Vernalization and photoperiodic regulation of flowering time and population genetic structure of Norwegian populations of *Arabidopsis thaliana*

Variasjon i krav til vernalisering og daglengde for blomstring samt populasjonsgenetisk struktur i norske populasjoner av *Arabidopsis thaliana*

Anna Lewandowska-Sabat

Norwegian University of Life Sciences • Universitetet for miljø- og biovitenskap
Department of Plant and Environmental Sciences

Philosophiae Doctor (PhD) Thesis 2009:44

Anna Lewandowska-Sabat

Philosophiae Doctor (PhD) Thesis 2009:44

ISBN 978-82-575-0902-6
ISSN 1503-1667

Norwegian University of Life Sciences
NO–1432 Ås, Norway
Phone +47 64 96 50 00
www.umb.no, e-mail: postmottak@umb.no

Vernalization and photoperiodic regulation of flowering time and population genetic structure of Norwegian populations of *Arabidopsis thaliana*

Variasjon i krav til vernalisering og daglengde for blomstring samt populasjonsgenetisk struktur i norske populasjoner av *Arabidopsis thaliana*

Anna M. Lewandowska-Sabat

Department of Plant and Environmental Sciences
Norwegian University of Life Sciences
P.O. Box 5003, 1432 Ås
Norway

Ås 2009

Thesis nr: 2009:44
ISSN: 1503-1667
ISBN: 978-82-575-0902-6

Table of contents

Acknowledgements I
Abbreviations III
Abstract V
Sammendrag VII
List of papers IX
Introduction 1
  The study organism – history and geographic distribution 1
  Genetic variation in natural populations 3
  Adaptive potential of natural variation 5
  Genetic structure and biogeography in *A. thaliana* 6
  Demographic processes and natural selection 8
  Studying of natural genetic variation for improving crop production 9
  Flowering time regulation in *A. thaliana* 11
  The vernalization pathway 13
  The photoperiodic pathway 16
  Photoreceptors 19
  Light spectrum at northern latitudes 20
  The circadian clock 20
  The gibberellin and autonomous pathways 22
Objectives 24
Material and methods 25
  Plant material 25
  Sequencing of candidate genes 26
  Real-time RT PCR 26
  Statistical analyses 27
Results and discussion 28
  Vernalization and photoperiodic responses are associated with winter duration at the site of origin of the populations 28
  FLC and *CO* sequence variation contributes to vernalization and photoperiodic sensitivity, respectively 29
  Variation in expression of *CRY2*, *GI* and *TOC1* contributes to photoperiodic phenotypes 30
  Population structure and genetic diversity 31
Conclusions and future perspectives 34
References 35
Paper I-III

Acknowledgements

The present work was performed from 2006-2010 in Plant Genetics group at Department of Plant and Environmental Sciences, Norwegian University of Life Sciences, and was supported by Norwegian Research Council: The Norwegian Arabidopsis Research Centre, project no. 151991/150.

Firstly, I would like to thank my supervisors Professor Odd Arne Rognli and Dr. Siri Fjellheim for scientific support and introducing me to the field of plant genetics and molecular ecology. Thank you for your guidance, fruitful discussions and always believing in me and trusting my skills. Thanks to Professor Jorunn Elisabeth Olsen for discussing difficulties of experimental design and to Dr. Carl Gunnar Fossdal for help to get started and guidance through the gene expression experiment.

I want to express my gratitude to Ragnhild Nestestog for her technical assistance and introducing me to the *Arabidopsis* world. I am also grateful to YeonKyeong Lee, Kjetil Fosnes, Magdalena Achrem, Lidia Skuza and Stine Indrelid for help with the lab work. This work could have never been performed without your excellent assistance.

I want to thank all my colleagues for your warm welcome, kindness, support and creating inspiring scientific atmosphere. Special thanks to Simen Rød Sandve for all scientific discussions, comments and interest.

Thanks to all my friends, especially Karolina, Wiebke, Signe, Anna, Natasa, YeonKyeong, Karen, Diane, Ellen – for your support and wonderful time together. I would like to show my gratitude to my parents and brothers for believing in me, support and encouragement. Finally, I wish to thank my husband Pawel for his patience, loving support and keeping the faith in the most difficult moments. Thanks for listening to my presentations, reading the papers and carrying my posters. Your help is very much appreciated.

Abbreviations

AFLP	Amplified fragment length polymorphism
AGL24	*AGAMOUS-LIKE 24*
AP1	*APETALA 1*
CAL	*CAULIFLOWER*
CC	companion cells
CCA1	*CIRCADIAN CLOCK ASSOCIATED 1*
CDF1	*CYCLING DOF FACTOR 1*
CK2	CASEIN KINASE 2
CLF	*CURLY LEAF*
CO	*CONSTANS*
CRY1	*CRYPTOCHROME 1*
CRY2	*CRYPTOCHROME 2*
CRY3	*CRYPTOCHROME 3*
Cvi	Cape Verde Islands
DOC1	*DELAY OF GERMINATION 1*
EFS	*EARLY FLOWERING IN SHORT DAYS*
FD	*FLOWERING LOCUS D*
FKF1	*FLAVIN-BINDING KELCH REPEAT*, *F-BOX*
FLC	*FLOWERING LOCUS C*
FLD	*FLOWERING LOCUS D* (autonomous pathway)
FLK	*FLOWERING LATE KH MOTIF*
FLM	*FLOWERING LOCUS M*
FLW1	*FLOWERING 1*
FM	floral meristem
FR	far red light
FRI	*FRIGIDA*
FRL1	*FRIGIDA-LIKE 1*
FRL2	*FRIGIDA-LIKE 2*
FT	*FLOWERING LOCUS T*
FUL	*FRUITFULL*
GA	gibberellins
GAI	*GIBBERELLIN-INSENSITIVE*
GI	*GIGANTEA*
Hd1	*Heading date 1*
Hd3a	*Heading date 3a*
Hd6	*Heading date 6*
HUA2	*ENHANCER OF AG (AGAMOUS)-4 2*
IBD	isolation-by-distance
IM	inflorescence meristem
LD	*LUMINIDEPENDENS*
LD	linkage disequilibrium
LDs	long days
LFY	*LEAFY*
LHP1	*LIKE HETEROCHROMATIN PROTEIN 1*
LHY	*LATE ELONGATED HYPOCOTYL*
LKP2	LOV KELCH PROTEIN 2
LOV	light, oxygen, or voltage

PHOT1	*PHOTOTROPIN 1*
PHOT2	*PHOTOTROPIN 2*
PHYA	*PHYTOCHROME A*
PHYB	*PHYTOCHROME B*
PHYC	*PHYTOCHROME C*
PHYD	*PHYTOCHROME D*
PHYE	*PHYTOCHROME E*
PRC2	POLYCOMB REPRESSION COMPLEX 2
PRR 3	*PSEUDO-RESPONSE REGULATOR 3*
PRR 7	*PSEUDO-RESPONSE REGULATOR 7*
PRR 9	*PSEUDO-RESPONSE REGULATOR 9*
QTL	QUANTITATIVE TRAIT LOCUS
R/FR	red/far red light
SAM	shoot apical meristem
SDG25	*SET DOMAIN GROUP 25*
SDs	short days
SE	sieve elements
SNP	single-nucleotide polymorphism
SOC1	*SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1*
SVP	*SHORT VEGETATIVE PHASE*
TOC1	*TIMING OF CAB EXPRESSION 1*
TOE1	*TARGET OF EAT 1*
TSF	*TWIN SISTER OF FT*
VIN3	*VERNALIZATION-INSENSITIVE 3*
VRN1	*VERNALIZATION 1*
VRN1	*Vernalization 1* (wheat and barley)
VRN2	*VERNALIZATION 2*
VRN2	*Vernalization 2* (wheat and barley)
VRN3	*Vernalization 3* (wheat and barley)
VRN5	*VERNALIZATION 5*
VRT2	*Vernalization gene* (wheat and barley)
ZTL	*ZEITLUPE*

Abstract

Natural environmental conditions vary across seasons and therefore timing of flowering is a major determinant of plant reproductive success and a trait with a potentially high adaptive value. Plants have developed mechanisms to incorporate external signals carrying information from the surrounding environment and ensure timing of flowering at a proper season.

In *Arabidopsis thaliana* vernalization is one of the pathways controlling flowering time in a response to environmental signals. Vernalization accelerates flowering after a period of low temperature. Two genes, *FLOWERING LOCUS C* (*FLC*) which is a repressor of flowering in the absence of vernalization and its activator *FRIGIDA* (*FRI*), play an important role in flowering time variation and appear as major targets for natural selection.

Photoperiod is an environmental cue perceived by plants in order to adjust to the seasonal changes. Adaptation of plants to photoperiods varying substantially with geographic location is essential for their reproductive success. Photoperiodic flowering in *A. thaliana* is controlled by a number of circadian clock regulated genes.

Genetic drift, gene flow and natural selection determine population genetic variation and structure and genetic characterization can be beneficial in understanding historical events, population dynamics and footprints of natural selection. During geographical range expansion genetic variation in populations is often reduced at the species distribution boundary due to bottlenecks.

Natural populations of *A. thaliana* from the most extreme northern distribution range of the species in Norway (59-68°N) have been used in this study. We have studied flowering time responses to vernalization and photoperiod in these populations. Geographical pattern of population genetic structure and diversity were also studied in order to elucidate the demographic history of populations within the northernmost range of the species. Furthermore, we have examined whether sequence variation at *FRI* and *FLC* is associated with vernalization responses. Moreover, association of *CONSTANS* (*CO*) sequence variation and transcript levels of key circadian clock regulated genes with photoperiodic responses was examined.

Coastal populations, from both arctic (65-68°N) and subarctic latitudes (61-62°N) showed lower vernalization sensitivity and higher photoperiodic sensitivity than populations from inland. Moreover, *FLC* sequence variation was associated with vernalization sensitivity

and *CO* sequence variation was associated with photoperiodic sensitivity. Furthermore, association of *CRYPTOCHROME 2* (*CRY2*), *GIGANTEA* (*GI*) and *TIMING OF CAB EXPRESSION 1* (*TOC1*) transcript levels and photoperiodic sensitivity was found. Genetic diversity in Northern (65-68°N) populations was lower than in Southern populations (59-62°N) and a high level of population subdivision (F_{ST}=0.85 ± 0.007) was found revealing that *A. thaliana* is highly structured at the regional scale.

These results suggest that selection favours highly photoperiod sensitive populations of *A. thaliana* at the coast by altering the mRNA level of circadian clock regulated genes. Moreover, *FLC* contributes to variation in vernalization sensitivity and may determine local adaptation of *A. thaliana* at its northernmost distribution boundary. A cryptic population structure was found, which may be a result of human-mediated seed dispersal. The depletion of genetic variation in the northernmost populations might be due to relatively few founder individuals and population bottlenecks during the northward expansion of the species.

Sammendrag

Miljøbetingelser i naturen varierer med skiftende sesong og derfor er blomstringstidspunkt en sentral faktor i planters reproduksjonsevne og et karaktertrekk med potensielt høy adaptiv verdi. Planter har utviklet mekanismer for å registrere informasjonsbærende signaler om plantenes miljø fra omgivelsene og utnytter disse for å sikre blomstring til riktig tidspunkt.

I *Arabidopsis thaliana* er vernalisering en av de genetiske ”pathways” som kontrollerer blomstringstid som en respons på ytre miljøfaktorer. Vernalisering fremmer blomstring etter en periode med lave temperaturer. To gener, *FLOWERING LOCUS C* (*FLC*) som undertrykker blomstring når vernalisering ikke har skjedd og *FRIGIDA* (*FRI*) som aktiverer *FLC*, spiller viktige roller i blomstringstidsvariasjon og fremstår som hovedmål for naturlig variasjon.

Fotoperiode er et miljøsignal som plantene oppfatter og anvender til å tilpasse seg skiftende sesonger. Planters tilpasning til ulike fotoperioder varierer med geografisk lokalitet og er avgjørende for deres reproduktive suksess. Fotoperiodisk blomstring i *A. thaliana* kontrolleres av en rekke gener som er regulert av den cirkadiske klokken.

Genetisk drift, genflyt og naturlig seleksjon er med på å forme populasjoners genetisk variasjon og struktur, og genetisk karakterisering kan være svært nyttig i forståelsen av historiske hendelser, populasjonsdynamikk og for å identifisere spor av naturlig variasjon. Ved planters spredning og utvidelse av geografisk utbredelse vil genetisk variasjon ofte bli redusert i kantene av utbredelsesområdet på grunn av genetiske ”bottlenecks”.

I denne studien har vi brukt naturlige populasjoner av *A. thaliana* fra den mest nordlige utbredelsen i Norge (59-68°N). I disse populasjonene har vi studert blomstringstid som respons på vernalisering og fotoperiode. Videre har vi studert geografiske mønster på populasjonsstruktur og genetisk diversitet for å belyse demografisk historie i populasjoner som befinner seg helt på nordgrensen av artens utbredelse. Vi har også undersøkt om sekvensvariasjon i *FRI* og *FLC* kan være assosiert med vernaliseringsrespons. Sekvensvariasjon i *CONSTANS* (*CO*) og transkripsjonsnivå i sentrale gener fra den cirkadiske klokken har blitt analysert for assosiasjon med fotoperiodisk respons.

Kystpopulasjoner både fra arktiske (65-68°N) og subarktiske (61-62°N) breddegrader viste lavere sensitivitet overfor vernalisering men høyere overfor fotoperiode enn populasjoner fra innlandet. Videre var sekvensvariasjon i *FLC* assosiert med vernaliseringssensitivitet mens sekvensvariasjon i *CO* var assosiert med fotoperiodisk

sensitivitet. Dessuten ble det funnet assosiasjon mellom fotoperiodisk sensitivitet og transkripsjonsnivå av *CRYPTOCHROME 2* (*CRY2*), *GIGANTEA* (*GI*) og *TIMING OF CAB EXPRESSION* (*TOC1*). Det ble funnet lavere nivå av genetisk diversitet i nordlige populasjoner (65-68°N) enn i sørlige populasjoner (59-62°N) og stor grad av inndeling i underpopulasjoner (F_{ST}=0.85 ± 0.007) viser at *A. thaliana* er strukturert genetisk på regionalt nivå.

Resultatene fra disse undersøkelsene indikerer at naturlig seleksjon favoriserer populasjoner som er sensitive til fotoperiode langs kysten ved å endre mRNA-nivået av gener som er regulert av den circadiske klokken. Videre er det vist at *FLC* bidrar til variasjon i vernaliseringssensitivitet og muligens også til lokal adaptasjon av *A. thaliana* ved dens mest nordlige utbredelsesområde. Populasjonene viste seg å ha en kryptisk populasjonsstruktur og dette er mest sannsynlig et resultat av menneskeskapt frøspredning. Mangel på genetisk diversitet i de nordlige populasjonene kan være et resultat av få ”founder” – individer samt genetiske ”bottlenecks” under nordlig ekspansjon av arten.

List of papers

The papers will be referred to by their Roman numerals throughout this thesis.

PAPER I

Lewandowska-Sabat AM, Fjellheim S, Rognli OA.

Differences between continental and coastal climate determine *FLC*-dependent variation in vernalization response of *Arabidopsis thaliana*.

Manuscript

PAPER II

Lewandowska-Sabat AM, Fjellheim S, Fossdal CG, Olsen JE, Rognli OA.

Photoperiodic sensitivity of flowering time in local populations of *Arabidopsis thaliana* is associated with proximity to the coast and altitude, and transcript levels of circadian clock regulated genes.

Manuscript

PAPER III

Lewandowska-Sabat AM, Fjellheim S, Rognli OA.

Genetic diversity and population structure in *Arabidopsis thaliana* from Northern Europe.

Manuscript

Introduction

The study organism – history and geographic distribution

Arabidopsis thaliana (L.) Heyhn (common name: thale cress) was discovered in the sixteenth century by Johannes Thal in the Harz Mountains and the first mutant was documented by Alexander Braun in 1873. It was first proposed as potential model organism for genetic studies by Freidrich Laibach in 1943 and since then, it has been largely implemented by the global scientific community (Somerville et al. 1985; Ausubel 2000; Federspiel 2000).

Arabidopsis thaliana is an almost completely self-fertilizing, small annual flowering plant with short life cycle (in a greenhouse 1.5-3 months), producing a large number of offspring each generation. Sequencing the full genome of *A. thaliana* in 2000 (220 Mb on 5 chromosomes) as the first one in plants put a new perspective into higher plants' genetics (The Arabidopsis Initiative 2000).

Arabidopsis thaliana belongs to the Brassicaceae family and is a wild relative of such species as cabbage, cauliflower, rapeseed and radish. Its genus comprises of nine species that can be mainly found in Europe. Only *A. thaliana* is distributed worldwide (Fig. 1), growing on open or disturbed habitat, on dry soil and tolerating little competition. *A. thaliana* may grow at sea level but also at high altitudes, up to 4000 m a.s.l. (Al-Shehbaz and O'Kane 2002). Populations grow at latitudinal range from 0° (mountains of Tanzania and Kenya) to 68°N (Northern Scandinavia; Fig. 2) which is believed to be at the northernmost distribution edge of the species.

Because of the variety of natural populations *A. thaliana* recently became a model species for the study of natural variation (Mitchell-Olds and Schmitt 2006). Plants collected in the wild represent nearly homozygous genotypes often called accessions or ecotypes but also populations. Although many collections of natural populations do not represent a strict definition of an ecotype (adapted to the specific environmental conditions), they are frequently referred to as ecotypes in the literature. The more neutral term accession does not suggest but also not exclude local adaptation and is recently preferred for the germplasm collection. Over 750 natural accessions of *A. thaliana* have been collected from around the world and are available from seed stock centers.

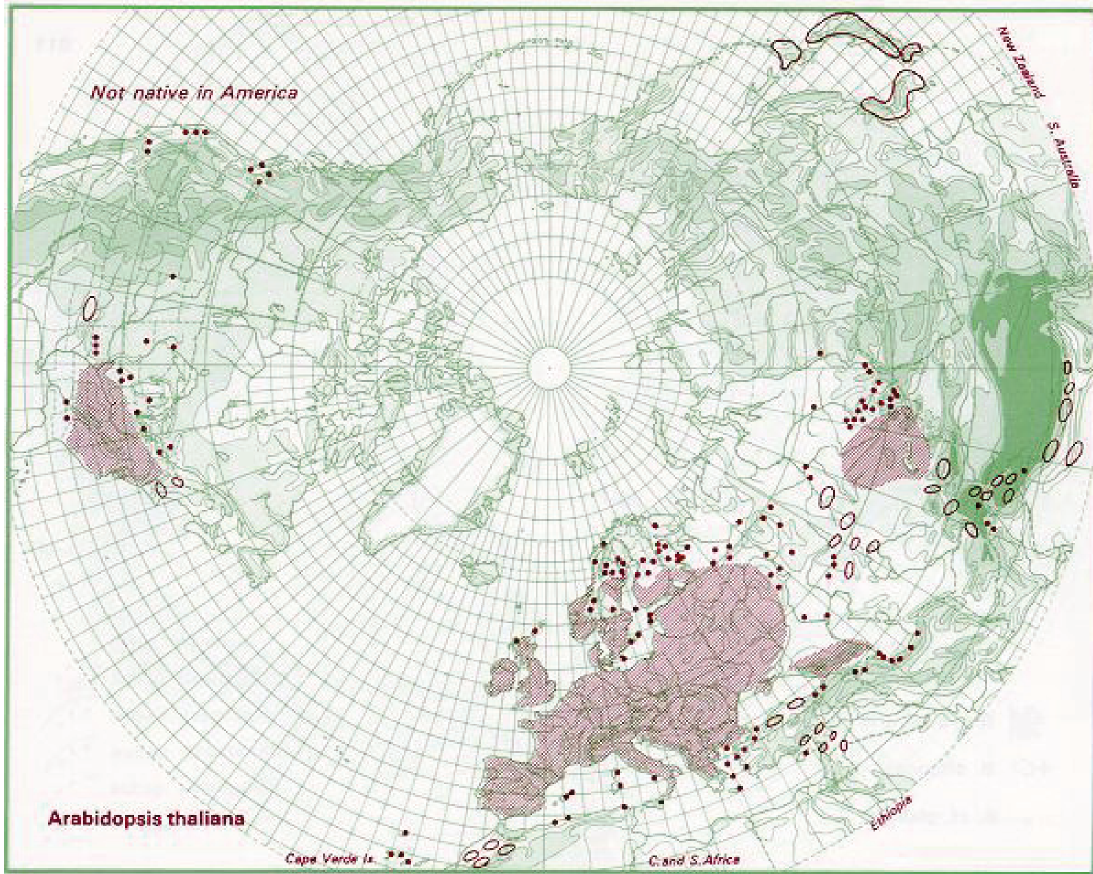


Fig. 1. Worldwide distribution of *Arabidopsis thaliana* (Naturhistoriska riksmuseet 1998).

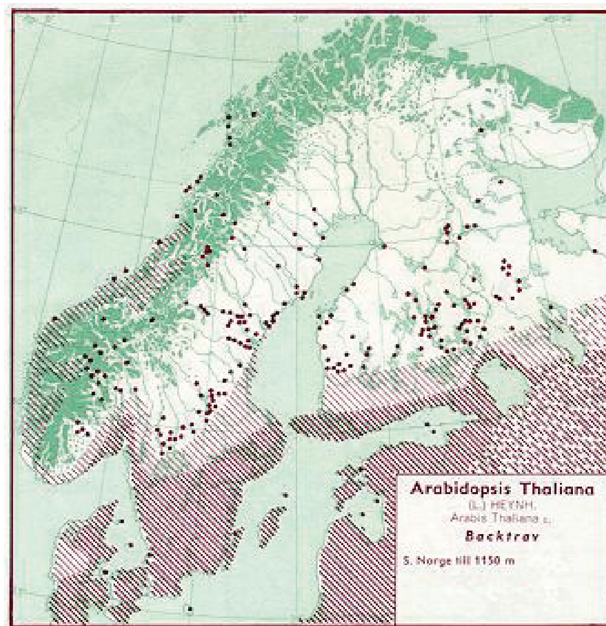


Fig. 2. Distribution of *Arabidopsis thaliana* in Scandinavia (Naturhistoriska riksmuseet 1998).

Genetic variation in natural populations

Natural variation among populations within species is believed to be a sign of evolutionary changes and adaptation to environmental conditions that vary in space and time. Natural populations of *A. thaliana* collected worldwide show a wide range of genetic and trait variation. A rapid postglacial expansion of *A. thaliana* and colonization of broad habitats imply that this species has a great ability to adapt to wide environmental conditions. Natural selection may have acted on naturally occurring mutations and therefore alleles that exist in a high frequency in populations are more likely adaptive.

Utilizing of natural variation contributes to discovery of gene function at population or species level, understanding genetic architecture of complex traits and integration of internal functions with environmental signals. Differences among natural populations are very often controlled by multiple genes and environmental factors have a great effect on a phenotype. Very often a single gene affects multiple traits (pleiotropy) and interacts with other genes (epistasis) thus some phenotypes appear only in certain genetic backgrounds. Mutant collections obtained in laboratory strain backgrounds harbor only a part of natural variation, since ～4% of the genome is highly dissimilar or deleted in wild populations of *A. thaliana* as compared to the reference sequence (Clark et al. 2007). Therefore, natural variation is a relevant resource to discover allelic variants interacting with genetic background and environment, and alleles showing small effects on traits important for plant adaptation (Benfey and Mitchell-Olds 2008). Natural variation is also an important tool in evolutionary biology and plant breeding.

Extensive molecular analyses of *A. thaliana* in the last decade has offered the largest of any plant species, number of genes and nucleotide polymorphisms underlying natural variation (Alonso-Blanco et al. 2005). Natural variation has been reported in many phenotypic traits in *A. thaliana* (reviewed by Koornneef et al. 2004; Alonso-Blanco et al. 2005) and the main interest is focused on flowering time and seed dormancy due to their agronomic importance.

Flowering time is one of the key adaptive traits enabling plants to flower at the most favorable conditions to reproduce. Several genes that contribute to *A. thaliana* natural variation in flowering time have been mapped to the corresponding QTLs (Quantitative Trait Locus) and isolated. Two well studied large effect loci *FRIGIDA* (*FRI*) (Johanson et al. 2000; Shindo et al. 2005) and *FLOWERING LOCUS C* (*FLC*) (Michaels et al. 2003; Werner et al. 2005a) account for natural variation in flowering time in response to vernalization in *A.*

thaliana. *FRI* was first identified by Napp-Zinn (1987) and his study of the cross between the early flowering population Limburg-5 and the late flowering Stockholm. Several loss-of-function alleles of *FRI* have been identified which are believed to be the genetic basis for the evolution of early flowering populations from a late flowering ancestor (Johanson et al. 2000; Le Corre et al. 2002; Shindo et al. 2005). It has been shown that loss-of-function alleles caused by insertions of transposon elements within the first intron of *FLC* reduce its expression and contribute to variation in flowering time (Gazzani et al. 2003; Michaels et al. 2003). Interestingly, allelic variation in *FLC cis*-regulatory sequences that changes the extent of *FLC* chromatin silencing might account for vernalization response (Shindo et al. 2006). Moreover, nonsense and splicing mutations in *FLC* encoding a putatively truncated protein were also found (Werner et al. 2005a). QTL mapping studies have also identified other loci contributing to vernalization response in *A. thaliana*. Two *FRI*-like genes, *FRIGIDA-LIKE 1* (*FRL1*) and *FRIGIDA-LIKE 2* (*FRL2*) (Schläppi 2006), as well as a *FLC*-like gene, *FLOWERING LOCUS M/FLOWERING 1* (*FLM/FLW1*) (Werner et al. 2005b) and *FLC* regulator *ENHANCER OF AG (AGAMOUS)-4 2* (*HUA2*) (Doyle et all. 2005; Wang et al. 2007) likely affects this variation and loss-of-function mutations in these genes contribute to early flowering in *A. thaliana*.

Natural variation for photoperiodic flowering was reported in *A. thaliana*, and is likely caused by three photoreceptor genes, *CRYPTOCHROME 2* (*CRY2*), *PHYTOCHROME C* (*PHYC*) and *PHYTOCHROME D* (*PHYD*). Loss-of-function alleles of *PHYC* and *PHYD* cause early flowering under non-inductive photoperiod (Aukerman et al. 1997; Balasubramanian et al. 2006). Single amino acid substitutions resulting in gain-of-function allele in Cape Verde Islands (Cvi) accession alters CRY2 light-induced regulation of flowering and cause rapid flowering under short days (El-Assal et al. 2001). Recent study has shown that sequence polymorphism in promoter region of *FLOWERING LOCUS T* (*FT*) affects *FT* expression and contributes to natural variation in flowering time in response to photoperiod and ambient temperature (Schwartz et al. 2009). These results suggest that *cis*-regulatory changes at *FT* underlie variation in flowering time responses in *A. thaliana*.

Two other *A. thaliana* photoreceptors have been identified to be involved in natural variation. A single amino acid substitution in *PHYTOCHROME A* (*PHYA*) decrease hypocotyl responses to far red (FR) light in the LM-2 accession (Maloof et al. 2001) and amino acid variation in *PHYTOCHROME B* (*PHYB*) cause differential hypocotyl responses to red light (Filiault et al. 2008). Moreover, change in the amino acid sequence of *FY*, a member of the autonomous pathway, mediates the late-flowering phenotype and reduced

flowering time responses to reduced ratio of red light to far-red light (R/FR) (Adams et al. 2009).

Germination time depending on seed dormancy can have strong influence on fitness. Some genotypes may stay dormant for multiple seasons while others germinate immediately. Thus seed dormancy decreases the risk of population extinction due to harsh environmental conditions that vary between seasons. QTL mapping studies have identified the locus *DOC1* (*DELAY OF GERMINATION 1*) which induce seed dormancy in *A. thaliana* and its functional nucleotide polymorphism is involved in natural variation (Betsnik et al. 2006).

Domestication has often led to a reduction in seed dormancy, an important trait for many crop species. Pre-mature break of dormancy can cause pre-harvest sprouting, a common problem in many regions of the world, and on the other hand increased dormancy, can lead to non-uniform germination in the field. In cereals, several studies have identified seed dormancy QTLs (Lin et al. 1998; Gu et al. 2005; Hori et al. 2007; Imtiaz et al. 2008).

Adaptive potential of natural variation

Molecular population analyses have estimated that 15-30% of *A. thaliana* genes have been subjected to natural selection (Nordborg et al. 2005; Wright and Gaut 2005; Schmid et al. 2006). However, these loci have not been analyzed in fitness studies and therefore it is not clear if natural selection acts on variation in these genes that contributes to adaptive phenotypes.

Flowering is an important life-history trait that contributes to plant fitness (Stearns, 1992) and therefore is a trait of potentially high adaptive value. Flowering time varies among *A. thaliana* accessions coming from very different habitats due to changes in temperatures and photoperiods varying with geographical location. Several genes regulating flowering time in response to environmental cues have been identified, however adaptive significance of natural variation has only been confirmed for *FRI*. Comparisons of statistical values of population differentiation at phenotypic traits (Q_{ST}) and at neutral molecular markers (F_{ST}) within and among populations have shown higher Q_{ST} for flowering time without vernalization and higher F_{ST} for functional and non-functional haplotypes than F_{ST} for neutral markers (Le Corre 2005). This suggests adaptive selection for rapid flowering of *A. thaliana* populations mediated by loss of *FRI* function. Moreover, haplotype sharing of regions flanking *FRI* locus and selective sweep suggest recent selection for early flowering (Hagenblad et al. 2004; Aranzana et al. 2005; Toomajian et al. 2006). Early flowering seems

to be more advantageous under habitats where stressful conditions such as drought or heat in the summer could damage a seed set. On the contrary, late flowering implies a longer vegetative growth phase and thus increased number of offspring due to larger accumulation of resources. It has been shown recently, that nonfunctional *FRI* alleles have negative pleiotropic effects on fitness in *A. thaliana* which may be a basis for natural selection to maintain both variants of *FRI* in natural populations (Scarcelli et al. 2007).

Although molecular data suggests adaptive significance of loss-of-function *FRI* alleles, latitudinal clines in flowering time in *A. thaliana* has rarely been found (Stinchcombe et al. 2004). The lack of apparent cline, which is a sign of plant flowering time adaptation to geographically varying conditions, may be caused by variation in local habitat conditions that often not result from geographic position (Shindo et al. 2005). Furthermore, *A. thaliana* often represent a mixture of native and newly introduced individuals due to human-mediated seed dispersal (Nordborg et al. 2005; Beck et al. 2008) which may also account for lack of the apparent cline.

Genetic structure and biogeography in A. thaliana

Population dynamics of *A. thaliana* is characterized by repeated extinctions and colonization events which likely result in structured distribution of genetic variation. Several molecular phylogenetic studies have been performed to examine relationship between genetic variation and geographical origin of accessions, i.e. isolation by distance (IBD). Genotyping of AFLP markers in 142 accessions was the first study that demonstrated significant IBD in *A. thaliana* (Sharbel et al. 2000). Later study of Nordborg et al. (2005) of 96 *A. thaliana* accessions based on genome-wide polymorphism genotyping suggested the IBD and existence of population structure at a global geographical scale. Further findings of Schmid et al. (2006) confirm this pattern in 351 accessions based on genome-wide SNP genotyping. Schmid et al. (2006) demonstrated also that central Asian accessions had lower level of polymorphism and increased level of genome-wide linkage disequilibrium (LD) as compared to Iberian Peninsula and central European accessions. This may suggest rapid postglacial colonization of Eurasia from glacial refugia. A recent study of 167 world-wide accessions based on genome-wide polymorphism at 10 loci (Beck et al. 2008) has also confirmed IBD in *A. thaliana*. However, reduced relationship between geographic and genetic distance was found in Europe relative to Asia. The authors explain this novel demonstration by potentially recent human-mediated dispersal of *A. thaliana* in Europe.

Arabidopsis thaliana distribution has also been analyzed with respect to climate as the main factor limiting the species distribution range (Hoffmann 2002). This study has shown that low temperature likely restricts the distribution range of *A. thaliana* in Northern Europe, while high temperature and low precipitation determines the range boundaries in North Africa, South-West Asia and in Middle Asia.

Repeated glaciations during Pleistocene had a vast impact on geographical range and genetic variation of many species. Several studies suggest northward species range expansion in Europe and North America (Hewitt 1996; Comes and Kadereit 1998). Most European species are believed to expand from glacial refugia in the peninsulas of Iberia, Italy, and the Balkans, as well as the Caucasus region and the Caspian Sea (Taberlet et al. 1998; Hewitt 2000; Petit et al. 2003).

An increasing interest in elucidating migration history of *A. thaliana*, guided several research groups to study molecular variation in order to detect biogeographic trends in Europe and Asia (e.g. Sharbel et al. 2000; François et al. 2008). Sharbel et al. (2000) proposed that *A. thaliana* re-colonized Europe from two postglacial refugia in Iberian Peninsula and central Asia, with a hybrid zone of two ancestral populations in central Europe (Fig. 3). However, this scenario is presently under discussion (Schmid et al. 2005; Bakker et al. 2006; Beck et al. 2008). In addition to the former scenario, François et al. (2008) proposed anthropogenic spread of *A. thaliana* in Europe, linked with spread of agriculture, as well as east-west migration pattern of re-colonization in Europe from an eastern refugium. A study of Beck et al. (2008) confirm these results and suggest Pleistocene range dynamics with IBD reduced in Europe, likely due to increased human activity.

Scandinavia contains the northern distribution limit of many European plant species. In these marginal areas, severe environmental conditions make reproduction and survival more challenging. Populations from these regions, where environmental gradients are steep, may experience instability in population size, increased drift and genetic divergence (Pamilo and Savolainen 1999). It has been suggested that this region has recovered the flora after last glaciations (~10,000 ya) and since Pleistocene it has been re-colonized by *A. thaliana* from southern and northeastern routes (Sharbel et al. 2000; François et al. 2008).

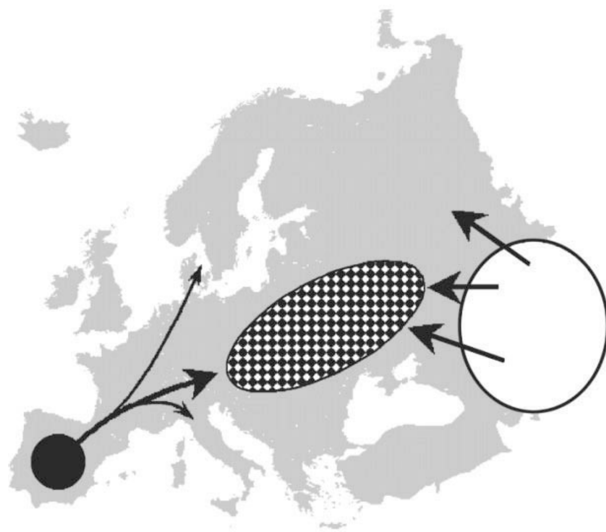


Fig. 3. Scenario for *Arabidopsis thaliana* postglacial colonization of Europe (from Sharbel et al. 2000).

Demographic processes and natural selection

Demographic factors such as population structure, gene flow and changes in effective population size contribute to a genome-wide departure from a neutral model of sequence polymorphism. These processes shape genetic variation across the genome simultaneously and independently from a natural selection and obscure the detection of gene-specific events. Therefore, identifying causative polymorphisms contributing to a given trait may be challenging due to demographic factors. Very often significant associations between phenotypic and genotypic variation are product of false positive associations caused by population structure (Pritchard and Rosenberg 1999; Caicedo et al. 2004; Zhao et al. 2007). Consequently, in *A. thaliana* but also in other organisms, linking naturally occurring variation with adaptive phenotypes requires precise knowledge of population genetic structure.

Linkage disequilibrium (LD) or the non-random pattern of association between alleles at different loci within a population is extensively studied in population genetics. LD-based association mapping is used to study the relationship between phenotypic variation and genetic polymorphism. The extent of LD determines how dense a map must be for association between markers and functionally important polymorphism to be detected, but it also limits the accuracy of the loci mapped. LD decays with genetic distance since distant loci are more likely to have recombined in the past than tightly linked loci.

The mating system, i.e. inbreeding vs. outcrossing, is one of the main factors that shape genetic variation in population. The selfing rate in *A. thaliana* has been estimated to be greater than 95% (Abbott and Gomes 1989; Stenøien et al. 2005) and therefore it was expected that *A. thaliana* show extensive levels of LD (Abbott and Gomes 1989; Nordborg et al. 2002). The anticipated high levels of LD in *A. thaliana* should make it an excellent model for LD-based association mapping.

