	_	ns	_	ns	_
(outgroup)	Divergence KA/KS	_	_	0.42	_	0.23	_

Table 5. Analysis of *DOG1* gene diversity. *DOG1* promoter and coding regions were compared. and the coding region was analyzed more in details. On the top of this table, general information about sequence size and number of SNPs are provided. Next, the diversity estimators are given for all *DOG1* fragments. Values of neutrality tests and significance are shown. The use or not of an outgroup (*A. lyrata*) is specified. "++", "+" and "–" in brackets reflects sensitivity of the test to population expansion, respectively: very sensitive, sensitive and not sensitive.

2. Identification and distribution of *DOG1* alleles

DOG1 promoter

Bentsink et al. (2006) suggests that variation of *DOG1* expression might influence primary dormancy degree. The responsible polymophisms in cis-regulatory region have not been clearly identified, nevertheless an insertion found in L*er* promoter might be associated with a decrease of expression level and then might be associated with low dormancy. The analysis of variation among 86 cis regulatory regions revealed that this insertion is not representative of natural variation and is specific to the L*er* genotype. Thus, in the *DOG1* promoter, no obvious sites potentially involved in phenotypic variation were identified.

Identification of *DOG1* alleles based on exon1

To analyze frequency and distribution of *DOG1* alleles, and to perform association genotype-phenotype, it is important to choose the most suitable polymorphisms based on neutrality tests and functional analyses. The two alleles of the *FRIGIDA* gene were identified by this way.

First, the molecular analysis of *DOG1* showed that exon1 shows singular evolutionary dynamics, with an excess of non-synonymous polymorphism that is not reflected in the pattern of interspecific divergence. More specifically, a peak of non synonymous substitution rate was observed at the beginning of exon1 at position 12, 13 and 14 of the amino acid sequence (Appendix 5).

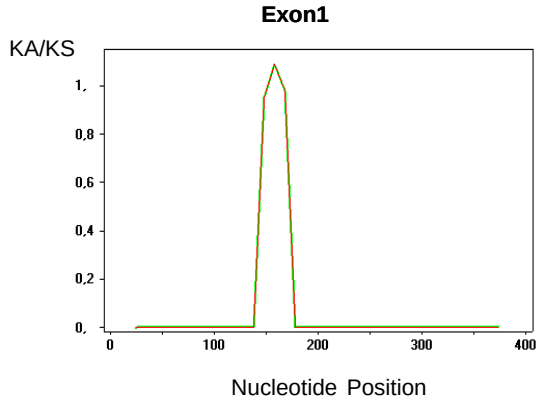


Figure 18. Peak of KA/KS in Exon 1.

Second, the molecular function of *DOG1* has not been discovered yet. Nevertheless, recent work in the department has established that self-binding of DOG1 is essential for function and requires a tyrosine residue at position 15 in exon 1 (Nakabayashi and Soppe, pers. Comm..). Interestingly, this residue is located immediately downstream of the amino-acid residue displaying polymorphism in *A. thaliana* (Appendix 5).

We therefore hypothesize that the three amino-acid of exon1 at position 12, 13 and 14 differentiate functionally different *DOG1* alleles. We sequenced exon 1 in a total 161 *A. thaliana* individuals collected worldwide to perform an analysis of association between the haplotypes at these three amino-acid residues and levels of dormancy.

We observed three alleles in high frequency: the first allele is called “one” (“DS_”, frequency: 0.47), the second is “two” (“DR_”, frequency: 0.21) and the third is “three” (“ECC”, frequency: 0.22). Three rare alleles were found as well: “rare-four” (“ECY”, frequency: 0.03) “rare-five” (“EFS”, frequency: 0.02) and rare-six (“ECS”, frequency: 0.01). The allele of *A. lyrata* sequence and three other sister species is “ESC” without deletion, thus allele “three” could be ancestral allele and “one” and “two” with a deletion are probably derived alleles. In addition, with allele “three” (ECC) *DOG1-DOG1* binding was observed whereas with allele “one” (DS_) *DOG1-DOG1* binding was not observed and for allele two it is not known. (Nakabayashi and Soppe, pers. Comm..)

Distribution of *DOG1* alleles in North/South and *FRI-F/FRI-N* subsamples

Different pattern of co-variation between traits were observed in northern and southern subsamples as well as in function of *FRIGIDA* functionality. However the distribution of the frequency of the three major alleles of *DOG1* gene (one, two and three) was equally distributed in all theses subsamples (Table 6).

***DOG1* alleles names**	***DOG1* alleles sequences**	***DOG1* alleles frequencies**	***DOG1* alleles (without rare alleles) frequencies**
one (derived)	DS_	0.47	0.52
two (derived)	DR_	0.21	0.23
three (ancestral)	ECC	0.22	0.25
rare-four	ECY	0.03	/
rare-five	EFS	0.02	/
rare-six	ECS	0.01	/
missing data	/	0.04	/
number of genotypes		161	145

***DOG1* alleles names**	***DOG1* alleles frequencies - North**	***DOG1* alleles frequencies - South**	***DOG1* alleles frequencies - *FRI-F***	***DOG1* alleles frequencies - *FRI-N***
one (derived)	0.53	0.5	0.48	0.55
two (derived)	0.2	0.27	0.25	0.25
three (ancestral)	0.27	0.23	0.27	0.2
rare-four	/	/	/	/
rare-five	/	/	/	/
rare-six	/	/	/	/
missing data	/	/	/	/
number of genotypes	71	74	67	40

Table 6. *DOG1* alleles frequencies in whole sample and subsamples.

Distribution of *DOG1* alleles in refugia and recolonization area

The allelic distribution of *DOG1* gene was influenced by demographic events. In the Mediterranean region the three alleles were in high frequency, these refugia includes several sub-refugia which is consistent with high diversity. For newly recolonized areas such as USA, Japan and Scandinavia, one allele among the three major alleles was not present or in low frequency consistent with the result of a recent bottleneck. Finally, only one allele of *DOG1* was found in central Asia, whereas everywhere else at least two alleles were present (Figure 19).

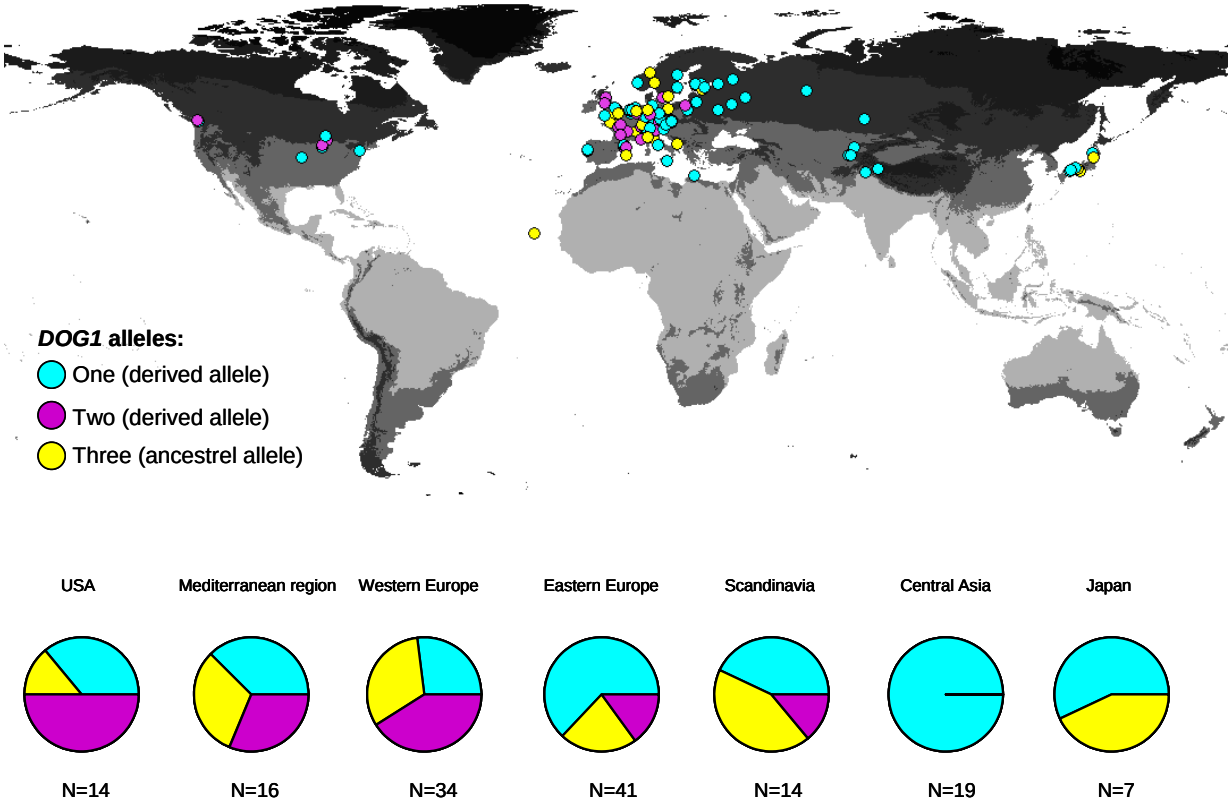


Figure 19. Geographical distribution of *DOG1* alleles. N: number of genotypes

III. Association mapping

1. Unified mixed model

Using the unified mixed model method initially developed by Yu et al. (2006), in collaboration with B. Stich, the three alleles in high frequency of *DOG1* gene (one, two and three) were not associated with Primary Dormancy. Nevertheless, when the rare alleles were included in the analysis the association between *DOG1* and germination percentage after 7 days of dry storage became significant and 7 % of the phenotypic variance was explained by *DOG1* variation. The five genotypes from northern Europe carrying the rare-fourth allele of *DOG1* (ECY) expressed a low dormancy. On contrary, the four genotypes carrying the rare-fith allele (EFS) which come from Spain and Italy were very dormant. It would be necessary to confirm those associations with a larger number of genotypes carrying these alleles. *DOG1* alleles based on the presence or not of a deletion at position 14 of AA sequence of exon1 (Appendix 5) were associated with germination at 4°C as well as with Flowering time with vernalization. As expected *FLC* and *FRIGIDA* genes, known as flowering time QTL were significantly associated with variation of flowering time. In addition, *FLC* was significantly associated with germination after 28 days of dry storage and *FRIGIDA* has associated with germination at 4°C (Table 7).

	Storage 7 days		**Storage 28 days**		**Storage 56 days**		**Primary Dormancy**	
gene (alleles)	p-value	expl-variance	p-value	expl-variance	p-value	expl-variance	p-value	expl-variance
FRIGIDA (FRI-F/FRI-N)	0.33141377	0.11715207	0.63360166	0	0.46145844	0	0.03813487	0
FLC (A/B)	0.00772616	4.16003997	0.00041255	91.4596272	0.00393865	4.85040345	0.09665635	87.4211779
DOG1 (with/without-deletion)	0.03639808	2.53330096	0.01701152	0	0.02108435	3.11261544	0.10246467	0
DOG1 (one/two/three)	0.02229447	3.57511872	0.03279972	0	0.06546932	2.31103386	0.34614669	90.680344
DOG1 (one/two/three/four/five/six)	0.00318577	7.12607095	0.00224142	0	0.01654255	5.0794202	0.02472652	0

	Secondary Dormancy		**Germination at 4°C**		**Flowering Time**	
gene (alleles)	p-value	expl-variance	p-value	expl-variance	p-value	expl-variance
FRIGIDA (FRI-F/FRI-N)	0.27518209	0.32333359	0.00378663	98.325224	5.49E-05	8.91566017
FLC (A/B)	0.54422678	0	0.84172236	98.1415197	0.00790421	4.09243992
DOG1 (with/without-deletion)	0.22994073	0.52232305	0.00579729	98.1749997	8.28E-08	14.5351949
DOG1 (one/two/three)	0.55442123	0	0.5167146	0	2.28E-06	12.6039978
DOG1 (one/two/three/four/five/six)	0.3136556	0.59886282	0.04356375	0	1.87E-06	14.6383198

Table 7. Results of unified mixed model.

2. Genome Wide Association mapping

The results of Genome Wide Association mapping are published in Atwell et al. (2010) and are shown in appendix 6, 7 and 8. In this study, the most significant associations are described for vegetative growth rate, flowering time, seed dormancy and germination at 4°C, with 111 genotypes among the 161 used for the phenotypic analyses (except flowering time).

SPATULA (*SPT*), candidate gene controlling natural variation of growth rate

SPT gene, a bHLH transcription factor, was associated with seedling growth rate, vegetative growth rate during vernalization and after vernalization, nevertheless it was significant only for growth rate during vernalization (Appendix 6 a, b and c). It has been shown that a *spt* loss-of-function mutation increased leaf size and total cell number within a leaf, and that *SPT* overexpression decreased leaf size and total cell number within a leaf. Moreover, genetic analysis suggested that *SPT* acts independently from another set of cell proliferation-dependent organ size regulators (Ichihashi et al**.** 2010). Thus, the detection of *SPT* gene as a vegetative growth rate QTL is consistent.

DOG1 detected as a flowering time QTL

With several independent phenotypic data of flowering time variation with or without vernalization, the dormancy gene *DOG1* was detected as a flowering time QTL. The peak around *DOG1* gene for a flowering time related trait is shown figure 20. It is possibly a false positive due to a simultaneous selection on primary dormancy and flowering time detected at the phenotypic level. As expected, *FLC* was detected as a flowering time QTL and surprisingly its expression was associated with *DOG1*.

.

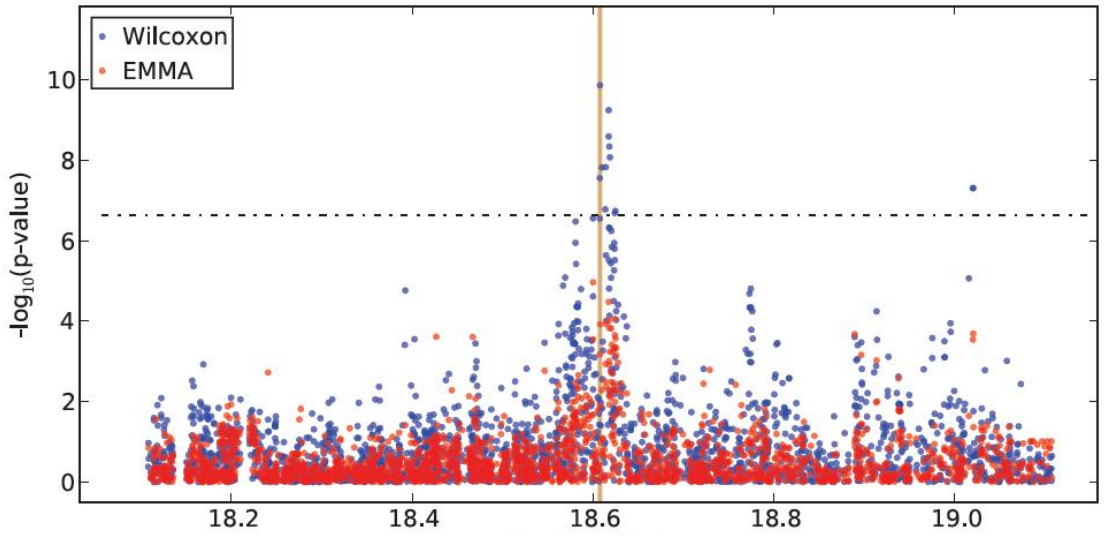


Figure 20. Peak of association around *DOG1* for leaf number at flowering time (10°C, 16 hours daylight ; Atwell et al. in press).