Nordborg et al. (2005) investigated the pattern of LD on a genomic scale and showed that in global samples of *A. thaliana* LD surprisingly enough decays within approximately 25-50 kb, a pattern typical for widely distributed outcrossing species and comparable to humans. Later Kim et al. (2007) showed that LD decay even faster, within 10 kb. In addition, the level of polymorphism in *A. thaliana* is not unusually low as expected for an inbreeder, it is lower than in *Drosophila melanogaster* but much higher than in humans. Low levels of LD in inbreeding species is not unusual and limited only to *A. thaliana* (Morrell et al. 2005; Song et al. 2009) and may result from ancient recombination that happened in its outcrossing ancestor (Bechsgaard et al. 2006) or selection that favors recombinant genotypes with new combinations of parental alleles (Bakker et al. 2006). Increased recombination levels in inbreeding species is less likely since the evolution of self-fertilization is associated with lower effective recombination rates and decreased effective population size. Thus, low LD in *A. thaliana* may be due to the inclusion in the studies of a global sample of populations that contain all historical polymorphisms and recombinations over thousands of generations (Kim et al. 2007). Although LD decays more rapidly in *A. thaliana* than it was anticipated still, it is higher than in other model organisms. Thus, due to its small genome size, comparably few markers are required for a whole-genome LD mapping.

Studying of natural genetic variation for improving crop production

Considerable conservation of the molecular mechanisms regulating flowering has been shown across distantly related plant species. Particularly, activation of *CONSTANS* (*CO*) and *FT* expression by photoperiod appears highly conserved among plants. The three major loci *Hd1*, *Hd3a*, and *Hd6* (*Heading date 1/3a/6*) involved in photoperiod sensitivity have been isolated in rice (Takahashi et al. 1998; Yano et al. 2000; Kojima et al. 2002). *Hd6* encodes a CK2 casein kinase and *Hd1* and *Hd3a* are orthologues of the *A. thaliana CO* and *FT* photoperiodic response genes. Particularly, the function of *FT* is highly conserved in many plant species. In rice, the induction of the *FT* homologue, *Hd3a* is mediated by a homologue

of *CO*, *Hd1*. *Hd1* promote flowering through elevated expression of *Hd3a* in short days (SDs) (Yano et al. 2000).

Analysis of the *CO*/*FT* system in poplar (*Populus trichocarpa*) has shown that the modification of the timing of *PtCO* and consequent *PtFT* expression is necessary for adaptation to different conditions that vary with latitude. Earlier peak in *CO* expression in Southern trees promotes growth at shorter days as compared to Northern trees. Bud dormancy and growth cessation in Northern trees is therefore induced much earlier in the autumn than in Southern trees since Northern trees recognize longer photoperiod as SDs (Böhlenius et al. 2006).

Since the process of vernalization is of great agronomic importance in grasses much has been learnt about genetic and molecular responses in these species. Four main genes; *VRN1*, *VRN2*, *VRN3* (*Vernalization 1-3*) and *VRT2* (*Vernalization gene*) are of great importance for vernalization in wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*). Grass *VRN1* is a homologue of meristem identity gene *APETALA 1* (*AP1*) (Schmitz et al. 2000; Yan et al. 2003), *VRN2* shares homology with the *CO* gene (Yan et al. 2004), and *VRN3* is a homologue of *FT* in *A. thaliana* (Yan et al. 2006). In grasses, *VRN2* represses *VRN3* and *VRN1* and vernalization or SDs inhibit *VRN2*. In turn, long days (LDs) activate *VRN3* that stimulate *VRN1* to induce flowering (Yan et al. 2004; Dubcovsky et al. 2006, Yan et al. 2006). It is believed that the vernalization response evolved separately in *A. thaliana* and grasses. In *A. thaliana* the MADS-box transcription factor FLC inhibits expression of *FT* and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) and flowering in the absence of vernalization. In grasses the VRN2 protein which shares no homology with FLC but with CO, evolved to repress VRN1/VRN3 in the lack of cold period. Therefore, even though *A. thaliana* provide insights into how flowering is regulated in cereals, this process is to a large extend controlled by different genes. This highlights the need for studying flowering time responses directly in crop species.

A QTL and an orthologue of the *A. thaliana CO* gene found in *Brassica nigra* is associated with flowering time variation (Kruskopf-Osterberg et al. 2002). Furthermore, orthologues of *FLC* in *Brassica napus* and *Brassica rapa* accounts for vernalization response in these species (Tadege et al. 2001; Schranz et al. 2002). Also the *DWARF8* gene accounts for flowering time variation in maize and is an orthologue of the *A. thaliana GIBBERELLIN-INSENSITIVE* (*GAI*) gibberellin signaling gene (Thornsberry et al. 2001).

The knowledge generated from studying *A. thaliana* facilitates the analyses of similar traits, pathways and genes in other species, particularly wild and domesticated crop plants.

Identification of not only QTLs having large effects on the trait variation, but also alleles of small effect will help to understand the regulation of plant gene networks.

Domestication and breeding activities have imposed genetic bottlenecks and reduced the genetic variation in cultivated crops. However, gene banks have been established in order to preserve domesticated and wild germplasm of crop plants. There is an increasing awareness of the importance of plant genetic diversity for the continued development of improved cultivars of crop plants, not least to be able to adapt cultivars to the rapidly changing climatic conditions. Good understanding and characterization of the allelic and phenotypic diversity present in germplasms is a prerequisite for a dynamic preservation and efficient utilization by conventional or transgenic breeding techniques. The development of efficient, high-throughput molecular marker techniques which will make it possible to undertake whole-genome selection is currently being undertaken in many crop species. These will be implemented for the improvement of traits like crop yield, disease resistance and food and feed quality. For many complex traits, improvements cannot be achieved through mutagenesis or transgenic approaches but will, for the foreseeable future, have to rely on the utilization of advanced germplasm and natural genetic variation through conventional and marker-assisted breeding.

Flowering time regulation in A. thaliana

Following seed germination, the young seedling grows during the vegetative phase. After receiving appropriate environmental and endogenous signals, the plant undergoes floral transition, i.e. the shift from vegetative to reproductive growth (Fig. 4).

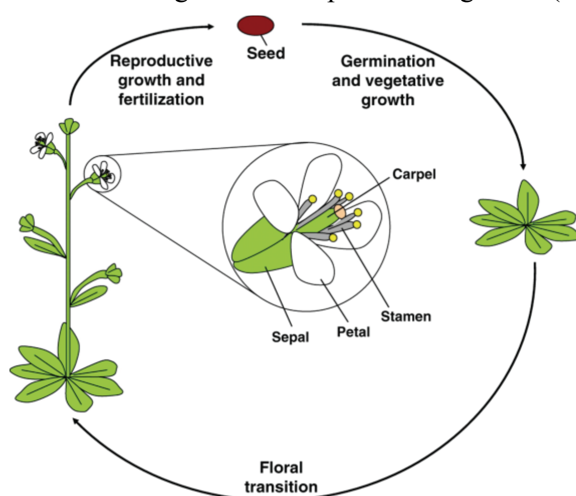


Fig. 4. *Arabidopsis thaliana* life cycle and flower architecture (from Liu et al. 2009).

Natural environmental conditions vary across seasons and therefore time of flowering is a major determinant of plant reproductive success. Flowering time is thus a major contributor to plant fitness (Stearns 1992). Two main life histories have been described in *A. thaliana* in Europe in regards to the variation in length and timing of natural life cycles. The summer annuals germinate in early spring or summer, and accomplish their life cycle during one growing season. The winter annuals germinate in the autumn, overwinter as rosettes and flower and set seeds in the spring or summer (Fig. 5). In general, the majority of Northern European populations are typically winter-annuals while populations from Southern Europe are either winter- or summer-annual types. However, this clasification is being questioned since very little is known about germination time of populations in their natural habitat (Wilczek et al. 2009).

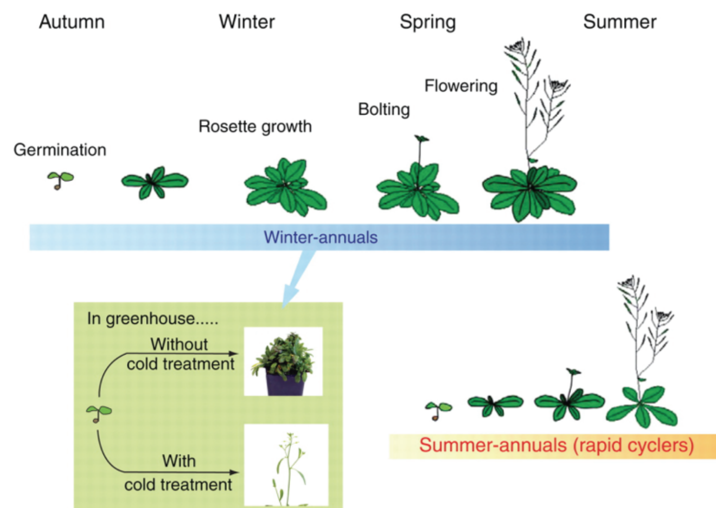


Fig. 5. *Arabidopsis thaliana* accessions are classified into summer-annual and winter-annual types (from Shindo et al. 2007).

Plants have developed mechanisms to incorporate external signals carrying information from the surrounding environment and ensure timing of flowering at a proper season. The transition from vegetative to reproductive stage is mediated by a complex of genetic pathways controlling flowering in response to developmental and environmental signals (Fig. 6). Four main interacting pathways controlling flowering time have been described; i.e. the vernalization response, photoperiodic response, autonomous and gibberellin pathways. Two of these pathways, the vernalization and photoperiodic pathways respond to environmental signals, a period of cold and day length, respectively, and regulate the expression of genes involved in flower development. The gibberellin pathway promotes

flowering in SDs through the action of hormonal inputs, and the autonomous pathway regulate flowering by monitoring internal developmental signals.

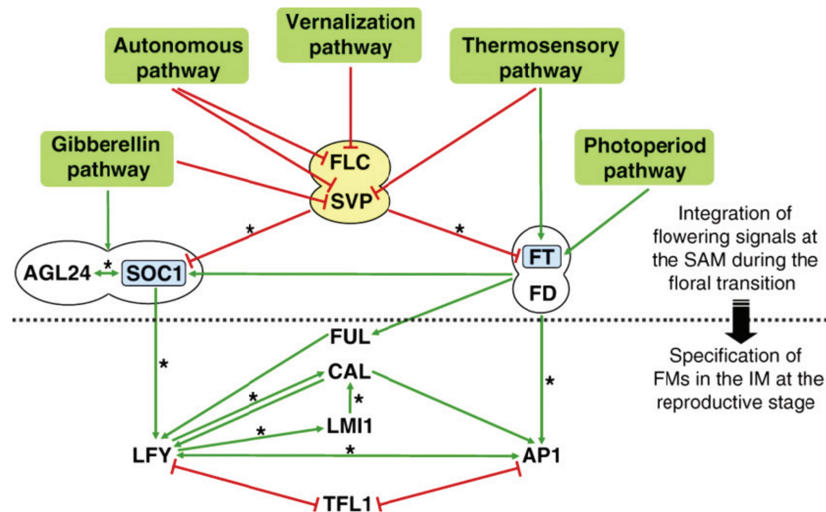


Fig. 6. Regulation of floral meristem identity. Two linked ellipses indicate protein-protein interactions. Asterisks indicate direct transcriptional regulation (from Liu et al. 2009).
AGL24 - AGAMOUS-LIKE 24
AP1 - APETALA1
CAL - CAULIFLOWER
FLC - FLOWERING LOCUS C
FM - floral meristem
FT - FLOWERING LOCUS T
FUL - FRUITFULL
IM - inflorescence meristem
LFY - LEAFY
SAM - shoot apical meristem
SOC1 - SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
SVP - SHORT VEGETATIVE PHASE

The vernalization pathway

Vernalization, i.e. prolonged period of non-freezing temperatures, accelerates flowering in winter-annual plants. Many plants from temperate climate flower only after they experience a period of cold temperatures such as those typically experienced in the winter. *FRI* and *FLC* are two major genes that confer vernalization requirement in *A. thaliana* (Shindo et al. 2005; Schmitz and Amasino 2007). *FRI* encodes a protein of unknown function and *FLC* encodes a transcription factor of the MADS box family (Michaels and Amasino 1999; Johanson et al. 2000). The role of *FRI* is to trigger *FLC* expression prior to vernalization to a level that strongly repress flowering until spring time. Repression of *FLC* in vernalized plants is stable throughout mitotical divisions i.e. even after exposure to cold has ended. The vernalized state in *A. thaliana* has epigenetic character. *FLC* chromatin in non-vernalized plants has features

of active genes, i.e. chromatin is enriched in acetylation of lysine 9 and 14 of histone 3 (H3K9, H3K14) and methylation of H3K4. Through vernalization *FLC* chromatin becomes enriched in methylation of H3K9 and H3K27, which is a feature of repressed chromatin regions (Bastow et al. 2004; Sung and Amasino 2004; Sung et al. 2006; Mylne et al. 2006; De Lucia et al. 2008). It has been shown that changes in *FLC* chromatin silencing contribute to natural variation in vernalization response (Shindo et al. 2006). *FLC* chromatin is fully silenced much earlier in the rapid responding to vernalization Edi-0 than in slowly responding Lov-1 (Fig. 7).

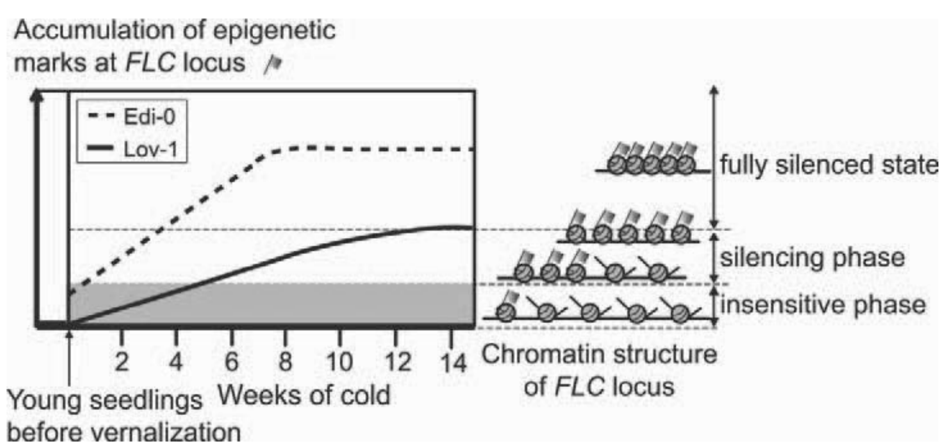


Fig. 7. Model proposing the quantitative accumulation of epigenetic modifications in *FLC* in two *Arabidopsis thaliana* accessions Edi-0 (UK) and Lov-1 (Sweden) (from Shindo et al. 2006).

Four genes, *VERNALIZATION 1* (*VRN1*), *VRN2*, *VRN5* and *VERNALIZATION-INSENSITIVE 3* (*VIN3*) that are responsible for vernalization response have been identified in *A. thaliana* (Gendall et al. 2001; Levy et al. 2002; Sung and Amasino 2004; Greb et al. 2007). In the autumn, components of the *FRI* complex maintain *FLC* chromatin in an active state that results in high level of *FLC* mRNA. The *FRI* complex overrides the repressing effect of the autonomous pathway on *FLC*. Subsequently during winter, the *VIN3* gene mediates the vernalization response by interacting with the PRC2-like complex (POLYCOMB REPRESSION COMPLEX 2) to epigenetically silence *FLC* (De Lucia et al. 2008). The inactive state of *FLC* is maintained by *VRN1*, *LHP1* (*LIKE HETEROCHROMATIN PROTEIN 1*) and the PRC2-like complex, even after plants are returned to warm conditions (Fig. 8). The *FLC* locus passing through meiosis to the next generation is reset to an active state that is necessary to maintain vernalization requirement in the following generation (reviewed in Schmitz and Amasino 2007). A recent study has shown that an early step in the

epigenetic silencing of *FLC* in response to cold may be the upregulation of long non-coding antisense transcripts covering the entire *FLC* locus (Swiezewski et al. 2009). This cold-induced process is linked to a reduction in sense transcription and is independent of the induction of VIN3, known to be the earliest step in the vernalization response (De Lucia et al. 2008).

It has also been shown recently that histon methyltransferase *SET DOMAIN GROUP 25* (*SDG25*) is a positive regulator of *FLC* expression since loss-of-function mutant *sdg25-1* had decreased the *FLC* transcript level and an early flowering phenotype. Moreover, *sdg25-1* mutants responded to long-day photoperiods and vernalization, suggesting that *SDG25* acts downstream of these signaling pathways (Berr et al. 2009). Moreover, another study has shown that *CURLY LEAF* (*CLF*) contributes to the *FLC* regulation since gain-of-function mutation in *CLF* reduces the *FLC* transcript levels and eliminates vernalization requirement in winter-annual accessions (Doyle and Amasino 2009).

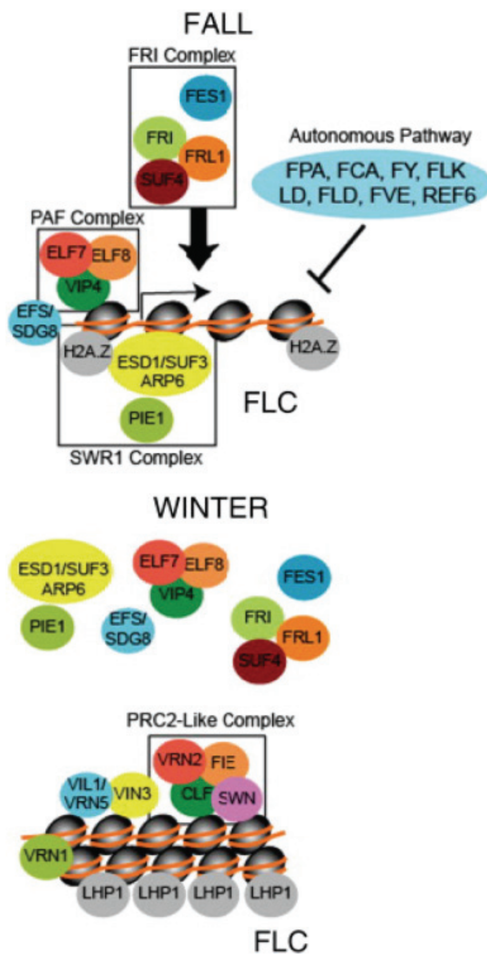


Fig. 8. Epigenetic model of vernalization response in *Arabidopsis thaliana* (from Schmitz and Amasino 2007).

The photoperiodic pathway

The day length (photoperiod) is the environmental cue that varies most systematically over seasons and years at a given geographic location. It is therefore perceived by living organisms and used to adjust their lifecycles to fit the seasonal changes in growth conditions. At higher latitudes the photoperiod varies more dramatic over seasons than at lower latitudes (Fig. 9). In temperate regions many plants flower in response to increasing day lengths in the spring in order to avoid damage by freezing temperatures in the winter. On the contrary, plants from the tropics flower when the day length decreases in the autumn to avoid extreme summer heat. Flowering in *A. thaliana* is accelerated by long days.

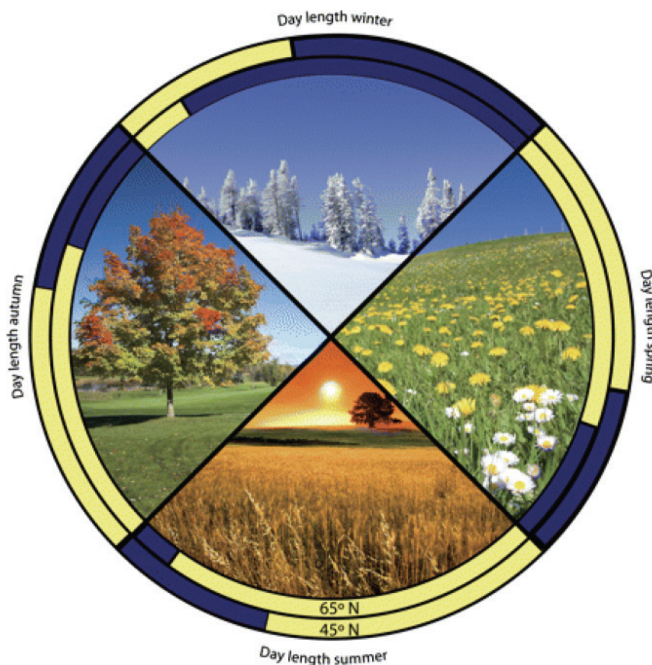


Fig. 9. Day length varies dramatically over season and latitudes. The figure depicts the day length at latitude 65°N and 45°N (from Lagercrantz 2009).

A key protein in photoperiodic sensing in *A. thaliana* is CO (Putterill et al. 1995). Expression of the *CO* gene is controlled by the circadian clock. Expression of *CO* peaks during the night in SDs and during the light period in LDs (Suarez-Lopez et al. 2001). The CO protein is degraded during the dark period and therefore it may function only when the days are long (Fig. 10). The expression peak of *CO* is broader in LDs than in SDs, due to the activity of FLAVIN-BINDING KELCH REPEAT, F-BOX (FKF1) that suppresses *CYCLING DOF FACTOR 1* (*CDF1*), a repressor of *CO* (Imaizumi et al. 2005). As a result CO protein

accumulates only during LDs and activates expression of the *FT* gene. The FT protein moves from the leaves to the shoot apical meristem (SAM), where it activates floral meristem identity genes and induce flowering (Corbesier et al. 2007; Jaeger et al. 2007; Mathieu et al. 2007). The stability of CO in light is controlled by a number of photoreceptors (Valverde et al. 2004). *PHYB* reduces the stability of CO protein in red light during the early day, and *PHYA* and *CRY2* stabilize CO protein in far-red and blue light, respectively, at the end of the day (Fig. 10). Transcription of *CO* is also controlled by several circadian clock output genes, e.g. *GIGANTEA* (*GI*), *FKF1* and *CDF1*. *CDF1* inhibits transcription of *CO* early in the day (Imaizumi et al. 2005) while *GI* and *FKF1* promote *CO* transcription and degradation of *CDF1* in the afternoon (Sawa et al. 2007) (Fig. 10). *GI* and *FKF1* form a complex which is stabilized in LDs by blue light and is a key component of photoperiodic flowering in *A. thaliana*. Moreover, *FKF1* acts as a blue light receptor in the control of *CO* transcription (Valverde et al. 2004; Sawa et al. 2007).

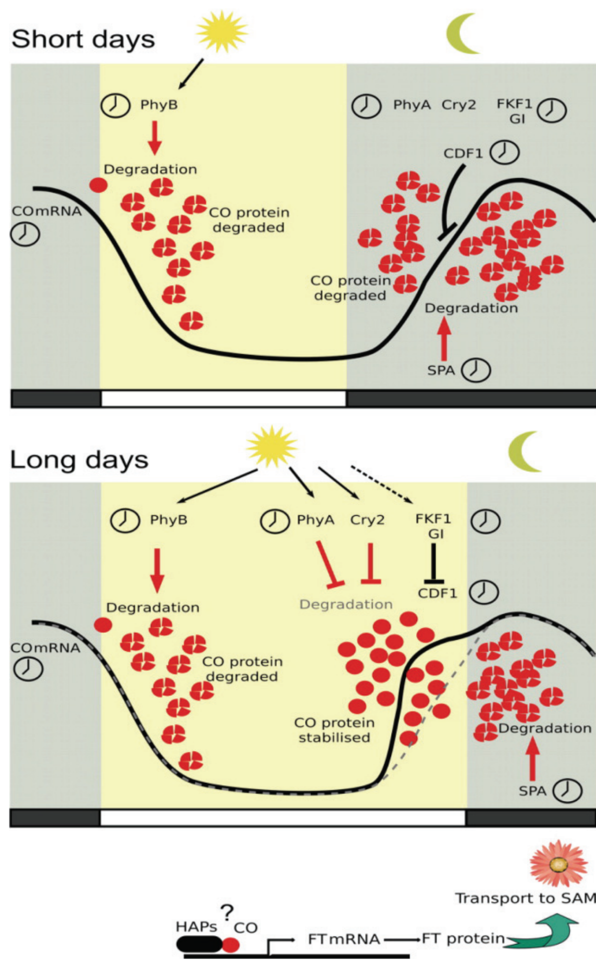


Fig. 10. Regulation of the *CONSTANS* gene at both the mRNA and protein levels (from Lagercrantz 2009).

Expression of *FT* in SAM induces early flowering (An et al. 2004) and the FT protein is a strong candidate for a mobile florigen signal (Corbesier et al. 2007; Jaeger et al. 2007; Mathieu et al. 2007). CO protein promotes transcription of *FT* in the leaves, and then FT protein moves towards the SAM. The FT protein forms a complex with FLOWERING LOCUS D (FD) in SAM and activates a number of floral meristem identity genes (Fig. 11) (Abe et al. 2005; Wigge et al. 2005; Turck et al. 2008).

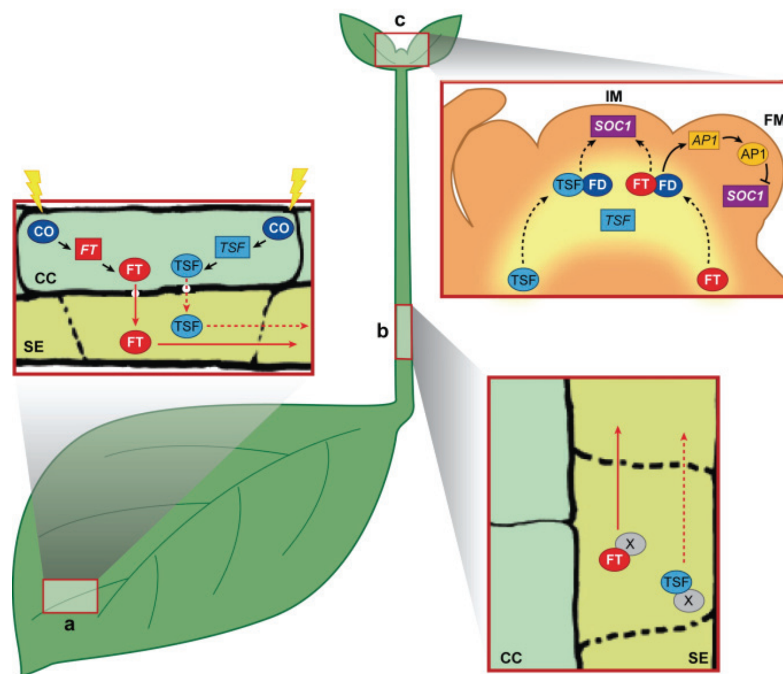


Fig. 11. FLOWERING LOCUS T (FT) as a systemic signal. Boxes, mRNA; circles, protein; solid black arrows, experimentally confirmed interconnection; dotted arrows, inferred interconnection (From Turck et al. 2008).
AP1 - APETALA 1
CC - companion cells
CO - CONSTANS
FD - FLOWERING LOCUS D
FM - floral meristem
IM - inflorescence meristem
SE - sieve elements
SOC1 - SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
TSF - TWIN SISTER OF FT

It has also been shown that the *GI*-regulated *microRNA172* promotes photoperiodic flowering through a *CO*-independent pathway (Jung et al. 2007). *microRNA172* controls the

expression of a number of *FT* repressor genes including *TARGET OF EAT 1* (*TOE1*) and its expression is higher in LDs than in SDs (Jung et al. 2007).

The link between the vernalization and the photoperiodic pathway occur at the *FLC* gene. *FLC* represses *FT* by binding directly to its promoter region (Helliwell et al. 2006). Vernalization suppresses *FLC* transcription and in turn *FT* can be activated by *CO* in LDs (Henderson et al. 2003).

Photoreceptors

Photoreceptors perceive and transfer light signals, play important roles in the photoperiodic response mechanism through regulation of CO stability, and are necessary for entraining the circadian clock to a 24 h cycle. Plants photoreceptors monitor the quality, quantity and direction of light. Photoreceptors include red (R)/far-red (FR) – absorbing phytochromes and the blue/UV-A – absorbing cryptochromes and phototropins (Chen et al. 2004; Jiao et al. 2007). Five phytochromes have been identified in *A. thaliana* (PHYA-E) (Sharrock and Quail 1989; Clack et al. 1994; Franklin et al. 2005). Three of them (PHYA-PHYC) are conserved among angiosperms (Mathews et al. 1995).

Phytochromes are synthesized in the inactive Pr form that absorbs red light, and are converted to the biologically active Pfr form. Pfr absorbs far-red light and is converted back to the inactive Pr form. Maximum light absorption of the phytochromes occurs at about 660 nm for the Pr form and at about 730 nm for the Pfr form. *PHYA* and *PHYB* play important roles in germination and seedling development (Nagatani et al. 1993; Reed et al. 1993; Shinomura et al. 1996; Franklin et al. 2007), whereas *PHYD* and *PHYE* are involved in internodal elongation and induction of flowering (Aukerman et al. 1997; Devlin et al. 1998; Devlin et al. 1999; Henning et al. 2002). Analyses of *phyc* mutants revealed that PHYC may function together with other phytochromes and cryptochromes in the control of plant development. *PHYC* acts redundantly with *PHYA* to regulate red light controlled hypocotyl growth and modulate *PHYB* function (Franklin et al. 2003).

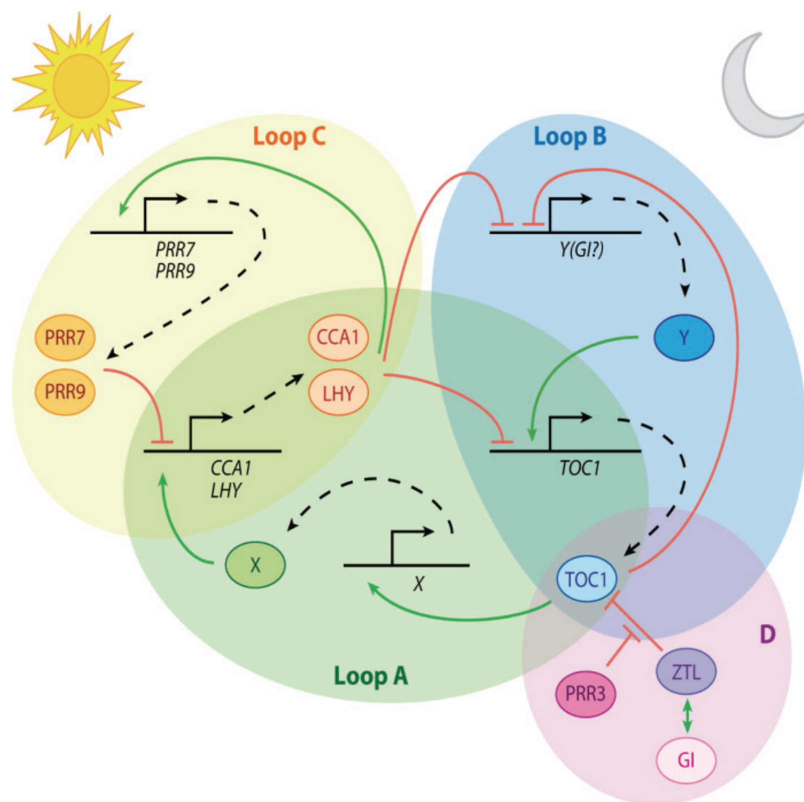
Three cryptochromes (CRY1-3), 2 phototropins (PHOT1-2), as well as ZEITLUPE (ZTL), FKF1, LOV KELCH PROTEIN 2 (LKP2) families mediate blue-light-induced (400-500 nm) responses in *A. thaliana* (Christie 2007; Li and Yang 2007; Jiao et al. 2007). Blue light photoreceptors promote entrainment of the circadian clock, seedling de-etiolation and flowering that are long-term responses, as well as stomatal opening and chloroplast movements which are rapid responses.

Light spectrum at northern latitudes

Not only annual photoperiod, but also light quality is different at higher latitudes (arctic and subarctic regions) compared to lower latitudes. The sun remains continuously above the horizon during the summer time at higher latitudes, and therefore the "end-of-day" period constitutes up to 20% of the day (Nilsen 1985). Light quality is influenced by the solar angle, which depends on the latitude and the time of the day. At higher latitudes or in the evening at moderate and lower latitudes, short wave blue light will be more scattered and refracted than long wave red light. Consequently, the spectrum will be rich in blue light and far-red light, since FR is little influenced by refraction. Thus, the period with light rich in blue and with low R/FR ratio last longer at higher latitudes than at moderate latitudes.

The circadian clock

Diurnal changes in light and temperature are caused by daily rotation of the earth around its axis. Many of the biological rhythms in most organisms are controlled by a circadian clock, an internal timer that keeps approximately 24 h periodicity. Circadian clocks control also seasonal processes such as flowering in plants and hibernation in mammals. Circadian rhythms are temperature compensated, i.e. they occur with the same periodicity across a range of different temperatures, and can be reset by changes in light or temperature levels. The plant circadian clock consists of three transcriptional feedback loops. Three genes form the first loop: *TIMING OF CAB EXPRESSION 1* (*TOC1*), the evening-phased clock regulated gene; *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) and *LATE ELONGATED HYPOCOTYL* (*LHY*). CCA1 and LHY are morning-phased transcription factors that negatively regulate expression of *TOC1* (Fig. 12) (Alabadi et al. 2001; Harmer 2009). The second loop comprises two evening-phased genes: *TOC1* and a yet unknown gene Y, which positively regulates expression of *TOC1*. Expression of Y is inhibited by TOC1, CCA1, and LHY (Locke et al. 2005). The third loop consists of the morning-phased genes *PSEUDO-RESPONSE REGULATOR 7* (*PRR7*), *PRR9*, *CCA1* and *LHY*. CCA1 and LHY promote the expression of *PRR7* and *PRR9* which in turn regulate *CCA1* and *LHY* expression negatively (Fig. 12) (Zeilinger et al. 2006; Locke et al. 2006; Harmer 2009).



Harmer SL. 2009.
Annu. Rev. Plant Biol. 60:357–77

Fig. 12. Model of the plant circadian clock (from Harmer 2009).
CCA1 - CIRCADIAN CLOCK ASSOCIATED 1
GI - GIGANTEA
LHY - LATE ELONGATED HYPOCOTYL
PRR3, *PRR7*, *PRR9 - PSEUDO-RESPONSE REGULATOR 3/7/9*
TOC1 - TIMING OF CAB EXPRESSION 1
ZTL - ZEITLUPE

The output from the circadian clock is also regulated at the posttranscriptional level. The ZTL protein, which contains the LOV (light, oxygen, or voltage) domain giving it ability to sense blue light (Imaizumi et al. 2003; Kim et al. 2007), degrades the TOC1 and PRR3 proteins (Mas et al. 2003; Kiba et al. 2007; Fujiwara et al. 2008). ZTL and GI interact with each other in a blue light-dependent manner and this interaction stabilizes and prevents degradation of both proteins in the dark (Fig. 12) (David et al. 2006; Kim et al. 2007). Balanced expression level of *GI*, *LHY* and *CCA1* accounts for the temperature compensation of the circadian clock (Gould et al. 2006), whereas the two clock components *PRR7* and *PRR9* integrate light and temperature signals (Salomé and McClung 2005).

Several circadian clock output genes control photoperiodic flowering through transcriptional regulation of *CO* in *A. thaliana*. The GI and FKF1 proteins follow the circadian expression pattern and FKF1 induce the LDs-specific daytime peak of *CO* transcripts, whereas GI is necessary for the *CO* transcription regardless of the photoperiod (Fig. 10) (Suárez-López et al. 2001; Imaizumi et al. 2003; Mizoguchi et al. 2005). Recent research have shown that interaction between *GI* and *TOC1* play an important role in photoperiodic control of flowering, however, TOC1 is not the only protein that negatively regulates *GI* expression (Ito et al. 2009). *toc1-2* mutants show an early flowering phenotype and *gi-2* mutation results in a late flowering phenotype. Moreover, *gi-2 toc1-2* double mutants are late flowering, since the GI-FKF1 complex is necessary for suppression of the floral repressor CDF1 (Fig. 10) in order to promote photoperiodic flowering (Ito et al. 2009).

It has recently been shown that *SHORT VEGETATIVE PHASE* (*SVP*), a MADS-box transcription factor that acts in a thermosensory pathway (Fig. 6), binds to *FT* suppressing its expression (Lee et al. 2007). *SVP* is partly regulated by the clock proteins CCA1 and LHY in the *GI*/*CO* independent manner (Fujiwara et al. 2008; Jang et al. 2009).

The gibberellin and autonomous pathways

Plants require not only environmental cues but also developmental signals in order to flower. Early study of autonomous pathway mutants revealed that these genes regulate *FLC* and act in parallel with the vernalization pathway (Koornneef et al. 1991). In the absence of the active *FRI* gene, autonomous pathway genes suppress expression of *FLC*. The autonomous pathway consists of genes *FCA*, *FY*, *FPA*, *FVE*, *LUMINIDEPENDENS* (*LD*), *FLOWERING LATE KH MOTIF* (*FLK*), and *FLOWERING LOCUS D* (*FLD*) (Marquardt et al. 2006). Although the autonomous pathway was first identified as a pathway that promote flowering through *FLC* repression, it is now well established that it also controls other targets in the *A. thaliana* genome (Bäurle et al. 2007; Veley and Michaels 2008). The FCA and FPA proteins play a broad role in development and in gene silencing through recognition of aberrant RNA and DNA methylation. This silencing process of target genes starts when low level of complementary siRNAs are present, making RNA aberrant (Bäurle et al. 2007). The work of Veley and Michaels (2008) demonstrates that autonomous pathway genes are crucial in silencing of some loci associated with high level of DNA methylation, but do not play crucial roles in genome-wide DNA methylation. This work has also shown that autonomous pathway

genes are crucial for growth and development independent of flowering in *A. thaliana* (Veley and Michaels 2008).

Gibberellins (GAs) play major roles in seed development, organ elongation and flowering time control (Yamaguchi 2008). The GAs function as mediators between environmental signals such as light and temperature, and developmental responses. The main target of GAs is a flower meristem identity gene *LEAFY* (*LFY*) (Blázquez and Weigel 2000). GA biosynthesis is required in order to initiate flowering under SDs conditions in *A. thaliana*. Moreover, cold treatment stimulates the expression of GA biosynthesis genes in *A. thaliana* seeds (Yamauchi et al. 2004) and breaks seed dormancy and induce germination.

Objectives

The objectives of this study were to characterize natural genetic variation that contributes to vernalization and photoperiod responses in natural populations of *A. thaliana*, investigate the relationship of these responses and climatic signatures at the site of origin of populations, and determine the geographic structure of genetic variation of *A. thaliana* at its northernmost global distribution range in Norway.

The variation in photoperiod during summer is very large at arctic as well as subarctic latitudes. Due to the influence of the North Atlantic Ocean, temperature and precipitation are rather similar along most of the coast. However, at a given latitude climatic transects are present due to distance from the ocean and coastal and inland populations are subjected to very different climates. To test whether these differential gradients in climate result in local adaptation we compared flowering time, vernalization sensitivity, vernalization saturation and photoperiodic sensitivity of coastal populations from both arctic and subarctic latitudes (62-68°N) with continental populations from subarctic latitudes (59-61°N). Association of sequence variation at the flowering time loci *FRI* and *FLC* with vernalization sensitivity, and association of *CO* sequence variation and transcript levels of flowering time genes with photoperiodic sensitivity, was investigated in addition.

In order to elucidate the evolutionary history of *A. thaliana* at its northernmost distribution boundary the genetic structure of natural populations in Norway was analyzed. *A. thaliana* in Norway is distributed in fragmented habitats and this will potentially lead to reduced gene flow and low genetic variation within populations relatively to regions with greater *A. thaliana* population density. Moreover, our strategy for sampling local populations focused on isolated locations with very limited impact by anthropogenic activity like agriculture. This sampling strategy should reduce bias in analyzing genetic diversity and increase our knowledge about demography and introduction history of *A. thaliana* at its northernmost range limit.

Material and methods

Plant material

Arabidopsis thaliana (L.) Heyhn is an annual self-fertilizing weed species which is distributed worldwide. It tolerates little competition and is preferentially growing on dryer lands. Populations growing in Norway (up to 68°N) are certainly at the northernmost distribution edge of the species. We have used populations of *A. thaliana* collected across latitudinal (59-68°N) and altitudinal (2-850 m a.s.l.) gradients in Norway (Fig. 13).

Fig. 13. *Arabidopsis thaliana* in Vega (65°N).

In order to determine vernalization responses, 27 populations were given vernalization treatments of 12, 9, 6, 3 and 0 weeks and 9 additional populations were exposed only to 12 weeks of vernalization. Screening of responses to 5 different photoperiods (8, 16, 19, 21 and 24 h) was performed in 10 of the populations (Fig. 14). Plant material from 31 populations was used in SNP genotyping in order to determine genetic structure of populations.

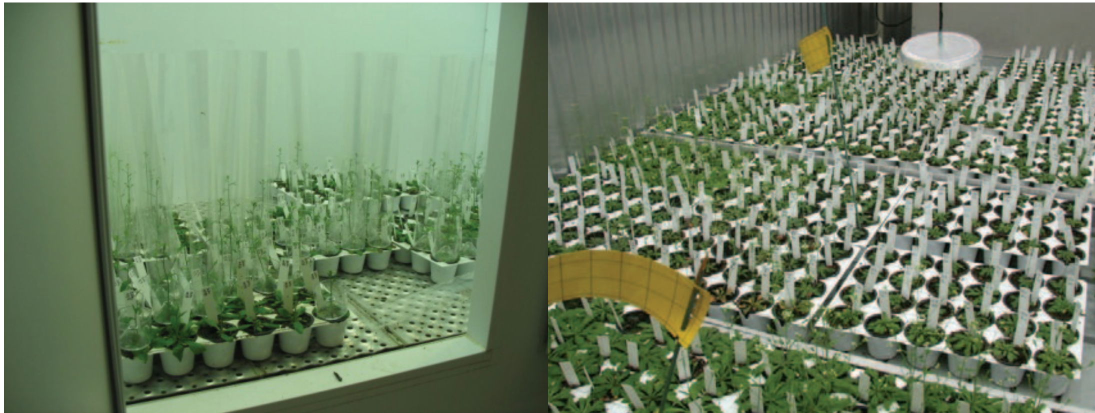


Fig. 14. Vernalization and photoperiod experiment at Norwegian University of Life Sciences.

Sequencing of candidate genes

Sequence variation of *FRI* (the promoter region and the first exon) and *FLC* (1.7 kb upstream of ATG, 1.7 kb of the first intron and the complete coding region) was determined in 29 local populations and in the laboratory strains Col-0 and L*er* in order to determine associations with vernalization sensitivity. Moreover, sequence variation was also determined in the promoter regions of *CRY2*, *CO* and *TOC1*, the coding regions of *PHYC*, *CO*, *SPINDLY* in 29 local populations, Col-0 and L*er*; and *CRY1* and *CRY2* in 14 populations and Col-0 in order to determine associations with photoperiodic sensitivity.

Real-time RT PCR

Transcript levels of *GI*, *FKF1*, *PHYA*, *CRY2*, *TOC1* and *CO* were analyzed at the end of the third day after plants were moved from vernalization to 8, 16 or 24 h photoperiods. For studying effects of SD vs. LD, transcript levels at the end of 8 and 16 h photoperiods were compared. For studying diurnal changes in transcript levels plants were harvested at the 1st, 9th, 13th, 17th and 20th h of the fifth day in 16 h photoperiod, and finally for studying transcript level changes over time, expression at the end of day 5 and 3 was compared. Transcript levels were calculated relative to *actin* using a comparative threshold cycle method (ΔCT method). Fold change in transcript levels over time and between different treatments was calculated using the $\Delta\Delta C_t$ method.

Statistical analyses

Regional differences in the climatic conditions at the site of origin of the populations were estimated by using mean January temperatures (T_{Jan}) as a proxy for winter temperature and winter duration, yearly precipitation (P_{Year}) as a proxy for distance from the ocean and the difference between January and July mean temperatures a measurement of the winter/summer temperature amplitude.

Population means for flowering time parameters for the 5 vernalization treatments as well as vernalization sensitivity and vernalization saturation were regressed against winter temperature, temperature amplitude and distance from the ocean at the site of origin of populations. Flowering time parameters for the 5 photoperiod treatments as well as photoperiodic sensitivity were regressed against winter temperature, distance from the ocean, altitude and latitude using individual data for each population. For association of *FLC* polymorphisms with vernalization sensitivity ANOVA was performed with coefficients that statistically describe a population's inferred ancestry as covariates (Korves et al. 2007; Samis et al. 2008). The same method was applied for association of polymorphism in *CO* with photoperiodic sensitivity. Simple linear regression and ANOVA analyses were performed using Minitab 15 Statistical Software (Minitab Inc. State College, PA, USA).

Molecular diversity indices and population genetic parameters were calculated using Arlequin version 3.1 (Excoffier et al. 2005). STRUCTURE 2.2.3 (Pritchard et al. 2000; Falush et al. 2003, accessed through the BIOPORTAL http://www.bioportal.uio.no/) was used to assign individuals to a number of putative clusters (K). Number of clusters was detected calculating the log probability of data between successive K clusters (Evanno et al. 2005) using STRUCTURE-SUM (Ehrich D 2006) in R 2.9.2 (R Development Core Team 2009). In order to determine isolation by distance, genetic distances (D_{SA}, Chakraborty and Jin 1993) were compared with geographical distances using a Mantel test (Mantel 1967).

Results and discussion

Vernalization and photoperiodic responses are associated with winter duration at the site of origin of the populations

Vernalization sensitivity and vernalization saturation were associated with distance from the ocean, winter temperature and winter/summer temperature amplitude (R^2=0.28, 0.30, 0.35, respectively, $P \leq 0.01$ for vernalization sensitivity, Fig. 4, Paper I; and R^2=0.16, 0.23, 0.18, respectively, $P \leq 0.05$ for vernalization saturation). Thus, lower vernalization sensitivity and longer time needed for saturation of the vernalization requirement were associated with proximity to the ocean, higher winter temperature and small temperature amplitude. These results indicate that climatic factors that are changing with distance from the ocean, such as winter duration put selection pressures on vernalization responses in *A. thaliana*. Natural selection favours populations with lower vernalization sensitivity in coastal regions, where winters are mild and short with unpredictable temperature fluctuations. Coastal populations needed longer time to fully saturate their vernalization requirement than continental populations. This indicates that populations from coastal regions use other mechanisms in order to determine proper timing for flowering in the spring, since unstable winter conditions could lead to premature flowering if populations were vernalization sensitive. Another mechanism could be increased sensitivity to photoperiods which is a more reliable indicator of the onset of spring. In inland regions with a stable continental climate, the rising temperatures in spring after long winters are sufficient signals for already induced populations to start growth and initiate flower development. In these regions selection favours populations which are highly vernalization sensitive and that need shorter time for vernalization saturation.

Photoperiodic sensitivity was highly associated with altitude, winter temperature and distance from the ocean (R^2=0.62, 0.85 and 0.91 respectively, $P<0.01$, Fig. 4, Paper II), but not with latitude. Thus, clinal variation in photoperiodic sensitivity was evidently following the gradients distance from the ocean, coastal populations being more sensitive to photoperiod than inland populations. This variation is likely a result of selection and is associated with climatic factors such as winter duration that vary depending on distance from the ocean. Although photoperiod varies along with latitude, climatic transects are present at the same latitude in Norway. Evidently, responses to photoperiod are very complex and

selection may create adaptation of flowering time to different photoperiods depending on the length of the winter.

The lack of apparent cline in vernalization sensitivity in *A. thaliana* in our (Fig. 4, Paper I) and other studies (e.g. Stinchcombe et al. 2004) could be due to the demographic history of the species. Since *A. thaliana* easily migrates as a weed, part of its genetic structure resulting from historical expansion of the species might be confounded by long distance human-mediated gene flow. Therefore, these populations often represent a mixture of native and newly introduced individuals (Nordborg et al. 2005; Beck et al. 2008; Paper III). Moreover, estimation of climatic parameters may be biased and incapable of capturing all differences in microclimatic factors that likely account for phenotypic variation within regions in our study. However, since strong association between photoperiodic sensitivity and distance from the ocean was found in a subsample of these populations (Fig. 4, Paper II), alternative explanations are more likely. The absence of a clear geographic gradient in vernalization sensitivity could be due to lack of tight genetic control of the vernalization response in these populations as compared to the control of photoperiodic response. Increasing photoperiod following cold periods i.e. spring conditions appear to be a more reliable factor for the plants to sense favourable conditions compared to duration of the winter that may vary greatly from year to year. Moreover, a longitudinal cline in photoperiodic plasticity of *A. thaliana* has been observed earlier on a larger geographic scale (Samis et al. 2008). Our results are consistent with these findings and confirm that environmental factors determining selection on flowering time in *A. thaliana* also vary longitudinally. Longitudinal clines in phenological traits have also been observed in other plant species (Chmura and Rozkowski 2002). These clines were explained by oceanic current oscillations, which determine seasonal onset dependent on the distance from the ocean. Our study confirms these findings and suggests that such clines exist on much smaller geographic scales.

FLC and CO sequence variation contributes to vernalization and photoperiodic sensitivity, respectively

Several studies have demonstrated that sequence variation at the *FRI* and *FLC* genes are affecting flowering time in natural populations of *A. thaliana* (Johanson et al. 2000; Le Corre et al. 2002; Gazzani et al. 2003; Michaels et al. 2003; Shindo et al. 2005; Shindo et al. 2006). Our study suggests that variation at the candidate locus *FLC* influence vernalization

sensitivity in *A. thaliana* with *FLC2* haplotypes being more sensitive to vernalization than *FLC1* haplotypes (Fig. 3, Paper I, P=0.002). The two novel *FLC* polymorphisms (DEL1 and SNP5) were significantly associated with vernalization sensitivity in the *A. thaliana* populations (Table 4, Paper I; P=0.038) while associations with the remaining polymorphisms were confounded by population structure. Therefore, functional differentiation of *FLC* haplotypes distinguished by these 2 polymorphisms is likely not due to population structure artefacts (Fig. 3, Paper I). These polymorphisms in the flanking region may have regulatory function in *FLC* expression similarly as demonstrated for promoter and the first *FLC* intron regions (Sheldon et al. 2002). Moreover, Shindo *et al*., (2006) shown that allelic variation in *FLC cis*-regulatory sequences might account for vernalization response through epigenetic maintenance of *FLC* chromatin silencing. Therefore, variation in vernalization sensitivity may be mediated by variation in silencing of *FLC* chromatin, however this hypothesis needs further study. Our work suggests that *FLC* plays an important role in natural variation in vernalization sensitivity.

Longitudinal differentiation at neutral loci has been demonstrated previously for Eurasian populations of *A. thaliana* (Beck et al. 2008) therefore candidate locus variation may be biased by population structure (Zhao et al. 2007). This should be taken into account when inference of causative effects of a locus is made. The longitudinal trend in *PHYC* variation observed by Samis et al. (2008) is not present in our dataset. However, sequence variation at *CO* appears to show a geographical pattern and SNP11 remained significantly associated with photoperiodic sensitivity after correction for population structure (P=0.049). Marginal significance of the association may be due to the small sample used (n=10). None of the polymorphisms present in the other genes demonstrated association with photoperiodic sensitivity.

Variation in expression of CRY2, GI and TOC1 contributes to photoperiodic phenotypes

Significant differences in fold change of transcript levels in SDs compared to LDs were found. *CRY2*, *TOC1* and *GI* transcript levels were much lower in coastal populations than in inland populations in SDs vs. LDs (Fig. 7, Paper II). Change in *CRY2*, *TOC1* and *GI* transcript levels in SDs compared to LDs was associated with photoperiodic sensitivity (P=0.006, 0.004, 0.024, respectively). Therefore, variation in expression of these genes may account for the photoperiodic sensitivity of these populations. The expression levels of *CRY2*, *TOC1* and *GI* may already be high in SDs in inland populations and therefore their response

to LDs is much smaller than in coastal populations. Thus, expression of these genes may reach the needed threshold already within 8 h photoperiod in inland populations, while in coastal populations the same threshold is reached when the day length is longer, probably near 16 h. Although, inland populations reach higher expression level at shorter days under laboratory conditions, their high vernalization sensitivity (Paper I) is likely sufficient to assure timing of flowering in proper seasons. As a result, tighter photoperiodic control in coastal populations might be a necessary adaptation for sensing the onset of spring in regions with relatively mild and unpredictable winter climates as opposed to continental climates with more stable winters.

Mimicking environment conditions found in natural habitats in the controlled growth rooms is very difficult, thus it is very likely that these populations respond to much longer photoperiods than what they did in this study. Furthermore, although we used a much lower ambient temperature (16°C), which reflects the natural growth conditions in Norway, than the standard temperature (21°C) used in most of the *A. thaliana* studies (Hoffmann 2002), these conditions may still be very artificial. Especially in the autumn short days are often accompanied by very cold weather. Therefore, either growing populations in a common garden experiment or using complete factorial photoperiod and temperature experiments would be useful for dissecting photoperiodic responses in these populations.

Very few studies of geographical trends in photoperiodic responses of natural *A. thaliana* populations have been performed (Samis et al. 2008). Our study is significant contribution to the current knowledge by providing detailed and fine-scale analyses of associations between climatic factors at site of origin and flowering time responses to photoperiod, as well as variation at the transcript levels of genes involved in genetic regulation of flowering time. Thus adaptation of flowering time of *A. thaliana* needs further investigations.

Population structure and genetic diversity

Gene diversity was negatively correlated with latitude ($r=-0.58$, $p\leq0.01$; Fig. 1, Paper III) indicating depletion in genetic diversity among Northern Norwegian populations of *A. thaliana* as compared to Southern populations. This depletion of genetic variation in marginal populations is probably due to relatively few founder individuals and population bottlenecks (Hewitt 2000). Our findings suggest that much of the genetic variability of *A. thaliana* might have been lost during the northward expansion of the species. Bottlenecks and relatively few

founder individuals will create high levels of LD (Fig. S2, Paper III). In addition, low estimated effective population size (Fig. 5, Paper III) and limited gene flow due to extremely isolated habitat result in increased genetic drift that genetically differentiates Northern populations from Southern populations. So far this prediction has been confirmed for many species at the margin of the species distribution (Eckert et al. 2008) and except for a few studies (Todokoro et al. 1995; Kuittinen et al. 1997), no reduction in genetic variation towards the range margin of *A. thaliana* has been found (Stenøien et al. 2005; Lundemo et al. 2009). Our study is the first to demonstrate genetic diversity reduction in *A. thaliana* in northernmost distribution range likely caused by geographic expansion of the species towards the range margin.

The number of clusters was detected in STRUCTURE at K=5, however due to lack of significance of the similarity coefficient, analyses of genetic structure were repeated after omitting populations from Central Norway that seem to represent admixed individuals originating from several founders. Consequently, the number of clusters was detected at K=3 and the similarity coefficient was significant (Fig. 3, Paper III). These results indicate that the genetic structure of Norwegian *A. thaliana* populations is complex. *A. thaliana* is often found within human-disturbed areas in its native worldwide distribution range (Beck et al. 2008). In Norway, its highly fragmented habitat results in extremely isolated populations with limited gene flow and human influence (Naturhistoriska riksmuseet 1998; Fremstad 1998) and these populations are believed to be a result of postglacial colonization. In spite of our sampling strategy focussing on extremely isolated populations from sites with possibly no major impact by human activity, we seem not to have been able to avoid recently introduced populations. The STRUCTURE analysis infers that Central populations represent mixture of genotypes from all geographic regions (Fig. 3, Paper III) and do not form one consistent cluster (Fig. 4, Paper III). Therefore, the cryptic population structure, particularly in this group seems to be a result of both natural colonization and human-mediated seed dispersal. This human-mediated gene flow may cause departure from historical stepping-stone model of genetic structure and isolation by distance (Beck et al. 2008). Moreover, LD decay in Central populations is significant (Fig. S2, Paper III), which indicates that recent admixture events and outcrossing with new genotypes may be related to their location along trade routes, but also present-day gene flow. Moreover, the highest effective population size was found in this group (Fig. 5, Paper III), which indicates more individuals contributing gametes to the next generation and possibly more founders in the Southern than in the Northern populations. Since historical times this geographic region is a traffical interchange in Norway and an

extensively agriculturally utilized region. Thus, human mediated transport of *A. thaliana* seeds to this region is highly likely.

The STRUCTURE analysis revealed that Northern populations appear remote and isolated from the remaining populations (Fig. 3, Paper III). Moreover, the highest level of LD was found among these populations (Table 3, Fig. S2, Paper III) which confirms that Northern populations are a homogenous group that has experienced less recombination with immigrant genotypes than Central populations. Moreover, this group of populations as well as Northwestern Coastal populations has the lowest effective population size (Fig. 5, Paper III) which indicates that very few individuals contribute gametes to the next generation in the North. In these small and isolated populations genetic drift will have a large effect and differentiate them genetically from larger populations further South in Norway. The present results suggest that there has been enough time for genetic drift and mutations to differentiate populations from the North from the remaining Norwegian populations while gene flow associated with human activity in Central Norway has resulted in haplotype sharing and cryptic genetic structure among these populations. Moreover, Northern populations originate from undisturbed regions where the chances for spreading by agricultural activity are limited.

Latitudinal and longitudinal clines in flowering time in *A. thaliana* have rarely been demonstrated (Stinchcombe et al. 2004; Samis et al. 2008). The composition of the population samples analyzed may influence the outcome of such analyses. Eurasian *A. thaliana* populations from the native distribution range of the species were studied in the above cited studies and no such cline was detected when accessions from both Europe and North America were studied (Shindo et al. 2005). Northern American populations are believed to be introduced recently (Bergelson et al. 1998) in contrary to Eurasian populations resulting from post-glacial colonization (Hoffmann 2002). Therefore, detecting geographical gradients in mixed population samples of native and introduced individuals having different demographic histories may be inclined. Our study on a regional scale in a restricted part of the native distribution of *A. thaliana* has shown that vernalization and photoperiodic responses are associated with distance from the ocean. Furthermore, latitudinal cline in hypocotyl elongation in response to light quality was previously demonstrated in a sample of Norwegian *A. thaliana* populations (Stenøien et al. 2002). These results taken together illuminate the importance of fine-scale regional analyses using populations sharing common demographic histories. Moreover, interacting effects of photoperiod and lower ambient temperature as being used in our study have the potential to reveal phenotypic responses involving genetic regulations that so far has gone undetected in *A. thaliana*.

Conclusions and future perspectives

Understanding the mechanisms underlying flowering time in model species is of great importance in the light of anticipated climate changes, which will have great impact on ecosystems worldwide. The isolation of QTLs and genes in *Arabidopsis* and their homologues in crops and wild plants allow identification of these mechanisms at the species level and improve our understanding of evolutionary forces shaping the genetic variation affecting plant adaptation to diverse environments. It is also of great importance to understand which abiotic factors exert selection pressure on the phenotypic trait and the underlying polymorphism that ensure plant adaptation, as well as to understand the mechanisms of this adaptation. In crops, adaptation to moderate changes in climate, like changes in temperature, season length and changes in water availability can be achieved by selecting cultivars carrying alleles responsible for appropriate flowering time. Knowledge gained from studying natural variation in *A. thaliana* can be used in breeding programs to successfully select or manipulate life cycle duration to maximize the range of environments in which crops will grow and give high yields.

Genetic diversity is one of the major components of species diversity. A species that harbor a great degree of genetic diversity will have more genotypes from which to choose the most superior alleles. Therefore, the monitoring of species diversity is an important task in studies of the ability of populations to meet predicted climate changes by adaptation.

Because of their sessile nature, plants have evolved numerous sensing systems to monitor the changing environment. The information obtained from external cues and its transmission to newly developing organs is essential for their adaptation and survival. The timing of flowering within a season is mainly determined by responses to temperature and photoperiod. Therefore, identifying genes/alleles and mechanisms of the vernalization and photoperiodic pathways fill some key knowledge gaps in current understanding of the mechanisms of adaptation to changing environments.

References

Abbott RJ, Gomes MF (1989) Population genetic structure and outcrossing rate of *Arabidopsis thaliana*. Heredity 42: 411-418.

Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A et al. (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309: 1052-1056.

Adams S, Allen T, Whitelam GC (2009) Interaction between the light quality and flowering time pathways in *Arabidopsis*. Plant Journal 60: 257-267.

Alabadi D, Oyama T, Yanovsky MJ, Harmon FG, Mas P, Kay SA (2001) Reciprocal regulation between *TOC1* and *LHY*/*CCA1* within the *Arabidopsis* circadian clock. Science 293: 880-83.

Alonso-Blanco C, Mendez-Vigo B, Koornneef M (2005) From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. International Journal of Developmental Biology 49: 717-732.

Al-Shehbaz IA, O'Kane SL (2002) Taxonomy and phylogeny of *Arabidopsis* (*Brassicaceae*). In: The Arabidopsis Book doi: 10.1199/tab.0001.

An H, Roussot C, Suarez-Lopez P, Corbesier L, Vincent C et al. (2004) CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of *Arabidopsis*. Development 131: 3615-3626.

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M et al. (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLOS Genetics 1:e60.

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

Ausubel FM (2000) *Arabidopsis* genome: A milestone in plant biology. Plant Physiology 124: 1451-1454.

Bakker EG, Stahl EA, Toomajian C, Nordborg M, Kreitman M et al. (2006) Distribution of genetic variation within and among local populations of *Arabidopsis thaliana* over its species range. Molecular Ecology 15: 1405-1418.

Balasubramanian S, Sureshkumar S, Agrawal M, Michael TP, Wessinger C et al. (2006) The *PHYTOCHROME C* photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. Nature Genetics 38: 711-715.

Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, Dean C (2004) Vernalization requires epigenetic silencing of *FLC* by histone methylation. Nature 427: 164-167.

Bechsgaard JS, Castric V, Charlesworth D, Vekemans X, Schierup MH (2006) The transition to self-compatibility in *Arabidopsis thaliana* and evolution within S haplotypes over 10 Myr. Molecular Biology and Evolution 23: 1741-1750.

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

Benfey PN, Mitchell-Olds T (2008) From genotype to phenotype: systems biology meets natural variation. Science 320: 495-497.

Bentsink L, Jowett J, Hanhart CJ, Koornneef M (2006) Cloning of *DOG1*, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. PNAS USA 103: 17042-17047.

Bergelson J, Stahl E, Dudek S, Kreitman M (1998) Genetic variation within and among populations of *Arabidopsis thaliana*. Genetics 148: 1311-1323.

Berr A, Xu L, Gao J, Cognat V, Steinmetz A et al. (2009) *SET DOMAIN GROUP25* encodes a histone methyltransferase and is involved in *FLOWERING LOCUS C* activation and repression of flowering. Plant Physiology 151: 1476-1485.

Blázquez MA, Weigel D (2000) Integration of floral inductive signals in *Arabidopsis*. Nature 404: 889-92.

Bäurle I, Smith L, Baulcombe DC, Dean C (2007) Widespread role for the flowering-time regulators *FCA* and *FPA* in RNA-mediated chromatin silencing. Science 318: 109-112.

Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S et al. (2006) *CO*/*FT* regulatory module controls timing of flowering and seasonal growth cessation in trees. Science 312: 1040-3.

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between the *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. PNAS USA 101: 15670-15675.

Chakraborty R, Jin L (1993) Determination of relatedness between individuals using DNA fingerprinting. Human Biology 65: 875-895.

Chen M, Chory J, Fankhauser C (2004) Light signal transduction in higher plants. Annual Review of Genetics 38: 87-117.

Chmura DJ, Rozkowski R (2002) Variability of beech provenances in spring and autumn phenology. Silvae Genetica 51: 123-127.

Christie JM (2007) Phototropin blue-light receptors. Annual Review of Plant Biology 58: 21-45.

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

Clack T, Mathews S, Sharrock RA (1994) The phytochrome apoprotein family in Arabidopsis is encoded by five genes: the sequences and expression of *PHYD* and *PHYE*. Plant Molecular Biology 25: 413-427.

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

Comes HP, Kadereit JW (1998) The effects of Quaternary climatic changes on plant distribution and evolution. Trends in Plant Science 3: 432-438.

Corbesier L, Vincent C, Jang S, Fornara F, Fan Q et al. (2007) FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. Science 316: 1030-1033.

David KM, Armbruster U, Tama N, Putterill J (2006) *Arabidopsis* GIGANTEA

protein is posttranscriptionally regulated by light and dark. FEBS Letters 580: 1193-97.

De Lucia F, Crevillena P, Jones AME, Greb T, Dean C (2008) A PHD-Polycomb repressive complex 2 triggers the epigenetic silencing of *FLC* during vernalization. PNAS USA 105: 16831-16836

Devlin PF, Patel SR, Whitelam GC (1998) Phytochrome E influences internode elongation and flowering time in *Arabidopsis*. Plant Cell 10: 1479-1487.

Devlin PF, Robson PRH, Patel SR, Goosey L, Sharrock R, Whitelam GC (1999) Phytochrome D acts in the shade-avoidance syndrome in *Arabidopsis* by controlling elongation growth and flowering time. Plant Physiology 119: 909-915.

Doyle MR, Amasino RM (2009) A single amino acid change in the Enhancer of Zeste ortholog *CURLY LEAF* results in vernalization independent, rapid flowering in *Arabidopsis*. Plant Physiology 151: 1688-1697.

Doyle MR, Bizzell CM, Keller MR, Michaels SD, Song J, Noh YS, Amasino RM (2005) HUA2 is required for the expression of floral repressors *in Arabidopsis thaliana*. Plant Journal 41: 376-385.

Dubcovsky J, Loukoianov A, Fu D, Valarik M, Sanchez A, Yan L (2006) Effect of photoperiod on the regulation of wheat vernalization genes *VRN1* and *VRN2*. Plant Molecular Biology 60: 469-480.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Molecular Ecology 17: 1170-1188.

Ehrich D (2006) AFLPdat: a collection of R functions for convenient handling of AFLP data. Molecular Ecology Notes 6: 603-604.

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

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

Excoffier L, Laval G, Schneider S (2005) Arlequin version 3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47-50.

Falush D, Stephens M, Pritchard J (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Federspiel N (2000) Deciphering a weed: genomic sequencing of *Arabidopsis*. Plant Physiology 124: 1456-1459.

Filiault DL, Wessinger CA, Dinneny JR, Lutes J, Borevitz JO et al. (2008) Amino acid polymorphisms in *Arabidopsis* phytochrome B cause differential responses to light. PNAS USA 105: 3157-3162.

Fjellheim S, Rognli O (2005) Molecular diversity of local Norwegian meadow fescue (*Festuca pratensis* Huds.) populations and Nordic cultivars: consequences for management and utilisation. Theoretical and Applied Genetics 111: 640-650.

François O, Blum MGB, Jakobsson M, Rosenberg NA (2008) Demographic

history of European populations of *Arabidopsis thaliana*. PLOS Genetics 4: e100075.

Franklin KA, Davis SJ, Stoddart WM, Vierstra RD, Whitelam GC (2003) Mutant analyses define multiple roles for phytochrome C in *Arabidopsis thaliana* photomorphogenesis. The Plant Cell 15: 1981-1989.

Franklin KA, Larner VS, Whitealm GC (2005) The signal transducing photoreceptors of plants. International Journal of Developmental Biology 49: 653-664.

Franklin KA, Whitelam G (2007) Red:far-red ratio perception and shade avoidance. In Light and Plant Development. ed. Whitelam G, Halliday K. pp. 211-34.

Fremstad E (1997) Vegetation types in Norway (Vegetasjonstyper i Norge). Norsk Institutt for Naturforskning Temahefte 12: 1-279.

Fujiwara S, Oda A, Yoshida R, Niinuma K, Miyata K et al. (2008) Circadian clock proteins LHY and CCA1 regulate SVP protein accumulation to control flowering in *Arabidopsis*. The Plant Cell 20: 2960-2971.

Gazzani S, Gendall AR, Lister C, Dean C (2003) Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. Plant Physiology 132: 1107-14.

Gendall AR, Levy YY, Wilson A, Dean C (2001) The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. Cell 107: 525-535.

Gould PD, Locke JC, Larue C, Southern MM, Davis SJ et al. (2006) The molecular basis of temperature compensation in the *Arabidopsis* circadian clock. Plant Cell 18: 1177-87.

Greb T, Mylne JS, Crevillen P, Geraldo N, An H et al. (2007) The PHD finger protein VRN5 functions in the epigenetic silencing of *Arabidopsis FLC*. Current Biology 17: 73-8.

Gu XY, Kianian SF, Hareland GA, Hoffer BL, Foley ME (2005) Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). Theoretical and Applied Genetics 110: 1108-1118.

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

Harmer SL (2009) The circadian system in higher plants. Annual Review of Plant Biology 60: 357-77.

Helliwell CA, Wood CC, Robertson M, Peacock WJ, Dennis ES (2006) The *Arabidopsis* FLC protein interacts directly in vivo with *SOC1* and *FT* chromatin and is part of a high-molecular-weight protein complex. The Plant Journal 46: 183-192.

Henderson IR, Shindo C, Dean C (2003) The need for winter in the switch to flowering. Annual Review of Genetics 37: 371-392.

Hennig L, Stoddart WM, Dieterle M, Whitelam GC, Schafer E (2002) Phytochrome E controls light-induced germination of *Arabidopsis*. Plant Physiology 128: 194-200.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature 405: 907-913.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in

divergence and speciation. Biological Journal of the Linnean Society 58: 247-276.

Hoffmann MH (2002) Biogeography of *Arabidopsis thaliana* (L.) Heynh. (*Brassicaceae*). Journal of Biogeography 29: 125-134.

Hori K, Sato K, Takeda K (2007) Detection of seed dormancy QTL in multiple mapping populations derived from crosses involving novel barley germplasm. Theoretical and Applied Genetics 115: 869-876.

Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA (2005). FKF1 F-box protein mediates cyclic degradation of a repressor of *CONSTANS* in *Arabidopsis*. Science 309: 293-297.

Imaizumi T, Tran HG, Swartz TE, Briggs WR, Kay SA (2003) FKF1 is essential for photoperiodic-specific light signaling in *Arabidopsis*. Nature 426: 302-6.

Imtiaz M, Ogbonnaya FC, Oman J, van Ginkel M (2008) Characterization of quantitative trait loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines. Genetics 178: 1725-1736.

Ito S, Kawamura H, Niwa Y, Nakamichi N, Yamashino T, Mizuno T (2009) A genetic study of the *Arabidopsis* circadian clock with reference to the *TIMING OF CAB EXPRESSION 1* (*TOC1*) gene. Plant Cell Physiology 50: 290-303.

Jaeger KE, Wigge PA (2007) FT protein acts as a long-range signal in *Arabidopsis*. Current Biology 17: 1050-1054.

Jang S, Torti S, Coupland G (2009). Genetic and spatial interactions between *FT*, *TSF* and *SVP* during the early stages of floral induction in *Arabidopsis*. Plant 60: 614-25.

Jiao Y, Lau OS, Deng XW (2007) Light-regulated transcriptional networks in higher plants. Nature Reviews Genetics 8: 217-230.

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

Jung JH, Seo YH, Seo PJ, Reyes JL, Yun J et al. (2007) The *GIGANTEA*-regulated microRNA172 mediates photoperiodic flowering independent of *CONSTANS* in *Arabidopsis*. The Plant Cell 19: 2736-2748.

Kiba T, Henriques R, Sakakibara H, Chua NH (2007) Targeted degradation of PSEUDO-RESPONSE REGULATOR5 by an SCFZTL complex regulates clock function and photomorphogenesis in *Arabidopsis thaliana*. Plant Cell 19: 2516-30.

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM et al. (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics 39: 1151-1155.

Kim WY, Fujiwara S, Suh SS, Kim J, Kim Y et al. (2007) ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. Nature 449: 356-60.

Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T et al. (2002) *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day condition. Plant Cell Physiology 43: 1096-1105.

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

Koornneef M, Hanhart CJ, Van der Veen JH (1991) A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. Molecular and General Genetics 229: 57-66.

Korves TM, Schmid KJ, Caicedo AL, Mays C, Stinchcombe JR et al. (2007) Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. American Naturalist 169: E141-E157.

Kruskopf-Osterberg M, Shavorskaya O, Lascoux M, Lagercrantz U (2002) Naturally occurring indel variation in the *Brassica nigra COL1* gene is associated with variation in flowering time. Genetics 161: 299-306.