COMATOSE, *NCED9* and *FLC* detected as primary dormancy QTL

COMATOSE and *NCED9* were detected as primary dormancy QTL (Appendix 8 a, b, c and d). These results are consistent. Indeed, a mutant analysis revealed that *COMATOSE* gene regulates germination potential (Russell et al. 2000) moreover, a functional analysis revealed that *NCED9* is required for ABA biosynthesis during seed development and is involved in seed dormancy induction (Lefebvre et al. 2006). In addition the flowering time QTL *FLC* was significantly associated with natural variation of primary dormancy. This result confirms the pleiotropic effects of *FLC* found in Chiang et al. (2009).

Discussion

I. Natural variation of Life history traits

Life-history traits controlling the duration and timing of life cycles tend to co-vary due to both natural selection and genetic or physiological constraints (Mitchell-Olds 1996; Donohue 2002; Donohue et al. 2005b; Donohue et al. 2005a; Rees et al. 2006; Roff and Fairbairn 2007; Montesinos-Navarro et al. 2010). Therefore, life-history traits studied in isolation provide an incomplete view on the relevance of life-cycle variation for adaptation. In this study, we examine genetic variation in several parameters covering the whole life-cycle across the distribution range of *Arabidopsis thaliana*.

1. Relevance of life-history trait variation studied in the greenhouse

Strong genetic variation and high heritability was found for primary dormancy, secondary dormancy, vegetative growth rate and flowering time in laboratory and greenhouse conditions among *Arabidopsis thaliana* genotypes collected in Europe. This dataset is thus ideal to analyze patterns of trait co-variation. Nonetheless, natural environments are more complex and often harsher than in the greenhouse and the traits characterized here may be expressed differently in the field (Metcalf and Mitchell-Olds 2009; Wilczek et al. 2009, Brachi et al. 2010).

Using laboratory and greenhouse conditions has the advantage to allow distinguishing between variation in primary dormancy and secondary dormancy, which might be difficult in the field. Moreover, several elements indicate that the variation revealed here for primary dormancy is helpful to understand germination as it is expressed in the field. Indeed, germination in laboratory conditions such as those used here, were reported to be under the control of the same QTLs as germination in the field (Alonso-Blanco et al. 2003; Huang et al. 2009; Atwell et al. 2010). Finally, levels of primary dormancy were correlated with those measured on seeds matured in the field (in Valencia, Spain) among few genotypes. Secondary dormancy in response to cold exposure was observed in the field as well as in laboratory conditions (Baskin and Baskin

1983; Cadman et al. 2006; Montesinos et al. 2009) however the genetic basis of this trait is not known and its association with population structure did not allow QTL detection via genome wide association mapping (Atwell et al. 2010). Our study is the first to report natural genetic variation for secondary seed dormancy.

Natural genetic variation of a large number of traits related to vegetative development was reported in *Arabidopsis thaliana* both in the greenhouse and in the field (Aarssen and Clauss 1992; Li et al. 1998; Korves et al. 2007; Hopkins et al. 2008). Growth rate, as a trait of major economical importance, has been studied extensively in crops but very few studies have examined genetic basis controlling growth rate in natural populations (McGraw and Garbutt 1990; Perez-Perez et al. 2002; El-Lithy et al. 2004; Meyer et al. 2007). In this study, we confirm the presence of extensive genetic variation controlling the rate of vegetative growth at rosette stage.

Genetic variation of flowering time has been extensively examined in greenhouse conditions and in the field (shindo et al. 2005; Caicedo et al. 2004; Wilczek et al. 2009; Brachi et al. 2010). Nevertheless the genetic basis controlling variation of flowering time differs in both environments but it differs as well in different field environments (Weinig et al. 2002). For both experimental conditions: greenhouse conditions and common garden, plants might grow in conditions very different from those of their original location, there is therefore no optimal setting for such analysis.

2. Strategies display continuous variation

Life-cycle strategies are determined by the suite of traits controlling plant development along the life cycle, i.e. by the vector of each trait value here. Strategies differ from each other essentially by their relative investments into survival or reproductive rate, the two major components of fitness. Our study reveals that they actually display continuous variation. There is no clustering around major strategies such as winter annuals (which favor a long period of vegetative growth to maximize reproductive rate) or rapid cyclers (which maximize the number of generations per year). Some strategies appear to be rare. For example, we do not observe genotypes displaying both strong dormancy and late flowering although this strategy is genetically possible, as it was observed in a local

population in Norway (I. Kronholm, unpublished). Instead, strategies with low dormancy, elevated growth rate and early flowering (i.e. putative rapid cyclers) are well represented (Figure 2).

3. Adaptive evolution of life-cycle traits

We find that overall, only life-cycle traits that describe similar life-history stages (such as germination or flowering) are highly correlated. Instead, germination, growth rate and flowering traits are essentially non-correlated (Table 2). Primary, secondary dormancy, seedling growth and stem leaf number are independent variables that together explain 20% of latitudinal variation (Table 1). Therefore, these traits that jointly define the life-cycle also jointly follow a latitudinal cline. Several traits, studied in isolation and generally with a smaller number of genotypes, have been shown to follow clines (Li et al. 1998; Caicedo et al. 2004; Schmuths et al. 2006). Only co-variation with latitudinal or climatic gradients controlling for co-variation with population structure provides unambiguous indication that the cline was caused by selection and not by population history (Mayr 1956; Endler 1977; Korves et al. 2007; Stillwell et al. 2007). In our sample, population structure was reduced and displayed the East-West gradient previously reported (Sharbel et al. 2000; Beck et al. 2008; Francois et al. 2008; Samis et al. 2008). We observe that at each scale and in each subsample, there are life-cycle traits that co-vary with environmental parameters, independent from population structure. Path analysis showed that the latitudinal gradients contributed more to variation in primary dormancy, seedling growth rate and flowering time than population structure (Figure 2). In addition, the contribution of latitudinal variation to secondary dormancy is of the same magnitude than the contribution of population structure. Thus, these four traits define a latitudinal cline that is largely independent from demographic history. Interestingly, dormancy seems to co-vary with the temperature gradient whereas stem leaf number with precipitation contrasts.

4. Selective pressures change across the species range

Adaptive evolution may also be itself the cause of population structure. Here, we believe that some of the life-history variation that correlates with population structure probably reflects true local adaptation. In addition, the fact that multiple regressions conducted independently on the Northern and Southern subsamples revealed associations with specific yet distinct ecological parameters provides a solid indication that some life-cycle traits undergo different and changing selective pressures across the species range. For example, primary dormancy explains a significant portion of temperature contrasts in the North. In the South, instead, germination at 4°C is associated with precipitation contrasts. This suggests that dormancy may be involved in limiting death caused by temperature variation along the year in the North whereas in the South germination at cold temperatures would rather be involved in avoiding death caused by precipitation variation (e.g. drought).

5. Antagonistic patterns of co-variation with growth rate

Patterns of growth rate variation further show that the life-cycle undergoes distinct evolutionary dynamics across the geographic range. And indeed, variation in growth rate may not follow simple environmental clines. Strategies delaying germination due to dormancy to ensure seedling survival might reduce fertility, indeed rosettes may have less time to reach large sizes at flowering timing (Mitchell-Olds 1996). Within such strategies, natural selection will favor the genotypes displaying enhanced growth rate to optimize plant size at reproduction timing. But enhanced growth rate may also be selected within strategies minimizing the length of the life-cycle so that the number of generations per year is maximized. Growth rate, however, cannot be indefinitely increased and the production of a large plant will always take some time. In addition, fast growth requires intense gas exchanges and may be subjected to trade-offs with water-use efficiency (Roff and Fairbairn 2007; Metcalf and Mitchell-Olds 2009; McKay et al. 2003). Our study reveals antagonistic patterns of growth rate across the species range, that are not visible across the full sample. In the North, growth rate is negatively

correlated with flowering time (Table 2), indicating that genotypes that flower early tend to grow fast. This relationship is not found in the South. It suggests that in the Northern area, increased growth rate has been selected to compensate for increased earliness. Instead, in the Southern area, growth rate evolution appears not to compensate for the correlated effect of early flowering on plant size at reproduction (Mitchell-Olds 1996; Griffith et al. 2004). A possible interpretation to this pattern is that in the North, there is comparatively stronger selection on reproductive rate, whereas in the South, selection on seedling survival may predominate. Some recent QTL studies suggest that correlations between growth rate and flowering time are the result of pleiotropic or closely linked loci (Tisne et al. 2008; Pieper 2009). QTLs promoting negatively correlated changes of growth rate and flowering time seem to be frequent, yet QTL correlating positively the two traits have also been observed, thereby confirming that the evolution of growth rate in response to flowering time changes is possible and may involve pleiotropic loci (B. Pieper and M. Reymond, unpublished).

II. Association / linkage mapping and genetic controlling life history traits variation

Association mapping is thought to be particularly efficient in inbreeding species, since genotyping needs to be done only once and immortal lines can be characterized for any phenotype (Keurentjes et al. 2008, there is also a paper by Nordborg and Weigel). Because several genes studied here have been extensively characterized by linkage mapping, this thesis offers a good opportunity to discuss the advantages and pitfalls of both methods.

1. *DOG1*, a major dormancy QTL not detected by GWA

Two major approaches of genotype-phenotype association with natural population were used. The first method was more precise and flexible concerning the haplotypic organization of *DOG1* which was defined in function of sites potentially under selection and the binding activity of this gene. With different allelelic organization of *DOG1* the

association was performed via unified mixed model method initially developed by Yu et al. (2006), in collaboration with B. Stich, with a relatively small number of SNPs to control for population structure. The second method was Genome Wide Association mapping method recently developed by Magnus Nordborg and his team (Atwell et al. 2010) using a large number of random SNPs covering the whole genome. Surprisingly, neither of the two methods could show a clear or significant association between *DOG1* and primary dormancy. Only rare alleles found in the sample of 161 genotypes seemed to associate with dormancy.

Only alleles with strong effect and in high frequency can be efficiently detected using Genome Wide Association mapping. Moreover, a QTL can be detected with a high resolution only if it was not subjected to a recent selection. Indeed, if the selection is too recent the linkage disequilibrium on both sides of the causal SNP is large due to a too low number of recombination events and in this case the resolution is low. By opposition, if the selection on a QTL is ancient the recombination events will allow the QTL detection with high resolution. Although the analysis of nucleotide diversity in *DOG1* revealed alleles in high frequency, there was no significant association with primary dormancy. It is possible that the allelic organization of *DOG1* based on exon1 was not appropriate to perform association. Indeed, the high diversity in exon 1 close to a binding site could the result of independent directional selection in populations isolated by glacial cycles which can mimic a balancing selection. Moreover, possibly other SNPs of *DOG1* influence the function and interact epistatically with polymorphisms in exon1. Preliminary studies of *DOG1* expression variation have shown that not only coding variation can influences the phenotype (Bentsink et al. (2006). Another aspect which plays an important role to detect efficiently a QTL is the complexity of the epistatic network in which the gene is involved. Indeed, the current GWA method does not include epistatic interactions, thus QTL which interacts with other loci could not be detected efficiently. It has been shown that *DOG1* interacts genetically with *DOG3* another dormancy QTL (Alonso-Blanco et al. 2003), however Bentsink et al. (2010) showed that natural variation for primary dormancy in *Arabidopsis thaliana* is mainly controlled by different additive genetic and molecular independent pathways rather than epistatic interactions. Therefore, it is probably not because of epistatic interactions of *DOG1* with other genes that *DOG1* was

not detected as dormancy QTL but maybe more because of the additive effects of other genes which might play a major role in the control of primary dormancy variation.

2. *DOG1* QTL and linkage mapping

A large number of linkage mapping analyses in laboratory and field conditions (Alonso-Blanco et al. 2003; Clerkx et al., 2004; Bentsink et al. 2010; Huang et al. 2010) detected *DOG1* as a major Dormancy QTL. Here I discuss the limits and advantages of linkage mapping trying to understand how a QTL could be detected via linkage mapping but not via GWA.

A limit of linkage mapping is that the two alleles involved are not necessarily representative of the natural population. If one of the parental lines carries a rare non-functional allele it could lead to an overestimation of the effect of a QTL compared to its real effect in natural population. It could be the case for *DOG1*. The genotype Landsberg *erecta* is the result of a strong human selection and is now adapted to laboratory conditions. L*er* was crossed with six different genotypes to create six RIL populations (Alonso-Blanco et al. 2003; Clerkx et al., 2004; Bentsink et al. 2010) and *DOG1* was detected as dormancy QTL in 5 of theses experimental populations. Interestingly, the genotype L*er* harboured a large insertion in the promoter of *DOG1* , which was not found in natural population. With an other RIL population created with Tac (from Washington) and Cal (from England) *DOG1* was detected as a QTL for field germination (Huang et al. 2009). Based on the sample of 161 genotypes Tac carries one of the allele in high frequency whereas Cal carries one of the rare alleles (ECY). Here again, one of the two alleles of *DOG1* gene is not present in high frequency in natural population. Thus, with all experimental populations mentioned previously, *DOG1* was detected as dormancy QTL with a rare allele carried by one of the parental line. The contribution of *DOG1* to natural variation of primary dormancy was possibly overestimated using linkage mapping method. Nevertheless, *DOG1* was also detected with the RIL population derived from the parental genotypes Bay-0 and Shakhdara, both of which carry a *DOG1* allele present in high frequency in natural population based on exon1. This finding contradicts the

previous conclusion, but a rare mutation could be present in other functional regions of the gene.

The most surprising result of this study is possibly that *DOG1* was not associated with natural variation of primary dormancy but with natural variation of flowering time, although no RIL population ever revealed a QTL for flowering time at this locus. In *Arabidopsis thaliana* outcrosses can occur naturally, but, in this highly selfing species, it is probable that genes involved in epistatic interactions could co-evolve independently through several generations and that recombinations between divergent alleles at these loci could lead to incompatibilities. An example is reported in Alcazar et al. 2009. It is therefore possible that disturbance of epistatic interactions could explain why *DOG1* was not associated with flowering time in RIL populations.