Kuittinen H, Sillanpãã MJ, Savolainen O (1997) Genetic basis of adaptation: flowering time in *Arabidopsis thaliana*. Theoretical and Applied Genetics 95: 573-583.

Lagercrantz U (2009) At the end of the day: a common molecular mechanism for photoperiod responses in plants? Journal of Experimental Botany 60: 2501-2515.

Laibach F (1943) *Arabidopsis thaliana* (L.) Heynh. als object fur genetische und entwicklungsphysiologische untersuchungen. Botanic Archives 44: 439-455.

Le Corre V (2005) Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: Comparison with markers and traits. Molecular Ecology 14: 4181-4192.

Le Corre V, Roux F, Reboud X (2002) DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. Molecular Biology and Evolution 19: 1261-1271.

Lee JH, Yoo SJ, Park SH, Hwang I, Lee JS, Ahn JH (2007) Role of *SVP* in the control of flowering time by ambient temperature in *Arabidopsis*. Genes and Development 21: 397-402.

Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C (2002) Multiple roles of *Arabidopsis VRN1* in vernalization and flowering time control. Science 297: 243-246.

Li QH, Yang HQ (2007) Cryptochrome signaling in plants. Photochemistry and Photobiology 83: 94-101

Lin SY, Sasaki T, Yano M (1998) Mapping quantitative trait loci controlling seed dormancy and heading date in rice, *Oryza sativa* L., using backcross inbred lines. Theoretical and Applied Genetics 96: 997-1003.

Liu C, Thong Z, Yu H (2009) Coming into bloom: the specification of floral meristems. Development 136: 3379-3391.

Locke JC, Kozma-Bognar L, Gould PD, Feher B, Kevei E et al. (2006) Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana*. Molecular Systems Biology 2: 59.

Locke JC, Southern MM, Kozma-Bognar L, Hibberd V, Brown PE et al. (2005) Extension of a genetic network model by iterative experimentation and mathematical analysis. Molecular Systems Biology 1:2005.0013.

Lundemo S, Falahati-Anbaran M, Stenøien HK (2009) Seed banks cause elevated generation times and effective population sizes of *Arabidopsis thaliana* in northern Europe. Molecular Ecology 18: 2798-811.

Maloof JN, Borevitz JO, Dabi T, Lutes J, Nehring RB et al. (2001) Natural variation

in light sensitivity of *Arabidopsis*. Nature Genetics 29: 441-446.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research 27: 209-220.

Marquardt S, Boss PK, Hadfield J, Dean C (2006) Additional targets of the *Arabidopsis* autonomous pathway members, *FCA* and *FY*. Journal of Experimental Botany 57: 3379-3386.

Mas P, Kim WY, Somers DE, Kay SA (2003) Targeted degradation of TOC1 by ZTL modulates circadian function in *Arabidopsis thaliana*. Nature 426: 567-70.

Mathews S, Lavin M, Sharrock RA (1995) Evolution of the phytochrome gene family and its utility for phylogenetic analyses of angiosperms. Annals of the Missouri Botanic Garden 82: 296-321.

Mathieu J, Warthmann N, Kuttner F, Schmid M (2007) Export of FT protein from phloem companion cells is sufficient for floral induction in *Arabidopsis*. Current Biology 17:1055-60.

Michaels SD, Amasino RM (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 11: 949-956.

Michaels SD, He Y, Scortecci KC, Amasino RM (2003) Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. PNAS USA 100: 10102-10107.

Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. Nature 441: 947-952.

Mizoguchi T, Wright L, Fujiwaraa S, Cremer F, Lee K et al. (2005) Distinct roles of *GIGANTEA* in promoting flowering and regulating circadian rhythms in *Arabidopsis*. Plant Cell 17: 2255-2270.

Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. PNAS USA 102: 2442-2447.

Mylne JS, Barrett L, Tessadori F, Mesnage S, Johnson L et al. (2006) *LHP1*, the *Arabidopsis* homologue of *HETEROCHROMATIN PROTEIN1*, is required for epigenetic silencing of *FLC*. PNAS USA 103: 5012-5017.

Nagatani A, Reed JW, Chory J (1993) Isolation and initial characterization of *Arabidopsis* mutants that are deficient in phytochrome A. Plant Physiology 102: 269-277.

Napp-Zinn K (1987) Vernalization: Environmental and genetic regulation. In: Manipulation of Flowering, ed. JG Atherton, pp. 123-32. London: Butterworths.

Nilsen J (1985). Light climate in northern areas. In: Å. Kaurin, O. Junttila and J. Nilsen, Editors, Plant Production in North, Norwegian University Press pp. 62-72.

Nordborg M, Borevitz JO, Bergelson J, Berry CC , Chory J et al. (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics 30: 190-193.

Nordborg M, Hu T, Ishino Y, Jhaveri J, Toomajian C et al. (2005) The pattern of polymorphism in *Arabidopsis thaliana*. PLOS Biol 3: e196.

Pamilo P, Savolainen O (1999) Post-glacial colonization, drift, local selection and conservation value of populations: a

northern perspective. Hereditas 130: 229-238.

Petit RJ, Aguinagalde I, De Beaulieu JL, Bittkau C, Brewer S et al. (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. Science 300: 1563-1565.

Pritchard J, Rosenberg N (1999). Use of unlinked genetic markers to detect population stratification in association studies. The American Journal of Human Genetics 65: 220-228.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Putterill J, Robson F, Lee K, Simon R, Coupland G (1995) The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80: 847-57.

R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reed JW, Nagpal P, Poole DS, Furuya M, Chory J (1993) Mutations in the gene for the red-far-red light receptor phytochrome B alter cell elongation and physiological responses throughout *Arabidopsis* development. Plant Cell 5: 147-157.

Salomé PA, McClung CR (2005) *PSEUDO-RESPONSE REGULATOR 7* and *9* are partially redundant genes essential for the temperature responsiveness of the *Arabidopsis* circadian clock. Plant Cell 17: 791-803.

Samis KE, Heath KD, Stinchcombe JR (2008) Discordant longitudinal clines in flowering time and *PHYTOCHROME C* in *Arabidopsis thaliana*. Evolution 62: 2971-2983.

Sawa M, Nusinow DA, Kay SA, Imaizumi T (2007) FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. Science 318: 261-265.

Scarcelli N, Cheverud JM, Schaal BA, Kover PX (2007) Antagonistic pleiotropic effects reduce the potential adaptive value of the *FRIGIDA* locus. PNAS 104: 16986-16991.

Schläppi MR (2006) *FRIGIDA LIKE 2* is a functional allele in Landsberg erecta and compensates for a nonsense allele of *FRIGIDA LIKE 1*. Plant Physiology 142: 1728-1738.

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

Schmid KJ, Torjek O, Meyer R, Schmuths H, Hoffmann MH, Altmann T (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.

Schmitz J, Franzen R, Ngyuen TH, Garcia-Maroto F, Pozzi C, Salamini F, Rohde W (2000) Cloning, mapping and expression analysis of barley MADS-box genes. Plant Molecular Biology 42: 899-913.

Schmitz RJ, Amasino R (2007). Vernalization: a model for investigating epigenetics and eukaryotic gene regulation in plants. Biochimica et Biophysica Acta 1769: 269-275.

Schranz ME, Quijada P, Sung S-B, Lukens L, Amasino R, Osborn TC (2002) Characterization and effects of the replicated flowering time gene *FLC* in *Brassica rapa*. Genetics 162: 1457-68.

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

Sharrock RA, Quail PH (1989) Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution and differential expression of a plant photoreceptor family. Genes and Development 3: 1745-1757.

Sheldon CC, Conn AB, Dennis ES, Peacock WJ (2002) Different regulatory regions are required for the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression. Plant Cell 14: 2527-2537.

Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C (2005) Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. Plant Physiology 138: 1163-1173.

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

Shindo C, Lister C, Crevillen P, Nordborg M, Dean C (2006) Variation in the epigenetic silencing of *FLC* contributes to natural variation in *Arabidopsis* vernalization response. Genes and Development 20: 3079-3083.

Shinomura T, Nagatani A, Hanzawa H, Kubota M, Watanabe M, Furuya M (1996) Action spectra for phytochrome A- and B-specific photoinduction of seed germination in *Arabidopsis thaliana*. PNAS USA 93: 8129-8133.

Somerville CR, McCourt P, Caspar T, Estelle M, Keith K (1985) *Arabidopsis thaliana* as a model system for plant genetics and molecular biology. Plant Genetics 651-660.

Song BH, Windsor AJ, Schmid KJ, Ramos-Onsins S, Schranzet ME al. (2009) Multilocus patterns of nucleotide diversity, population structure and linkage disequilibrium in *Boechera stricta*, a wild relative of *Arabidopsis*. Genetics 181: 1021-1033.

Stearns SC. The evolution of life histories (1992) Oxford: Oxford University Press.

Stenøien HK, Fenster CB, Kuittinen H, Savolainen O (2002) Quantifying latitudinal clines to light responses in natural populations of *Arabidopsis thaliana* (*Brassicaceae*). American Journal of Botany 89: 1604-1608.

Stenøien HK, Fenster CB, Tonteri A, Savolainen O (2005) Genetic variability in natural populations of Arabidopsis thaliana in northern Europe. Molecular Ecology 14: 137-148.

Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C et al. (2004) A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. PNAS USA 101: 4712-4717.

Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G (2001) *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. Nature 410: 1116-20.

Sung S, Amasino RM (2004) Vernalization in *Arabidopsis thaliana* is

mediated by the PHD finger protein VIN3. Nature 427: 159-164.

Sung S, He Y, Eshoo TW, Tamada Y, Johnson L (2006) Epigenetic maintenance of the vernalized state in *Arabidopsis thaliana* requires LIKE HETEROCHROMATIN PROTEIN 1. Nature Genetics 38: 706-710.

Swiezewski S, Liu F, Magusin A, Dean C (2009) Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. Nature 462 doi:10.1038/nature08618.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7: 453-464.

Tadege M, Sheldon CC, Helliwell CA, Stoutjesdijk P, Dennis ES et al. (2001) Control of flowering time by *FLC* orthologues in *Brassica napus*. Plant Journal 28: 545-53.

Takahashi Y, Shomura A, Sasaki T, Yano M (1998) *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. PNAS USA 98: 7922-7927.

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

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D et al. (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics 28: 7922-27.

Tian D, Araki H, Stahl E, Bergelson J, Kreitman M (2002) Signature of balancing selection in *Arabidopsis*. PNAS USA 99: 11525-11530.

Todokoro S, Terauchi R, Kawano S (1995) Microsatellite polymorphisms in natural populations of *Arabidopsis thaliana* in Japan. Japanese Journal of Genetics 70: 543-554.

Toomajian C, Hu TT, Aranzana MJ, Lister C, Tang C et al. (2006) A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. PLOS Biology 4: e137.

Turck F, Fornara F, Coupland G (2008) Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annual Review of Plant Biology 59: 573-594.

Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G (2004) Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science 303: 1003-1006.

Veley KM, Michaels SD (2008) Functional redundancy and new roles for genes of the autonomous floral-promotion pathway. Plant Physiology 147: 682-695.

Wang Q, Sajja U, Rosloski S, Humphrey T, Kim MC et al. (2007) HUA2 caused natural variation in shoot morphology of *A. thaliana*. Current Biology 17: 1513-1519.

Werner JD, Borevitz JO, Uhlenhaut NH, Ecker JR, Chory J, Weigel D (2005a) *FRIGIDA*-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. Genetics 170: 1197-1207.

Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR et al. (2005b) Quantitative trait locus mapping and DNA array hybridization identify an *FLM* deletion as a cause for natural flowering-time variation. PNAS USA 102: 2460-2465.

Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M et al. (2005) Integration of spatial and temporal information during floral induction in *Arabidopsis*. Science 309: 1056-1059.

Wilczek AM, Roe JL, Knapp MC, Cooper MD, Lopez-Gallego C et al (2009) Effects of genetic perturbation on seasonal life history plasticity. Science 323: 930-4.

Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. Molecular Biology and Evolution 22: 506-519.

Yamaguchi S (2008) Gibberellin metabolism and its regulation. Annual Review of Plant Biology 59: 225-251.

Yamauchi Y, Ogawa M, Kuwahara A, Hanada A, Kamiya Y, Yamaguchi S (2004) Activation of gibberellin biosynthesis and response pathways by low temperature during imbibition of *Arabidopsis thaliana* seeds. Plant Cell 16: 367-78.

Yan L, Fu D, Li C, Blechl A, Tranquilli G et al. (2006) The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. PNAS USA 103: 19581-19586.

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W et al. (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science 303: 1640-1644.

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene VRN1 (2003). PNAS USA 100: 6263-6268.

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L et al. (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. Plant Cell 12: 2473-2484.

Zeilinger MN, Farre EM, Taylor SR, Kay SA, Doyle FJ (2006) A novel computational model of the circadian clock in *Arabidopsis* that incorporates *PRR7* and *PRR9*. Molecular Systems Biology 2: 58.

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C et al. (2007) An *Arabidopsis* example of association mapping in structured samples. PLOS Genetics 3: e4.

PAPER I

Differences between continental and coastal climate determine *FLC*-dependent variation in vernalization response of *Arabidopsis thaliana*

Anna M. Lewandowska-Sabat, Siri Fjellheim, Odd Arne Rognli

Department of Plant and Environmental Sciences, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway

Corresponding author

E-mail: odd-arne.rognli@umb.no

Telephone: +47 64 96 55 78

Fax: +47 64 96 59 01

Abstract

Flowering is a major determinant of plant reproductive success and a trait with a potentially high adaptive value. Vernalization, i.e. period of low temperature, is one of the pathways controlling flowering time in response to environmental signals. Two main regulators of the vernalization pathway, *FLOWERING LOCUS C* (*FLC*) and its activator *FRIGIDA* (*FRI*) account for natural variation in flowering time in response to vernalization in *Arabidopsis thaliana* and appear as major targets for natural selection. This study attempts to comprehend flowering time and vernalization responses in natural populations of *A. thaliana* from higher latitudes and their significance for local adaptation. *A. thaliana* populations were collected from a range of high-latitude (59-68°N) locations in Norway and phenotypic screening for response to 5 different vernalization treatments as well as sequence analyses of *FRI* and *FLC* were performed. The results showed that flowering time, vernalization sensitivity and vernalization saturation were associated with distance from the ocean, which suggests that climatic factors such as length of the winter put selection pressures on these traits. Selection favours populations with higher vernalization sensitivity in continental climates and populations with lower vernalization sensitivity in coastal environments. Moreover, variation in vernalization sensitivity was associated with *FLC* polymorphisms. The results indicate that *FLC* contributes to variation in vernalization sensitivity and may determine local adaptation in *A. thaliana* at the northernmost limit of its distribution.

Key words: *Arabidopsis thaliana*, *FLC*, *FRI*, natural populations, vernalization.

Introduction

Variation in the length of the growing season is on a large scale associated with latitude and therefore phenological traits are expected to vary along latitudinal clines. Latitudinal variation in adaptive traits has been demonstrated in many organisms e.g. plants, trees and *Drosophila* (Kalisz and Wardle 1994; Hurme *et al*., 1997; Verrelli and Eanes 2001, Olsson and Ågren 2002). Flowering time is one of the key adaptive traits enabling plants to flower at the most advantageous time for pollination, seed set and dispersal. This timing of reproduction is critical for plant fitness, and is a particular challenge for annual plants that reproduce only once during their life time. In the widely studied annual model plant *Arabidopsis thaliana*, latitudinal clines in flowering time in response to vernalization have rarely been demonstrated (Stinchcombe *et al*., 2004), and were only found among accessions with functional alleles at the *FRIGIDA* (*FRI*) locus.

The vast variation in environmental factors should impose uneven selection pressure across broad geographic areas. Ecologically essential factors vary latitudinally on the continental-wide scale and therefore latitudinal coordinates may be a significant proxy for climatic factors such as temperature and photoperiod. However, it remains unclear whether the differences in selection pressures will be evident on a small geographic scale, where variation in environmental cues is less marked. Moreover, when the climatic transect is present at the same latitude e.g. due to differences in the distance from the ocean or altitude, latitude cannot be used as a rough substitute for ecologically relevant cues.

Arabidopsis thaliana is a self-compatible weedy species with a worldwide distribution. It grows within a broad climatic and latitudinal range from far North (68°N) to equatorial locations, which makes it an excellent model for studying natural variation in adaptive features (Koornneef *et al*., 2004). Variation in flowering time in *A. thaliana* has been shown to be associated with variation in genes affecting flowering in response to abiotic signals (Johanson *et al*., 2000; El-Assal *et al*., 2001; Shindo *et al*., 2005; Shindo *et al*., 2006; Balasubramanian *et al*., 2006; Korves *et al*., 2007). Studying natural variation not only enables the identification of the function of particular genes but also makes it possible to infer which evolutionary mechanisms that maintain this variation and which alleles confer adaptation to particular climatic conditions (Koornneef *et al*., 2004). However, some of the association between traits and variation at the candidate locus may be confounded due to population structure (Zhao *et al*., 2007).

A process that is central in timing of flowering in temperate regions is vernalization.

Vernalization is defined as an extended period of cold that triggers transition from vegetative to reproductive growth in plants (Simpson and Dean 2002). Two main groups related to vernalization requirements are found among *A. thaliana* natural populations. The winter annuals require vernalization in order to complete their life cycle and summer annuals, also called rapid cyclers, do not have vernalization requirement in order to flower (Simpson and Dean 2002). The germination and flowering time of these populations under natural conditions is, however, not well understood (Wilczek *et al*., 2009). Under constant growth chamber conditions both groups of populations flower without vernalization but flowering time of winter annuals is strongly delayed (Karlsson *et al*., 1993; Nordborg and Bergelson 1999; Koornneef *et al*., 2004). The majority of Northern European accessions display late flowering phenotype and require vernalization in order to flower, whereas both winter and summer annuals are found among Southern and Western European accessions (Lawrence 1976; Johanson *et al*., 2000; Le Corre *et al*., 2002).

In recent years, rapid progress has been made in elucidating the main pathways mediating signals from the environment that control flowering in *A. thaliana*. *FRI* confers a vernalization requirement by upregulating expression of flowering repressor *FLOWERING LOCUS C* (*FLC*) (Bastow *et al*., 2004). In *A. thaliana* with functional *FRI* alleles, the vernalization mediating genes *VERNALIZATION 1* (*VRN1*), *VRN2, VRN5,* and *VERNALIZATION-INSENSITIVE 3* (*VIN3*) inhibit expression of *FLC*, which remains at low level throughout subsequent development (Gendall *et al*., 2001; Levy *et al*., 2002; Sung and Amasino 2004; Sung *et al*., 2006; Greb *et al*., 2007). Thus, both *FRI* and *FLC* play crucial roles in mediating vernalization response. However, adaptive significance has only been demonstrated for variation at *FRI* (Le Corre 2005). Accessions without vernalization requirement often carry non-functional *FRI* alleles and under laboratory conditions display early-flowering phenotypes (Johanson *et al*., 2000; Le Corre *et al*., 2002; Shindo *et al*., 2005).

Natural populations of *A. thaliana* in Norway covering geographic variation both in terms of latitude, longitude and altitude was used in this study. These populations cover the northernmost distribution of *A. thaliana* worldwide. This sampling allows the detection of regional differences in phenotypic traits between coastal and inland populations, which are subjected to very different climates. The most relevant factor that may account for variation in responses to vernalization is winter duration. We tested this hypothesis by comparing flowering time, vernalization sensitivity and vernalization saturation of coastal populations from both arctic and subarctic latitudes (62-68°N) with continental populations from

subarctic latitudes (59-61°N). Furthermore, we have examined whether sequence variation in *FRI* and *FLC* is associated with flowering time and/or vernalization responses.

Material and methods

Plant materials and phenotypic screening of vernalization response

Thirty-six populations of *A. thaliana* were collected during summer 2003 from locations in Norway varying in latitude and altitude (Table 1, Fig. 1, Table S1). Individuals from each population were collected from site-specific environment as adult plants with siliques. Twenty-seven populations were given a treatment of 12, 9, 6, 3 and 0 weeks of vernalization using 5-12 lines per population (depending on plant survival). Nine additional populations were exposed only to 12 weeks of vernalization. Original seeds collected from the field were sown in standard soil (Hasselfors Garden AB, Örebro, Sweden) in 6.5 cm pots. The seeds were stratified for 3 days in darkness at 4°C to synchronize their germination and then left to germinate at 23°C and 8 h of light with a photon flux density of 150 ± 10 µmol m^{-2} s^{-1} at 400-700 nm followed by 16 h of dark period for 12 days. Then plants were transferred to 4°C and 8 h light (50 ± 5 µmol m^{-2} s^{-1}) followed by 16 h of dark period until vernalization treatment was completed. By the end of the vernalization treatment the conditions in the growth rooms were changed to 23°C and 16 h of light (150 ± 10 µmol m^{-2} s^{-1}) followed by an 8 h dark period. The red/far-red ratio was 1.8 ± 0.2. The light was produced by HQI lighting systems (LU400/XO/T/40 Philips Osram, General Electric, Hungary). Plants were watered twice a week or once a week during the vernalization treatment and rotated within the growth room to avoid position effect. Seeds used for the five vernalization treatments were planted in time intervals so that all treatments terminated at the same date. Flowering time was measured daily as days to bolting (DTB), days to flowering (DTF) (both recorded after condition in the growth chambers had been changed to 23°C/16 h of light) and number of rosette leaves at flowering (RLN, no data for 12 weeks of vernalization). All individuals from 3 and 0 weeks of vernalization were kept for 100 days in 23°C/16 h of light, i.e. to the end of the experiment. DTB or DTF was set to 100 days for individuals that did not flower after 6, 3 and 0 weeks of vernalization.

Sequencing of FRI and FLC

To address whether sequence variation in *FLC* is associated with any geographic pattern in flowering time and vernalization response we genotyped one individual from each of 29 Norwegian populations, the laboratory strains Columbia (Col-0) and Landsberg *erecta* (L*er*). For *FLC* we sequenced 1.7 kb upstream of ATG, 1.7 kb of the first intron and the complete coding region. The promoter region and the first exon of *FRI* were sequenced to determine the presence or absence of the two non-functional alleles typical of L*er* and Col-0. Genomic DNA was extracted using Aqua Pure Genomic DNA Isolation Kit (Bio-Rad, Hercules, CA, USA) and PCR was performed with DyNAzyme II DNA Polymerase (Finnzymes, Espoo, Finland) using standard procedure. Primers used for the *FLC* gene (Table S2) were design based on the reference sequence published in TAIR (http://www.arabidopsis.org/, locus AT5G10140.1) using Primer3 v. 0.4.0 (Rozen and Skaletsky 2000). Col-like and L*er*-like deletions in *FRI* were genotyped using primers published by Johanson *et al*. (2000). We sequenced PCR products directly using BigDye terminator v.3.1 Cycle Sequencing reactions kit (Applied Biosystems, Foster City, California, USA). Sequences were run on ABI Prism 3100 Automated Capillary DNA Sequencer (Applied Biosystems). All sequences were aligned and analyzed in Sequencher 4.7 DNA sequence analysis software (Gene Codes, Ann Arbor, Michigan, USA).

Climate data

Climate data for the collection sites were obtained from The Norwegian Meteorological Institute (http://met.no/English/ and http://www.senorge.no/startpage.aspx) using data from the meteorological stations closest to the site of origin of the populations. These are mean January and July temperatures and yearly precipitation (for years 1961-1990) (Table S1). Regional differences in the climatic conditions at the site of origin of populations were estimated using mean January temperatures as a proxy for winter temperature and winter duration and yearly precipitation as a proxy for distance from the ocean. The difference between mean January and July temperature was used as a proxy for temperature amplitude between winter and summer seasons at the site of origin of populations. Thus, lower January temperature will be associated with longer winters and longer lasting snow cover, and larger temperature amplitude will be associated with harsh winters and warm summers, typical for inland climates. Large amount of precipitation will be characteristic for coastal areas. To

facilitate interpretation, reciprocal of log transformed precipitation data was used in analyses.

Statistical analyses

Vernalization sensitivity (V_{sens}) was estimated as the area under the regression line of each population mean DTB and DTF, respectively, regressed against weeks of vernalization. For each population only vernalization weeks until saturation were used. Thus, larger area under regression plot will indicate slower response i.e. lower vernalization sensitivity. To facilitate interpretation reciprocal values of V_{sens} were used. As a measure of vernalization saturation (V_{sat}) the time (weeks) when there was no further decrease in DTB and DTF, respectively, was used (observed with increasing period of vernalization). The statistical significance between DTB and DTF, respectively, in each population in response to different vernalization periods was estimated using ANOVA. Vernalization requirement (V_r) was defined by flowering or lack of flowering in the absence of vernalization. Differences in V_{sens} and V_{sat} between coastal and inland populations as well as *FLC* haplotypes (defined by significantly associated SNPs) were determined using ANOVA. Relationship between coastal and inland populations and *FLC* haplotypes was determined by correlation analysis.

Population mean for DTB, DTF and RLN for 5 vernalization treatments as well as V_{sens} and V_{sat} were regressed against winter temperature, temperature amplitude and distance from the ocean at the site of origin of populations. Regression analyses were performed after omitting populations with putative non-functional *FRI* alleles that flowered without vernalization.

Haplotype diversity (Hd) and nucleotide diversity for *FRI* and *FLC* were estimated using DnaSP v.4.50.3 software (http://www.ub.es/dnasp/). For association of *FLC* polymorphisms with DTB, DTF, V_{sens} and V_{sat} only polymorphisms present in more than one population was used. Association test was performed with ANOVA according to Korves *et al*. (2007) using coefficients that statistically describe a population's inferred ancestry as covariates (Samis *et al*., 2008). These coefficients were estimated using STRUCTURE 2.2.3 (Pritchard *et al*., 2000; Falush *et al*., 2003, through the BIOPORTAL http://www.bioportal.uio.no/) and SNP data on 31 Norwegian populations of *A. thaliana* (including all populations analysed in this study, Lewandowska-Sabat *et al*., unpublished results, Paper III). Including ancestry coefficients in association tests will determine if variation in the traits is due to population structure. In *FLC* association tests populations with lack of vernalization requirement were not included in the analyses, since ecological factors

impose natural selection on flowering time responses probably only in populations with putatively functional *FRI* alleles (Stinchcombe *et al.*, 2004). Moreover, the epistatic interaction of *FRI* and *FLC* must be present in order to observe *FLC*-trait association (Caicedo *et al.*, 2004). Simple linear regression and ANOVA analyses were performed using Minitab 15 Statistical Software (Minitab Inc. State College, PA, USA).

Results

Phenotypic variation

Variation in flowering time among populations and among individuals within populations was high (data not shown). Flowering of all populations was strongly accelerated by vernalization for more than 6 weeks (Fig. 2). All plants, except a few individuals from population Ors-1, flowered after 6 weeks of vernalization. Three weeks of vernalization was sufficient to trigger flowering in populations Nfro-1, Lom-4, Had-1, Had-2, Had-3, Lod-1, Eid-1, Kon-2 and Kon-3. Surprisingly, three populations, Nfro-1, Sk-2, and Lod-1 flowered without vernalization. Flowering of Sk-2 and Lod-1 was however, much delayed compared with Nfro-1. Coastal populations flowered significantly later than inland populations across all vernalization treatments (mean 15.2, SE=0.25 and mean 11.5, SE=0.16, respectively, $P<0.001$ data for DTF12w). Inland populations were more sensitive to vernalization and required less time to saturate their vernalization requirement than coastal populations ($P<0.001$, $P=0.012$, respectively, Fig. 3).

Regression analyses

Regression analyses revealed that variation in DTB and DTF in response to all vernalization periods is highly associated with distance from the ocean, winter temperature and temperature amplitude and weaker association between RLN and these parameters were also found (Table 2). Later flowering was associated with proximity to the ocean, higher winter temperature (Fig. 4) and lower temperature amplitude. However, coefficients of determination (R^2) and significance decreased with decreasing period of vernalization (Table 2).

V_{sens} and V_{sat} were associated with distance from the ocean, winter temperature and temperature amplitude (Table 2). Lower vernalization sensitivity and longer time until

saturation of vernalization requirement was associated with proximity to the ocean, higher winter temperature (Fig. 4) and lower temperature amplitude. There was no apparent relationship between V_r and climatic factors or distance from the ocean (see Table 3 for V_{sens}, V_{sat} and V_r estimates).

Due to few observations, multivariate analyses did not give the best models for prediction. Therefore, simple linear regression was considered the most suitable method also considering that the climatic predictors are highly collinear.

Variation in FRI and FLC

Four different haplotypes were found among the 29 populations based on polymorphism in *FRI* (in total 928 bp) and the haplotype diversity (Hd) was 0.480. The Col-type and the L*er*-type deletions in *FRI* were not present in any of the populations. There were 10 non-synonymous and 2 synonymous SNPs. One insertion resulted in an additional amino acid and one deletion resulted in deletion of the last amino acid of exon 1 (Table S3). Nucleotide diversity for the entire *FRI* region studied was 0.00260 and for the coding region 0.00303. One indel (IN2) and 2 SNPs (SNP-176 and SNP-142, Table 4) in the 5'flanking region of *FRI* have been found previously in the Cvi and H51 accessions from Cape Verde Islands and Sweden, respectively (Le Corre *et al.,* 2002; Gazzani *et al.,* 2003).

Fourteen different haplotypes were found among the 29 populations based on polymorphism in *FLC* (in total 4436 bp, Hd=0.872). Col-0 and L*er* were present as 2 separate haplotypes. Only 1 SNP (synonymous) was found in the coding region (Table S4). Nucleotide diversity for the entire *FLC* region was 0.00199 and for the coding region 0.00087. Most of the polymorphisms in *FLC* found in Norwegian populations have been identified earlier (Gazani *et al.,* 2003; Caicedo *et al.,* 2004; Lempe *et al.,* 2005; as well as in GBrowse of TAIR, http://www.arabidopsis.org/). A few polymorphisms in the 5'flanking region of *FLC* has not been described before (Table S4).

After accounting for population structure only DEL1 and SNP5 in *FLC* (in linkage disequilibrium) remained significantly associated with V_{sens} (Table 4). Based on these two SNPs, the material was divided into two haplogroups.

Association of all *FLC* SNPs and indels with DTB and DTF, respectively, across all vernalization treatments and V_{sat} was not significant after accounting for population structure (data not shown).

The differences in V_{sens} and V_{sat} between *FLC* haplotypes determined based on DEL1

and SNP5 were significant (P=0.001, P=0.021, respectively, Fig. 3). Lower V_{sens} and longer time to V_{sat} were common in *FLC1* haplotype (Fig. 3). *FLC1* haplotype was prevalent in coastal populations and *FLC2* haplotype in inland populations (r=0.60, P=0.002).

Discussion

Coastal to continental climate gradients shape variation in flowering time and vernalization sensitivity

In this study we have shown that climatic factors that are changing with distance from the ocean put selection pressures on flowering time and vernalization responses in *A. thaliana* (Fig. 4). The strong influence of the Gulf Stream creates a mild winter climate along the Norwegian coast far north above the Arctic Circle. Natural selection favours late flowering populations with lower vernalization sensitivity in coastal regions, where winters are mild, shorter and with possibly unpredictable temperature fluctuations (Fig. 3, Fig. 4). Coastal populations need also longer time to fully saturate their vernalization requirement (Fig. 3). This suggests that unstable winter conditions at the coast could lead to premature flowering if populations were vernalization sensitive. These populations must have other mechanism securing proper timing of reproduction, e.g. photoperiod perception (Lewandowska-Sabat *et al*., unpublished results, Paper II). On the contrary, in a continental climate, selection favours rapid flowering populations with higher vernalization sensitivity and shorter time efficient for vernalization saturation (Fig. 3). Long and harsh winters at inland locations but also rapid spring temperatures increase due to solar energy warming up land area faster than the sea, suggest that increasing temperatures and photoperiod following cold period are reliable signals for inland populations to flower. We have demonstrated that winter duration as defined by winter temperatures and temperature amplitudes is a critical factor that impose natural selection on variation in vernalization responses in natural populations of *A. thaliana*. Earlier study of flowering time in *A. thaliana* has not found correlation with winter temperature (Stinchcombe *et al.*, 2004) and this suggests that the regional scale of present study is well suited for detecting spatial heterogeneity in ecologically relevant cues.

The lack of an apparent latitudinal cline in flowering time of *A. thaliana* as compared to other plant species may be due to their dissimilar demographic histories. The postglacial colonization pattern of *A. thaliana* may be associated with past and present human migration (François *et al*., 2008; Beck *et al*., 2008; Lewandowska-Sabat *et al*., unpublished results,

Paper III). Therefore, even if populations are sampled from presumably isolated and undisturbed sites, these populations may represent a mixture of adapted and newly introduced genotypes. Furthermore, variation in flowering time even between populations originating from presumably similar environments may result from differences in biotic and abiotic factors at their local sites (Olsson and Ågren 2002, Shindo *et al*., 2005). The typical examples are local temperatures, light quality and quantity in open habitats versus forests or at the South versus North-facing slopes, soil composition, but also pathogens and herbivores exposure. These factors may create local microclimate and therefore part of the flowering time variation cannot be elucidated using rough climatic proxies, which probably are more realistic on regional scales. Thus, local microclimatic conditions and, to some extent, admixture of genotypes sharing different demographic histories, are likely accounting for variation in flowering time within regions in our study.

Polymorphism at the FRI gene

Polymorphism associated with vernalization requirement has been demonstrated previously in many plant species (Osborn *et al*., 1997; Yan *et al*., 2006; Reeves *et al*., 2007). In *A. thaliana*, loss-of-function polymorphisms at the *FRI* gene have been shown to affect flowering time in response to vernalization in natural populations (Johanson *et al*., 2000; Le Corre *et al.,* 2002; Schindo *et al*., 2005). The adaptive significance of loss-of-function *FRI* alleles and selection for early flowering in natural populations has been demonstrated by Le Corre (2005). In our study, no indels typical for Col-0 and L*er* were found in any of the populations. These *FRI* mutations seem to be restricted to Germany and Central/Eastern Europe (Hagenblad *et al.,* 2004). Populations carrying non-functional *FRI* alleles may have been eliminated by natural selection since absolute vernalization requirement in severe environment is critical for survival. Lack of vernalization requirement in Nfro-1 may be due to several polymorphisms and indels in *FRI*. Moreover, most of the polymorphisms present in population Nfro-1 have been identified in population Jl-1 from the Czech Republic (Lempe *et al.,* 2005). This may suggest a recent introduction of this population to Norway by human-mediated seed flow (Lewandowska-Sabat *et al*., unpublished results, Paper III). No differences between Sk-2 and Lod-1 and other populations were found in *FRI*, but lack of vernalization requirement in these populations may still be due to putatively loss-of-function mutations in other *FRI* regions than examined here, as well as due to allelic changes at additional loci. One of the candidate loci may be *EARLY FLOWERING IN SHORT DAYS*

(*EFS*). Mutations in *EFS* result in low level of *FLC* prior to vernalization, even in the presence of the functional FRI gene (Kim *et al*., 2005). Moreover, it has been shown recently by Wilczek *et al.,* (2009) that loss-of-function *FRI* mutations convert winter to summer annuals only in milder climates and that repression of flowering by *FLC* in winter annuals may be counteracted by ambient temperature and photoperiod in natural habitats. This might well be the case in our study, where the existence of summer and winter annuals among populations from the same area may result from their capability of switching between two life histories and differences in sensitivity to environmental cues.

Variation at FLC is associated with vernalization sensitivity

The two novel *FLC* polymorphisms (DEL1 and SNP5) were significantly associated with vernalization sensitivity in the *A. thaliana* populations (Table 4, P=0.038) but associations with the remaining polymorphisms were confounded by population structure. Therefore, functional differentiation of *FLC* haplotypes distinguished by these polymorphisms is likely not due to population structure artefact (Fig. 3). These significant polymorphisms in the 5' flanking region may have *cis*-regulatory function in *FLC* expression similarly as demonstrated for regions of the promoter and the first *FLC* intron (Sheldon *et al*., 2002). Moreover, Shindo *et al*., (2006) showed that allelic variation in the *FLC cis*-regulatory sequences might account for vernalization response through epigenetic maintenance of *FLC* chromatin silencing. This may be the case in our study, where variation in the epigenetic maintenance of *FLC* silencing may contribute to vernalization sensitivity. Vernalization inhibits expression and induces the epigenetic silencing of *FLC* (Bastow *et al*., 2004; Sung and Amasino 2004), and possibly in populations adapted to mild and variable winters, i.e. in the coastal regions, short periods of vernalization are not efficient in inducing stable *FLC* silencing and rapid vernalization response. On the contrary, in populations adapted to the cold and stable continental climates, shorter periods of vernalization are required to stable silencing of the *FLC* chromatin and induction of rapid vernalization response. However, this hypothesis needs further study. Most of the polymorphisms were in LD with each other, thus their associations with vernalization sensitivity could be removed due to lack of statistical power.