3. GWA and genetic control of co-variation between traits

Linkage disequilibrium (non random association between genes) between non causal markers and causal genes can lead to false positive. The linkage can occur at a large scale including the whole genome due to population structure (geographical isolation) or familial relatedness (from recent co-ancestry). Linkage disequilibrium can be local as well: between two genes physically close, or between two genes physically very distant due to simultaneous selection on the two genes. Such simultaneous selection can occur more easily in highly selfing species because of a lower recombination rate compared to obligate outcrossing species. Therefore, the association between *DOG1* and natural variation of flowering time could be explained by a linkage disequilibrium between *DOG1* alleles and allelic variation at flowering time genes due to selection. Indeed, our study of phenotypic variation has shown that primary dormancy and flowering time are correlated, especially in a genetic background carrying a functional *FRIGIDA* allele. Because flowering time variation is controlled by epistatic interactions, it would be interesting to study association between *DOG1* alleles and allelic variation of all flowering time genes including epistatic interaction between them.

FLC was associated with natural variation of primary dormancy and flowering time. It indicates that this gene might have pleiotropic effects, which confirms the results of Chiang et al. (2009). Thus, *FLC* is probably one of the gene controlling correlation between primary dormancy and flowering time.

Summary

Natural variation of life history strategies is modulated by both natural selection and physiological constraints. In this study, variation and co-variation of life history traits at the phenotypic and genomic level were studied in details in the model plant species *Arabidopsis thaliana* in order to elucidate the genetic and environmental factors controlling life-cycle evolution.

Natural variation for growth rate, flowering time, primary seed dormancy and secondary seed dormancy were measured in a common environment across a set of 161 *A. thaliana* genotypes isolated from various locations throughout the species' range. The results show that natural variation of life-cycle strategies was continuous. An antagonistic pattern of co-variation between flowering time and growth rate was observed. These two traits were negatively or positively correlated respectively depending on the latitude of origin of the genotypes. Functionality of *FRIGIDA*, a gene previously identified as a major contributor to flowering time variation, was associated with distinct evolutionary trajectories characterized by different patterns of co-variation between life history traits. Indeed, two negative correlations were observed specifically among genotypes carrying functional *FRIGIDA* allele: between flowering time and primary dormancy as well as between primary dormancy and secondary dormancy. These results indicate that selection on and trade-offs among traits controlling life-cycle strategies change along the distribution range of *A. thaliana* and are influenced by *FRIGIDA* functionality. Moreover, natural variation of four life history traits, primary dormancy, secondary dormancy, seedling growth rate and stem leaf number at flowering timing, followed a latitudinal gradient indicating that these traits are involved in local adaptation. Climatic parameters related to temperature and precipitation influenced different life history traits across the distribution range of the species or depending on the functionality of *FRIGIDA*.

The genetic basis of life history variation in *Arabidopsis thaliana* was studied at fine and large scales, first focusing specifically on the nucleotidic diversity of a dormancy QTL (*DOG1*), second at the genomic level using the newly developed genome-wide association mapping method (GWA) developed by M. Nordborg and his team (Atwell et al. 2010). For exon1 of *DOG1*, non synonymous substitution rate was higher than synonymous substitution rate which is rare. It could be the result of independent selective events in populations isolated by glacial cycles. Only rare alleles based on non synonymous substitutions in exon1 were associated with primary dormancy, alleles in high frequency were significantly associated with flowering timing. This probable false positive association was first detected using the unified mixed model method initially developed by E.S. Buckler (2006), in collaboration with B. Stich and then confirmed with GWA method using a larger sample. The detection of this false positive revealed linkage disequilibrium between *DOG1* and flowering time genes which is most likely the result of a simultaneous selection on these genes. *FLC* was significantly associated with primary dormancy and flowering time indicating pleiotropic effect. These results provided some light concerning genetic control of co-variation between life history traits. *DOG1* was not significantly associated with primary dormancy probably because expression level of this gene influences as well phenotypic variation. Finally, GWA method allowed detection of new candidate genes controlling life history traits. *SPT* gene which controls final leaf size was found as a vegetative growth rate QTL.

Zusammenfassung

Natürliche Variation in Lebenszyklus-Strategien wird sowohl durch natürliche Selektion als auch durch physiologische Beschränkungen moduliert. In dieser Arbeit wurde eine detaillierte Analyse der Variation und Kovariation von Lebenszyklus-Merkmalen auf phänotypischer und genomischer Ebene in der Modellpflanze *Arabidopsis thaliana* durchgeführt, um die genetischen und umweltbedingten Faktoren zu erhellen, welche die Evolution von Lebenszyklen lenken.

Die natürliche Variation in Wachstumsrate, Blütezeit, primärer Samen-Dormanz und sekundärer Samen-Dormanz wurde in einer gemeinsamen Umgebung für ein Set von 161 *A. thaliana* Genotypen untersucht, die von diversen Standorten aus dem Verbreitungsgebiet der Art stammten. Die Ergebnisse zeigen, dass die natürliche Variation von Lebenszyklus-Strategien kontinuierlich ist. Ein entgegengesetztes Kovariationsmuster wurde zwischen Blütezeit und Wachstumsrate beobachtet. Diese beiden Merkmale waren – abhängig vom Herkunftsbreitengrad der Genotypen – entweder negative oder positiv korreliert. Die Funktionalität des Gens *FRIGIDA*, für das bereits ein bedeutender Beitrag zur Blütezeitvariation beschrieben wurde, zeigt eine Assoziation mit getrennten evolutionären Entwicklungslinien, die durch unterschiedliche Muster der Kovariation zwischen Lebenszyklus-Merkmalen gekennzeichnet sind. So wurden zwei negative Korrelationen spezifisch in Genotypen mit einem funktionellen *FRIGIDA*-Allel gefunden: zwischen Blütezeit und primärer Dormanz, sowie zwischen primärer und sekundärer Dormanz.

Diese Ergebnisse deuten darauf hin, dass sich die Selektion, die auf Eigenschaften wirkt, welche Lebenszyklus-Strategien bestimmen, sowie Abwägungen zwischen diesen Merkmalen, entlang des Verbreitungsgebietes von *Arabidopsis thaliana* ändern und durch die Funktionalität von *FRIGIDA* beeinflusst werden.

Darüberhinaus zeigte die natürliche Variation in vier Lebenszyklus-Eigenschaften – in primärer Dormanz, sekundärer Dormanz, Keimlingswachstumsrate und Anzahl der Stengelblätter zum Zeitpunkt der Blüte – einen Gradienten über die unterschiedlichen geographischen Breitengrade, was darauf hindeutet, dass diese Merkmale in die Anpassung an lokale Gegebenheiten involviert sind. Klimatische Parameter in Zusammenhang mit Temperatur und Niederschlag beeinflussten verschiedene Lebenszyklus-Eigenschaften, variierend über das Verbreitungsgebiet der Art oder in Abhängigkeit von der Funktionalität von *FRIGIDA*.

Die genetische Grundlage der Variation im Lebenszyklus in *Arabidopsis thaliana* wurde sowohl im feinen als auch im großen Maßstab untersucht. Dabei wurde im Rahmen des Ersteren gezielt die Nukleotid-Diversität eines mit Dormanz assoziierten QTLs (*DOG1*) untersucht. Für die großmaßstäbliche Analyse auf genomischer Ebene wurde die neue Methode der Genom-weiten Assoziations-Kartierung (Genome-wide association mapping, GWA) genutzt - entwickelt von M. Nordborg und seinem Team (Atwell et al. 2010). Für das erste Exon von *DOG1* war die Rate nicht-synonymer Substitutionen höher als die Rate synonymer Substitutionen, was selten ist. Es könnte das Ergebnis unabhängiger Selektionsereignisse in durch Eiszeiten isolierten Populationen sein. Lediglich seltene Allele basierend auf nicht-synonymen Substitutionen in Exon I waren mit primärer Dormanz assoziiert. Allele von größerer Häufigkeit waren signifikant mit Blütezeit assoziiert. Diese wahrscheinlich falsch-positive Assoziation wurde zuerst in

Zusammenarbeit mit B. Stich mittels der ursprünglich von E. S. Buckler (2006) entwickelten Unified-Mixed-Model-Methode detektiert und dann mittels der GWA-Methode mit einer größeren Stichprobe bestätigt. Die Detektion dieser falsch-positiven Assoziation offenbarte ein Kopplungsungleichgewicht zwischen *DOG1* und Blütezeit-Genen, was höchstwahrscheinlich das Ergebnis von simultan auf diese Gene wirkender Selektion ist. *FLC* zeigte eine signifikante Assoziation mit primärer Dormanz und Blütezeit, was pleiotrope Effekte andeutet. Diese Ergebnisse werfen mehr Licht auf die genetische Steuerung der Korrelation von Lebenszyklus-Merkmalen. *DOG1* war nicht signifikant mit primärer Dormanz assoziiert. Wahrscheinlich ist dies der Fall, weil auch das Expressionslevel dieses Gens phänotypische Variation beeinflusst.
Schließlich ermöglichte die GWA-Methode die Identifizierung neuer Gen-Kandidaten für die Regulation von Lebenszyklus-Eigenschaften. Das Gen *SPT*, welches die Blattgröße kontrolliert, wurde als ein QTL für die vegetative Wachstumrate gefunden.

Appendix

Appendix 1: stock centers, names and geographical location of the 161 genotypes used in this study.

Stock Number (CS)	Name	Country	Latitude	Longitude
6600	Aa-0	Germany	50.90	9.6
22630	Ag-0	France	45.00	1.5
10224	AK	Japan	39.70	140.1
1656	Alc-0	Spain	40.31	-3.2
22526	Amel-1	Netherlands	53.10	5.8
22626	An-1	Belgium	51.40	4.5
22519	Ang	France	47.50	0.6
22520	Ann-1	France	45.90	6.1
22529	Baa-1	Netherlands	51.30	6.1
22633	Bay-0	Germany	49.00	11.0
6613	Be-0	Germany	50.00	9.0
22579	Bil-7	Sweden	63.20	18.4
10184	Bl-1	Italy	44.00	11.0
10185	Bla-10	Spain	41.50	2.6
6645	Blh-1	Czech Republic	48.60	16.5
22551	Boot	United Kingdom	54.40	-3.3
22590	Bor-1	Czech Republic	49.20	16.5
22591	Bor-4	Czech Republic	49.20	16.5
22628	Br-0	Czech Republic	49.00	16.5
6627	Bs-1	Switzerland	47.40	7.4
6094	Bso-1a	France	52.30	4.6
22656	Bur-0	Ireland	53.50	-8.0
10257	Byn	Norway	63.30	10.1
22620	C24	Portugal	40.20	-8.3
6659	Cal-0	United Kingdom	53.30	-1.6
10271	Calamin	Belgium	50.00	
1065	Can-0	Spain	28.00	-15.5
6042	Car-1	Italy	41.90	12.5
22523	Cerv-1	Italy	41.90	12.5
22521	Chat-1	France	48.10	1.3
1072	Chi-0	Russia	54.00	34.0
22603	CIBC-17	United Kingdom	51.40	0.6
22602	CIBC-5	United Kingdom	51.40	0.6
22625	Col-0	United States	38.30	-92.3
22621	CS 22491	Russia	61.50	34.2
22639	Ct-1	Italy	37.50	15.0
22614	Cvi-0	Cape Verde	16.00	-24.0
10296	Daejon	Korea	36.00	
1116	Dra-0	Czech Republic	49.40	16.3
10038	Driel	Netherlands	52.00	5.8
22572	Eden-1	Sweden	62.60	18.1
22573	Eden-2	Sweden	62.60	18.1
22657	Edi-0	United Kingdom	56.00	-3.0

6693	Eil-0	Germany	51.50	12.6
6088	Ely-1a	United Kingdom	52.40	0.3
1637	Ema-1	United Kingdom	51.30	-0.2
1138	En-2	Germany	50.20	8.8
22548	Eri	Sweden	56.50	15.4
6699	Es-0	Finland	60.00	25.0
6700	Est-0	Estonia	59.00	26.0
22629	Est-1	Estonia	58.50	25.5
22645	Fei-0	Portugal	40.55	-8.3
6705	Fi-1	Germany	50.30	8.0
10158	Fuk	Japan	34.50	133.4
1184	Gd-1	Germany	53.60	10.8
22608	Got-7	Germany	51.40	9.8
1211	Gre-0	United States	43.20	-85.2
22534	Hey	Netherlands	51.30	5.9
6179	Hog	Tajikistan	39.00	68.3
22597	HR-10	United Kingdom	51.40	0.6
22596	HR-5	United Kingdom	51.40	0.6
10223	IK	Japan	35.30	135.1
10043	Jea	France	43.60	7.3
6752	Ka-0	Austria	46.70	13.9
22638	Kas-1	India	35.00	77.0
1264	Kas-2	India	35.00	77.0
6754	Kil-0	United Kingdom	55.40	-5.5
22654	Kin-0	United States	44.60	-85.5
6045	KL-PW-1	Germany	50.90	7.0
1287	Kn-0	Lithuania	54.70	23.7
22566	Knox-10	United States	41.20	-86.5
22567	Knox-18	United States	41.20	-86.5
22491	Konchezero	Russia	61.50	34.2
22651	Kondara	Tajikistan	38.90	69.0
10372	Kyo-1	Japan	35.00	135.8
22606	Kz-1	Kazakhstan	49.80	73.2
22607	Kz-9	Kazakhstan	49.80	73.2
22618	Ler-1	Poland	52.70	15.2
8070	Lim	United States	40.50	-75.4
22650	LL-0	Spain	41.70	2.6
6784	Lm-2	France	48.00	0.2
22594	Lp2-2	Czech Republic	49.30	16.5
22595	Lp2-6	Czech Republic	49.30	16.5
22615	Lz-0	France	46.00	3.5
6793	Mh-1	Poland	53.60	20.3
22640	Mr-0	Italy	44.50	9.5
22635	Mrk-0	Germany	49.00	9.5
22655	Ms-0	Russia	56.00	38.0
22642	Mt-0	Libya	33.00	23.0
22636	MZ-0	Germany	50.50	8.5
22619	Nd-1	Germany	51.00	10.0
22599	NFA-10	United Kingdom	51.40	0.6
22598	NFA-8	United Kingdom	51.40	0.6