FLC haplotypes were significantly different in vernalization saturation (Fig. 3) however, this diversity was removed after correcting for population structure. Although *FLC* sequence variation do not contribute to vernalization saturation, still variation in the

epigenetic *FLC* silencing may mediate this response. It has been demonstrated recently that long antisense transcripts covering the entire *FLC* locus and likely reducing a sense transcripts level are an early step in vernalization-mediated epigenetic silencing of *FLC* (Swiezewski *et al*., 2009). It might be that the variation in antisense *FLC* transcripts level contributes to variation in vernalization saturation in our study. However, this remains unresolved and needs further investigation.

All associations between flowering time and polymorphisms in *FLC* were confounded by population structure. It appears therefore that flowering time may be determined by variation at other candidate loci (e.g. Balasubramanian *et al*., 2006) and other environmental cues e.g. photoperiod may further account for this variation.

Nucleotide diversity of *FRI* and *FLC* was lower than in other studies (Nordborg *et al.*, 2005; Schmid *et al.*, 2005) and may be due to the relatively limited geographical range represented by these *A. thaliana* populations (Le Corre *et al.*, 2002) as well as demographic processes such as bottlenecks. Moreover, these populations represent the boundary of a species geographic distribution, where genetic variation in populations is expected to be reduced (Eckert *et al*., 2008).

Understanding the genetic and physiological processes that underlie climatic adaptation in plants is important in the face of predicted rapid climate changes. These changes are predicted to be most extensive in arctic and subarctic regions which will experience extended growth seasons and more unstable winter climates (Høgda *et al*., 2007). Earlier spring growth and prolonged autumns with relatively higher temperatures and lower light intensities will be challenging for locally adapted annuals and perennials. This study contributes to the current knowledge on diversity and dynamics of vernalization responses and elucidates the significance of climatic factors that impose natural selection on these responses in natural populations of *A. thaliana*. Moreover, our sampling strategy provided an unbiased sample of isolated, presumably locally adapted populations from sites with robustly limited human-mediated gene flow, which allowed us to detect association of *FLC* variation with vernalization sensitivity.

Acknowledgments

We are grateful to Jorunn Elisabeth Olsen for her discussion on the experimental design and the results. We thank Caroline Dean for providing helpful comments on a draft of this manuscript. We also thank Ragnhild Nestestog and YeonKyeong Lee for technical assistance and Simen Rød Sandve for discussion and comments on the results. This study was supported by a grant from the Research Council of Norway: The Norwegian Arabidopsis Research Centre, project no. 151991/150, and a PhD scholarship to Anna Lewandowska-Sabat from the Norwegian University of Life Sciences.

References

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

Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, Dean C. 2004. Vernalization requires epigenetic silencing of *FLC* by histone methylation. *Nature* **427**: 164-167.

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

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD. 2004. Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *PNAS* **101**: 15670-15675.

Eckert CG, Samis KE, Lougheed SC. 2008. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology* **17**: 1170-1188.

El-Assal SE-D, Alonso-Blanco C, Peeters AJM, Raz V, Koornneef M. 2001. A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nature Genetics* **29**: 435-40.

Falush D, Stephens M, Pritchard J. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**: 1567-1587.

François O, Blum MGB, Jakobsson M, Rosenberg NA. 2008. Demographic history of European populations of *Arabidopsis thaliana*. *PLOS Genetics* **4**: e100075.

Gazzani S, Gendall AR, Lister C, Dean C. 2003. Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiology* **132**: 1107-1114.

Gendall AR, **Levy YY**, **Wilson A**, **Dean C. 2001.** The *VERNALIZATION 2* gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. *Cell* **107**: 525-35.

Greb T, Mylne JS, Crevillen P, Geraldo N, An H, Gendall AR, Dean C. 2007. The PHD finger protein *VRN5* functions in the epigenetic silencing of *Arabidopsis FLC*. *Current Biology* **17**: 73-8.

Hagenblad J, Tang C, Molitor J, Werner J, Zhao K, Zheng H, Marjoram P, Weigel D, Nordborg M. 2004. Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. *Genetics* **168**: 1627-1638.

Høgda KA, Karlsen SR, Tømmervik H. 2007. Changes in growing season in Fennoscandia 1982–1999. In: Ørbæk JB *et al.*, eds. *Arctic alpine ecosystems and people in a changing environment.* Springer Berlin Heidelberg, ISBN 978-3-540-48512-4.

Hurme P, Repo T, Pääkönen T, Savolainen O. 1997. Climatic adaptation of bud set and frost hardiness in Scots pine (*Pinus sylvestris* L.). *Canadian Journal of Forest Research* **27**: 716-723.

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

Kalisz S, Wardle GM. 1994. Life history variation in *Campanula americana* (*Campanulaceae*): population differentiation. *American Journal of Botany* **81**: 521-527.

Karlsson BH, Sills GR, Nienhuis J. 1993. Effects of photoperiod and vernalization on the number of leaves at flowering in 32 *Arabidopsis thaliana (Brassicaceae)* ecotypes. *American Journal of Botany* **80**: 646-648.

Kim, SY, He Y, Jacob Y, Noh Y-S, Michaels S, Amasino R. 2005. Establishment of the vernalization-responsive, winter-annual habit in *Arabidopsis* requires a putative histone H3 methyl transferase. *Plant Cell* **17**: 3301-3310.

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

Korves TM, Schmid KJ, Caicedo AL, Mays C, Stinchcombe JR, Purugganan MD, Schmitt J. 2007. Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. *American Naturalist* **169**: E141-E157.

Lawrence MJ. 1976. Variations in natural populations of *Arabidopsis thaliana* (L.) Heynh. In: Vaughan JG, Macleod AJ, eds. *The biology and chemistry of the Cruciferae*. BMG Jones 167–90 London/New York/San Francisco: Academic.

Le Corre V. 2005. Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. *Molecular Ecology* **14**: 4181-4192.

Le Corre V, Roux F, Reboud X. 2002. DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Molecular Biology and Evolution* **19**: 1261-1271.

Lempe J, **Balasubramanian S**, **Sureshkumar S**, **Singh A**, **Schmid M**, **Weigel D. 2005.** Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLOS Genetics* **1**: 109-18.

Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C. 2002. Multiple roles of *Arabidopsis VRN1* in vernalization and flowering time control. *Science* **297**: 243-246.

Nordborg M, Bergelson J. 1999. The effect of seed and rosette cold treatment on germination and flowering time in some *Arabidopsis thaliana (Brassicaceae)* ecotypes. *American Journal of Botany* **86**: 470-475.

Nordborg M, Hu T, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R et al. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLOS Biology* **3**: e196.

Olsson K, Ågren J. 2002. Latitudinal population differentiation in phenology, life history and flower morphology in the perennial herb *Lythrum salicaria*. *Journal of Evolutionary Biology* **15**: 983-996.

Osborn TC, Kole C, Parkin IAP, Sharpe AG, Kuiper M, Lydiate DJ, Trick M. 1997. Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. *Genetics* **146**: 1123-1129.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.

Reeves PA, He Y, Schmitz RJ, Amasino RM, Panella LW, Richards CM. 2007. Evolutionary conservation of the *FLOWERING LOCUS C* mediated vernalization response: evidence from the sugar beet (*Beta vulgaris*). *Genetics* **176**: 295-307.

Rozen S, Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S. *Bioinformatics methods and protocols: methods in molecular biology.* Totowa, USA, Humana Press 365-386.

Samis KE, Heath KD, Stinchcombe JR. 2008. Discordant longitudinal clines in flowering time and *PHYTOCHROME C* in *Arabidopsis thaliana*. *Evolution* **62**: 2971-2983.

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

Sheldon CC, Conn AB, Dennis ES, Peacock WJ. 2002. Different regulatory regions are required for the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression. *Plant Cell* **14**: 2527-2537.

Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005. Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. *Plant Physiology* **138**: 1163-1173.

Shindo C, Lister C, Crevillen P, Nordborg M, Dean C. 2006. Variation in the epigenetic silencing of *FLC* contributes to natural variation in *Arabidopsis* vernalization response. *Genes and Development* **20**: 3079-3083.

Simpson GG, Dean C. 2002. *Arabidopsis*, the Rosetta stone of flowering time? *Science* **296**: 285-289.

Stinchcombe JR, **Weinig C**, **Ungerer M**, **Olsen KM**, **Mays C**, **Halldorsdottir SS**, **Purugganan MD**, **Schmitt J. 2004.** A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *PNAS* **101**: 4712-4717.

Sung S, Amasino RM. 2004. Vernalization in *Arabidopsis thaliana* is mediated by the PHD

finger protein VIN3. *Nature* **427**: 159-164.

Sung S, Schmitz RJ, Amasino RM. 2006. A PHD finger protein involved in both the vernalization and photoperiod pathways in *Arabidopsis*. *Genes and Development* **20**: 3244-3248.

Swiezewski S, Liu F, Magusin A, Dean C. 2009. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* **462** doi:10.1038/nature08618.

Verrelli BC, Eanes WF. 2001. Clinal variation for amino acid polymorphisms at the *PGM* locus in *Drosophila melanogaster*. *Genetics* **157**: 1649-1663.

Wilczek AM, Roe JL, Knapp MC, Cooper MD, Lopez-Gallego C, Martin LJ, Muir CD, Sim S, Walker A, Schmitt J *et al.* 2009. Effects of genetic perturbation on seasonal life history plasticity. *Science* **323**: 930-4.

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J. 2006. The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. *PNAS USA* **103**: 19581-19586.

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M. 2007. An *Arabidopsis* example of association mapping in structured samples. *PLOS Genetics* **3**: e4.

Figure legends

Figure 1. Collection sites of Norwegian *Arabidopsis thaliana* populations investigated in this study.

Figure 2. Time to flowering (days) plotted against number of weeks of vernalization. The date when temperature and day length were changed from vernalizing conditions to normal (23°C and 16 h day length) is set to day 1. Day 100 indicates that populations did not flower. Populations were assembled into seven groups: red – West Coast, dashed line – coastal, dark blue – Northern Norway, light blue – Southern Norway, green – Central Norway, dotted line – inland, lilac – Southeastern Norway. Assignment of the populations to the group is listed in Table 1.

Figure 3. Mean vernalization sensitivity (V_{sens}) and vernalization saturation (V_{sat}) in coastal and inland population groups and in the two *FLC* haplotypes. ***P≤0.001; *P≤0.05. *FLC1* haplotype was prevalent in coastal populations and *FLC2* haplotype was prevalent in inland populations (P=0.002, r=0.60).

Figure 4. Regression analyses in *Arabidopsis thaliana*. (a) Vernalization sensitivity (V_{sens}) regressed into distance from the ocean (precipitation), R^2=0.28, P=0.005 and (b) winter temperature (^{o}C), R^2=0.30, P=0.004; (c) days to flowering (DTF) after 12 weeks of vernalization regressed into distance from the ocean (precipitation), R^2=0.46, P<0.001 and (d) winter temperature (^{o}C), R^2=0.45, P<0.001. Open diamonds indicate inland populations and filled diamonds indicate coastal populations. Yearly mean precipitation (mm, 1961-1990) was used as proxy for distance from the ocean.

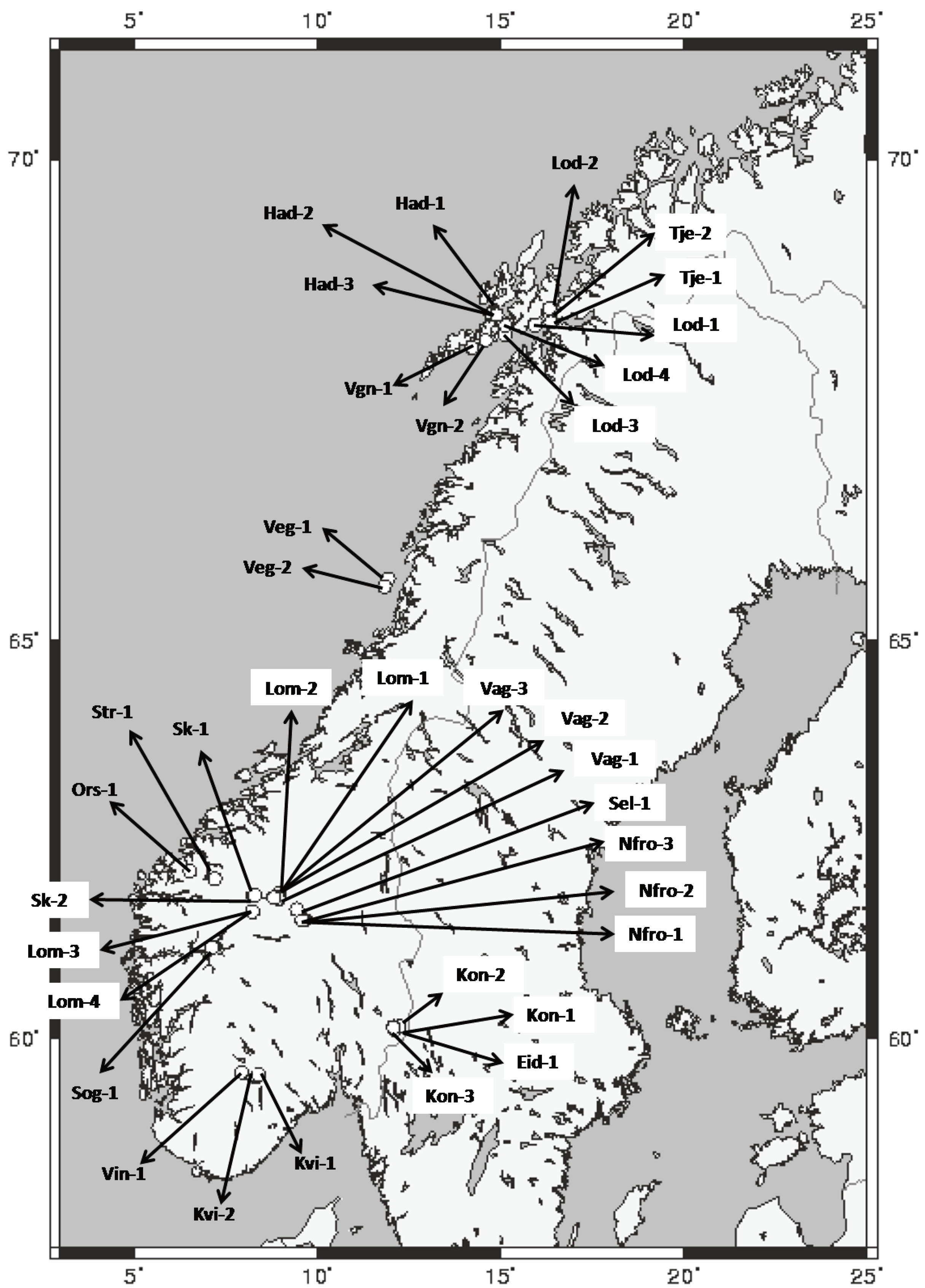


Figure 1

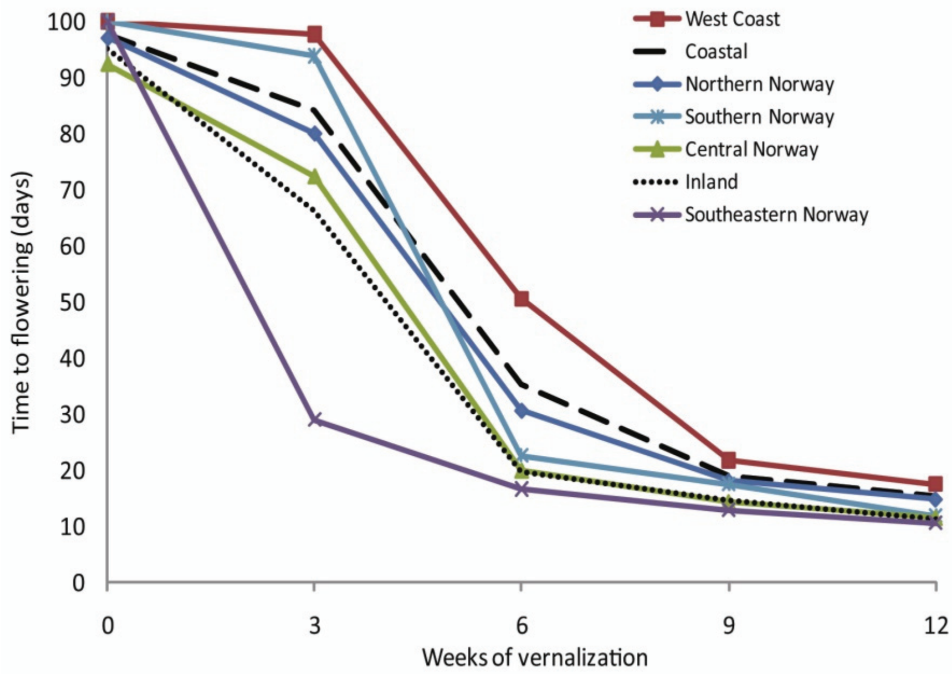


Figure 2

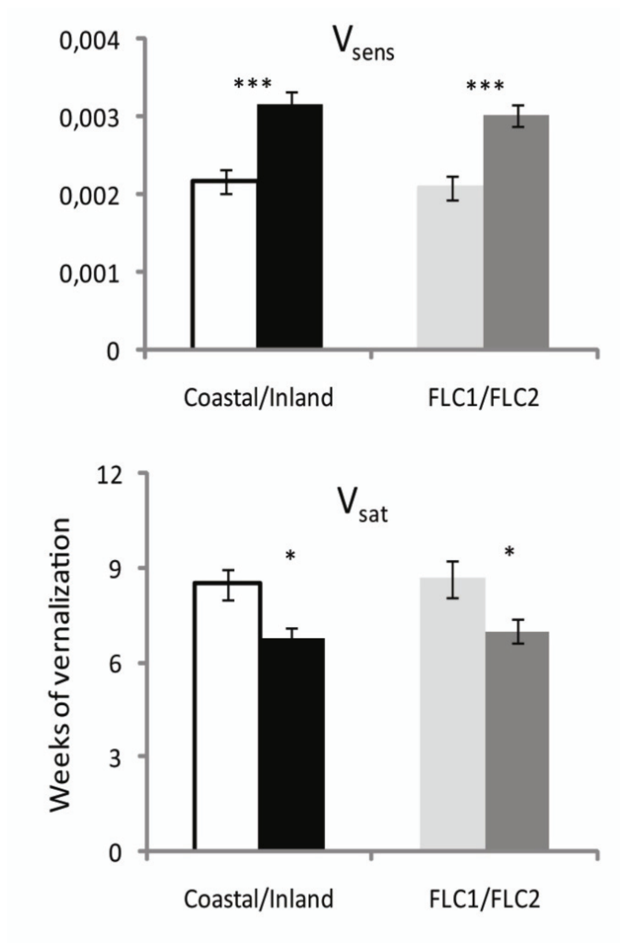


Figure 3

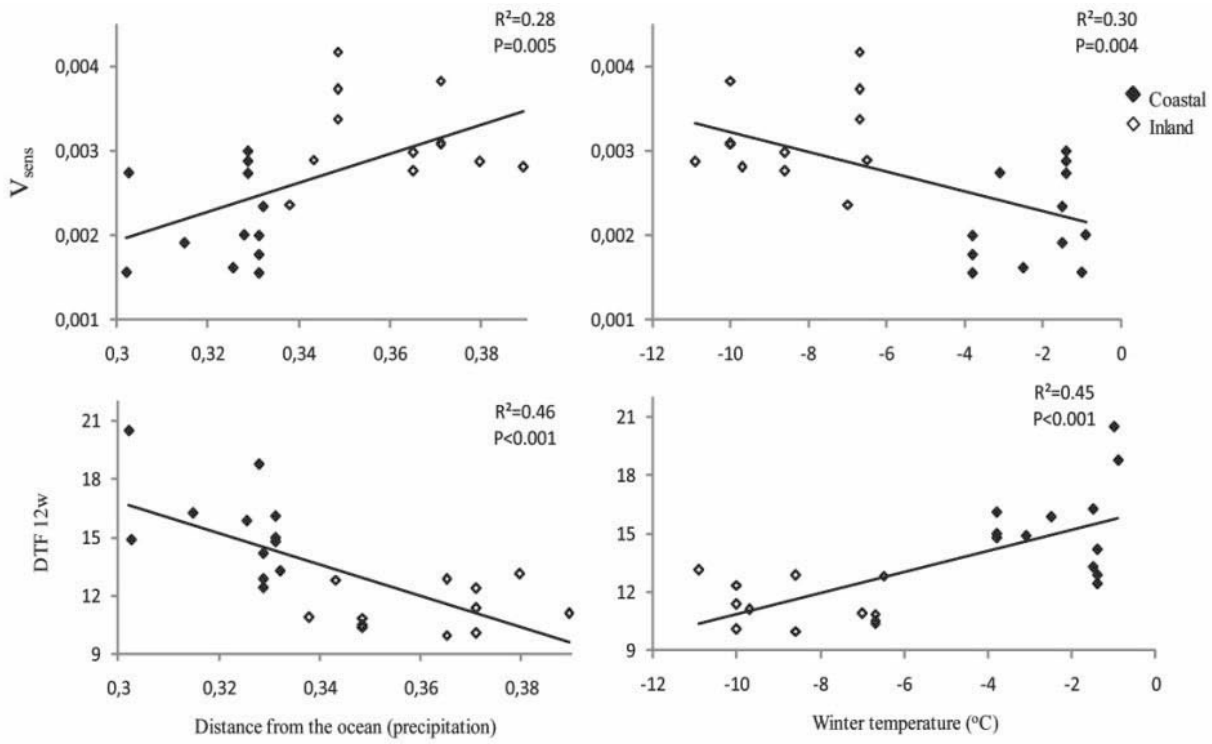


Figure 4

Table 1. Collection data for Norwegian populations of *Arabidopsis thaliana* used in this study. Additional information can be found in Supplementary material (Table S1).

Population name	Latitude(°N)	Longitude(°E)	Altitude (m.a.s.l)	Region of origin	
Vgn-1	68°10'11"	14°13'15"	3	Northern Norway	Coastal
Had-1	68°30'34"	14°43'57"	50	Northern Norway	Coastal
Had-2	68°30'42"	14°45'44"	60	Northern Norway	Coastal
Had-3	68°30'27"	14°53'32"	60	Northern Norway	Coastal
Lod-1	68°23'51"	15°56'6"	100	Northern Norway	Coastal
Lod-2	68°33'56"	16°19'7"	100	Northern Norway	Coastal
Tje-1	68°25'29"	16°22'9"	110	Northern Norway	Coastal
Tje-2	68°25'29"	16°22'9"	80	Northern Norway	Coastal
Lod-3	68°18'17"	15°06'51"	82	Northern Norway	Coastal
Lod-4	68°23'59"	15°08'23"	80	Northern Norway	Coastal
Vgn-2	68°14'38"	14°35'46"	97	Northern Norway	Coastal
Veg-1	65°42'24"	11°56'4"	3	West Coast	Coastal
Veg-2	65°36'45"	11°50'18"	5	West Coast	Coastal
Sog-1	61°12'33"	7°07'39"	8	West Coast	Coastal
Str-1	62°6'12"	7°12'11"	10	West Coast	Coastal
Ors-1	62°12'19"	6°32'1"	2	West Coast	Coastal
Nfro-1	61°34'38"	9°39'42"	735	Central Norway	Inland
Nfro-2	61°34'44"	9°38'28"	816	Central Norway	Inland
Nfro-3	61°34'23"	9°35'44"	579	Central Norway	Inland
Sel-1	61°42'58"	9°26'26"	687	Central Norway	Inland
Vag-1	61°55'36"	8°57'51"	770	Central Norway	Inland
Vag-2	61°51'47"	8°58'43"	640	Central Norway	Inland
Vag-3	61°51'28"	8°56'35"	400	Central Norway	Inland
Lom-1	61°51'42"	8°49'30"	525	Central Norway	Inland
Lom-2	61°51'49"	8°49'2"	700	Central Norway	Inland
Sk-1	61°53'56"	8°15'14"	604	Central Norway	Inland
Sk-2	61°53'18"	8°18'43"	598	Central Norway	Inland
Lom-3	61°40'52"	8°13'51"	850	Central Norway	Inland
Lom-4	61°40'52"	8°13'51"	850	Central Norway	Inland
Eid-1	60°6'44"	12°7'28"	200	Southeastern Norway	Inland
Kon-1	60°9'48"	12°14'27"	400	Southeastern Norway	Inland
Kon-2	60°9'10'	12°5'6"	350	Southeastern Norway	Inland
Kon-3	60°8'53"	12°5'6"	300	Southeastern Norway	Inland
Kvi-1	59°29'38"	8°24'55"	703	Southern Norway	Inland
Kvi-2	59°29'25"	8°24'5"	680	Southern Norway	Inland
Vin-1	59°32'9"	7°56'58"	524	Southern Norway	Inland

Table 2. The relationship between climatic factors at the site of origin of populations, distance from the ocean and flowering time, vernalization sensitivity (V_{sens}) and vernalization saturation (V_{sat}) in *Arabidopsis thaliana*. Determination coefficients of simple linear regression are presented with significance at *=P≤0.05; **=≤0.01; ***=P≤0.001. DTB – days to bolting, DTF – days to flowering, RLN – number of rosette leaves at flowering at 3 vernalization treatments (9, 6 and 3 weeks). Numbers in *italic* indicate negative correlation.

	Distance from the ocean	**Winter temperature**	**Temperature amplitude**
DTB 12w	*0.25***	0.17*	*0.15**
DTF 12w	*0.46****	0.45***	*0.45****
DTB 9w	*0.33***	0.21*	*0.19**
DTF 9w	*0.40****	0.26**	*0.21**
RLN 9w	*0.23**	0.14*	*0.10*
DTB 6w	*0.25***	0.19*	*0.18**
DTF 6w	*0.27***	0.18*	*0.15**
RLN 6w	*0.05*	0.03	*0.01*
DTB 3w	*0.24***	0.22*	*0.28***
DTF 3w	*0*	0	*0.06*
RLN 3w	*0.08*	0.05	*0.03*
V_{sens}DTB	0.28**	*0.30***	0.35***
V_{sens}DTF	0.11	*0.08*	0.15*
V_{sat}DTB	*0.16**	0.23**	*0.18**
V_{sat}DTF	*0.10*	0.08	*0.03*

Table 3. Vernalization sensitivity, saturation and requirement in Norwegian populations of *Arabidopsis thaliana*. V_{sens} – vernalization sensitivity, V_{sat} – vernalization saturation, V_r – vernalization requirement, days to bolting was used to estimate the data, shadowed fields show coastal populations.

Population	$\mathbf{V_{sens}}$	$\mathbf{V_{sat}}$	$\mathbf{V_r}$
Vgn-1	0.00192	9	+
Had-1	0.00274	9	+
Had-2	0.00288	6	+
Had-3	0.00300	6	+
Lod-1	0.00350	9	–
Lod-2	0.00178	9	+
Tje-1	0.00200	9	+
Tje-2	0.00156	9	+
Lod-3	0.00162	12	+
Lod-4	0.00274	6	+
Veg-1	0.00201	9	+
Sog-1	0.00234	9	+
Ors-1	0.00157	9	+
Nfro-1	0.00465	12	–
Sel-1	0.00288	6	+
Vag-2	0.00281	6	+
Vag-3	0.00310	6	+
Lom-1	0.00383	6	+
Lom-2	0.00309	6	+
Sk-1	0.00277	6	+
Sk-2	0.00387	6	–
Lom-4	0.00299	9	+
Eid-1	0.00373	9	+
Kon-2	0.00338	9	+
Kon-3	0.00418	6	+
Kvi-2	0.00290	6	+
Vin-1	0.00237	6	+

Table 4. Associations of *FLC* polymorphisms with vernalization sensitivity (V_{sens}). Only SNPs and indels present in more than one population were tested. Polymorphisms in bold were significantly associated with phenotype after accounting for population structure, polymorphisms underlined are those that were associated with V_{sens} without accounting for population structure at $^{*}P\leq0.05$, $^{**}P\leq0.01$, $^{***}P\leq0.001$. LD = polymorphisms in Linkage Disequilibrium at $P<0.001$ (Fisher's test). Days to bolting were used to estimate V_{sens}.

Polymorphism	F	P	Position	
SNP1**	3.16	0.090	-1751	5'upstream of ATG
SNP2^{LD}	0.02	0.899	-1395	
SNP3*	2.80	0.109	-1394	
SNP4*	2.80	0.109	-1382	
DEL1^{LD}	**4.92**	**0.038**	**-1293**	
SNP5^{LD}	**4.92**	**0.038**	**-1189**	
SNP6^{LD}	0.13	0.726	-1046	
DEL2	1.33	0.261	-790	
SNP7	0.13	0.726	-736	
SNP8^{LD}**	3.93	0.061	-735	
SNP9	1.33	0.261	217	intron1
SNP10^{LD}**	3.93	0.061	259	
DEL3^{LD}**	3.93	0.061	349	
SNP11	1.33	0.261	471	
SNP12	1.33	0.261	515	
SNP13^{LD}**	3.93	0.061	552	
DEL4	0.13	0.726	576	
SNP14*	2.80	0.109	1281	
SNP15	0.13	0.718	1390	
DEL5*	2.80	0.109	1468	
SNP16^{LD}**	3.93	0.061	1747	
SNP17*	2.80	0.109	1790	
SNP18	0.02	0.878	3661	
SNP19^{LD}**	3.93	0.061	4325	intron5
SNP20^{LD}**	3.93	0.061	5535	ex7

Supplementary materials

Table S1. Climate and collection data for populations. * = refered in report of Anders Often: Botaniske undersøkelser av sørberg i Østerdalene. Hedmark. 1997. Rapport nr. 10/97. Fylkesmannen i Hedmark. Miljøvernavdelingen. Mean January and July temperatures (°C) and yearly precipitation (mm) for years 1961-1990. SF – Siri Fjellheim, AB – Anders Bryn, RN – Ragnhild Nestestog.

Pop.	Temperature January	Temperature July	Precipitation yearly	Coll. date	Collector	Notes
Nfro-1	-11.5	15	430	30.06.2003	SF	Southwest slope, in pasture landscape; 10x3m.
Nfro-2	-11.5	15	430	30.06.2003	SF	Southern rock, in pasture landscape; 3x2m.
Nfro-3	-11.5	15	430	30.06.2003	SF	Southern slope, roadside; 10x6m.
Sel-1	-10.9	15	430	01.07.2003	SF	Southwest crag in enclosed pasture by the old garden; 20x10m.
Vag-1	-9.7	14	370	01.07.2003	SF	Southwest slope, roadside; 5x2m.
Vag-2	-9.7	14	370	02.07.2003	SF	Southwest slope, roadside; 20x2m.
Vag-3	-10.0	12	495	02.07.2003	SF	Steep southern slope with crags; scree and rocky ground; 100x50.
Lom-1	-10.0	12	495	02.07.2003	SF	Southern slope, roadside; 50x5m.
Lom-2	-10.0	12	495	02.07.2003	SF	Steep southwest slope with crags, scree; on a shelf of dry ground.
Sk-1	-8.6	12	548	02.07.2003	SF	Southwest cliff in old cultivated landscape.
Sk-2	-9.4	14	317	02.07.2003	SF	Steep southern rocky slope with crags, scree.
Lom-3	-8.6	12	548	03.07.2003	SF	Cultivated landscape, sheep walk, dry ground; 50x50m.
Lom-4	-8.6	12	548	03.07.2003	SF	On the small rock shelf, in the area of small mountainside, rocks.
Vgn-1	-1.5	13	1500	07.07.2003	SF	Southern slope between the road and the sea, 15m from the sea; 10x15m.
Had-1	-1.4	12	1100	07.07.2003	SF	Grass-covered, southwestern slope with cliffs and scree surrounded by sheep pasture.
Had-2	-1.4	12	1100	07.07.2003	SF	Southern pasture; 20x20m.
Had-3	-1.4	12	1100	07.07.2003	SF	Southern, abrupt rocks in small, dry patches of cultivated landscape.
Lod-1	-3.1	13	1600	08.07.2003	SF	Under precipitous southern slope, craggy. Individuals 1-23: 10x5m.
Lod-2	-3.8	12	1045	08.07.2003	SF	On cliffs in precipitous slope.
Tje-1	-3.8	12	1045	08.07.2003	SF	On the dry ground, small shelf under a, southern, precipitous cliff.
Tje-2	-3.8	12	1045	08.07.2003	SF	30m below population Tje-1, rocky ground; 2x2m.
Lod-3	-2.5	13	1180	09.07.2003	SF	Southern cliff, small shelf on the top of the rocky ground; 5x3m.
Lod-4	-3.1	13	2015	10.07.2003	SF	Southern sheep pasture/slope (quite steep), dry rocky ground; 8x3m.
Vgn-2	-1.5	13	1500	10.07.2003	SF	Small shelf on precipitous scree under southern cliff; 30x20m.
Sog-1	-1.5	15	1025	15.07.2003	SF	The edge of the road/parking place; 15x2m.
Str-1	-0.8	13	1351	16.07.2003	SF	Crag at the edge of road in Geiranger center.
Ors-1	-1.0	14	2040	15.07.2003	SF	Roadsige, by the ferry landing site in Lekneset; 10x2m.
Veg-1	-0.9	13	1120	24.06.2003	AB	Dry slope with shallow mulch
Veg-2	-0.9	13	1120	26.06.2003	AB	On the beach rock.
Eid-1	-6.7	15	740	16.06.2003	SF	*Southern rock, on the ground under precipitous, partly projecting shale cliff.
Kon-1	-6.7	15	740	16.06.2003	SF	*Southern rock, rocky ground in the mixed forest.
Kon-2	-6.7	15	740	27.06.2003	SF	*Southern rock, rocky ground, scree, surrounded by residential settlement.
Kon-3	-6.7	15	740	16.06.2003	SF	*5-20m high crag surrounded by natural forest.
Kvi-1	-6.5	16	820	09.07.2003	RN	Moss-covered rock near coniferous forest, pasture-land near the farm.
Kvi-2	-6.5	16	820	09.07.2003	RN	Roadside.
Vin-1	-7.0	14	910	09.07.2003	RN	Scree in the pasture area; 3x5m.

Table S2. Primers used for amplification and sequencing of different parts of *FLC*.

PCR/Sequencing Primer name	Sequence (5' →3')
FLC-PROM2F	TGATAATAGTAAGGCTTGAACTTGG
FLC-PROM2R	CGAAAAGAAGTGGGTGAAACT
FLC-PROM3F	AGTTTCACCCACTCCGAGTG
FLC-PROM3R	TTCGTCGGGCCAGATATTTT
FLC-5'UTR-F	GCCCGACGAAGAAAAAGTAG
FLC-5'UTR-R	GGATGCGTCACAGAGAACAG
FLC-EX1F	AAAACAGAAGATAAAAGGGGGAAC
FLC-EX1R	GCGACTTGAACCCAAACCT
FLC-INTR1-1F	CTGTTCTCTGTGACGCATCC
FLC-INTR1-1R	GCACACGACGATTGTGATTC
FLC-INTR1-2F	GAATCACAATCGTCGTGTGC
FLC-INTR1-2R	TTTGTTTTGTTAATAGATATCAGTCC
FLC-EX2-6F	CACACAACACGCAGTGCTTA
FLC-EX2-6R	GTTTCCAGTGGCCTTTTCAA
FLC-EX7F	TACAAACGCTCGCCCTTATC
FLC-EX7R	TGGTTGTTATTTGGTGGTGTG

	5'upstream of ATG																				exon1																intron1				
Nt. Position	-368	-367	-329	-328	-323	-311->65	-296	-286	-260	-247->-183	-177	-176	-144	-142	-115	-96->-92	-61	-53	-52	-18	-11	4	18	60	65	101	109	125	193	-//-	933	934	946*	949*	951*	963*	966*	998	1000	1001	1010	1046
																						Ser/Ala	NAC		Arg/Gln	Lys/Thr	Glu/Gln	Ser/Thr	Phe/Ile		NAC	6aa	Met*/Val	Val*/Ile	Val*/Ile	Ile*/Met	Phe*/del					
Col-0	T	A	C	A	T	–	G	A	T	ND	A	C	T	A	–	T	G	T	A	T	G	T	G	–	G	A	G	G	T		G	DEL 4	–	–	–	A	C	C	T	C	A	–
L*er*						IND1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–						ND	A	G	C							–
Nfro-1	G	C	G	G	C	–	A	T	C	DEL2	–	T	A		–	DEL3	A	C	T	C	T	G	C	IN3	A	C	C	C	A		A	ND	G	A	T	G	DEL5	T	G	A	G	INS4
Sel-1						–						T			–									–								ND	A	G	C							–
Vag-2						–						T			–									–								ND	A	G	C							–
Vag-3						–						T			–									–								ND	A	G	C							–
Lom-1						–						T			–									–								ND	A	G	C							–
Lom-2						–						T			–									–								ND	A	G	C							–
Sk-1						–						T			–									–								ND	A	G	C							–
Sk-2						–						T			–									–								ND	A	G	C							–
Lom-4						–						T			–									–								ND	A	G	C							–
Vgn-1						–						T			–									–								ND	A	G	C							–
Had-1						–						T			IN2									–								ND	A	G	C							–
Had-2						–						T			IN2									–								ND	A	G	C							–
Had-3						–						T			IN2									–								ND	A	G	C							–
Lod-1						–						T			–									–								ND	A	G	C							–
Lod-2						–						T			–									–								ND	A	G	C							–
Tje-1						–						T			–									–								ND	A	G	C							–
Lod-3						–						T			–									–								ND	A	G	C							–
Lod-4						–						T			–									–								ND	A	G	C							–
Vgn-2						–						T			–									–								ND	A	G	C							–
Sog-1						–						T			IN2									–								ND	A	G	C							–
Ors-1						–						T			–									–								ND	A	G	C							–
Veg-1						–						T			IN2									–								ND	A	G	C							–
Eid-1						–						T			–									–								ND	A	G	C							–
Kon-1						–						T			–									–								ND	A	G	C							–
Kon-2						–						T			–									–								ND	A	G	C							–
Kon-3						–						T			–									–								ND	A	G	C							–
Kvi-1						–						T			IN2									–								ND	A	G	C							–
Kvi-2						–						T			IN2									–								ND	A	G	C							–
Vin-1						–						T		C	–									–								ND	A	G	C							–

Table S3. Polymorphisms in the *FRI* gene. IND1 = deletion of 376 bp combined with insertion of 31 bp, DEL2 = deletion of 65 bp, IN2 = insertion of TAG, DEL3 = deletion of 5 bp, IN3 = insertion CAC, DEL4 = deletion of 16 bp, DEL5 = deletion of 9 bp (span exon 1 and intron 1), INS4 = insertion of 20 bp, – = absence of the corresponding indel or nucleotide, open fields = nucleotides identical to Col-0, NAC = no amino acid change, * = nucleotide and amino acid in L*er* due to deletion in Col-0; 6aa = amino acids not present in Col-0 , ND = no deletion.

	5'upstream of ATG													intron1															ex4	intron5		ex7	3'UTR
Nt. Position	-1751	-1501	-1483	-1395	-1394	-1382	-1364	-1293	-1189	-1046	-790	-736	-735	217	259	349	471	489	515	552	576	1281	1390	1468	1747	1790	- //-	3661	4130	4325	- //-	5535	5670
																													NAC			NAC	
Col-0	G	A	G	A	T	T	T	C	G	T	A	A	A	G	G	-	T	A	A	G	C	A	G	ND	G	G		T	C	T		G	-
Ler				T		A				C		C		A					C			G		DEL30					T				IN
Nfro-1					A								C	A	T	T			C	C		G		DEL30								A	IN
Sel-1					A						-						A									A							
Vag-2				T		A				C		C		A					C		-	G	T	DEL30									IN
Vag-3				T		A				C		C		A					C		-	G	T	DEL30									N
Lom-1				T		A				C		C		A					C		-	G	T	DEL30									IN
Lom-2					A					C		C		A					C		-					A		G					
Sk-1				T		A				C		C		A					C		-					A							IN
Sk-2						A				C		C		A					C		-	G	T	DEL30				G					N
Lom-4					A								C	A	T	T			C	C		G		DEL30	A					G		A	IN
Vgn-1	A	G				A		-	C				C	A	T	T			C	C		G		DEL30	A					G		A	IN
Had-1				T		A				C		C		A					C		-	G	T	DEL30									N
Had-2				T		A				C		C		A					C		-	G	T	DEL30									IN
Had-3				T		A				C		C		A					C		-	G	T	DEL30									IN
Lod-1	A					A		-	C				C	A	T	T			C	C		G		DEL30	A					G		A	N
Lod-2	A					A		-	C				C	A	T	T			C	C		G		DEL30	A					G		A	N
Tje-1	A					A		-	C				C	A	T	T			C	C		G		DEL30	A					G		A	IN
Lod-3	A					A		-	C				C	A	T	T			C	C		G		DEL30	A					G		A	N
Lod-4	A					A		-	C				C	A	T	T			C	C		G		DEL30	A					G		A	IN
Vgn-2	A					A		-	C				C	A	T	T			C	C		G		DEL30	A					G		A	N
Sog-1	A					A		-	C				C	A	T	T			C	C		G		DEL30	A					G		A	IN
Ors-1				T		A				C		C		A					C		-	G	T	DEL30									IN
Veg-1						A		-	C				C	A	T	T			C	C		G		DEL30	A					G		A	IN
Eid-1					A						-						A									A							N
Kon-1			A									C		A				T				G		DEL30									N
Kon-2					A		G				-						A									A							N
Kon-3					A						-						A									A							
Kvi-1	A					A		-	C				C	A	T	T			C	C		G		DEL30	A					G		A	N
Kvi-2	A					A		-	C				C	A	T	T			C	C		G		DEL30	A			G		G		A	
Vin-1					A						-						A									A							
	*	*	*	*			*	*	*		*		*					*					*					*		*			

Table S4. Polymorphisms in the *FLC* gene. Open fields = nucleotides identical to Col-0, DEL30 = deletion of 30 bp, ND = no deletion, N = no data, IN = insertion AG, * = new polymorphisms.

PAPER II

Photoperiodic sensitivity of flowering time in local populations of *Arabidopsis thaliana* is associated with proximity to the coast and altitude, and transcript levels of circadian clock regulated genes

Anna M. Lewandowska-Sabat[1], Siri Fjellheim[1], Carl Gunnar Fossdal[2], Jorunn Elisabeth Olsen[1], Odd Arne Rognli[1]

[1]Norwegian University of Life Sciences, Department of Plant and Environmental Sciences, P.O. Box 5003, NO-1432 Ås, Norway

[2]Norwegian Forest and Landscape Institute, P.O. Box 115, NO-1431 Ås, Norway

Corresponding author

E-mail: odd-arne.rognli@umb.no

Telephone: +47 64 96 55 78

Fax: +47 64 96 59 01

Abstract

Adaptation of plants to photoperiods that vary substantially within their geographic range is essential for flowering at the proper season. The photoperiodic response in *Arabidopsis thaliana* is controlled by a number of photoreceptors and circadian clock regulated genes. The objective of this study was to examine variation in photoperiodic control of flowering time among natural populations of *A. thaliana* from high-latitude and high-altitude locations in Norway, i.e. from the northernmost distribution range of the species. We found that photoperiodic sensitivity of flowering time is not a result of a latitudinal cline but is associated with distance from the ocean. Coastal populations showed stronger photoperiodic sensitivity than continental populations. This is likely an adaptation for correctly sensing the onset of spring in regions with relatively mild and unpredictable winter climates as opposed to continental climates with more stable winter conditions regarding temperature fluctuations. Variation in transcript levels of *CRYPTOCHROME 2* (*CRY2*), *TIMING OF CAB EXPRESSION 1* (*TOC1*), *GIGANTEA* (*GI*) and *CONSTANS* (*CO*) among populations and photoperiods was associated with flowering time and photoperiodic sensitivity. Our results suggest that natural selection favors highly photoperiod sensitive genotypes of *A. thaliana* in coastal regions by altering the mRNA level of circadian clock regulated genes.

Key words: *Arabidopsis thaliana*, *CO*, *CRY2*, circadian clock, flowering time, *GI*, natural populations, photoperiodic sensitivity, *TOC1*.

Introduction

Photoperiod is an environmental cue that many organisms use to regulate seasonal changes in behaviour, migration and reproduction (Kardailsky *et al*., 1999; Yano *et al*., 2000; Bradshaw and Holzapfel 2001; O'Malley and Banks 2008; Zhang *et al*., 2008). At temperate and subarctic latitudes, photoperiod, light quality and temperature are major environmental signals that plants sense in order to synchronize their flowering time with the changing seasons (Bradshaw *et al*., 2004; Putterill *et al*., 2004). Variation in flowering time of plants contributes to local adaptation to different growth conditions and hence clinal variation in flowering time is commonly believed to be a sign of adaptive evolution.

Arabidopsis thaliana is a facultative long-day plant and flowering time in response to photoperiod has been intensively studied by many research groups (Lee and Amasino 1995; Onouchi and Coupland 1998; Simpson and Dean 2002; El-Assal *et al*., 2003; Balasubramanian *et al*., 2006; Schwartz *et al*., 2009). Despite the large number of molecular studies of flowering time in *A. thaliana*, there is still a lack of knowledge about the geographical patterns of photoperiodic responses. It has been hypothesized that populations originating from higher latitudes and adapted to harsh climate will be more sensitive to changes in photoperiod than populations from lower latitudes, since it is crucial for them to reproduce at a proper time of the year. However, this hypothesis has not been thoroughly tested. Banta *et al*. (2007) observed that the fitness of accessions from Northern and Southern Europe did not differ under different photoperiods. Moreover, Samis *et al*. (2008) observed that photoperiodic responses in Eurasian *A. thaliana* accessions are correlated with longitude, and not with latitude.

Control of development and flowering time involves complex interactions between the circadian clock and the photoperiodic pathway genes. Photoperiodic control of flowering in *A. thaliana* is regulated by light signals perceived by phytochrome (PHYA-E) and cryptochrome photoreceptors (CRY1 and CRY2) sensing light of red-far-red and blue wavelengths, respectively (Sharrock and Quail 1989; Franklin *et al*., 2005; Li and Yang 2007). Also ZEITLUPE (ZTL), FLAVIN-BINDING KELCH REPEAT F-BOX 1 (FKF1) and LOV KELCH PROTEIN2 (LKP2) families mediate blue-light-induced responses in *A. thaliana* (Li and Yang 2007; Jiao *et al*., 2007). The photoreceptors are involved in entraining the circadian clock. The circadian clock is a daily rhythmic mechanism and a part of a large network that enables living

organisms to synchronize their endogenous responses with the diurnal changes in light and temperature conditions. The circadian oscillator gives adaptive advantage by providing a time measurement regardless of ambient temperature (Michael *et al.*, 2003; Edwards *et al.*, 2005). The fundamental feedback loops of the circadian oscillator consist of *TIMING OF CAB EXPRESSION 1* (*TOC1*), *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), *LATE ELONGATED HYPOCOTYL* (*LHY*), *GIGANTEA* (*GI*) and *LUX ARYTHMO* (*LUX*) (Alabadi *et al.*, 2001; Gardner *et al.*, 2006; Harmer 2009).

Flowering in response to long days (LD) in *A. thaliana* is controlled by two main loci acting downstream of the light receptors and the circadian clock, the B-box-type zinc finger protein CONSTANS (CO) and FLOWERING LOCUS T (FT), a strong candidate for a mobile florigen signal (Putterill *et al.*, 1995; Kardailsky *et al.*, 1999; Corbesier *et al.*, 2007; Turck *et al.*, 2008). Photoperiod controls both the mRNA and the protein levels of CO, which is regulated by upstream genes that follow a circadian expression pattern. Flowering is promoted under LD when high levels of *CO* mRNA coincides with light periods (Suárez-López *et al.*, 2001; Yanovsky and Kay 2002) and the CO protein, which is stabilized in light, then directly induces expression of the *FT* gene (Samach *et al.*, 2000; Valverde *et al.*, 2004).

It has previously been shown that phenotypic variation in the flowering time and hypocotyl elongation, as well as photoperiod pathway activation are associated with amino acid substitution at *PHYA*, *PHYB*, *PHYC, PHYD* and *CRY2* in *A. thaliana* (Aukerman *et al.*, 1997; Maloof *et al.*, 2001; El-Assal *et al.*, 2001; Balasubramanian *et al.*, 2006; Filiault *et al.*, 2008). Loss-of-function alleles of *PHYC* and *PHYD* cause early flowering under non-inductive photoperiod (Aukerman et al. 1997; Balasubramanian et al. 2006), and high levels of *CRY2* protein in short days (SD) appear to cause day length insensitivity in the accession from Cape Verde Islands (Cvi) compared to L*er* (El-Assal *et al.*, 2001; Mockler *et al.*, 2003). Attribution of adaptive features to photoreceptors and their crucial role in entrainment of the circadian clock and control of photoperiodic flowering make these genes excellent candidates for studying variation also at the mRNA level.

The day length in the growing season in Norway varies between 20-24 h at northern latitudes (68°N) compared to 17-19 h at more southern, subarctic latitudes (59-62°N). However, along the coast the seasonal pattern of temperature and precipitation are rather similar irrespective of latitude in these regions. Due to temperature fluctuations during winter season in

coastal regions, the precipitation comes not only in the form of snow but more often in the form of rain, sleet and hail. Therefore, very often these conditions mimic those in spring. On the contrary, in inland regions stable low temperatures during winter season result in long lasting snow cover. Thus, at a given latitude climatic transects are present and populations from coastal and inland regions are subjected to very different climates. In this study we have screened *A. thaliana* populations from the northern distribution boundary of the species (59-68°N) for photoperiodic flowering time responses and investigated which environmental cues determine local adaptation of *A. thaliana.* Furthermore, we try to explain variation in photoperiodic flowering time responses at the molecular level by transcript profiling of key circadian clock regulated genes involved in the photoperiodic pathway.

Material and methods

Plant materials and photoperiodic response experiment

Twenty-nine local populations of *Arabidopsis thaliana* (L.) Heyhn, collected from continental and coastal locations in Norway, were used in this study (Table 1). Details about the populations can be found in Lewandowska-Sabat et al. (unpublished results, Paper I). Of these, 10 populations from 5 geographically distant regions were used for screening photoperiodic responses while candidate genes were sequenced from individuals from all populations (Fig. 1, Table 1).

Screening of response to 5 different photoperiods (8, 16, 19, 21 and 24 h) was performed using 3 maternal families per population and 5 individual plants per family in each treatment. Seeds multiplied through a previous vernalization experiment (Lewandowska-Sabat *et al.*, unpublished results, Paper I) were sown in soil (Hasselfors Garden AB, Örebro, Sweden) in 6.5 cm pots and stratified at 4°C for 4 days in darkness to equalize germination time. After stratification the plants were placed in a chamber at 23°C and an 8 h photoperiod with a photon flux density of 150 ± 10 µmol m^{-2} s^{-1} at 400-700 nm for 13 days. Plants were then vernalized at 4°C under an 8 h photoperiod with a photon flux density of 50 ± 5 µmol m^{-2} s^{-1} for 14 weeks. The light was produced by HQI lighting systems (LU400/XO/T/40 Philips General Electric,

Hungary). The cold period was applied in order to saturate the vernalization requirement of all genotypes and isolate the effect of photoperiod *per se*. After vernalization the plants were transferred to 5 growth chambers at 16°C with 5 different day/night regimes: short day (SD): 8 h PAR light (Photosynthetically Active Radiation)/16 h dark, 16 h PAR light/8 h dark (LD), 19 h light (16 h PAR light and 1.5 h incandescent light before and after PAR light period)/5 h dark, 21 h light (16 h PAR light and 2.5 h incandescent light before and after PAR light period)/3 h dark and 24 h light (16 h PAR light followed by 8 h incandescent light). The irradiance of the PAR and incandescent light was 150 ± 10 µmol m^{-2} s^{-1} and 10 ± 2 µmol m^{-2} s^{-1}, respectively, and light was produced by HQI lighting systems and incandescent bulbs (LU400/XO/T/40 Philips General Electric, Hungary and Osram, Munich, Germany, respectively). Flowering time was scored daily as days to bolting (DTB, date when the length of the bolting stem was approximately 1 cm) and days to flowering (DTF, date of first flower to open). Plants were watered every week during vernalization and twice a week during germination and exposure to the different photoperiods.

Sequencing of the candidate genes

Genomic DNA was extracted from leaf tissue harvested in liquid nitrogen and stored in a -70°C using Aqua Pure Genomic DNA Isolation Kit (Bio-Rad, Hercules, CA, USA) and PCR was performed with DyNAzyme II DNA Polymerase (Finnzymes, Espoo, Finland) using a standard procedure. Promoter regions of *CRY2*, *CO* and *TOC1*, and the coding regions of *PHYC*, *CO*, *SPINDLY* (*SPY*, part of the functional domain from exon 1 to exon 5) were sequenced from 29 populations (Table 1) as well as Col-0 and L*er*. *CRY1* (exon 1 to exon 3) and *CRY2* (exon 1 to exon 3) were sequenced from 14 populations and Col-0. One individual plant per population was used. Primers were designed using Primer3 (Rozen and Skaletsky 2000) and reference sequences published in TAIR (http://www.arabidopsis.org/) (Table S1). PCR products were directly sequenced using BigDye terminator v.3.1 Cycle Sequencing reactions kit (Applied Biosystems, Foster City, California, USA). Sequences were run on ABI Prism 3100 Automated Capillary DNA Sequencer (Applied Biosystems Foster City, California, USA). All sequences were aligned and analysed in Sequencher 4.7 DNA sequence analysis software (Gene Codes, Ann Arbor, Michigan, USA).

Real-time RT-PCR

For Real-time RT-PCR analyses of transcript levels of *GI*, *FKF1*, *PHYA*, *CRY2*, *TOC1* and *CO*, 2-3 young leaves was harvested from each plant at the end of the third day after plants were returned from vernalization to an 8, 16 or 24 h photoperiod. For studying effects of SD vs. LD, transcript levels at the end of the 8 and 16 h photoperiods were compared. For studying diurnal changes in transcript levels plants were harvested at 1st, 9th, 13th, 17th and 20th h of the fifth day in 16 h photoperiod, and finally for studying transcript level changes over time, expression at the end of day 5 and 3 was compared. Transcript levels were analysed in 6 representative populations; Lod-1, Tje-1, Ors-1, Sk-2, Eid-1, and Kvi-1 (Table 1) and in Col-0. SD vs. LD transcript levels and transcript level changes over time were not analyzed for population Lod-1. Total RNA was extracted using RNeasy Mini Kit (Qiagen, Hilden, Germany) and purified using RNeasy MinElute Cleanup Kit (Qiagen). RNA quality was analyzed using Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and RNA quantity was measured tree times by a NanoDrop® ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and averaged for real-time PCR quantification. 500 ng of total RNA was reverse transcribed, cDNA was then diluted 10 times and 5 µl were used as a template for each qPCR reaction. First strand cDNA synthesis and qPCR was performed using SuperScript III Platinum SYBR Green Two-Step qRT-PCR Kit with ROX (Invitrogen, Carlsbad, CA, USA) according to the manufacture recommendations using 25 µl reaction volumes. Transcript levels were analyzed using a AB7500 Real-time PCR machine (Applied Biosystems) according to the 7500 System software: 50°C 2 min, 95°C 2 min, and 40 cycles of 95°C 15 s, and 60°C 1 min. All primers were used at a concentration of 0.2 µM and a validation experiment for each product was conducted. Target specificity of each primer pair was verified by dissociation curve analyses (product melting curve). As an endogenous control we used transcript analyses of *βtubulin* and *actin*. Primer sequences are presented in Table S2. Each RNA sample was assayed in duplicate and each experiment was repeated twice with independent biological material. Transcript levels were calculated relative to *actin* using a comparative threshold cycle method (ΔCT method). Fold change in transcript levels over time and between different treatments was calculated using the $\Delta\Delta C_t$ method.

Statistical and genetic analyses

Climate data for the collection sites were obtained from The Norwegian Meteorological Institute (http://met.no/English/ and http://www.senorge.no/startpage.aspx). Yearly means of precipitation (P_{year}) and monthly means of temperature from January (T_{Jan}) for years 1961-1990 were used as a measure for coastal and inland climate differences. Reciprocal log transformed P_{year} was used as a proxy for distance from the ocean and T_{Jan} was used as estimator of winter temperature and winter duration. The photoperiodic parameter was calculated as the mean day length (DL) for the first 30 days in the spring with average daily temperature above 5°C and no snow cover (Table S3).

Flowering time parameters (DTB and DTF) for the 5 photoperiod treatments were regressed against altitude, winter temperature, distance from the ocean, latitude and DL using individual data for each population. Photoperiodic sensitivity (P) for each population was estimated by regressing mean values for DTB and DTF on hours of photoperiod, i.e. $P_{8/16}$ denotes the regression coefficient for response to 8 and 16 h; $P_{8/24}$ the regression coefficient for response to 8, 16, 19, 21, and 24 h; and $P_{16/24}$ the regression coefficient for response to 16 and 24 h photoperiod. Populations with steeper slopes have largest photoperiodic sensitivity. The regression coefficients were multiplied by (-1) to facilitate interpretation. Photoperiodic sensitivity was regressed against altitude, winter temperature, distance from the ocean and latitude. In all analyses simple linear regression was used by using phenotypic data as response variables and climatic factors as predictors.

Populations were classified into different groups based on latitude: i) Northern – populations from Lofoten, and Southern – populations from West Coast, Gudbrandsdalen, Kongsvinger and Telemark; and ii) distance from the ocean: coastal including populations both from subarctic and arctic latitudes (from West Coast and Lofoten) and inland including populations from Gudbrandsdalen, Kongsvinger and Telemark (see Fig.1 and Table 1 for geographic origin of populations). Phenotypic differences between populations, treatments and geographical groups as well as differences between populations in transcript levels were tested statistically using ANOVA for DTB, DTF, $P_{8/16}$, $P_{8/24}$, and $P_{16/24}$. All regressions and ANOVA analyses were performed using Minitab 15 Statistical Software (Minitab Inc. State College, PA, USA).

DNA sequence polymorphism analyses of *PHYC*, *CRY2*, *CO*, *TOC1*, and *SPY* were performed using DnaSP v.4.50.3 software (http://www.ub.es/dnasp/). Due to lack of variation, *CRY1* was not included in the analyses. Phylogenetic analyses of *CO* were conducted using neighbor-joining in MEGA version 4 (Tamura *et al*., 2007) and maximum likelihood in TREEFINDER (Jobb 2008).

For association tests of SNPs with mean flowering time and photoperiodic sensitivity only polymorphisms present in more than one population were used. Association tests were performed using the general linear model incorporated in ANOVA and coefficients that statistically describe a population's inferred ancestry as covariates (Korves *et al*., 2007; Samis *et al*., 2008). One inferred ancestry coefficient at the time was used in the model. These coefficients were estimated using STRUCTURE 2.2.3 (Pritchard *et al*., 2000; Falush *et al*., 2003, through the BIOPORTAL http://www.bioportal.uio.no/) and SNP data on 31 Norwegian populations of *A. thaliana* (Lewandowska-Sabat *et al*., unpublished, Paper III). STRUCTURE was run for 1,000,000 iterations after a burn-in length of 100,000 iterations for deduction of each of K populations. Five runs were considered for each K, with K ranging from 1 to 15. The admixture model was run with allele frequencies uncorrelated. Number of hypothetical ancestral populations was estimated at K=5 (Evanno *et al*., 2005) using STRUCTURE-SUM (Ehrich 2006) in R 2.9.2 (R Development Core Team 2009). Mean over 5 runs was taken in order to find coefficients describing each population. Ancestry coefficients describe the proportion of a population's genome that comes from each of these inferred ancestral populations. Including ancestry coefficients in association tests will determine if variation in the traits is due to population structure.

Results

Variation in flowering time and photoperiodic sensitivity

All of the populations flowered in SD (8 h photoperiod); however, flowering was delayed by 2.7-12.2 days as compared to longer photoperiods (Fig. 2, 3). Since DTB and DTF were highly correlated only data for DTB are presented. Significant variation in DTB was found between populations in all treatments (Fig. 3, statistical data not shown). The four inland populations,

Nfro-1, Lom-4, Eid-1 and Kon-3 and all coastal populations, Vgn-1, Lod-1, Tje-1, Ors-1 flowered significantly later in SD than in LD (16 h photoperiod) (2.0-11.1 days, $P<0.001$). Also, three inland populations, Nfro-1, Sk-2 and Kvi-1, flowered significantly later in 16 than 24 h of photoperiod (1.4-4.3 days, $P<0.001$, Fig. 3). Coastal populations flowered significantly later than inland populations in all photoperiods (Fig. 2). The difference in mean flowering time between Northern and Southern Norwegian populations was small but significant (16.0±2.04 and 12.0±0.99, respectively, $P<0.001$). Photoperiodic sensitivity was significantly higher in coastal than in inland populations ($P=0.009$, $F=11.51$ for $P_{8/16}$ and $P=0.003$, $F=18.15$ for $P_{8/24}$). No significant difference in photoperiodic sensitivity was found between Northern and Southern Norwegian populations and no difference in $P_{16/24}$ was found between any groups.

Regression analyses revealed that flowering time was highly associated with altitude, winter temperature and distance from the ocean at the collection sites in all treatments whereas association with latitude and day length in spring (DL) was much lower (Table S4).

Photoperiodic sensitivity ($P_{8/24}$) was highly associated with altitude, winter temperature and distance from the ocean (R^2=0.62, 0.85 and 0.89 respectively, $P<0.01$), but not with latitude (see Fig. 4, Table 2 for coefficients of determination and Table 3 for photoperiodic sensitivity estimates). There was no association of $P_{16/24}$ with any of the variables tested.

CO sequence variation is associated with photoperiodic sensitivity

Three groups of *CO* haplotypes were distinguished (based on phylogenetic tree, Fig. S1). The first and second group mostly consists of coastal populations while the third group consists of only inland populations (Table 4). Populations Nfro1, Sk-2, Sog-1 and Kvi-2 remained ungrouped. However, only SNP11 in *CO* (position 951, Table 4) was associated with DTB 8 h ($P=0.024$) and photoperiodic sensitivity ($P=0.013$ for $P_{8/16}$ and $P=0.020$ for $P_{8/24}$). After including inferred ancestry coefficients in the model, associations of DTB 8 h with *CO* polymorphism became non significant ($P=0.164$). $P_{8/16}$ remained marginally significant ($P=0.049$) and $P_{8/24}$ marginally insignificant ($P=0.074$). No association between haplotypes/polymorphisms and flowering time or photoperiodic sensitivity was observed for the remaining candidate genes.

The highest haplotype diversity (Hd) was observed in *PHYC* (0.818) and *CO* (0.739) and the lowest in *CRY2* (0.133). Nucleotide diversity was highest in *CRY2* (0.00445), lower in *TOC1*

(0.00147), *PHYC* (0.00106) and lowest in *CO* (0.00086) and *SPY* (0.00070). Very little variation (most of them are known polymorphisms) were found in *CRY1* and only population Eid-1 has one silent mutation and one amino acid change in exon 2 of *CRY2* (Val_{126} and Gln_{127}/Ser). Therefore these regions were not included in polymorphism and phylogenetic analyses. Overview of SNPs and indels detected for *TOC1*, *CRY2*, *SPY* and *PHYC* in these populations are presented in Tables S5-S8.

Expression of GI and CO is associated with flowering time under long days

Among six genes tested in the expression study *CRY2, TOC1*, *GI* and *CO* showed significant differences in diurnal transcript levels (Fig. 5), in fold change of transcript levels over time (Fig. 6) and in transcript levels in 8 h compared to 16 h photoperiod (Fig. 7).

Coastal populations (Lod-1, Tje-1 and Ors-1) had significantly lower transcript levels of *GI* and *CO* than inland populations (Sk-2, Eid-1 and Kvi-1) at the end of the day (P<0.001; 13th h, Fig. 5b; *CO* also at 9th h, Fig. S2). *GI* and *CO* transcript levels were associated with DTB in 16 h (P=0.008, P=0.001, respectively). Populations from the Northern Norway had significantly lower transcript levels of *CRY2* and *TOC1* than populations from the Southern Norway at the end of the day (P<0.001; 13th h, Fig. 5b; *CRY2* also at 9th h, *TOC1* also at 1st and 9th h; Fig. S2). No association between *CRY2* and *TOC1* transcript levels and DTB in 16 h was found.

Significant differences in fold change of transcript levels over time were found (Fig. 6). Coastal populations (Tje-1 and Ors-1) showed significantly less change in transcript levels of *GI* and *CO* over time than inland populations (Sk-2, Eid-1 and Kvi-1; P<0.001). Change in transcript levels of *GI* and *CO* was associated with DTB in 16 h (P=0.014, P=0.003, respectively). Transcript level changes of *CRY2* and *TOC1* were lower in the Northern Norwegian Tje-1 population than in Southern Norwegian populations (P<0.001). No association between *CRY2* and *TOC1* transcript level changes over time and DTB in 16 h was found.

Expression of CRY2, TOC1 and GI is associated with photoperiodic response

Significant differences in fold changes of transcript levels in SD compared to LD were found (Fig. 7). Coastal populations (Tje-1 and Ors-1) had significantly lower transcript levels of *CRY2*,

TOC1 and *GI* in SD compared to LD than inland populations (Sk-2, Eid-1 and Kvi-1; $P<0.001$). Fold changes of *CRY2*, *TOC1* and *GI* transcript levels between SD and LD were associated with $P_{8/16}$ ($P=0.006$, 0.004, 0.024, respectively). The Northern Norwegian population Tje-1 had lower *CO* transcript level in SD than in LD compared to Southern Norwegian populations ($P<0.001$). No association between changes of *CO* transcript levels between SD and LD with $P_{8/16}$ was detected. No significant differences in transcript levels in response to 24 h photoperiod were found. No significant differences among population groups and no association with flowering and photoperiodic sensitivity were found in transcript levels of *FKF1* and *PHYA*. The results of the gene expression studies as related to flowering time and photoperiodic sensitivity are summarized in Table 5.

Discussion

Photoperiodic sensitivity in A. thaliana *populations is associated with proximity to the coast*

We have demonstrated that photoperiodic sensitivity in Norwegian populations of *A. thaliana* originating from its northernmost distribution range was associated with distance from the ocean and duration of the winter (Fig. 4); coastal populations being more sensitive to photoperiod than inland populations (Fig. 2, 3). Flowering was markedly promoted by increasing the photoperiod from 8 to 16 h, which indicates that critical photoperiod for flowering in populations originating from these latitudes, is between these day lengths (critical photoperiod defined by Roberts and Summerfield 1987 and used in study of Pouteau *et al*., 2008 on *A. thaliana* mutants).

Contradictory results on latitudinal and longitudinal clines in flowering time of *A. thaliana* have been reported earlier and indeed very few studies have demonstrated clines in flowering time which can be explained by photoperiod (Stinchcombe *et al*., 2004; Shindo *et al*., 2005; Samis *et al*., 2008). Shindo *et al*. (2005) failed to detect a latitudinal cline in flowering time in a sample of accessions both from Northern America and Europe. This may be due to the geographic composition of the sample and differences in demographic histories between Northern American and European accessions. Latitudinal clines in flowering time under field conditions were demonstrated by Stinchcombe *et al.* (2004) and the recent study of Samis *et al*. (2008) found a longitudinal cline in photoperiodic sensitivity. This longitudinal variation in flowering time among *A. thaliana* populations from Eurasia might reflect climatic differences due to varying distance from the ocean (Samis *et al*., 2008). In our study mean yearly precipitation was used as a proxy for distance from the ocean. We demonstrate that under laboratory conditions clinal variation in flowering time related to distance from the ocean was observed even in a sample of populations originating from a relatively restricted geographic region. This variation is likely a result of natural selection and is associated with climatic factors such as duration of winter that vary with distance from the ocean. Coastal regions of Norway are characterized by high amount of yearly precipitation and higher winter temperatures as compared to inland locations; these factors are excellent predictors of differential climates. Although plants at the same latitudes are subjected to similar photoperiods, natural selection, caused by differences in the length of the winter may create adaptation to different photoperiods. Moreover,

populations with low vernalization sensitivity i.e. those originating from coastal regions at both arctic and subarctic latitudes (Lewandowska-Sabat *et al.*, unpublished results, Paper I) were very responsive towards differences in photoperiods. Thus it seems likely that highly photoperiodic sensitive populations are favored by natural selection at coastal locations where photoperiod is a more reliable indicator of the changing seasons than temperature. This gives insurance against damages to floral organs caused by sudden spring frost.

Clines in *A. thaliana* flowering time have only been demonstrated in accessions with functional *FRIGIDA* (*FRI*) alleles (Stinchcombe *et al*., 2004). Epistatic interaction of *FRI* and *FLOWERING LOCUS C* (*FLC*) contributing to a latitudinal cline as demonstrated by Caicedo *et al*. (2004), is likely confounded with population structure (Zhao *et al*., 2007). However, the epistatic interaction of *FRI* and *FLC* has been demonstrated in studies manipulating photoperiod and temperature conditions simultaneously (Li *et al*., 2006; Scarcelli *et al*., 2007). This indicates how demanding it is to detect such effects using controlled growth conditions with fixed photoperiods and temperatures as well as uncontrolled conditions typical of field studies, both of which may not be the ecologically relevant. A latitudinal cline in hypocotyl length in response to red and far-red light in Norwegian *A. thaliana* populations has been demonstrated (Stenøien *et al*., 2002) suggesting that this variation may be mediated by the phytochrome pathway. No latitudinal cline in flowering time was detected in these populations and the authors have suggested that other floral transition pathway components than those acting downstream in the seedling developmental pathway may be responsible for this variation.

In most of the published studies latitude is used as a proxy for unknown environmental variables that are believed to vary with latitude. Thus, the lack of evident clines in *A. thaliana* is not unexpected since other environmental factors than photoperiod, not associated with latitude also affect survival, flowering time and seed production. The geographic distribution of the population samples that are being investigated appears to be decisive for detecting signs of local adaptation resulting from differential climatic gradients. Frequently, populations from broad geographic regions with different climates and demographic histories are studied. It is assumed that environmental cues that vary across broad geographic scales will impose natural selection on flowering time. Accordingly, we would expect that populations originating from restricted geographical scales (Stenøien *et al*., 2002), which still experience wide geographic variation in

environmental cues but also share common demographic history, will be ideal for detecting signs of local adaptation. This is indeed the case in the present study.

Variation in the transcript levels of GI and CO may account for flowering time under long days

Swindell *et al.* (2007) found that genes differentially expressed among populations under temperature stress were associated with geographical temperature gradients. In our study, the transcript levels of *GI* and *CO* were much lower in coastal populations than in inland populations at the end of the long day (Fig. 5) and over time (Fig. 6). Moreover, coastal populations flower significantly later than inland populations in LD (Fig. 2, 3). This indicates that variation in *GI* and *CO* expression level may account for flowering time variation under long days in *A. thaliana* populations. Therefore, longer episodes with increasing photoperiod certainly is perceived by the coastal populations as the onset of spring and late flowering may be a mechanism preventing against spring frost damage. It has been shown earlier that *GI* mediated peak expression of *CO* at the end of long days is crucial in photoperiodic flowering, however, the role of these genes in determining natural variation in flowering time has not yet been demonstrated (Sawa *et al*., 2007; Fornara *et al*., 2009). Only one recent study has shown that sequence polymorphism in promoter region of *FLOWERING LOCUS T* (*FT*) affects *FT* expression and contributes to natural variation in flowering time in response to photoperiod and ambient temperature (Schwartz *et al*., 2009).

CRY2 and *TOC1* transcript levels were lower in Northern than in Southern Norwegian populations at the end of the long day (Fig. 5) and over time (Fig. 6), and was not associated with flowering time. This indicates that variation in *CRY2* and *TOC1* expression levels may account for variation in other traits than flowering time, e.g. hypocotyl elongation. It may also reflect adaptation to light quality that varies with latitude, i.e. the daily light period enriched in blue and far-red light lasts longer at arctic than at subarctic latitudes.