1394	No-0	Germany	51.00	13.0
22643	Nok-3	Netherlands	52.50	4.0
22584	Omo2-1	Sweden	56.20	15.3
22585	Omo2-3	Sweden	56.20	15.3
22658	Oy-0	Norway	60.30	6.2
10212	Pak-1	Pakistan	33.90	73.4
10214	Pak-3	Pakistan	33.90	73.4
1445	Per-1	Russia	58.00	56.2
926	Pet-0	Russia	59.90	29.9
22571	Pna-10	United States	42.10	-86.4
22570	Pna-17	United States	42.10	-86.4
22649	Pro-0	Spain	43.20	-6.0
22593	Pu2-23	Croatia	42.50	18.1
22592	Pu2-7	Croatia	42.50	18.1
22632	Ra-0	France	46.00	3.4
22610	Ren-1	France	48.70	-1.6
22611	Ren-11	France	48.70	-1.6
6849	Ri-0	Canada	49.10	-122.7
10232	RIB-1	Japan	34.60	133.8
913	RLD-1	Russia	56.60	35.9
22568	Rmx-A02	United States	42.10	-86.4
22569	Rmx-A180	United States	42.10	-86.4
22524	Rome-1	Italy	41.90	12.5
22565	RRS-10	United States	41.50	-86.3
22564	RRS-7	United States	41.50	-86.3
6848	Rsh-0	Russia	56.60	35.9
6917	Sah-0	Spain	39.00	3.1
22652	Sakhdara	Tajikistan	38.90	69.0
1514	Sav-o	Czech Republic	49.80	13.0
22646	Se-0	Spain	41.50	2.5
6077	Sed-1 (Sid-1)	United Kingdom	50.70	-3.2
1504	Sei-0	Italy	46.50	11.6
1516	Sf-2	Spain	42.00	2.2
10281	Sij-1	Uzbekistan	41.44	70.1
22653	Sorbo	Tajikistan	38.90	69.0
22582	Spr1-2	Sweden	56.40	14.4
22600	Sq-1	United Kingdom	51.40	0.6
22601	Sq-8	United Kingdom	51.40	0.6
1535	St-0	Sweden	59.00	18.0
10256	Strand	Norway	60.30	11.2
6865	Stw-0	Russia	52.70	36.1
10297	Suwon	Korea	34.00	
22604	Tamm-2	Finland	59.70	23.3
22605	Tamm-27	Finland	59.70	23.3
6918	Te-0	Finland	60.10	23.2
10039	Terlet	Netherlands	52.10	6.0
22537	Tha-1	Netherlands	52.10	4.3
22647	Ts-1	Spain	41.50	3.0
22648	Ts-5	Spain	41.50	3.0
22518	Tschag	Austria	47.10	9.9

1565	Tsu-0	Japan	34.40	136.3
10191	Uk-2	Germany	48.00	7.6
22587	Ull2-3	Sweden	56.10	13.6
22612	Uod-1	Austria	48.10	14.7
22613	Uod-7	Austria	48.10	14.7
22627	Van-0	Canada	49.50	-123.0
10172	Vil-0	Spain	42.80	3.8
22644	Wa-1	Poland	52.50	21.0
22540	Wag-1	Netherlands	52.00	5.7
22622	Wei-0	Switzerland	47.40	8.4
10371	Wha-2	United Kingdom	53.00	-0.4
22623	Ws-0	Ukraine	52.30	30.0
22659	Ws-2	Ukraine	52.50	30.0
22637	Wt-5	Germany	52.50	9.5
10215	Yam	Japan	38.30	140.3
22588	Zdr-1	Czech Republic	49.20	16.5
22589	Zdr-6	Czech Republic	49.20	16.5
1629	Zu-1	Switzerland	47.30	8.5

Appendix 2: Names and sequences of Primers used for amplification and sequencing of *DOG1* gene.

Amplification		
PCR name	**Primers name**	**5´-Sequences-3´**
PCR1	prom1F	ACC ATG AAC AAG AAC GAT TC
	9164F	TGA AAG TTG GAA GAT TAG TAC GTG C
PCR2	prom5F	GGA ACA ACA ACT CGC ACT CGC ACT CTC
	end-exon2R	TTC GGT GTC AAC CAT ACA GTC
PCR3	end-exon1F	CGA CGG CTA CGA ATC TTC AG
	5799F	GAA TGT GTT TCC CAT GGC TTA G
PCR4	D1E1	AAA CAC AAA CAC GCA AAC CA
	intron1-rev	GCC GCA CCG TAC TGA CTA CC

Sequencing		
PCR name	**Primers name**	**5´-Sequences-3´**
PCR1	prom1F	ACC ATG AAC AAG AAC GAT TC
	10737R	ACC GGA TAG GTG GCC AAA G
	prom2FR	CAT TAA CGG GTA ATT TGC
	prom3F	GTG TCG AAC TAT CCT CAT AC
	prom3FR	CCT CAA AAT CAC GAG GTC G
	prom4F	GTA CAA TCC GCT GTC TCA GGA CAT C
	9789F	CCT GAG ACA GCG GAT TGT ACA G
	9164F	TGA AAG TTG GAA GAT TAG TAC GTG C
PCR2	prom5F	GGA ACA ACA ACT CGC ACT CGC ACT CTC
	8501short	CAT CGG TGA GCA AGA TCA GC
	intron1-rev	GCC GCA CCG TAC TGA CTA CC
	end-exon2R	TTC GGT GTC AAC CAT ACA GTC
PCR3	end-exon1F	CGA CGG CTA CGA ATC TTC AG
	end-exon2F	ACT GTA TGG TTG ACA CCG
	6957F-R	GTG GCA GCA ACA TGA TCT CGT C
	6957F	GAC GAG ATC ATG TTG CTG CCA C
	J6260F	CCT AAG TAA CCA CTT CAT GGT GG
	6314F	GGG GCA TCT AGA ATT GTG TCA TCT
	5799F	GAA TGT GTT TCC CAT GGC TTA G
PCR4	intron1-rev	GCC GCA CCG TAC TGA CTA CC

Appendix 3: Population structure and detection of the true number of clusters

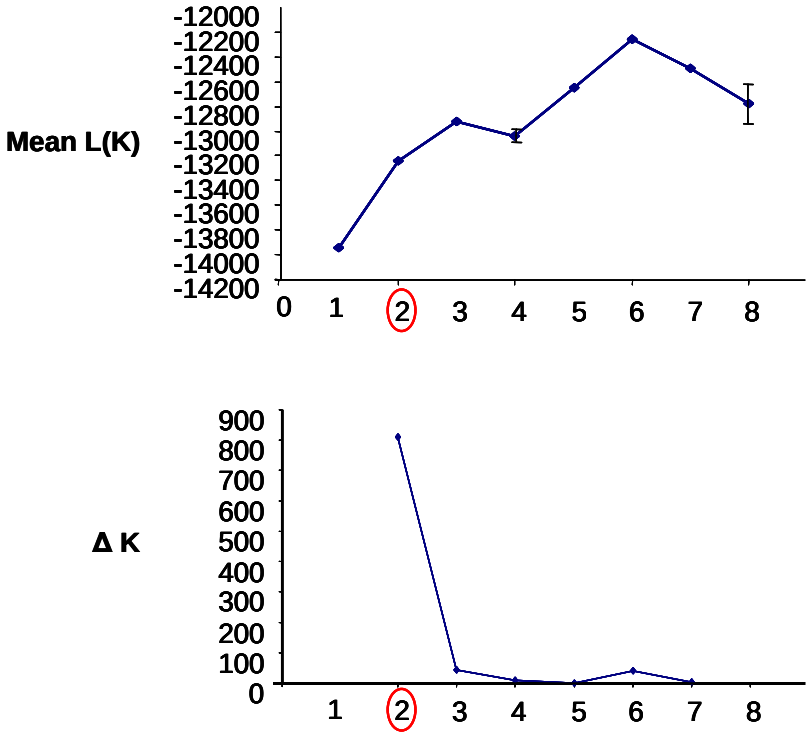
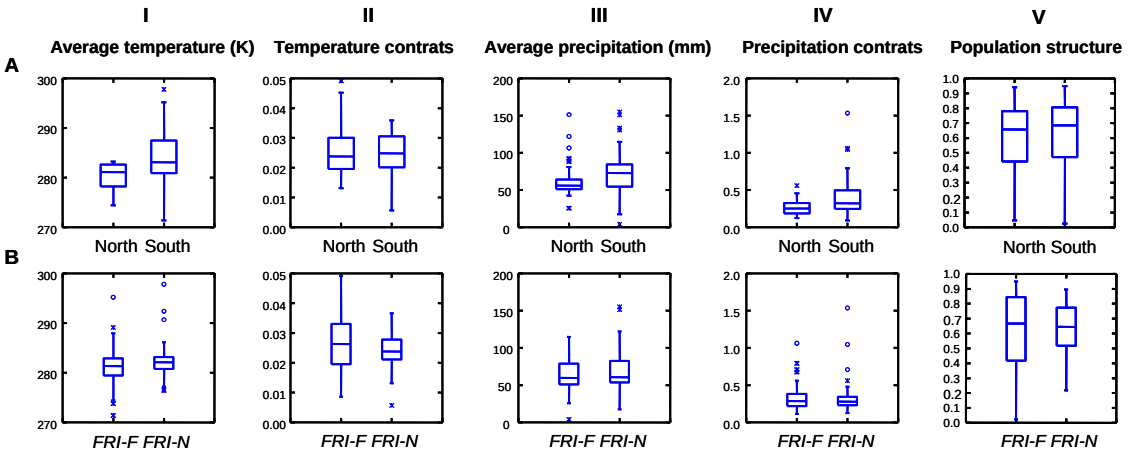


Fig. Graphical method allowing detection of the true number of groups K (Evanno et al. 2005). (**A**) Mean L(K) over 5 runs for each K with 160 genotypes using 139 SNPs as information Basis. (**B**) Δ K calculated as
Δ K = | L(K+1) – 2L(K) + L(K-1)| / sdL(K)
The modal value of this distribution is the true K (red circle) or the uppermost level of structure, here 2 clusters.

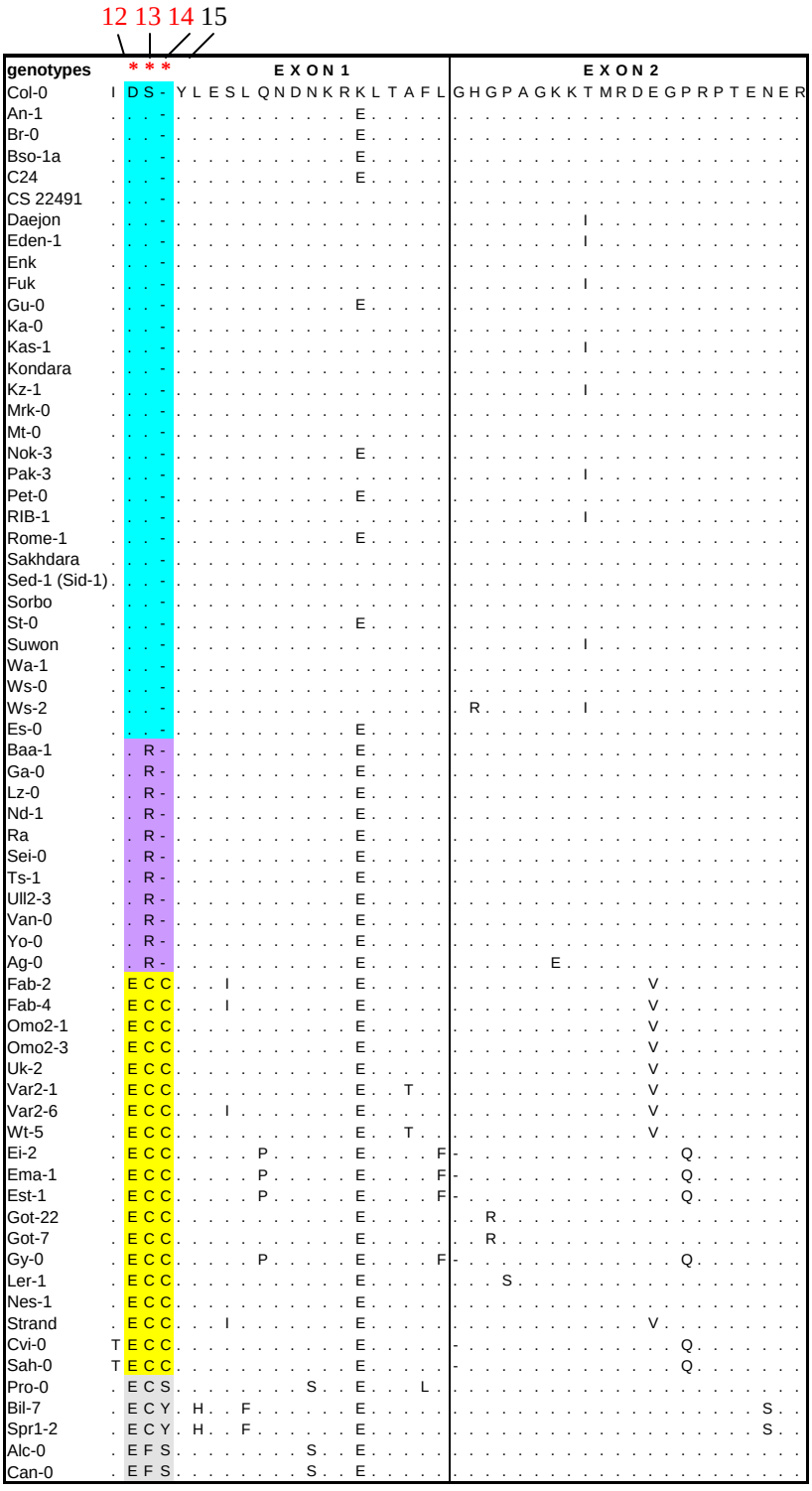
Appendix 4: Distribution of climatic variables and population structure in geographical and *FRIGIDA* groups.



		I	II	III	IV	V
		Average temperature	Temperature contrasts	Average precipitation	Precipitation contrasts	Population structure
A	North / South	*** (79/78)	ns (79/78)	*** (79/78)	*** (79/78)	ns (80/80)
B	*FRI-F / FRI-N*	ns (77/42)	ns (77/42)	ns (77/42)	ns (77/42)	ns (77/41)

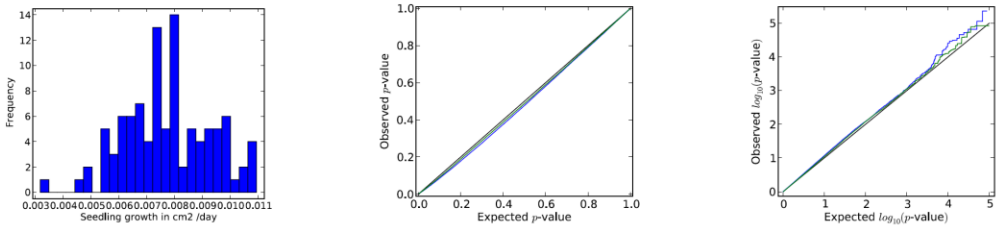
These boxplots show the distribution of climatic variables and genetic structure in geographical (**A**) and *FRIGIDA* (**B**) groups. **I.** Annual average of temperature in kelvin (K). **II.** Annual average of coefficient of variaiton of temperature. **III.** Annual average of precipitation in millimeters (mm). **IV.** Annual average of coefficient of variaiton of precipitation. **V.** Genetic structure based on 139 SNPs estimated by the proportional assignment to the western and eastern clusters between 0 and 1. **A**. Genotypes in North and South were collected respectively at a latitude above 49.2 and below 49.2. **B.** Genotypes with functionnal *FRIGIDA* allele (*FRI-F*), genotypes with non functional *FRIGIDA* allele (*FRI-N*). The climatic and genetic (population structure) differences between groups were tested using Mann-Whitney test. The results (p-values) and the number of genotypes are shown in brackets in the table.

Appendix 5. Polymorphisms of Amino Acid (AA) for exon1 and exon2 of *DOG1*gene among 73 genotypes. *DOG1* alleles were defined based on polymorphisms at position 12, 13 and 14, which are associated with a peak of non synonymous substitution rate. Tyrosine residue at position 15 in exon 1 is required for self-binding of DOG1. (Nakabayashi and Soppe, pers. Comm..)

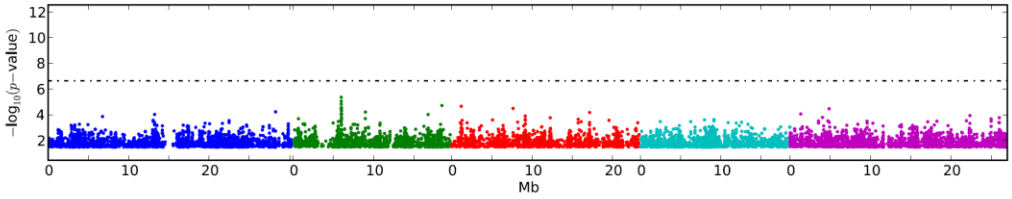


Appendix 6a. Results of Genome Wide Association mapping for Seedling growth rate (Atwell et al. 2010).

Phenotype histogram and quantile-quantile plots of p-values

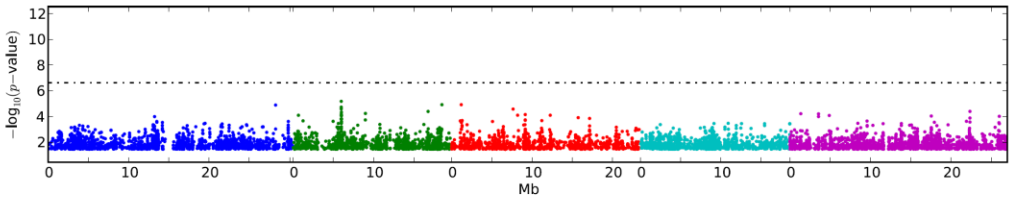


Wilcoxon results



Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
7	4.49477	TWD1	AT3G21640	3	7634734	-13250
56	3.47779	ABH1	AT2G13540	2	5625884	-18082
61	3.4256	IAA7	AT3G23050	3	8198919	1759
62	3.41825	SHD	AT4G24190	4	12549174	2553
62	3.41825	SLY1	AT4G24210	4	12549174	-14389
79	3.3408	ABA2	AT1G52340	1	19498608	-3291
95	3.25878	BIN4	AT5G24630	5	8419676	12850
146	3.0556	SPT	AT4G36930	4	17413399	-725
173	3.00128	CIP1	AT5G41790	5	16738374	-6384
194	2.93296	DAG1	AT3G61850	3	22904359	-1884

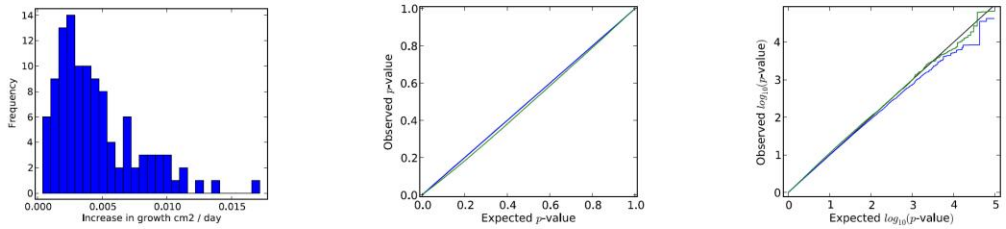
EMMA results



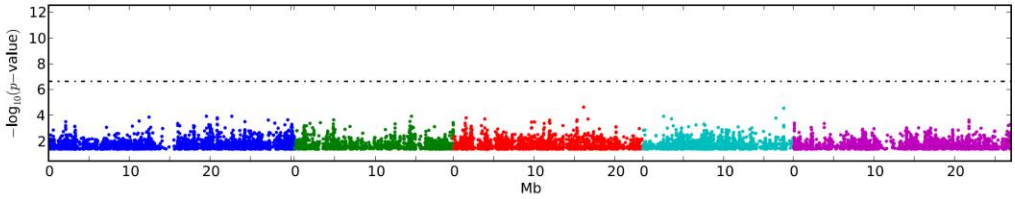
Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
17	4.10221	IAA7	AT3G23050	3	8198919	1759
19	4.0981	CAND1*	AT2G02560	2	669884	-19903
25	4.01706	AGL31*	AT5G65050	5	25992654	-6826
46	3.61854	NUA	AT1G79280	1	29839534	-1663
56	3.54362	similar to LSH1	AT3G04510	3	1212312	3330
58	3.52199	LIP1	AT5G64813	5	25949715	19593
62	3.4735	BIN4	AT5G24630	5	8419676	12850
64	3.47115	ABA2	AT1G52340	1	19498608	-3291
66	3.46134	ABH1	AT2G13540	2	5625884	-18082
69	3.45162	SLY1	AT4G24210	4	12549174	-14389

Appendix 6b. Results of Genome Wide Association mapping for growth rate during vernalization (Atwell et al. 2010).

Phenotype histogram and quantile-quantile plots of p-values

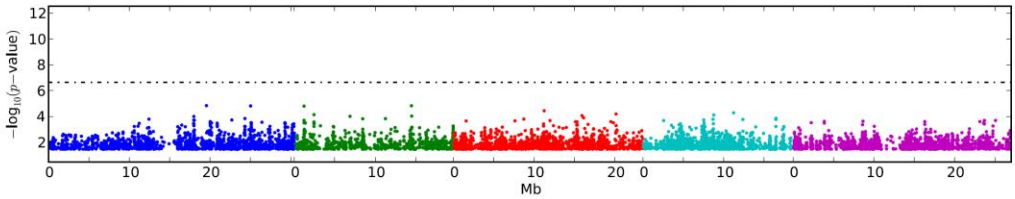


Wilcoxon results



Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
2	4.55017	SPT	AT4G36930	4	17416904	613
23	3.51761	similar to DRM1	AT2G33830	2	14311215	5345
96	3.0418	IMB1	AT2G34900	2	14736609	-3880
132	2.93044	AT1G54070	AT1G54070	1	20183306	-3001
140	2.89996	ABI1	AT4G26080	4	13215741	4229
153	2.8754	EMS1	AT5G07280	5	2277218	-7823

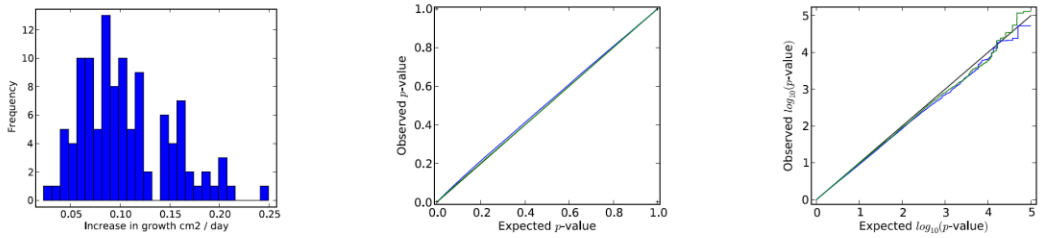
EMMA results



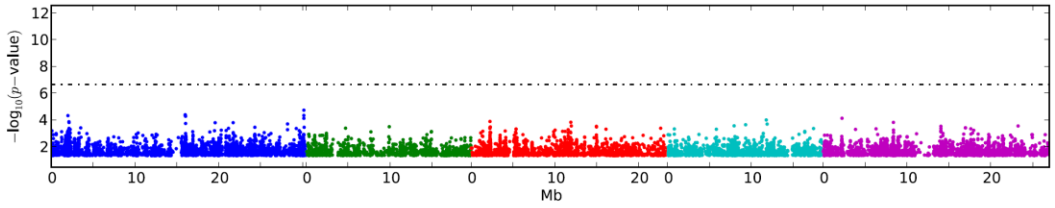
Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
1	6.87733	SPT	AT4G36930	4	17416904	613
9	4.12721	ELF8	AT2G06210	2	2426982	1918
19	3.80692	EER5	AT2G19560	2	8460380	12470
37	3.64544	TWD1	AT3G21640	3	7606169	12768
42	3.59043	EIR1	AT5G57090	5	23108371	-9548
50	3.50462	YAP169*	AT5G07200	5	2235834	7720
50	3.50462	ATS3*	AT5G07190	5	2235834	1674
60	3.4876	AT1G54070	AT1G54070	1	20183306	-3001
64	3.47149	MSI1	AT5G58230	5	23572897	-341
80	3.4151	ZFP1	AT1G80730	1	30355812	-10673

Appendix 6c. Results of Genome Wide Association mapping for growth rate after vernalization (Atwell et al. 2010).

Phenotype histogram and quantile-quantile plots of p-values

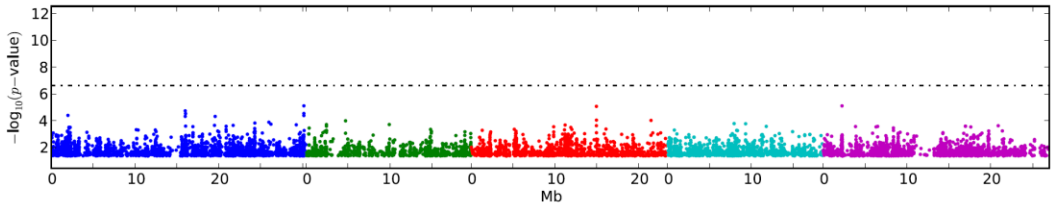


Wilcoxon results



Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
7	4.11622	ATS3*	AT5G07190	5	2235834	-1674
23	3.53972	VIN3*	AT5G57380	5	23263587	34
29	3.44251	CLV2	AT1G65380	1	24306481	13569
30	3.44061	GA4H	AT1G80340	1	30195155	10430
36	3.36984	FUS6	AT3G61140	3	22631510	-5727
42	3.34915	SPT	AT4G36930	4	17416904	613
60	3.19901	AHG2	AT1G55870	1	20879615	-19636
64	3.18473	ABA2	AT1G52340	1	19491510	1953
74	3.11408	FUS3	AT3G26790	3	9856956	0
114	2.93266	CIP1	AT5G41790	5	16769529	19454

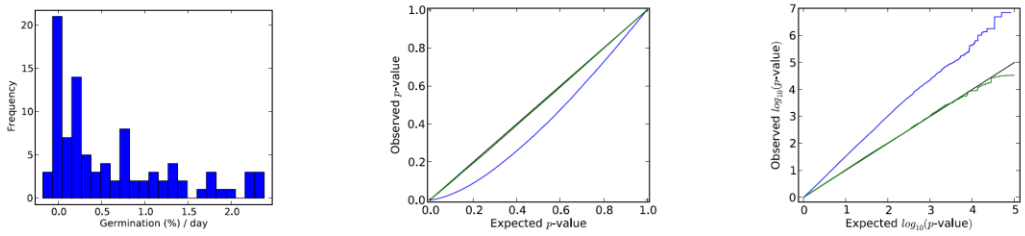
EMMA results



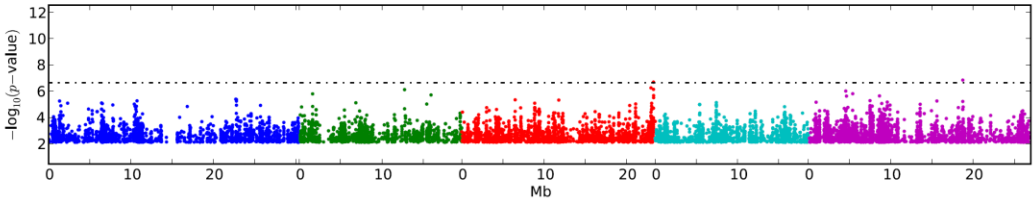
Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
2	5.10822	ATS3*	AT5G07190	5	2235834	-1674
4	4.77153	SPT	AT4G36930	4	17416904	613
15	3.88694	AP1	AT1G69120	1	25983643	2350
21	3.7115	ELF8	AT2G06210	2	2426982	1918
26	3.65243	ABA2	AT1G52340	1	19491510	1953
35	3.58327	AHG2	AT1G55870	1	20879615	-19636
40	3.54669	FUS3	AT3G26790	3	9856956	0
48	3.45903	APRR5	AT5G24470	5	8377696	-18820
53	3.35717	SPA3*	AT3G15354	3	5156199	12903
53	3.35717	SPL5*	AT3G15270	3	5156199	-14844

Appendix 7a. Results of Genome Wide Association mapping for secondary dormancy (Atwell et al. 2010).