Variation in the transcript levels of CRY2, TOC1 and GI may account for photoperiodic response

CRY2, *TOC1* and *GI* transcript levels were much lower in coastal populations than in inland populations in SD compared to LD (Fig. 7). The variation in expression of these genes was

associated with and may account for the photoperiodic response in these populations. The expression levels of these genes may already be high in SD in inland populations and therefore their response to LD is much smaller than in coastal populations. However, expression of *CO* is lower only in Northern Norwegian population in SD compared to LD. This indicates that photoperiodic response is not regulated at the *CO* mRNA level in these populations; rather it might be the case that the CO protein level as well as other downstream genes of *CRY2*, *TOC1* and *GI* account for this variation.

Thus, expression of *CRY2*, *TOC1* and *GI* may reach the needed threshold already within 8 h photoperiod in inland populations, while in coastal populations the same threshold is reached when the day length is longer, probably near 16 h. However, mimicking environmental conditions found in natural habitats in controlled growth rooms is very difficult, thus it is very likely that these populations respond to much longer photoperiods than what they did in this study. This may also explain why flowering time at 16 and 24 h photoperiod did not differ under the conditions used here. Furthermore, we used a much lower ambient temperature (16°C), which reflects the natural growth conditions in Norway, than the standard temperature (21°C) used in most of the *A. thaliana* studies (Hoffmann 2002). Therefore, interacting effects of photoperiod and lower ambient temperature (e.g. Halliday *et al*., 2003) as being used in our study have the potential to reveal phenotypic responses involving genetic regulations that so far has gone undetected in *A. thaliana*. On the other hand, these conditions may still be very artificial, especially in the autumn when short days are often accompanied by very cold weather. Although inland populations reach higher expression level earlier under laboratory conditions, their high vernalization sensitivity (Lewandowska-Sabat *et al*., unpublished results, Paper I) is likely sufficient to assure timing of flowering in proper seasons. As a result, tighter photoperiodic control in coastal populations might be a necessary adaptation for sensing the onset of spring in regions with relatively mild and unpredictable winter climates as opposed to continental climates with more stable winters.

Col-0 shows expression pattern surprisingly similar to coastal populations, however flowering time and photoperiodic sensitivity in this reference population is rather similar to inland populations. This may imply that the expression pattern of the genes studied here does not reflect the phenotype of Col-0 since this population is extremely different from the Norwegian populations at the sequence level (Table 4, Table S5-S8). Moreover, the loss-of-function *FRI*

allele results in early flowering phenotype in Col-0 (Johanson *et al*., 2000) and possibly overrides the transcriptional changes observed in our study. In addition, a long vernalization treatment could enhance flowering time under SD in both reference populations Col-0 and L*er*.

It has been shown for *Populus* trees from northern Europe that variation in *CO* and *FT* expression is adaptive. Earlier peak in *CO* expression in southern trees (51°N) promotes growth at shorter days as compared to northern trees (63°N) (Böhlenius *et al*., 2006). In Norway spruce, showing distinct latitudinal and altitudinal clines and found over much of the geographic range studied here for *A. thaliana,* the *FT*-like gene *PaFT4* is believed to act as a key integrator of photoperiodic and thermal signals in the control of growth rhythms (Gyllenstrand *et al*., 2007). However the relationship of the expression of *CONSTANS*-like genes and such adaptive traits is less clear (Holefors *et al*., 2009) while the expression of spruce phytochromes are correlated with adaptive traits and also there is growing evidence that epigenetic mechanisms may well be fine-tuning the distinct clines observed in nature (Johnsen *et al*., 2005). Also in our study, despite the vernalization requirement having been fully met, we cannot rule out that other epigenetic-like effects working through molecular mechanism such as chromatin remodeling and microRNAs might affect the expression patterns and phenotypes observed. We also recognize that for some of the gene products studied here the differences may be larger at the protein level than at the nucleotide and transcript level, and quantitative proteome studies may reveal differences not yet elucidated.

Since the winter temperature along the Atlantic coastline of Norway varies a lot with repeating freezing-thawing cycles due to the influence of the Gulf Stream, tight expression control of *TOC1*, *GI* and *CRY2* may play a major role in photoperiodic control of flowering in coastal populations. A combination of signals sent by rising ambient temperature and more importantly increasing day length in spring might be crucial for coastal populations to flower at the right time. Notably, *GI* plays a critical role in temperature compensation of the circadian clock in *Arabidopsis* (Gould *et al*., 2006). Also, other clock components like *PRR7* and *PRR9* integrate light and temperature signalling (Salomé and McClung, 2005) and *GI* and *TOC1* play roles in photoperiodic control of flowering time (Ito *et al.*, 2009). Therefore an interaction of *GI*, *TOC1* and *CRY2* is likely fundamental for day length and temperature sensing and proper timing of flowering in *A. thaliana* populations. However, *GI* plays several distinct roles in flowering time and other loci may also contribute to variation in photoperiodic flowering (Ito *et al.*, 2009).

CO sequence variation may account for photoperiodic sensitivity

Nucleotide variation in *CO* appears to show a geographical pattern. Three haplogroups were distinguished based on sequence variation (Table 4, Fig. S1); two of the haplogroups include most of the coastal populations while the third haplogroup includes the inland populations. No significant association of *CO* haplotypes with photoperiodic sensitivity was found, but SNP11 in *CO* was associated with photoperiodic sensitivity ($P_{8/16}$) and remained marginally significant after correction for population structure. SNP11 was not associated with flowering time under SD after accounting for population structure. This suggests that polymorphisms in *CO* are associated with photoperiodic sensitivity that differs with distance from the ocean, even after accounting for population structure. It is noteworthy that association of *CO* polymorphism with flowering time under SD was confounded with population structure and that no association with *CO* expression pattern and DTB in 8 h was found. Thus, variation at other loci may account for the variation in flowering time under short days. Furthermore, association between DTB in 16 h and the *CO* expression pattern and lack of association with the *CO* polymorphism indicates that selection acts at the *CO* mRNA level to modulate flowering time responses under LD. Conversely, association of *CO* polymorphism and lack of association of the *CO* expression pattern with $P_{8/16}$ indicates that variation at the mRNA level at other loci may be responsible for photoperiodic response (i.e. *CRY2*, *GI* and *TOC1*) and likely these genes may regulate CO at the protein level. It is clear that photoperiodic response is a very complex process and the combined effect of these genes is crucial for survival in the high-latitude conditions that these populations are adapted to.

The low level of nucleotide diversity found here as compared to other studies (Nordborg *et al.*, 2005, Schmid *et al.*, 2005) might be due to the relatively limited geographical range represented by these *A. thaliana* populations (Le Corre *et al.*, 2002), demographic processes such as bottlenecks, as well as positive selection towards advantageous mutations. It may also be due to the fact that at the boundary of a species distribution, genetic variability may be reduced due to extreme ecological conditions and strong selection pressures acting on populations.

In conclusion, we have shown that photoperiodic sensitivity of *A. thaliana* populations originating from its northernmost species boundary is associated with climatic factors that vary

depending on distance from the ocean and altitude. Coastal populations were more responsive towards changes in photoperiod than inland populations. Therefore, the length of the winter imposes natural selection that creates clinal variation in photoperiodic sensitivity and most likely determines local adaptation in *A. thaliana*. This natural variation in photoperiodic sensitivity may be mediated by variation at *CRY2*, *GI* and *TOC1* mRNA level. Furthermore, *GI* and *CO* transcript levels may determine flowering time variation under long days in *A. thaliana* populations. These results suggest that in plants adapted to unstable weather conditions photoperiod may play a major role in determining survival as natural selection tends to favour populations with greater sensitivity to photoperiod in mild winter climates.

Acknowledgments

We thank two graduate students Asma Naurin Malik and Anusa Balasingham for help with RNA isolation for real-time RT-PCR and sequencing of *CO*. We are also grateful to Magdalena Achrem and Lidia Skuza for help with the real-time RT-PCR, and YeonKyeong Lee and Kjetil Fosnes for assistance in sequencing. This study was supported by a grant from the Research Council of Norway: The Norwegian Arabidopsis Research Centre, project no. 151991/150, and a PhD scholarship to Anna Lewandowska-Sabat from the Norwegian University of Life Sciences.

References

Alabadi D, Oyama T, Yanovsky MJ, Harmon FG, Más P, Kay SA. 2001. Reciprocal regulation between *TOC1* and *LHY/CCA1* within the *Arabidopsis* circadian clock. *Science* **293**: 880-3.

Aukerman MJ, Hirschfeld M, Wester L, Weaver M, Clack T Amasino RM, Sharrock RA. 1997. A deletion in the *PHYD* gene of the *Arabidopsis* Wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing. *Plant Cell* **9**: 1317-1326.

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

Banta JA, Dole J, Cruzan MB, Pigliucci M. 2007. Evidence of local adaptation to coarse-grained environmental variation in *Arabidopsis thaliana*. *Evolution* **61**: 2419-2432.

Bradshaw WE, Holzapfel CM. 2001. Phenotypic evolution and the genetic architecture underlying photoperiodic time measurement. *Journal of Insect Physiology* **47**: 809-820.

Bradshaw WE, Zani PA, Holzapfel CM. 2004. Adaptation to temperate climates. *Evolution* **58**: 1748-1762.

Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O. 2006. *CO*/*FT* regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* **312**: 1040-3.

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD. 2004. Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *PNAS* **101**: 15670-15675.

Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, Coupland G. 2007. FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science* **316**: 1030-1033.

Edwards KD, Lynn JR, Gyula P, Nagy F, Millar AJ. 2005. Natural allelic variation in the temperature-compensation mechanisms of the *Arabidopsis thaliana* circadian clock. *Genetics* **170**: 387-400.

Ehrich D. 2006. AFLPdat: a collection of R functions for convenient handling of AFLP data. *Molecular Ecology Notes* **6**: 603-604.

El-Assal SE, Alonso-Blanco C, Peeters AJM, Raz V, Koornneef M. 2001. A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nature Genetics* **29**: 435-440.

El-Assal SE, Alonso-Blanco C, Peeters AJM, Wagemaker C, Weller JL, Koornneef M. 2003. The role of *CRYPTOCHROME 2* in flowering in *Arabidopsis*. *Plant Physiology* **133**: 1504-1516.

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

Falush D, Stephens M, Pritchard J. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**: 1567-1587.

Filiault DL, Wessinger CA, Dinneny JR, Lutes J, Borevitz JO, Weigel D, Chory J, Maloof JN. 2008. Amino acid polymorphisms in *Arabidopsis PHYTOCHROME B* cause differential responses to light. *PNAS* **105**: 3157-3162.

Fornara F, Panigrahi KC, Gissot L, Sauerbrunn N, Rühl M, Jarillo JA, Coupland G. 2009. Arabidopsis DOF transcription factors act redundantly to reduce *CONSTANS* expression and are essential for a photoperiodic flowering response. *Developmental Cell* **17**: 75-86.

Franklin KA, Larner VS, Whitealm GC. 2005. The signal transducing photoreceptors of plants. *International Journal of Developmental Biology* **49**: 653-664.

Gardner MJ, Hubbard KE, Hotta CT, Dodd AN, Webb AA. 2006. How plants tell the time. *Biochemical Journal* **397**: 15-24.

Gould PD, Locke JC, Larue C, Southern MM, Davis SJ, Hanano S, Moyle R, Milich R, Putterill J, Millar AJ, *et al.* 2006. The molecular basis of temperature compensation in the *Arabidopsis* circadian clock. *Plant Cell* **18**: 1177-87.

Gyllenstrand N, Clapham D, Källman T, Lagercrantz U. 2007. A Norway spruce *FLOWERING LOCUS T* homolog is implicated in control of growth rhythm in conifers. *Plant Physiology* **144**: 248-257.

Halliday KL, Salter MG, Thingnæs E, Whitelam GC. 2003. Phytochrome control of flowering is temperature sensitive and correlates with expression of the floral integrator *FT*. *Plant Journal* **33**: 1-11

Harmer SL. 2009. The circadian system in higher plants. *Annual Review of Plant Biology* **60**: 357-77.

Hoffmann MH. 2002. Biogeography of *Arabidopsis thaliana* (L.) Heynh. (*Brassicaceae*). *Journal of Biogeography* **29**: 125-134.

Holefors A, Opseth L, Rosnes AKR, Ripel L, Snipen L, Fossdal CG, Olsen JE. 2009. Identification of *PaCOL1* and *PaCOL2*, two *CONSTANS*-like genes showing decreased transcript levels preceding short day induced growth cessation in Norway spruce. *Plant Physiology and Biochemistry* **47**: 105-115.

Ito S, Kawamura H, Niwa Y, Nakamichi N, Yamashino T, Mizuno T. 2009. A genetic study of the *Arabidopsis* circadian clock with reference to the *TIMING OF CAB EXPRESSION 1* (*TOC1*) gene. *Plant Cell Physiology* **50**: 290-303.

Jiao Y, Lau OS, Deng XW. 2007. Light-regulated transcriptional networks in higher plants. *Nature Reviews Genetics* **8**: 217-230.

Jobb G. 2008. TREEFINDER version of October 2008. Munich, Germany. Distributed by the author at www.treefinder.de

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

Johnsen Ø, Fossdal C.G, Nagy NE, Mølmann J, Dæhlen OG, Skrøppa T. 2005. Climatic adaptation in *Picea abies* progenies is affected by the temperature during zygotic embryogenesis and seed maturation. *Plant, Cell and Environment* **28**: 1090-1102.

Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D. 1999. Activation tagging of the floral inducer *FT*. *Science* **286**: 1962-5.

Korves TM, Schmid KJ, Caicedo AL, Mays C, Stinchcombe JR, Purugganan MD, Schmitt J. 2007. Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. *American Naturalist* **169**: E141-E157.

Le Corre V, Roux F, Reboud X. 2002. DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Molecular Biology and Evolution* **19**: 1261-1271.

Lee I, Amasino RM. 1995. Effect of vernalization, photoperiod, and light quality on the flowering phenotype of *Arabidopsis* plants containing the *FRIGIDA* gene. *Plant Physiology* **108**: 157-162.

Li QH and Yang HQ. 2007. Cryptochrome signaling in plants. *Photochemistry and Photobiology* **83**: 94-101.

Li Y, Roycewicz P, Smith E, Borevitz JO. 2006. Genetics of local adaptation in the laboratory: flowering time Quantitative Trait Loci under geographic and seasonal conditions in *Arabidopsis*. *PLOS ONE* 1: e105.

Maloof JN, Borevitz JO, Dabi T, Lutes J, Nehring R, Redfern JL, Trainer GT, Wilson JM, Asami T, Berry CC, *et al.* 2001. Natural variation in light sensitivity of *Arabidopsis*. *Nature Genetics* **29**: 441-446.

Michael TP, Salomé PA, Yu HJ, Spencer TR, Sharp EL, McPeek MA, Alonso JM, Ecker JR, McClung CR. 2003. Enhanced fitness conferred by naturally occurring variation in the circadian clock. *Science* **302**: 1049-1053.

Mockler TC, Yang H, Yu X, Parikh D, Cheng Y-C, Dolan S, Lin C. 2003. Regulation of photoperiodic flowering by *Arabidopsis* photoreceptors. *PNAS* **100**: 2140-2145.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal1 R, *et al.* 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology* **3**: e196.

O'Malley KG, Banks MA. 2008. A latitudinal cline in the Chinook salmon (*Oncorhynchus tshawytscha*) clock gene: evidence for selection on *PolyQ* length variants. *Proceedings Biological Sciences* **275**: 2813-21.

Onouchi H, Coupland G. 1998. The regulation of flowering time of *Arabidopsis* in response to daylength. *Journal of Plant Research* **111**: 271-275.

Pouteau S, Carré I, Gaudin V, Ferret V, Lefebvre D, Wilson M. 2008. Diversification of photoperiodic response patterns in a collection of early-flowering mutants of *Arabidopsis*. *Plant Physiology* **148**: 1465-1473.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.

Putterill J, Laurie R, Macknight R. 2004. It's time to flower: the genetic control of flowering time. *Bioessays* **26**: 363-373.

Putterill J, Robson F, Lee K, Simon R, Coupland G. 1995. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**: 847-57.

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Roberts EH, Summerfield RJ. 1987. Measurement and prediction of flowering in annual crops. In: Atherton JG, ed. *Manipulation of Flowering*. Butterworths, London, 17-50.

Rozen S, Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S. *Bioinformatics methods and protocols: methods in molecular biology*. Totowa, USA, Humana Press, 365-386.

Salomé PA, McClung CR. 2005. *PSEUDO-RESPONSE REGULATOR 7* and *9* are partially redundant genes essential for the temperature responsiveness of the *Arabidopsis* circadian clock. *Plant Cell* **17**: 791-803.

Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G. 2000. Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science* **288**: 1613-1616.

Samis KE, Heath KD, Stinchcombe JR. 2008. Discordant longitudinal clines in flowering time and *PHYTOCHROME C* in *Arabidopsis thaliana*. *Evolution* **62**: 2971-2983.

Sawa M, Nusinow DA, Kay SA, Imaizumi T. 2007. FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. *Science* **318**: 261-5.

Scarcelli N, Cheverud JM, Schaal BA, Kover PX. 2007. Antagonistic pleiotropic effects reduce the potential adaptive value of the *FRIGIDA* locus. *PNAS* **104**: 16986-16991.

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

Schwartz C, Balasubramanian S, Warthmann N, Michael TP, Lempe J, Sureshkumar S, Kobayashi Y, Maloof JN, Borevitz JO, Chory J, Weigel D. 2009. *Cis*-regulatory changes at *FLOWERING LOCUS T* mediate natural variation in flowering responses of *Arabidopsis thaliana*. *Genetics* **183**: 723-732

Sharrock RA, Quail PH. 1989. Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution, and differential expression of a plant regulatory photoreceptor family. *Genes and Development* **3**: 1745-57.

Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005. Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. *Plant Physiology* **138**: 1163-1173.

Simpson GG, Dean C. 2002. *Arabidopsis*, the rosetta stone of flowering time? *Science* **296**: 285-289.

Stenøien HK, Fenster CB, Kuittinen H, Savolainen O. 2002. Quantifying latitudinal clines to light responses in natural populations of *Arabidopsis thaliana* (*Brassicaceae*). *American Journal of Botany* **89**: 1604-1608.

Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS, Purugganan MD, Schmitt J. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *PNAS* **101**: 4712-4717.

Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G. 2001. *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* **410**: 1116-20.

Swindell WR, Huebner M, Weber AP. 2007. Plastic and adaptive gene expression patterns associated with temperature stress in *Arabidopsis thaliana*. *Heredity* **99**: 143-50.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**: 1596-1599.

Turck F, Fornara F, Coupland G. 2008. Regulation and identity of florigen: *FLOWERING LOCUS T* moves center stage. *Annual Review of Plant Biology* **59**: 573-594.

Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. 2004. Photoreceptor regulation of *CONSTANS* protein in photoperiodic flowering. *Science* **303**: 1003-1006.

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y. 2000. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *Plant Cell* **12**: 2473-2484.

Yanovsky MJ, Kay SA. 2002. Molecular basis of seasonal time measurement in *Arabidopsis*. *Nature* **419**: 308-312.

Zhang Q, Li H, Li R, Hu R, Fan C, Chen F, Wang Z, Liu X, Fu Y, Lin C. 2008. Association of the circadian rhythmic expression of *GmCRY1a* with a latitudinal cline in photoperiodic flowering of soybean. *PNAS* **105**: 21028-33.

Figure legends

Figure 1. Collection sites of Norwegian *Arabidopsis thaliana* populations investigated in this study. Orange symbols denote populations used in the photoperiodic response experiment.

Figure 2. Average photoperiodic flowering responses in populations of *Arabidopsis thaliana* from coastal and inland locations in Norway. Results are means of 60 plants (coastal populations) ±SE and 90 plants (inland populations) ±SE.

Figure 3. Flowering time among *Arabidopsis thaliana* populations in response to 8, 16 and 24 h photoperiod. Results are means of 15 plants ±SE.

Figure 4. Regression analyses of photoperiodic sensitivity ($P_{8/24}$) in *Arabidopsis thaliana* populations with distance from the ocean and mean winter temperature. Open diamonds indicate inland populations and filled diamonds coastal populations. Yearly mean precipitation (mm, 1961-1990) was used as proxy for distance from the ocean.

Figure 5. Transcript levels of *CRY2*, *TOC1*, *GI* and *CO* normalized to *actin* in 6 Norwegian populations and Col-0 of *Arabidopsis thaliana*. A) Diurnal transcript levels in 16 h photoperiod. B) Transcript level at the 13th h of the 16 h photoperiod. *GI* and *CO* transcript levels was associated with DTB in 16 h photoperiod (P=0.008, P=0.001, respectively). White and dark bars show light and dark periods, respectively, filled squares (■) indicate one biological replicate, ***=P≤0.001. Y-axis denotes ΔCt values. Results are mean of 2 biological replicates analysed in duplicate ±SE.

Figure 6. Fold change of *CRY2*, *TOC1*, *GI* and *CO* transcript levels normalized to *actin* over time in 5 Norwegian populations and Col-0 of *Arabidopsis thaliana*. Transcript levels at the end of the 5th and the 3rd day were compared. Changes in transcript levels of *GI* and *CO* were associated with DTB in 16 h photoperiod (P=0.014, P=0.003, respectively). ***=P≤0.001, **=P≤0.01. Results are mean of 2 biological replicates analysed in duplicate ±SE.

Figure 7. Fold change of *CRY2*, *TOC1*, *GI* and *CO* transcript levels normalized to *actin* in SD as compared to LD in 5 Norwegian populations and Col-0 of *Arabidopsis thaliana*. Changes in *CRY2*, *TOC1* and *GI* transcript levels were associated with $P_{8/16}$ (photoperiodic sensitivity; P=0.006, 0.004, 0.024, respectively). SD – short day, LD – long day, ***=P≤0.001. Results are mean of 2 biological replicates analysed in duplicate ±SE.