Phenotype histogram and quantile-quantile plots of p-values

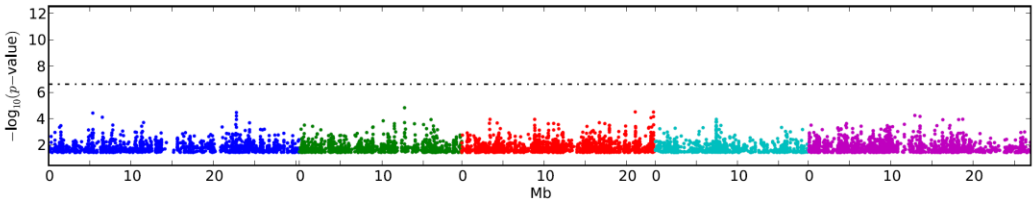


Wilcoxon results



Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
1	6.84223	CUL4	AT5G46210	5	18737357	11288
7	5.80704	PIN5	AT5G16530	5	5381631	-19107
17	5.32313	AT3G29970	AT3G29970	3	11764053	15004
50	4.91147	SMZ	AT3G54990	3	20368973	15722
68	4.79768	RGA1	AT2G01570	2	260584	-3035
74	4.73996	VRN5*	AT3G24440	3	8862085	13949
113	4.5054	MAF5*	AT5G65080	5	26015363	0
163	4.33035	EMS1	AT5G07280	5	2298420	9733
195	4.24459	GA1	AT4G02780	4	1225277	12490
210	4.21064	CIPK3	AT2G26980	2	11526552	-969

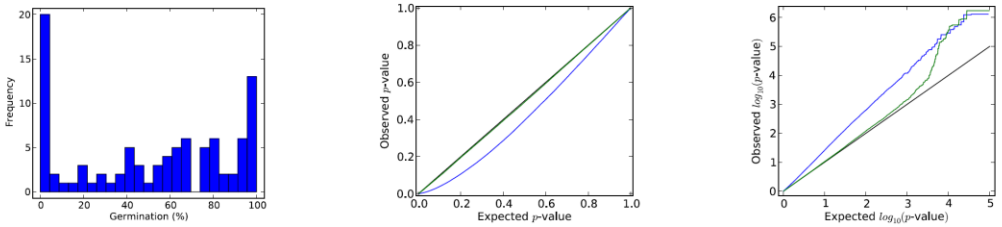
EMMA results



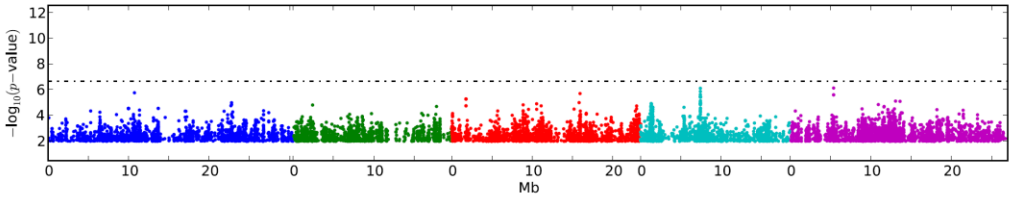
Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
6	4.42936	GA4	AT1G15550	1	5361649	-15488
15	3.95779	CUL4*	AT5G46210	5	18737357	11288
40	3.62182	CIPK3	AT2G26980	2	11526552	-969
58	3.47661	similar to FRL2*	AT5G27230	5	9582938	-1157
78	3.37052	AT1G56220	AT1G56220	1	21055754	7351
89	3.32306	SEPALLATA3	AT1G24260	1	8579221	14410
93	3.30501	HY5	AT5G11260	5	3588592	4789
96	3.29316	VRN5*	AT3G24440	3	8859072	16962
101	3.26845	SPA1*	AT2G46340	2	19039722	-5236
110	3.2306	AT3G29970	AT3G29970	3	11764053	15004

Appendix 7b. Results of Genome Wide Association mapping for germination at 4 °C (Atwell et al. 2010).

Phenotype histogram and quantile-quantile plots of p-values

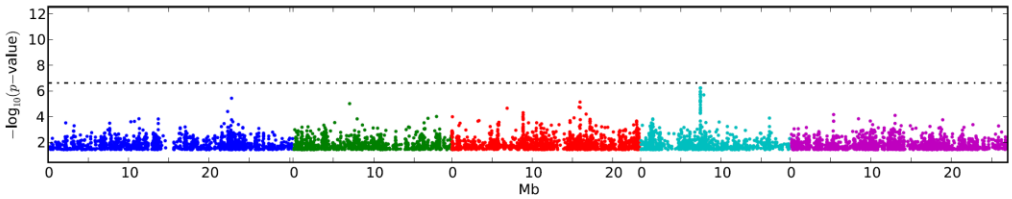


Wilcoxon results



Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
18	5.25646	RCI2B	AT3G05890	3	1764835	-6337
42	4.80407	VRN5*	AT3G24440	3	8862085	13949
43	4.79027	ELF8	AT2G06210	2	2416665	12235
49	4.70829	GA1	AT4G02780	4	1260361	-15548
52	4.67725	SHP2*	AT2G42830	2	17808221	-19111
60	4.62212	DFL2	AT4G03400	4	1495176	2359
81	4.45089	TEL1	AT3G26120	3	9527950	-19685
95	4.34536	PIN3	AT1G70940	1	26758582	8525
121	4.19497	HAB1	AT1G72770	1	27398919	1976
121	4.19497	HAP2C	AT1G72830	1	27398919	10199

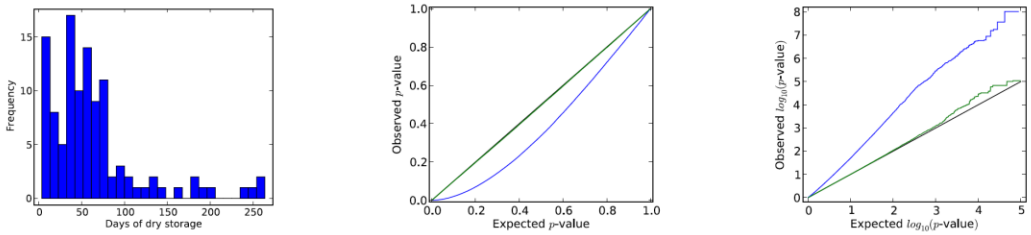
EMMA results



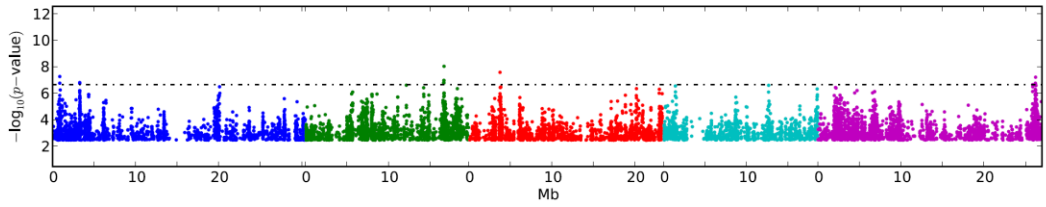
Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
30	4.65885	DWF1*	AT3G19820	3	6874323	5300
42	4.01552	SHP2*	AT2G42830	2	17808221	-19111
47	3.89569	VRN5*	AT3G24440	3	8862517	13517
49	3.87479	ELF4	AT2G40080	2	16747199	-5209
55	3.82547	DFL2	AT4G03400	4	1494311	3224
57	3.80704	VIN3*	AT5G57380	5	23244201	19420
93	3.52172	TEL1	AT3G26120	3	9527950	-19685
159	3.20315	PIF3	AT1G09530	1	3091677	12136
162	3.1966	AT2G33830	AT2G33830	2	14326088	-8661
172	3.1607	DOG1	AT5G45830	5	18625091	-16464

Appendix 8a. Results of Genome Wide Association mapping for primary dormancy (DSDS50) (Atwell et al. 2010).

Phenotype histogram and quantile-quantile plots of p-values

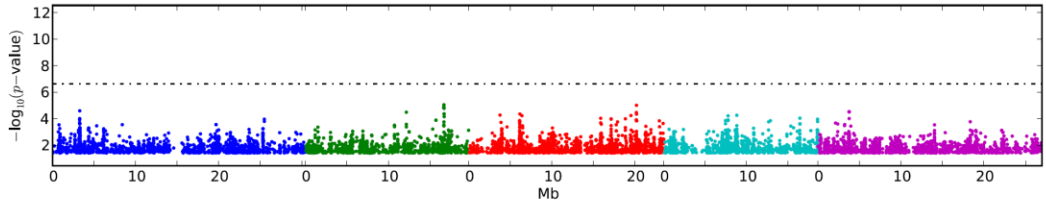


Wilcoxon results



Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
32	6.31068	COMATOSE	AT4G39850	4	18491140	0
53	6.00096	ELF4	AT2G40080	2	16729405	11967
62	5.92629	HBT*	AT2G20000	2	8642534	0
71	5.88552	CYCD4;1	AT5G65420	5	26167140	-5921
107	5.65344	DFL2	AT4G03400	4	1517051	-17187
111	5.62377	ATS3*	AT5G07190	5	2241523	2851
151	5.46186	EMS1	AT5G07280	5	2297704	9017
156	5.46076	ATMYB65	AT3G11440	3	3621028	15918
170	5.37576	ATRAB28	AT1G03120	1	770282	17142
190	5.2813	FLC*	AT5G10140	5	3188328	-8879

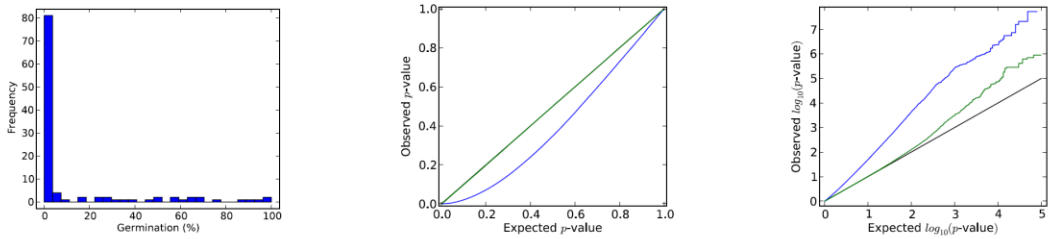
EMMA results



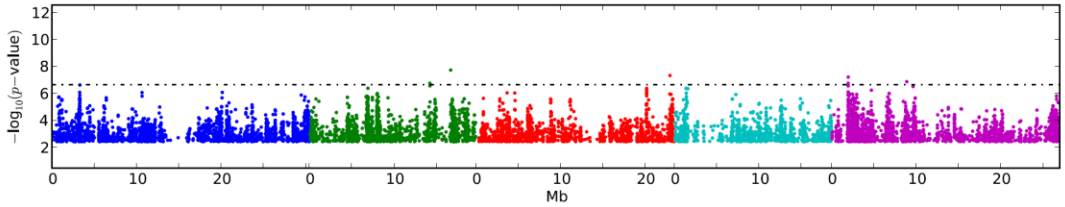
Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
18	4.35971	ELF4	AT2G40080	2	16729405	11967
91	3.54974	IAA6	AT1G52830	1	19690968	-13656
98	3.50137	PKL	AT2G25170	2	10749896	18831
111	3.39456	HY2	AT3G09150	3	2797374	-6159
129	3.27598	FRS1*	AT4G19990	4	10830760	-2098
158	3.15411	COMATOSE	AT4G39850	4	18491434	0
173	3.12085	DFL1	AT5G54510	5	22135025	13294
188	3.07406	UVI4	AT2G42260	2	17626506	-15795
197	3.04985	DOG1	AT5G45830	5	18599415	7294

Appendix 8b. Results of Genome Wide Association mapping for germination percentage after 7 days of dry storage (Atwell et al. 2010).

Phenotype histogram and quantile-quantile plots of p-values

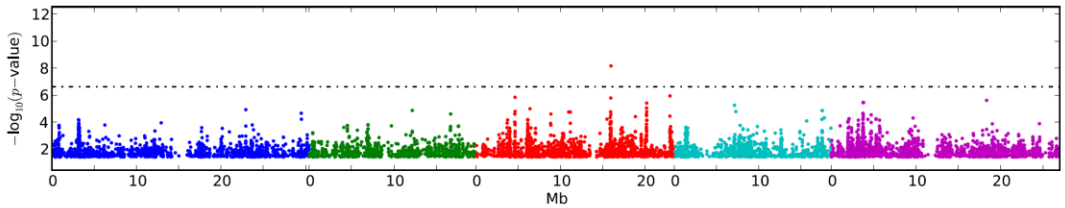


Wilcoxon results



Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
41	5.87448	NCED9	AT1G78390	1	29479269	16519
54	5.73754	PHYB	AT2G18790	2	8165576	14064
132	5.47282	SHY2	AT1G04240	1	1145606	-16056
137	5.46925	EMS1	AT5G07280	5	2297704	9017
142	5.4688	COMATOSE	AT4G39850	4	18491140	0
154	5.40765	LKP2	AT2G18915	2	8185020	16634
155	5.40389	ZFP1	AT1G80730	1	30338768	5447
157	5.39549	DFL2	AT4G03400	4	1517051	-17187
164	5.35569	ATS3*	AT5G07190	5	2222946	-14562
173	5.31955	ABA1	AT5G67030	5	26793918	-19456

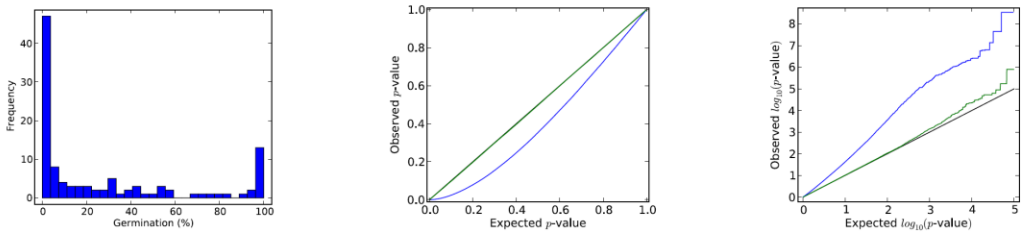
EMMA results



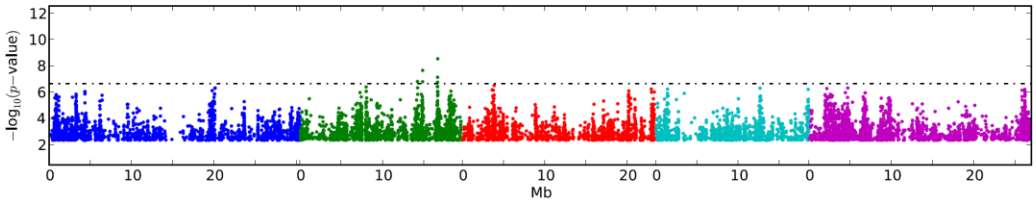
Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
32	4.66173	NCED9	AT1G78390	1	29479269	16519
33	4.65738	EMF1	AT5G11530	5	3719911	18362
34	4.60894	VRN5*	AT3G24440	3	8866118	9916
61	4.21009	FY	AT5G13480	5	4308421	18110
63	4.18653	HYL1	AT1G09700	1	3138580	0
66	4.15523	FLC*	AT5G10140	5	3178232	0
134	3.73271	TFL1	AT5G03840	5	1027938	-2127
177	3.55875	COMATOSE	AT4G39850	4	18516649	19519
232	3.39237	AT5G44300	AT5G44300	5	17862574	850
367	3.07196	DOG1	AT5G45830	5	18599101	7608

Appendix 8c. Results of Genome Wide Association mapping for germination percentage after 28 days of dry storage (Atwell et al. 2010).

Phenotype histogram and quantile-quantile plots of p-values

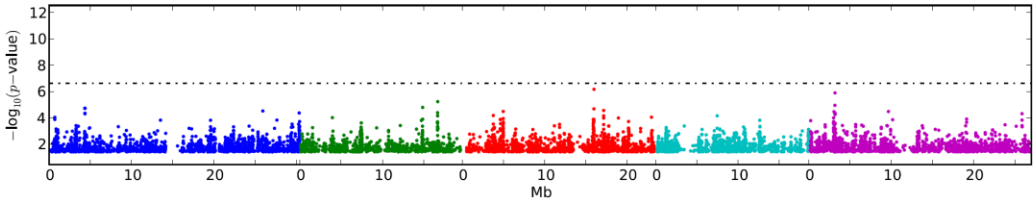


Wilcoxon results



Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
21	6.20501	COMATOSE	AT4G39850	4	18491140	0
23	6.17584	ATMYB65	AT3G11440	3	3621028	15918
26	6.10511	SPA4*	AT1G53090	1	19780965	-6055
37	6.01469	ELF4	AT2G40080	2	16729405	11967
38	6.00612	ATGID1B	AT3G63010	3	23290445	-9955
41	5.9587	AGL15	AT5G13790	5	4429786	19231
56	5.82535	IMB1	AT2G34900	2	14735537	-2808
70	5.7262	DFL2	AT4G03400	4	1517051	-17187
75	5.69726	PGI1	AT4G24620	4	12723011	-10176
100	5.62778	EMS1	AT5G07280	5	2297704	9017

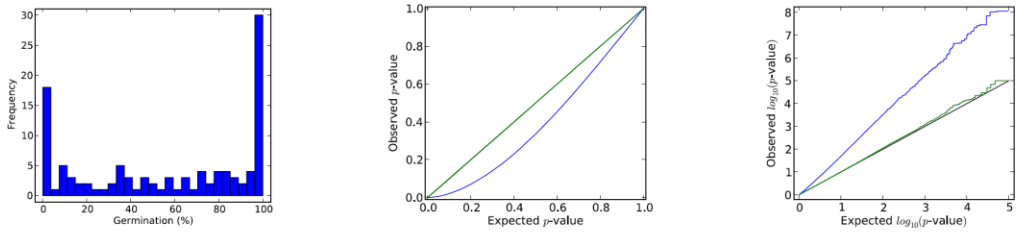
EMMA results



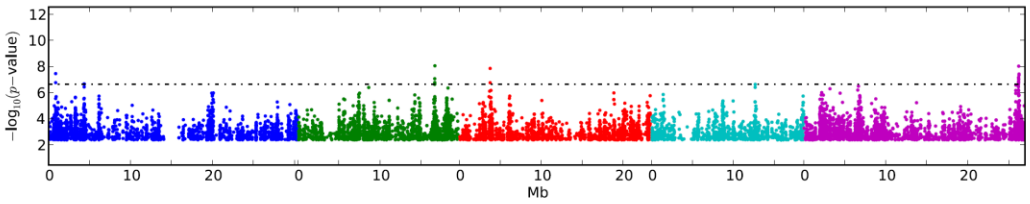
Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
2	5.89834	FLC*	AT5G10140	5	3171211	2287
15	4.46862	ACT7	AT5G09810	5	3050873	-1225
17	4.37073	ZFP1	AT1G80730	1	30338768	5447
50	3.80453	IMB1	AT2G34900	2	14735537	-2808
52	3.78542	AT5G01560	AT5G01560	5	219459	0
79	3.52273	NCED9	AT1G78390	1	29479269	16519
87	3.47038	GIL1	AT5G58960	5	23804153	-18790
109	3.36139	ELF5	AT5G62640	5	25154213	12446
149	3.20885	PKS2	AT1G14280	1	4873342	-4280
153	3.20207	PSD	AT1G72560	1	27348820	-15691

Appendix 8d. Results of Genome Wide Association mapping for germination percentage after 56 days of dry storage (Atwell et al. 2010).

Phenotype histogram and quantile-quantile plots of p-values

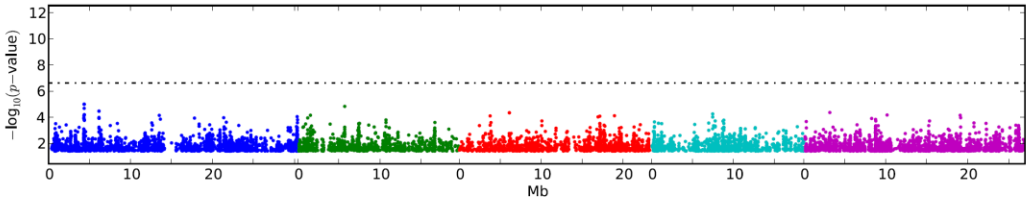


Wilcoxon results



Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
31	6.39096	HBT	AT2G20000	2	8642534	0
33	6.29298	FLC*	AT5G10140	5	3188328	-8879
46	5.95613	AGL15*	AT5G13790	5	4429786	19231
60	5.76202	ATGID1B	AT3G63010	3	23290445	-9955
62	5.73333	COMATOSE	AT4G39850	4	18491140	0
67	5.69039	ATMYB65	AT3G11440	3	3621028	15918
77	5.57797	IMB1	AT2G34900	2	14735537	-2808
86	5.56102	ELF4	AT2G40080	2	16729405	11967
94	5.50112	AGL20*	AT2G45660	2	18826719	-8598
95	5.49346	ATS3*	AT5G07190	5	2241523	2851

EMMA results



Rank	Score	Gene	Gene ID	Chr	SNP pos (bp)	Distance to gene (bp)
8	4.37042	FLC*	AT5G10140	5	3171211	2287
16	4.15566	LBA1	AT5G47010	5	19083198	-6038
22	4.05406	ZFP1	AT1G80730	1	30338768	5447
29	3.95268	DDF1*	AT1G12610	1	4304481	-13466
37	3.84108	KNAT3	AT5G25220	5	8757437	18850
41	3.80086	PKL	AT2G25170	2	10749896	18831
67	3.51887	GA1	AT4G02780	4	1227355	10412
76	3.42835	IAA6	AT1G52830	1	19660176	15988
79	3.40602	HBT*	AT2G20000	2	8642534	0
88	3.36822	COL9	AT3G07650	3	2453015	8477

References

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng DZ, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JDG, Michael T, Nemri A, Roux F, Salt DE, Tang CL, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465(7298): 627-631.

Abbott, R. J., and M. F. Gomes. 1989. Population genetic-structure and outcrossing rate of *Arabidopsis-thaliana* (L) Heynh. Heredity 62:411-418.

Achard, P., H. Cheng, L. De Grauwe, J. Decat, H. Schoutteten, T. Moritz, D. Van der Straeten, J. R. Peng, and N. P. Harberd. 2006. Integration of plant responses to environmentally activated phytohormonal signals. Science 311:91-94.

Alcazar R, Garcia AV, Parker JE, Reymond M. 2009. Incremental steps toward incompatibility revealed by *Arabidopsis* epistatic interactions modulating salicylic acid pathway activation. *Proceedings of the National Academy of Sciences of the United States of America* 106(1): 334-339.

Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M. 2009. What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell* 21(7): 1877-1896.

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

Aranzana, M. J., S. Kim, K. Y. Zhao, E. Bakker, M. Horton, K. Jakob, C. Lister, J. Molitor, C. Shindo, C. L. Tang, C. Toomajian, B. Traw, H. G. Zheng, J. Bergelson, C. Dean, P. Marjoram, and M. Nordborg. 2005. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. Plos Genetics 1:531-539.

Balasubramanian, S., S. Sureshkumar, M. Agrawal, T. P. Michael, C. Wessinger, J. N. Maloof, R. Clark, N. Warthmann, J. Chory, and D. Weigel. 2006. The *PHYTOCHROME C* photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. Nature Genetics 38:711-715.

Baskin, C. C., and J. M. Baskin. 1998. Ecology, biogeography, and evolution of dormancy and germination - Introduction. Seeds:1-3.

Baskin, J. M., and C. C. Baskin. 2004. A classification system for seed dormancy. Seed Science Research 14:1-16.

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

Bentsink, L., J. Jowett, C. J. Hanhart, and M. Koornneef. 2006. Cloning of DOG1, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. Proc Natl Acad Sci U S A 103:17042-7.

Bentsink L, Hanson J, Hanhart CJ, Blankestijn-de Vries H, Coltrane C, Keizer P, El-Lithy M, Alonso-Blanco C, de Andres MT, Reymond M, van Eeuwijk F, Smeekens S, Koornneef M. 2010. Natural variation for seed dormancy in

Arabidopsis is regulated by additive genetic and molecular pathways. *Proceedings of the National Academy of Sciences of the United States of America* 107(9): 4264-4269.

Beveridge, C. A., J. L. Weller, S. R. Singer, and J. M. I. Hofer. 2003. Axillary meristem development. Budding relationships between networks controlling flowering, branching, and photoperiod responsiveness. Plant Physiology 131:927-934.

Bewley, J. D. 1997. Seed germination and dormancy. Plant Cell 9:1055-1066.

Bewley, J. D., M. Black, J. D. Bewley, and M. Black. 1994. Seeds: Physiology of development and germination, Second edition. Seeds: Physiology of development and germination, Second edition:xv+445p.

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

Cadman, C. S. C., P. E. Toorop, H. W. M. Hilhorst, and W. E. Finch-Savage. 2006. Gene expression profiles of *Arabidopsis* Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism. Plant Journal 46:805-822.

Caicedo, A. L., J. R. Stinchcombe, K. M. Olsen, J. Schmitt, and M. D. Purugganan. 2004. Epistatic interaction between *Arabidopsis* FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. Proceedings of the National Academy of Sciences of the United States of America 101:15670-15675.

Chiang, G. C. K., D. Barua, E. M. Kramer, R. M. Amasino, and K. Donohue. 2009. Major flowering time gene, FLOWERING LOCUS C, regulates seed germination in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America 106:11661-11666.

Chibani, K., S. Ali-Rachedi, C. Job, D. Job, M. Jullien, and P. Grappin. 2006. Proteomic analysis of seed dormancy in Arabidopsis. Plant Physiology 142:1493-1510.

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

Donohue K. 2005. Seeds and seasons: Interpreting germination timing in the field. *Seed Science Research* 15(3): 175-187.

Donohue K, Dorn D, Griffith C, Kim E, Aguilera A, Polisetty CR, Schmitt J. 2005. Niche construction through germination cueing: Life-history responses to timing of germination in *Arabidopsis thaliana*. *Evolution* 59(4): 771-785.

Donohue K, Dorn L, Griffith C, Kim E, Aguilera A, Polisetty CR, Schmitt J. 2005. Environmental and genetic influences on the germination of *Arabidopsis thaliana* in the field. *Evolution* 59(4): 740-757.

Donohue K, Dorn L, Griffith C, Kim E, Aguilera A, Polisetty CR, Schmitt J. 2005. The evolutionary ecology of seed germination of *Arabidopsis thaliana*: Variable natural selection on germination timing. *Evolution* 59(4): 758-770.

Dorn, L. A., and T. Mitchellolds. 1991. Genetics of Brassica-Campestris .1. Genetic Constraints on Evolution of Life-History Characters. Evolution 45:371-379.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620.

Fay, J. C., and C. I. Wu. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.

Finch-Savage, W. E., C. S. C. Cadman, P. E. Toorop, J. R. Lynn, and H. W. M. Hilhorst. 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. Plant Journal 51:60-78.

Finch-Savage, W. E., and G. Leubner-Metzger. 2006. Seed dormancy and the control of germination. New Phytologist 171:501-523.

Francois, O., M. G. B. Blum, M. Jakobsson, and N. A. Rosenberg. 2008. Demographic history of European populations of *Arabidopsis thaliana*. Plos Genetics 4

Fu, Y. X. 1995. Statistical Properties of Segregating Sites. Theoretical Population Biology 48:172-197.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.

Geber, M. A. 1990. The Cost of Meristem Limitation in Polygonum-Arenastrum - Negative Genetic Correlations between Fecundity and Growth. Evolution 44:799-819.

Gonzalez, N., G. T. S. Beemster, and D. Inze. 2009. David and Goliath: what can the tiny weed Arabidopsis teach us to improve biomass production in crops? Current Opinion in Plant Biology 12:157-164.

Griffith C, Kim E, Donohue K. 2004. Life-history variation and adaptation in the historically mobile plant *Arabidopsis thaliana* (brassicaceae) in north america. *American Journal of Botany* 91(6): 837-849.

Hilhorst, H. W. M. 1995. A Critical Update on Seed Dormancy .1. Primary Dormancy. Seed Science Research 5:61-73.

Hirschhorn, J. N., and M. J. Daly. 2005. Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics 6:95-108.

Hopkins, R., J. Schmitt, and J. R. Stinchcombe. 2008. A latitudinal cline and response to vernalization in leaf angle and morphology in *Arabidopsis thaliana* (Brassicaceae). New Phytologist 179:155-164.

Horvath, D. 2009. Common mechanisms regulate flowering and dormancy. Plant Science 177:523-531.

Horvath, D. P., J. V. Anderson, W. S. Chao, and M. E. Foley. 2003. Knowing when to grow: signals regulating bud dormancy. Trends in Plant Science 8:534-540.

Huang XQ, Schmitt J, Dorn L, Griffith C, Effgen S, Takao S, Koornneef M, Donohue K. 2010. The earliest stages of adaptation in an experimental plant population: Strong selection on QTLs for seed dormancy. *Molecular Ecology* 19(7): 1335-1351.

Ichihashi, Y., G. Horiguchi, S. Gleissberg, and H. Tsukaya. The bHLH Transcription Factor SPATULA Controls Final Leaf Size in *Arabidopsis thaliana*. Plant and Cell Physiology 51:252-261.

Johanson, U., J. West, C. Lister, S. Michaels, R. Amasino, and C. Dean. 2000. Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science 290:344-347.

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. 2008. Efficient control of population structure in model organism association mapping. *Genetics* 178(3): 1709-1723.

Keurentjes JJB, Koornneef M, Vreugdenhil D. 2008. Quantitative genetics in the age of omics. *Current Opinion in Plant Biology* 11(2): 123-128.

Kim, S., V. Plagnol, T. T. Hu, C. Toomajian, R. M. Clark, S. Ossowski, J. R. Ecker, D. Weigel, and M. Nordborg. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana.* Nature Genetics 39:1151-1155.

Koornneef M, Alonso-Blanco C, Vreugdenhil D. 2004. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology* 55: 141-172.

Korves, T. M., K. J. Schmid, A. L. Caicedo, C. Mays, J. R. Stinchcombe, M. D. Purugganan, and J. Schmitt. 2007. Fitness effects associated with the major flowering time gene FRIGIDA in *Arabidopsis thaliana* in the field. American Naturalist 169:E141-E157.

Krizek, B. A. 2009. Making bigger plants: key regulators of final organ size. Current Opinion in Plant Biology 12:17-22.

Lang, G. A., J. D. Early, G. C. Martin, and R. L. Darnell. 1987. Endodormancy, Paradormancy, and Ecodormancy - Physiological Terminology and Classification for Dormancy Research. Hortscience 22:371-377.

Laserna, M. P., R. A. Sanchez, and J. F. Botto. 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.

Lefebvre, V., H. North, A. Frey, B. Sotta, M. Seo, M. Okamoto, E. Nambara, and A. Marion-Poll. 2006. Functional analysis of *Arabidopsis* NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. Plant Journal 45:309-319.

Lempe, J., S. Balasubramanian, S. Sureshkumar, A. Singh, M. Schmid, and D. Weigel. 2005. Diversity of flowering responses in wild *Arabidopsis thaliana* strains. Plos Genetics 1:109-118.

Loudet, O., S. Chaillou, A. Krapp, and F. Daniel-Vedele. 2003. Quantitative trait loci analysis of water and anion contents in interaction with nitrogen availability in *Arabidopsis thaliana*. Genetics 163:711-722.

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

McKay, J. K., J. H. Richards, and T. Mitchell-Olds. 2003. Genetics of drought adaptation in *Arabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. Molecular Ecology 12:1137-1151.

McKay JK, Richards JH, Nemali KS, Sen S, Mitchell-Olds T, Boles S, Stahl EA, Wayne T, Juenger TE. 2008. Genetics of drought adaptation in *Arabidopsis thaliana* ii. QTL analysis of a new mapping population, kas-1 x tsu-1. *Evolution* 62(12): 3014-3026.

Metcalf, C. J. E., and T. Mitchell-Olds. 2009. Life history in a model system: opening the black box with *Arabidopsis thaliana*. Ecology Letters 12:593-600.

MitchellOlds, T. 1996a. Genetic constraints on life-history evolution: Quantitative-trait loci influencing growth and flowering in Arabidopsis thaliana. Evolution 50:140-145.

MitchellOlds, T. 1996b. Pleiotropy causes long-term genetic constraints on life-history evolution in *Brassica rapa*. Evolution 50:1849-1858.

Montesinos-Navarro A, Wig J, Pico FX, Tonsor SJ. 2010. *Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude. *New Phytologist* 189(1): 282-294.

Montesinos, A., S. J. Tonsor, C. Alonso-Blanco, and F. X. Pico. 2009. Demographic and genetic patterns of variation among populations of *Arabidopsis thaliana* from contrasting native environments. PLoS One 4:e7213.

Nordborg, M., J. O. Borevitz, J. Bergelson, C. C. Berry, J. Chory, J. Hagenblad, M. Kreitman, J. N. Maloof, T. Noyes, P. J. Oefner, E. A. Stahl, and D. Weigel. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics 30:190-193.

Ostrowski, M. F., J. David, S. Santoni, H. McKhann, X. Reboud, V. Le Corre, C. Camilleri, D. Brunel, D. Bouchez, B. Faure, and T. Bataillon. 2006. Evidence for a large-scale population structure among accessions of *Arabidopsis thaliana*: possible causes and consequences for the distribution of linkage disequilibrium. Molecular Ecology 15:1507-1517.

Penfield, S., and J. King. 2009. Towards a systems biology approach to understanding seed dormancy and germination. Proc Biol Sci 276:3561-9.

Pico, F. X., B. Mendez-Vigo, J. M. Martinez-Zapater, and C. Alonso-Blanco. 2008. Natural Genetic Variation of *Arabidopsis thaliana* Is Geographically Structured in the Iberian Peninsula. Genetics 180:1009-1021.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38:904-909.

Roff, D. 2002. Life history evolution. Life history evolution:i-vii, 1-527.

Roff, D. A., and D. J. Fairbairn. 2007. The evolution of trade-offs: where are we? Journal of Evolutionary Biology 20:433-447.

Russell, L., V. Larner, S. Kurup, S. Bougourd, and M. Holdsworth. 2000. The *Arabidopsis* COMATOSE locus regulates germination potential. Development 127:3759-3767.

Schmid, K., O. Torjek, R. Meyer, H. Schmuths, M. H. Hoffmann, and T. Altmann. 2006. Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. Theoretical and Applied Genetics 112:1104-1114.

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(4): 623-634.

Sharbel, T. F., B. Haubold, and T. Mitchell-Olds. 2000. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. Molecular Ecology 9:2109-2118.

Simpson, G. G., and C. Dean. 2002. Flowering - *Arabidopsis*, the rosetta stone of flowering time? Science 296:285-289.
Stearns, S. C., and S. C. Stearns. 1992. The evolution of life histories. The evolution of life histories.:i-xii, 1-249.
Stich, B., A. E. Melchinger, M. Heckenberger, J. Mohring, A. Schechert, and H. P. Piepho. 2008a. Association mapping in multiple segregating populations of sugar beet (*Beta vulgaris* L.). Theoretical and Applied Genetics 117:1167-1179.
Stich, B., J. Mohring, H. P. Piepho, M. Heckenberger, E. S. Buckler, and A. E. Melchinger. 2008b. Comparison of mixed-model approaches for association mapping. Genetics 178:1745-1754.
Stich, B., H. P. Piepho, B. Schulz, and A. E. Melchinger. 2008c. Multi-trait association mapping in sugar beet (*Beta vulgaris* L.). Theoretical and Applied Genetics 117:947-954.
Stinchcombe, J. R., A. L. Caicedo, R. Hopkins, C. Mays, E. W. Boyd, M. D. Purugganan, and J. Schmitt. 2005. Vernalization sensitivity in *Arabidopsis thaliana* (brassicaceae): The effects of latitude and FLC variation. American Journal of Botany 92:1701-1707.
Stinchcombe, J. R., C. Weinig, M. Ungerer, K. M. Olsen, C. Mays, S. S. Halldorsdottir, M. D. Purugganan, and J. Schmitt. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. Proceedings of the National Academy of Sciences of the United States of America 101:4712-4717.
Stratton, D. A. 1998. Reaction norm functions and QTL-environment interactions for flowering time in *Arabidopsis thaliana*. Heredity 81:144-155.
Tajima, F. 1989. Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics 123:585-595.
Vleeshouwers, L. M., H. J. Bouwmeester, and C. M. Karssen. 1995. Redefining seed dormancy: An attempt to integrate physiology and ecology. Journal of Ecology 83:1031-1037.
Weinig C, Ungerer MC, Dorn LA, Kane NC, Toyonaga Y, Halldorsdottir SS, Mackay TFC, Purugganan MD, Schmitt J. 2002. Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. *Genetics* 162(4): 1875-1884.
Weinreich, D. M. 2005. The rank ordering of genotypic fitness values predicts genetic constraint on natural selection on landscapes lacking sign epistasis. Genetics 171:1397-1405.
Weinreich, D. M., and L. Chao. 2005. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. Evolution 59:1175-1182.
Weinreich, D. M., R. A. Watson, and L. Chao. 2005. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. Evolution 59:1165-1174.
Werner, J. D., J. O. Borevitz, N. H. Uhlenhaut, J. R. Ecker, J. Chory, and D. Weigel. 2005. FRIGIDA-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. Genetics 170:1197-1207.
Wilczek, A. M., J. L. Roe, M. C. Knapp, M. D. Cooper, C. Lopez-Gallego, L. J. Martin, C. D. Muir, S. Sim, A. Walker, J. Anderson, J. F. Egan, B. T. Moyers, R. Petipas, A. Giakountis, E. Charbit, G. Coupland, S. M. Welch, and J. Schmitt. 2009.

Effects of Genetic Perturbation on Seasonal Life History Plasticity. Science 323:930-934.
Yu, J. M., J. B. Holland, M. D. McMullen, and E. S. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178:539-551.
Yu, J. M., G. Pressoir, W. H. Briggs, I. V. Bi, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich, and E. S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38:203-208.
Zeng, K., Y. X. Fu, S. H. Shi, and C. I. Wu. 2006. Statistical tests for detecting positive selection by utilizing high-frequency variants. Genetics 174:1431-1439.
Zhao, K. Y., M. J. Aranzana, S. Kim, C. Lister, C. Shindo, C. L. Tang, C. Toomajian, H. G. Zheng, C. Dean, P. Marjoram, and M. Nordborg. 2007. An *Arabidopsis* example of association mapping in structured samples. Plos Genetics 3

Acknowledgements

I would like to thank Juliette de Meaux and Maarten Koornneef.
They gave me the chance to work on the topic I love the most.
They gave me the chance to learn English and to meet wonderful people.

Thank you to Annie Schmitt, Kathleen Donohue and George Chiang for the discussions about ecology and germination.

Thank you to Magnus Nordborg.

Thank you to Benjamin stich who helpt me for statistics and to be more logic.

Thank you to Hugues Barbier who helpt me at the beginning of my thesis and suppoted me.

Thank you to Ute Tartler for her help in the lab and her support.

I would like to thank deeply, Derry Voisin, Daniela Liebsch, Michiel Kwantes and Bjorn Pieper for their support.

I would like to thank again Derry Voisin who has always supported me.

Thank you to my parents.

My sister who always knows exactly what to tell me especially when I start to do stupid things.

Her husband (Fred) as my private strategy consultant and art critic.

And one person I met during my PhD and who has changed my life…

Lebenslauf

Name: **DEBIEU**
Vorname: **Marilyne**
Institut: Max-Planck-Institut für Züchtungsforschung
Adresse: Carl-von-Linné-Weg 10, 50829 Köln, Deutschland
Telefon: ++49 (221) 5062 467
E-mail: debieu@mpiz-koeln.mpg.de
Geburtsdatum: 13. April, 1981
Nationalität: Französisch

DERZEITIGE POSITION

2006-derzeit: **MAX-PLANCK-INSTITUT FÜR ZÜCHTUNGSFORSCHUNG, KÖLN, DEUTSCHLAND**
- Ph.D.-Projekt: "Life history strategies in *Arabidopsis thaliana*"
- Betreuer: Juliette de Meaux und Maarten Koornneef
- Verteidigung der Dissertation: April 2010

AUSBILDUNG

2004-2005: **UNIVERSITÄT PARIS SUD-XI, ORSAY, FRANKREICH**
Master in Biologie, 2. Jahr (Abschluß cum laude, Schwerpunkt Evolutionsbiologie.

2003-2004: **UNIVERSITÄT PARIS SUD-XI, ORSAY, FRANKREICH**
Master in Biologie, 1. Jahr, Schwerpunkt Molekular-Biologie, Genetik und Evolution.

2002-2003: **UNIVERSITÄT PARIS SUD-XI, ORSAY, FRANKREICH**
Bachelor in Biologie, Schwerpunkt Biologie, Chemie und Genetik.

2000-2002: **UNIVERSITÄT PARIS SUD-XI, ORSAY, FRANKREICH**
Allgemeines Universitätsdiplom (DEUG) in Biowissenschaften (cum laude), Schwerpunkt Biologie, Chemie, Mathematik und Physik.

FORSCHUNGS-PRAKTIKA

2005: **UMR DE GENETIQUE VEGETALE, GIF-SUR-YVETTE, FRANKREICH**
6 Monate
- Projekt: "Evolution einer Multigen-Familie während der Mais-Domestizierung."
- Betreuer: Maud Tenaillon und Domenica Manicacci

2003-2004: **LABORATOIRE ECOLOGIE, SYSTEMATIQUE & EVOLUTION, UNIVERSITÄT PARIS SUD-XI, ORSAY, FRANKREICH**
4 Monate
- Projekt: "Phylogeographische Analyse von *Fraxinus excelsio*r und *Fraxinus angustifolia."*
- Betreuer: Juan Fernandez-Manjarres und Nathalie Frascaria-Lacoste

2002: **HOPITAL HOTEL DIEU, PARIS, FRANKREICH**
2 Monate
Praktikum als medizinische Assistentin im Zentrum für Schlafstörungen.

PUBLIKATIONEN

1: S. Atwell, Y. Huang, B.J. Vilhjalmsson, G. Willems, M. Horton, Y. Li, D. Meng, A. Platt, A. Tarone, T. T.Hu, R. Jiang, N. W. Muliyati, X. Zhang, M. A. Amer, I. Baxter, B. Brachi, J. Chory, C. Dean, **M. Debieu**, J. de Meaux, J. R. Ecker, N. Faure, J. M. Kniskern, J. D. G. Jones, T. Michael, A. Nemri, F. Roux, D. E. Salt, C. Tang, M. Todesco, M. B. Traw, D. Weigel, P. Marjoram, J. Borevitz, J. Bergelson, M. Nordborg (2010). Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* inbred lines. *Nature* 465(7298): 627-631.

2: Chiang, George; Bartsch, Melanie; Barua, Deepak; Nakabayashi, Kazumi; **Debieu, Marilyne**; Kronholm, Ilkka; Koornneef, Maarten; Soppe, Wim; Donohue, Kathleen; De Meaux, Juliette. *DOG1* expression predicts maternal effects and geographic variation in germination in *Arabidopsis thaliana*. Molecular Ecology. *submitted*

3: **M. Debieu**, C. Tang, B. Stich, S. Effgen, E. Josephs, J. Schmitt, M. Nordborg, M. Koornneef and J. de Meaux. Co-variation of germination and growth related traits across the European latitudinal range of *Arabidopsis thaliana. In preparation*

4: J. Corbi, **M. Debieu**, A. Rousselet, P. Montalent, M. Le Guilloux, D. Manicacci and M.Tenaillon (2011). Contrasted patterns of selection since maize domestication on duplicated genes encoding a starch pathway enzyme. Theoretical and Applied Genetics 122(4): 705-722.

Eidesstattliche Erklärung

Köln, 2011

Hiermit versichere ich, daß die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfstmittel vollständig angegeben und die Stellen der Arbeit- einschließlich Tabellen, Karten und Abbildungen-, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde.

Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Dr. Juliette de Meaux und Prof. Dr. Maarten Koornneef betreut worden.

Teile dieser Arbeit sind veröffentlicht in:

Atwell et al. 2010