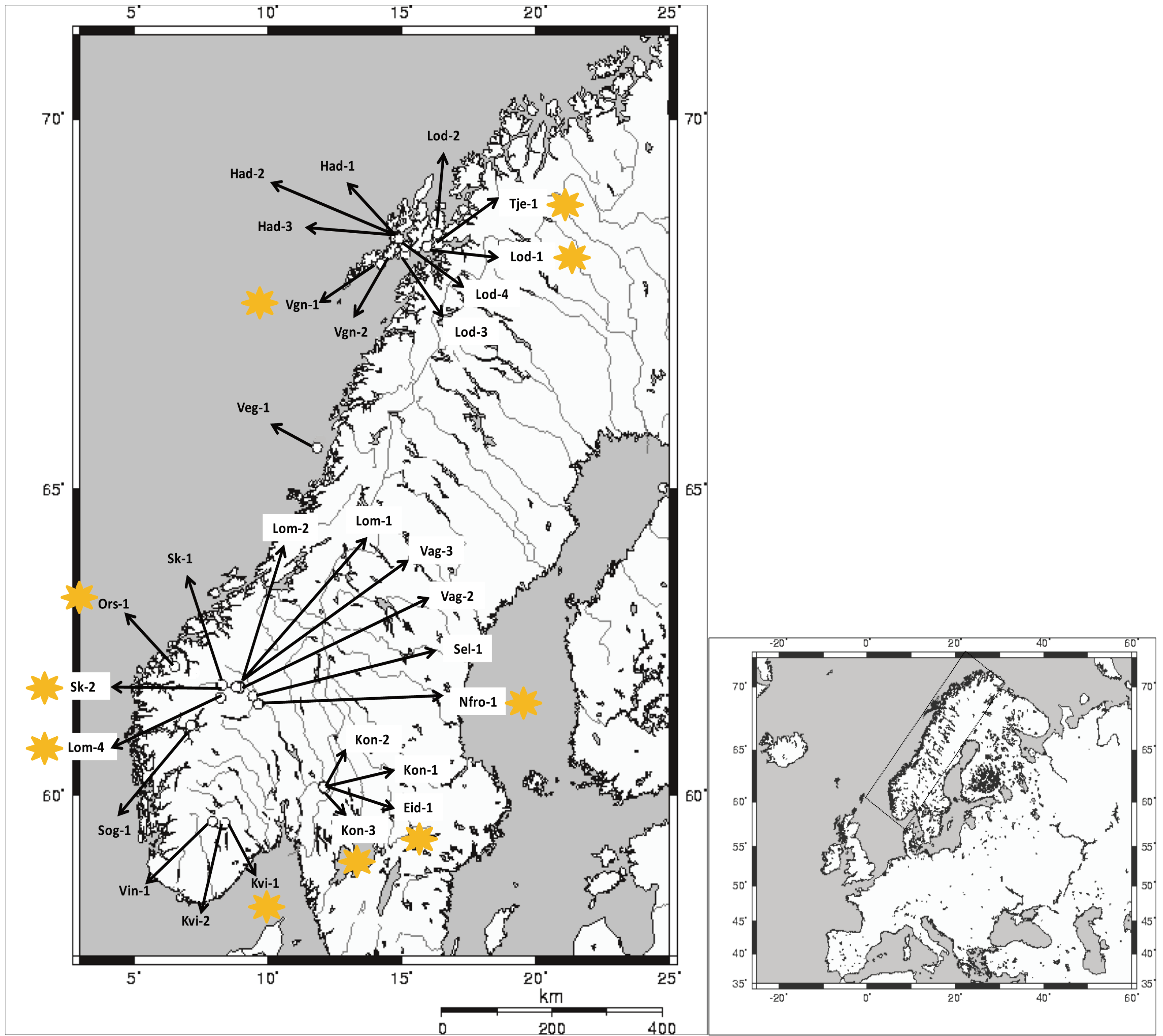


Figure 1

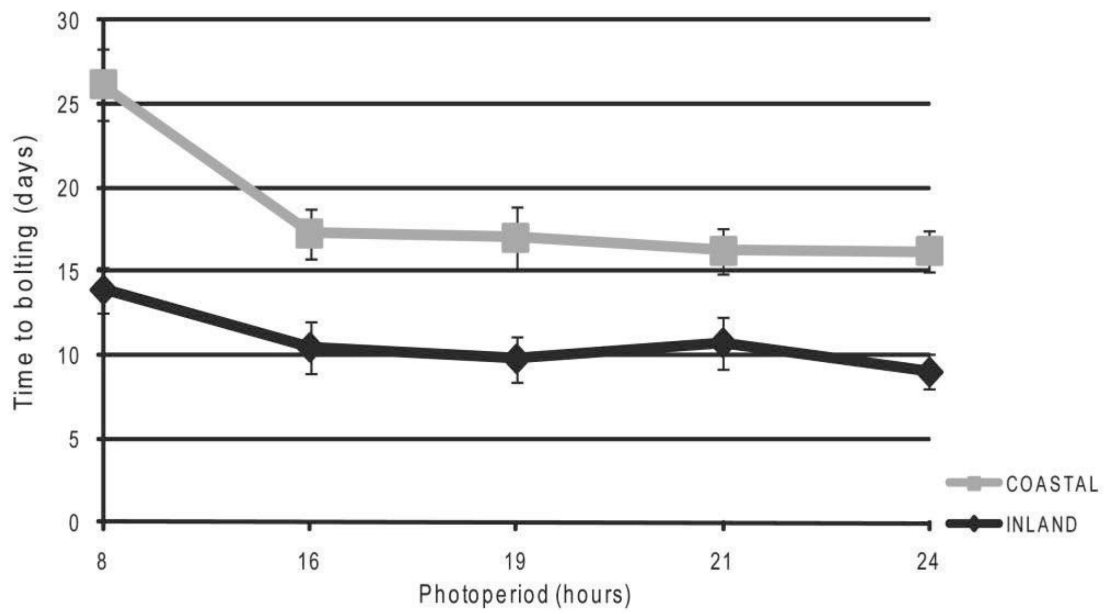


Figure 2

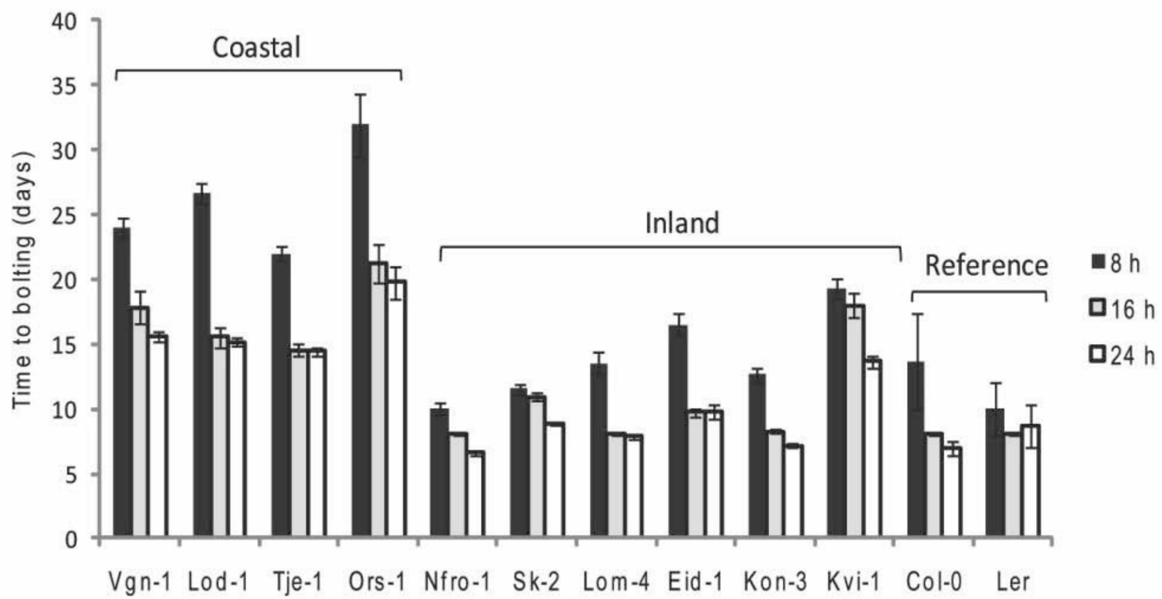


Figure 3

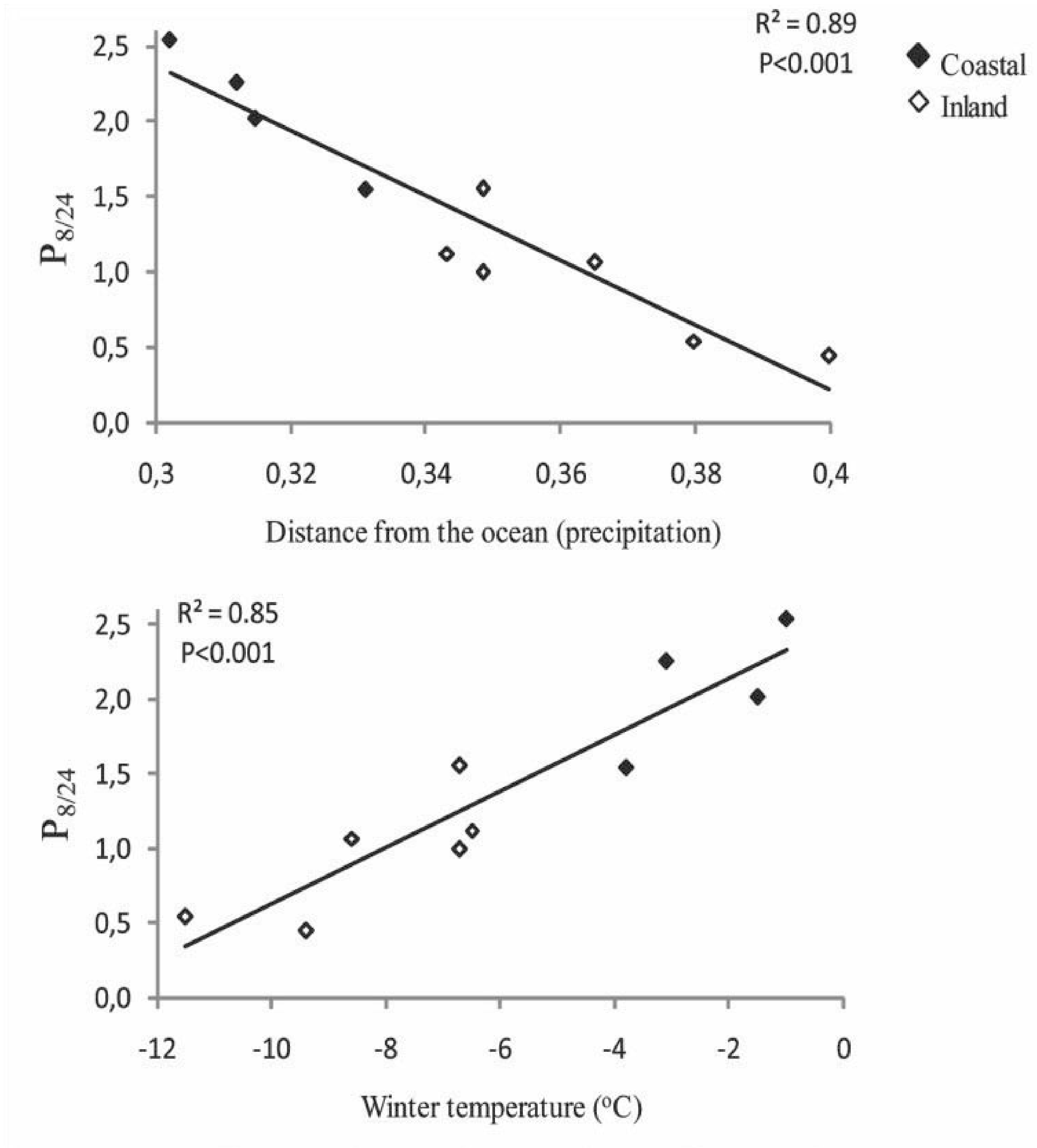


Figure 4

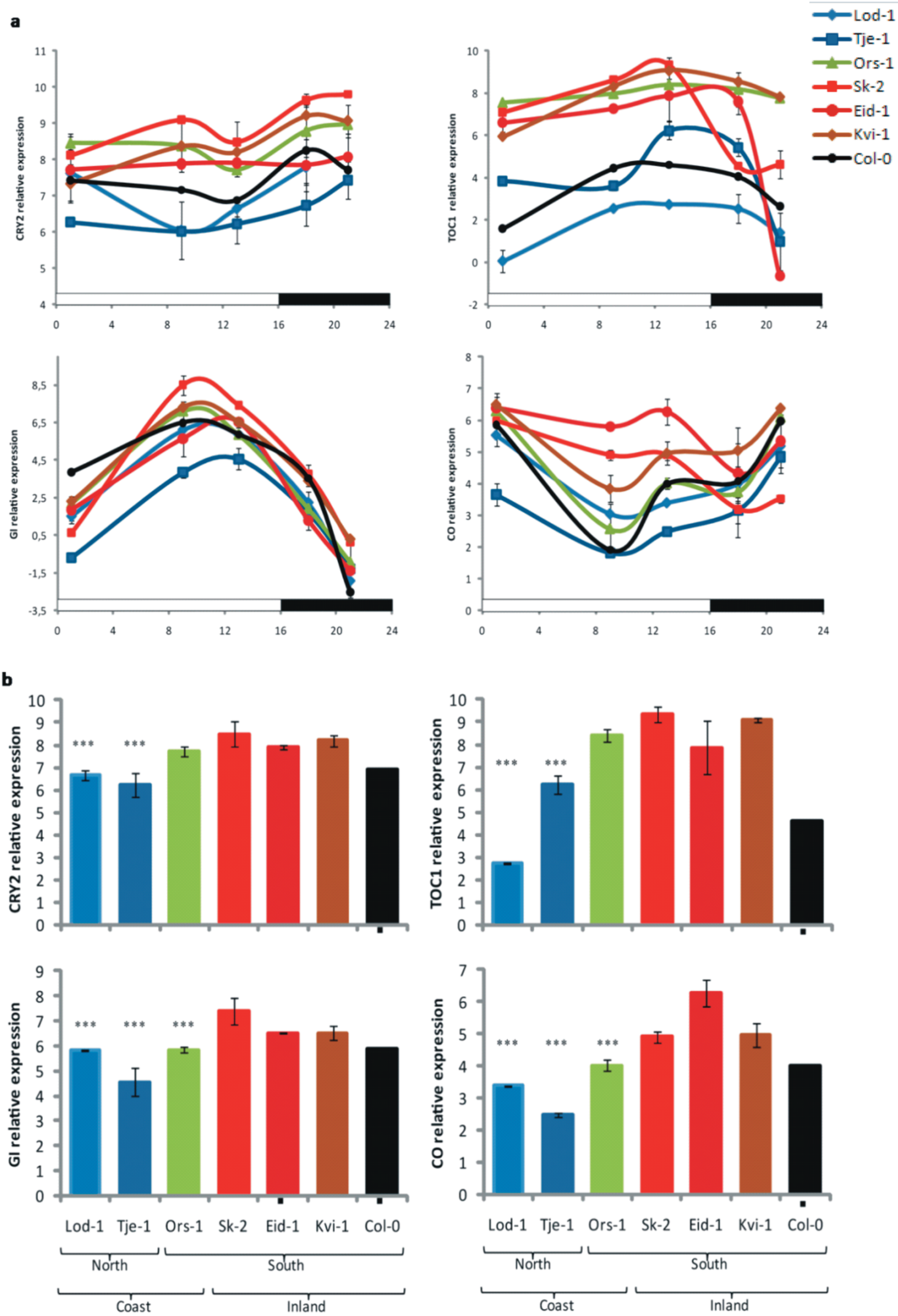


Figure 5

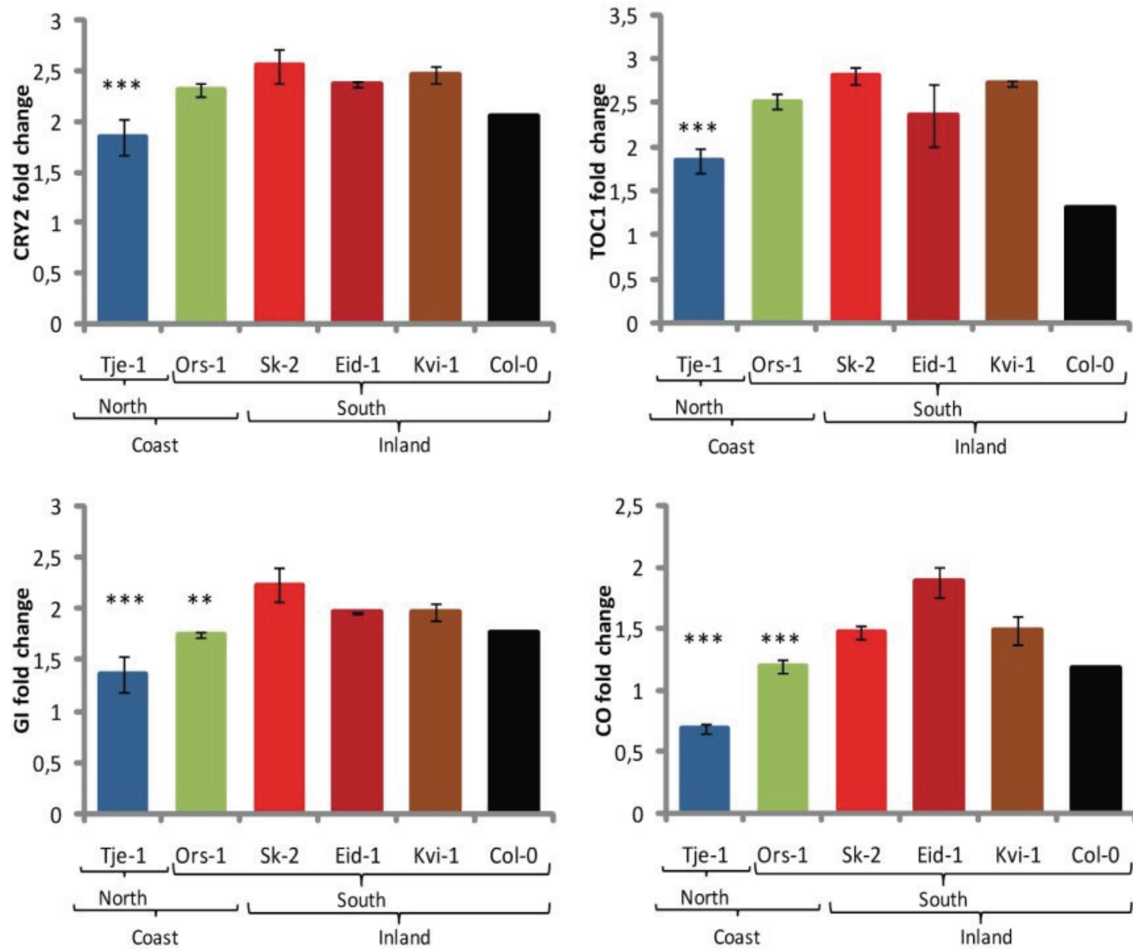


Figure 6

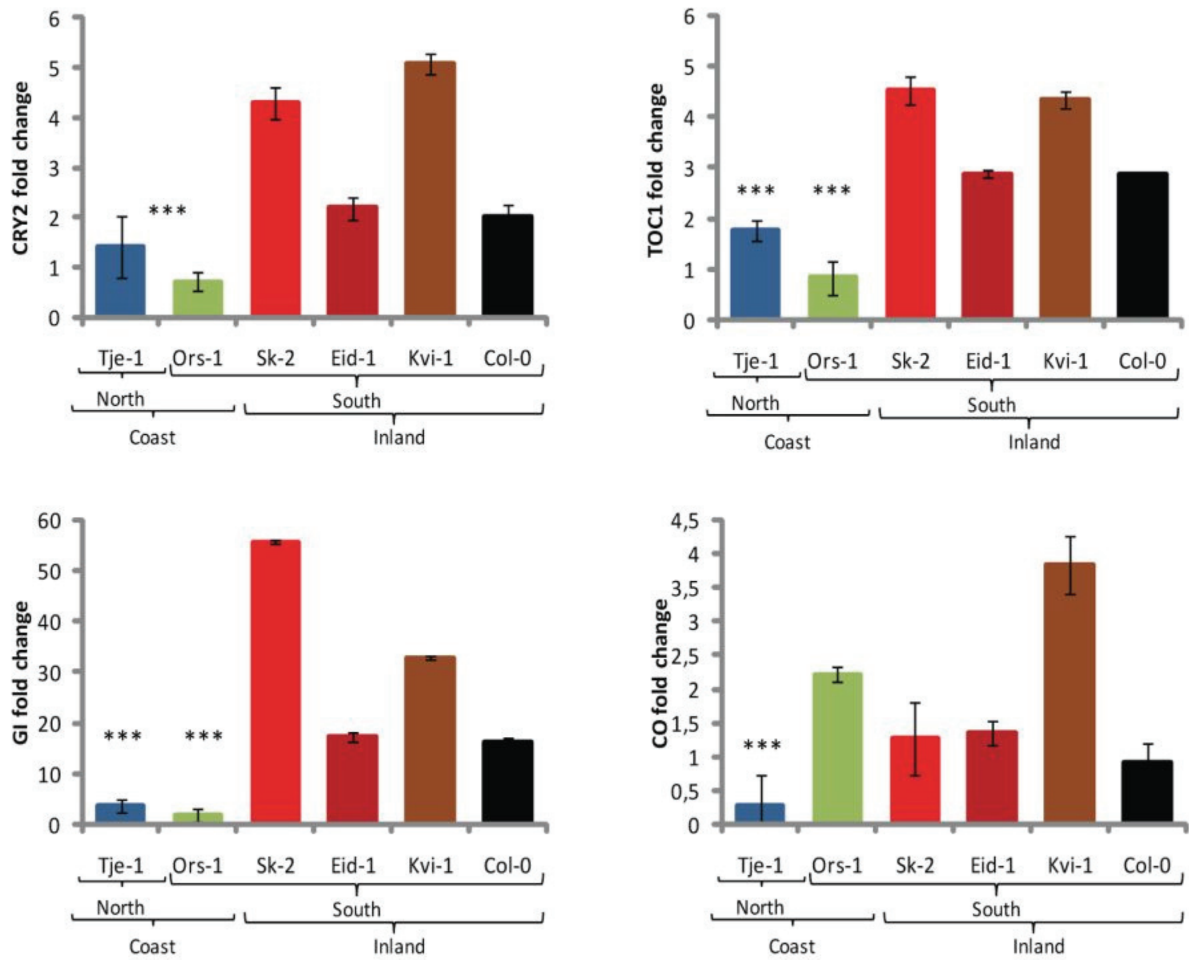


Figure 7

Table 1. Collection data for 29 Norwegian populations of *Arabidopsis thaliana*. *=populations used in screening for photoperiodic response. Flowering – days to bolting in response to 12 weeks of vernalization and under long day conditions; Rapid (5-10 days), Slow (10-15 days), Very slow (>15 days), and vernalization requirement as described in Lewandowska-Sabat *et al*., unpublished (Paper I). Different shading specifies coastal and inland populations.

Population	Latitude	Longitude	Altitude	Flowering	Vernalization	Region	Groups	Groups
Vgn-1*	68°10'11"	14°13'15"	3	Slow	Yes	Lofoten	Coastal	Northern
Had-1	68°30'34"	14°43'57"	50	Rapid	Yes	Lofoten	Coastal	Northern
Had-2	68°30'42"	14°45'44"	60	Rapid	Yes	Lofoten	Coastal	Northern
Had-3	68°30'27"	14°53'32"	60	Rapid	Yes	Lofoten	Coastal	Northern
Lod-1*	68°23'51"	15°56'6"	100	Slow	No	Lofoten	Coastal	Northern
Lod-2	68°33'56"	16°19'7"	100	Slow	Yes	Lofoten	Coastal	Northern
Tje-1*	68°25'29"	16°22'9"	110	Slow	Yes	Lofoten	Coastal	Northern
Lod-3	68°18'17"	15°06'51"	82	Slow	Yes	Lofoten	Coastal	Northern
Lod-4	68°23'59"	15°08'23"	80	Slow	Yes	Lofoten	Coastal	Northern
Vgn-2	68°14'38"	14°35'46"	97	Slow	Yes	Lofoten	Coastal	Northern
Sog-1	61°12'33"	7°07'39"	8	Slow	Yes	West Coast	Coastal	Southern
Ors-1*	62°12'19"	6°32'1"	2	Very Slow	Yes	West Coast	Coastal	Southern
Veg-1	65°42'24"	11°56'4"	3	Slow	Yes	West Coast	Coastal	Southern
Nfro-1*	61°34'38"	9°39'42"	735	Rapid	No	Gudbrandsdalen	Inland	Southern
Sel-1	61°42'58"	9°26'26"	687	Rapid	Yes	Gudbrandsdalen	Inland	Southern
Vag-2	61°51'47"	8°58'43"	640	Rapid	Yes	Gudbrandsdalen	Inland	Southern
Vag-3	61°51'28"	8°56'35"	400	Rapid	Yes	Gudbrandsdalen	Inland	Southern
Lom-1	61°51'42"	8°49'30"	525	Rapid	Yes	Gudbrandsdalen	Inland	Southern
Lom-2	61°51'49"	8°49'2"	700	Rapid	Yes	Gudbrandsdalen	Inland	Southern
Sk-1	61°53'56"	8°15'14"	604	Rapid	Yes	Gudbrandsdalen	Inland	Southern
Sk-2*	61°53'18"	8°18'43"	598	Rapid	No	Gudbrandsdalen	Inland	Southern
Lom-4*	61°40'52"	8°13'51"	850	Rapid	Yes	Gudbrandsdalen	Inland	Southern
Eid-1*	60°6'44"	12°7'28"	200	Rapid	Yes	Kongsvinger	Inland	Southern
Kon-1	60°9'48"	12°14'27"	400	Rapid	Yes	Kongsvinger	Inland	Southern
Kon-2	60°9'10'	12°5'6"	350	Rapid	Yes	Kongsvinger	Inland	Southern
Kon-3*	60°8'53"	12°5'6"	300	Rapid	Yes	Kongsvinger	Inland	Southern
Kvi-1*	59°29'38"	8°24'55"	703	Slow	Yes	Telemark	Inland	Southern
Kvi-2	59°29'25"	8°24'5"	680	Rapid	Yes	Telemark	Inland	Southern
Vin-1	59°32'9"	7°56'58"	524	Rapid	Yes	Telemark	Inland	Southern

Table 2. Coefficients of determination for photoperiodic sensitivity, altitude, climatic factors and latitude among Norwegian populations of *Arabidopsis thaliana*.
*=P≤0.05, **=P≤0.01, ***=P≤0.001, ns=non significant. Distance from the ocean was estimated based on yearly mean precipitation (1961-1990). Winter temperature was estimated based on mean temperatures for January (1961-1990). Mean days to bolting were used for photoperiodic response estimates.

Photoperiodic sensitivity	**Distance from the ocean**	**Winter temperature**	**Altitude**	**Latitude**
$\mathbf{P_{8/16}}$	0.64^{**}	0.53^{**}	0.51^{*}	0.20^{ns}
$\mathbf{P_{8/24}}$	0.89^{***}	0.85^{***}	0.62^{**}	0.18^{ns}
$\mathbf{P_{16/24}}$	0	0	0	0

Table 3. Photoperiodic sensitivity for Norwegian populations of *Arabidopsis thaliana*. $P_{8/16}$ = mean DTB (days to bolting) regressed on 8 and 16 h photoperiod; $P_{8/24}$ = mean DTB regressed on 8, 16, 19, 21 and 24 h photoperiod; $P_{16/24}$ = mean DTB regressed on 16 and 24 h photoperiod. For all analyses simple linear regression was used.

	Population	$P_{8/16}$	$P_{8/24}$	$P_{16/24}$
COASTAL	Vgn-1	6.14	2.02	2.32
	Lod-1	11.13	2.26	0.39
	Tje-1	7.47	1.55	0.10
	Ors-1	10.67	2.54	1.48
INLAND	Nfro-1	2.00	0.54	1.36
	Sk-2	0.60	0.45	2.07
	Lom-4	5.53	1.07	0.14
	Eid-1	6.80	1.56	0.07
	Kon-3	4.36	1.00	1.09
	Kvi-1	1.29	1.12	4.33

Table 4. Nucleotide and deletion polymorphisms in *CO* in Norwegian populations, Col-0 and L*er* of *Arabidopsis thaliana*. Numbers indicate positions relatively to start codon ATG (2313 bp in total). DEL1=deletion of 7 bp, DEL2=deletion of 14 bp, DEL3=deletion of 3 bp, NA=no amino acid change. * represents polymorphisms reported by others. Shaded boxes indicate 3 haplogroups.

CO		promoter							exon 1				intron 1	
Position		-567	-549	-366	-360	-228	-39	-10	297	435	495	680	813	951
									NA	NA	NA	His/Arg		
Col-0		A	C	A	T	-	C	-	C	T	C	A	C	T
L*er*						DEL1	A						A	
Nfro-1	Inland		A	G	C	DEL1	A						A	
Sk-2	Inland					DEL1	A						A	
Sog-1	Coastal					DEL1	A		T				A	
Kvi-2	Inland					DEL1	A						A	
Vag-2	Inland					DEL1	A						A	C
Vag-3	Inland					DEL1	A						A	C
Lom-1	Inland					DEL1	A						A	C
Lom-4	Inland					DEL1	A						A	C
Vgn-1	Coastal					DEL1	A						A	C
Lod-1	Coastal					DEL1	A						A	C
Lod-2	Coastal					DEL1	A						A	C
Tje-1	Coastal					DEL1	A						A	C
Lod-3	Coastal					DEL1	A						A	C
Lod-4	Coastal					DEL1	A						A	C
Vgn-2	Coastal					DEL1	A						A	C
Ors-1	Coastal					DEL1	A						A	C
Veg-1	Coastal					DEL1	A						A	C
Had-1	Coastal	G				DEL2	A			C	T		A	
Had-2	Coastal	G				DEL2	A			C	T		A	
Had-3	Coastal	G				DEL2	A			C	T		A	
Lom-2	Inland	G				DEL2	A			C	T		A	
Eid-1	Inland						A	DEL3				G	A	
Kon-1	Inland					DEL1	A	DEL3				G	A	
Kon-2	Inland					DEL1	A	DEL3				G	A	
Kon-3	Inland					DEL1	A	DEL3				G	A	
Kvi-1	Inland					DEL1	A	DEL3				G	A	
Vin-1	Inland					DEL1	A	DEL3				G	A	
Sk-1	Inland					DEL1	A	DEL3				G	A	
Sel-1	Inland					DEL1	A	DEL3				G	A	
			*	*			*						*	

Table 5. Summary of the transcript level investigations in Norwegian populations of *Arabidopsis thaliana*. F – transcript level association with flowering time under long days, S – transcript level association with photoperiodic sensitivity.

		Population	CRY2		TOC1		GI		CO	
North	Coast	Lod-1	–		–		F		F	
North	Coast	Tje-1	–	S	–	S	F	S	F	–
South	Coast	Ors-1	–	S	–	S	F	S	F	–
South	Inland	Sk-2	–	S	–	S	F	S	F	–
South	Inland	Eid-1	–	S	–	S	F	S	F	–
South	Inland	Kvi-1	–	S	–	S	F	S	F	–

Suplementary materials

Figure S1. Phylogenetic tree based on *CONSTANS* (maximum likelihood). The numbers indicate bootstrap values, the number of bootstrap replications was 1000.

Figure S2. Relative transcript levels of *CRY2* at the 9th h, *TOC1* at the 1st and the 9th h, and *CO* at the 9th h of the 16 h photoperiod in 6 Norwegian populations and Col-0 of *Arabidopsis thaliana*. Transcript level of *CO* was associated with DTB at 16 h photoperiod ($P \leq 0.001$). Transcript levels were normalized to *actin*. Filled squares (■) indicate one biological replicate. ***=$P \leq 0.001$. Results are mean of 2 biological replicates analysed in duplicate ±SE.

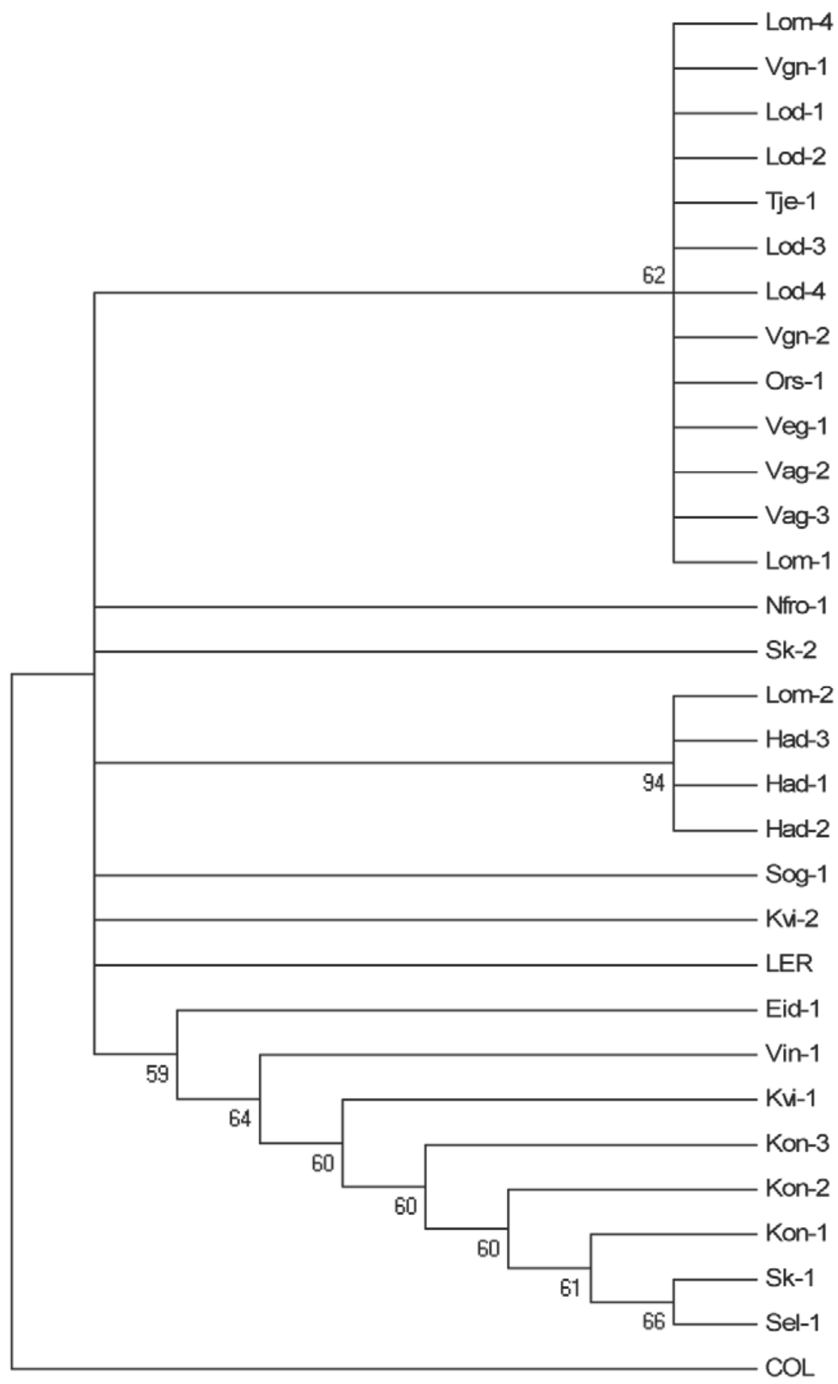


Figure S1

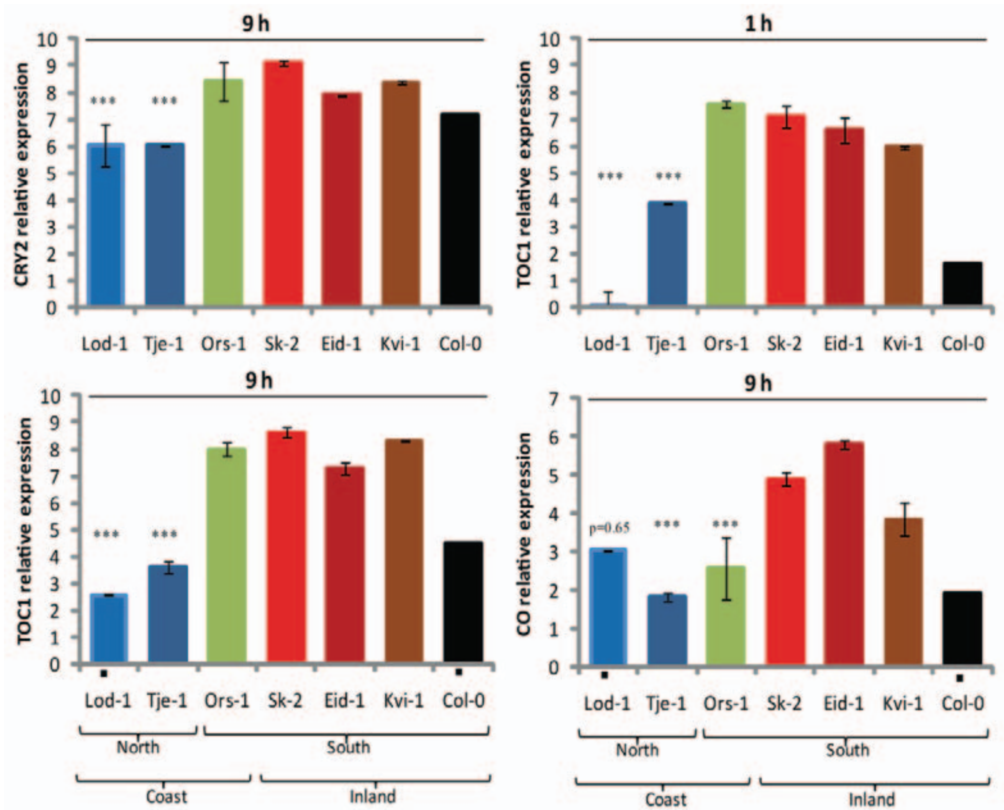


Figure S2

Table S1. Primers used for amplification and sequencing.

PCR/Sequencing Primer name	Sequence (5' →3')
PHYC-1 ex1 F	GGAGAATCATTCGGATCCTTG
PHYC-1 ex1 R	AAAGCATCACACAACAACAACA
PHYC-2 ex1 F	CTGCCGGGGCTTTAAGAT
PHYC-2 ex1 R	GATCCATTATATTTGGTGATTGAGTG
PHYC-3 ex1 F	TTCCGCAATGCACCAATAG
PHYC-3 ex1 R	CCACAGAAAAGATATTTGCTTCC
PHYC ex2 F	CTGTTTTCGCAGATTTTCATC
PHYC ex2 R	GCGAATGAGAAGAATAGAATAATCC
PHYC ex3 F	TCTAAATCTTCTGTTCTAATCCTGTGC
PHYC ex3 R	TCTTGTAGACAGGTCTTCAGCTTC
CRY 1 ex1&2 F	CTGGTTGTGGTTCTGGTGGT
CRY 1 ex1&2 R	TGGAAGAAGAGGAGACTCAGG
CRY 1 ex3 F	GGATCCATTGGTGTTTGAGG
CRY 1 ex3 R	GAGAGTCGGATTTGCGAGTT
CRY 2 ex1&2 F	GCTGCTCACGAAGGATCTGT
CRY 2 ex1&2 R	TGCAGTTATTGGCATCAACC
CRY 2 ex3 F	GAGAATGAGGCCGAGAAACC
CRY 2 ex3 R	ATACTGCCAGCCAAGGATGT
CRY 2 ex4 F	CGCCAAATATGACCCAGAAG
CRY 2 ex4 R	TCTTCCCTTCTGATGCCAAC
CRY2 promF	AAGTGGGCATGACCAGAATC
CRY2 promR	AGCTGCTGCTAATGCAGGAT
CO-1 F	CCCCTCTACTTATTCTCTCACACC
CO-1 R	AAAGCCTTACGTTATGGTTAATGG
CO-2 F	CCTCAGGGACTCACTACAACG
CO-2 R	TGGTTTTGTTTAACTAATGCAGAGA
CO promF	CATGTGGACTCCAAAATGC
CO promR	CCACTACCTATGTCGTTACTCTCTTG
TOC1 promF	TTTCTTTCTGTCGACCTTTTGA
TOC1 promR	TCAGATTAACAACTAAACCCACACA

Table S2. Primers used for real-time RT-PCR.

Real-time PCR Primer name	**Sequence (5' →3')**	**Ref.**
primerF CRY2-RT	AATCCCGCGTTACAAGGC	Endo *et al*., 2007
primerR CRY2-RT	TTCCGAGTTCCACACCAG	
primerF GI-RT	TTCTCGCAAGTGCATCAGATCT	Paltiel *et al*., 2006
primerR GI-RT	CCGTGGCTTCAAGTAGCTCAA	
primerF FKF1-RT	GGACGTTGCTTCAATTGGTT	Primer3
primerR FKF1-RT	AACCGACCCCATCTTAGCTT	
primerF CO-RT	ATGGCCGGTTCGCAAA	Primer3
primerR CO-RT	TGTTGTACATTAGCATCGTGTTGAA	
primerF TOC1-RT	TCACCATGAGCCAATGAAAA	Ding *et al*., 2007
primerR TOC1-RT	TTGAAACTTCTCCGCCAAAC	
primerF TUBβ-RT	AGGGAATGGACGAGATGGAAT	Primer3
primerR TUBβ-RT	TCCGCAGTTGCATCTTGGT	
PrimerF ACT-RT	AACTCGAGACAGCCAAAACCA	Primer3
PrimerR ACT-RT	ACGGAATCGCTCTGATCCAAT	
PrimerF PHYA-RT	GGAACTCGCTGCAGCAAACAAGTA	Primer3
PrimerR PHYA-RT	AGCGATAAGGAAAGTTACAAGCAAAG	

Ding Z, Millar AJ, Davis AM, Davis SJ. 2007. *TIME FOR COFFEE* encodes a nuclear regulator in the *Arabidopsis thaliana* circadian clock. *Plant Cell* **19**: 1522-1536.

Endo M, Mochizuki N, Susuki T, Nagatini A. 2007. *CRYPTOCHROME2* in vascular bundles regulates flowering in *Arabidopsis*. *The Plant Cell* **19**: 84-93.

Paltiel J, Amin R, Gover A, Ori N, Samach A. 2006. Novel roles for *GIGANTEA* revealed under environmental conditions that modify its expression in *Arabidopsis* and *Medicago truncatula*. *Planta* **224**: 1255-1268

Table S3. Climate and day length (DL) data for the sites of origin of the 10 populations of *Arabidopsis thaliana* used in the photoperiodic response experiment.
DL- mean day length for the first 30 days in the spring with average daily temperature above 5°C and no snow cover. T_{Jan} – monthly means of temperature from January, P_{year} – yearly means of precipitation; for years 1961-1990.

Population	Latitude	Longitude	T_{Jan}	P_{year}	DL
Vgn-1	68°10'11"	14°13'15"	-1.5	1500	21.69
Lod-1	68°23'51"	15°56'6"	-3.1	1600	20.11
Tje-1	68°25'29"	16°22'9"	-3.8	1045	23.69
Ors-1	62°12'19"	6°32'1"	-1.0	2040	18.66
Nfro-1	61°34'38"	9°39'42"	-11.5	430	16.74
Sk-2	61°53'18"	8°18'43"	-9.4	317	17.04
Lom-4	61°40'52"	8°13'51"	-8.6	548	18.78
Eid-1	60°6'44"	12°7'28"	-6.7	740	17.20
Kon-3	60°8'53"	12°5'6"	-6.7	740	17.20
Kvi-1	59°29'38"	8°24'55"	-6.5	820	16.88

Table S4. Coefficients of determination for flowering time and altitude, climatic factors, latitude and day length (DL) in 10 populations of *Arabidopsis thaliana* in 5 photoperiod treatments. *P≤0.05, **P≤0.01, ***P≤0.001; DTB = days to bolting, DTF = days to flowering; T_{Jan} – monthly means of temperature from January, P_{year} – yearly means of precipitation; for years 1961-1990.

	DTB24	**DTF24**	**DTB21**	**DTF21**	**DTB19**	**DTF19**	**DTB16**	**DTF16**	**DTB8**	**DTF8**
Altitude	0.389^{***}	0.445^{***}	0.102^{***}	0.104^{***}	0.271^{***}	0.325^{***}	0.220^{***}	0.264^{***}	0.439^{***}	0.475^{***}
T_{Jan}	0.681^{***}	0.720^{***}	0.326^{***}	0.290^{***}	0.555^{***}	0.622^{***}	0.485^{***}	0.513^{***}	0.670^{***}	0.652^{***}
P_{year}	0.697^{***}	0.735^{***}	0.379^{***}	0.365^{***}	0.626^{***}	0.680^{***}	0.494^{***}	0.526^{***}	0.720^{***}	0.740^{***}
Latitude	0.222^{***}	0.256^{***}	0.077^{***}	0.060^{*}	0.124^{***}	0.166^{***}	0.104^{***}	0.154^{***}	0.221^{***}	0.190^{***}
DL	0.235^{***}	0.253^{***}	0.064^{**}	0.041^{**}	0.119^{***}	0.159^{***}	0.105^{***}	0.138^{***}	0.215^{***}	0.165^{***}

Table S5. Polymorphisms table of *TOC1* promoter region in *Arabidopsis thaliana*.
Numbers indicate positions relatively to start codon ATG (780 bp in total). Asterisk (*) indicates polymorphisms reported by others.

TOC1 promoter								
Position	-659	-658	-607	-593	-519	-502	-501	-227
Col-0	T	T	C	A	T	T	C	A
L*er*	-	-				-	-	G
Nfro-1		-	T	C				
Sel-1		-						
Vag-2		-	T	C				
Vag-3		-	T	C				
Lom-1		-	T	C				
Lom-2		-	T	C				
Sk-1		-	T	C				
Sk-2		-	T	C				
Lom-4								G
Vgn-1		-	T	C				
Had-1		-	T	C				
Had-2		-	T	C				
Had-3		-	T	C				
Lod-1		-	T	C				
Lod-2		-	T	C				
Tje-1		-	T	C				
Lod-3		-	T	C				
Lod-4		-	T	C				
Vgn-2		-	T	C	C			
Sog-1								G
Ors-1		-	T	C				
Veg-1		-	T	C				
Eid-1								G
Kon-1								G
Kon-2		-	T	C				
Kon-3		-	T	C				
Kvi-1		-						
Kvi-2		-						
Vin-1		-						
	*	*		*				*

CRY2 5' flanking region

Col-0	C	A	G	-	T	C	T	G	A	C	T	G	A	C	T	T	A	A	A	C	C	-	T	-	A	C	C	G	G	A
L*er*				-	G																									
Nfro-1				-	G																									
Sel-1				-	G																									
Vag-2	T	T	A	IN1		G	A	C	DEL1	A	C	A	C	T	A	A	T	-	G	T	T	T	A	IN2	T	A	A	C	C	T
Vag-3	T	T	A	IN1		G	A	C	DEL1	A	C	A	C	T	A	A	T	-	G	T	T	T	A	IN2	T	A	A	C	C	T
Lom-1				-	G																									
Lom-2				-	G																									
Sk-1				-	G																									
Sk-2				-	G																									
Lom-4				-	G																									
Vgn-1				-	G																									
Had-1				-	G																									
Had-2				-	G																									
Had-3				-	G																									
Lod-1				-	G																									
Lod-2				-	G																									
Tje-1				-	G																									
Lod-3				-	G																									
Lod-4				-	G																									
Vgn-2				-	G																									
Sog-1				-	G																									
Ors-1				-	G																									
Veg-1				-	G																									
Eid-1				-	G																									
Kon-1				-	G																									
Kon-2				-	G																									
Kon-3				-	G																									
Kvi-1				-	G																									
Kvi-2				-	G																									
Vin-1				-	G																									
		*	*		*	*	*	*					*	*		*			*	*	*		*			*			*	

Table S6. Polymorphisms table of *CRY2* 5' flanking region in *Arabidopsis thaliana*. Numbers indicate positions relatively to start codon ATG (748 bp in total). IN1=insertion of 2 bp (AA), DEL1=deletion of 12 bp, IN2=insertion of 4 bp. Asterisk (*) represents polymorphisms reported by others.

Table S7. Polymorphisms table of *SPY* in *Arabidopsis thaliana*. Numbers indicate positions relatively to start codon ATG (1413 bp in total). Asterisk (*) represents polymorphisms reported by others.

SPY	exon 1	intron 2	intron 3		intron 4
Position	326	790	1008	1046	1246
	His/Arg(109)				
Col-0	A	A	T	A	G
Ler			G	T	
Nfro-1		T		T	A
Sel-1	G			T	
Vag-2	G			T	
Vag-3	G			T	
Lom-1	G			T	
Lom-2				T	
Sk-1	G			T	
Sk-2	G	T		T	
Lom-4				T	
Vgn-1				T	
Had-1		T		T	
Had-2		T		T	
Had-3		T		T	
Lod-1	G			T	
Lod-2	G			T	
Tje-1	G			T	
Lod-3	G			T	
Lod-4	G			T	
Vgn-2	G			T	
Sog-1				T	
Ors-1		T		T	
Veg-1	G			T	
Eid-1	G			T	
Kon-1	G			T	
Kon-2	G			T	
Kon-3	G			T	
Kvi-1		T		T	
Kvi-2		T		T	
Vin-1				T	
			*	*	

PHYC — exon 1 (positions 26–162), exon 2 (positions 231–289), exon 3 (positions 320–328)

Position	26	136	147	182	192	409	525	546	609	688	702	757	111	121	124	132	132	132	133	138	139	141	148	150	151	152	154	157	157	158	161	162	┤├	231	234	234	239	259	260	269	289	┤├	320	324	327	328	328	328
	Cys/Tyr	Ile/Val	NAC	Ala/Val	NAC	Gln/Glu	NAC	NAC	NAC	Gly/Ser	NAC	NAC	NAC	NAC	NAC	NAC	NAC	NAC	Asn/Lys	NAC	NAC	NAC	NAC	NAC	Ser/Thr	NAC	Ser/Cys	NAC	NAC	NAC	NAC	NAC		Thr/Lys	Lys/Gln/A	Lys/Gln/A	Met/Ile	NAC	Glu/Lys	NAC	NAC		NAC	NAC	NAC	NAC	Leu/Trp	Arg/Lys
Col-0	G	A	C	C	G	G	A	T	T	G	G	C	A	T	C	A	T	A	C	T	C	A	A	C	T	C	C	T	A	C	A	T		C	A	A	A	A	G	C	C		T	T	G	A	T	G
Ler	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Nfro-1									C		T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C		G	G	A
Sel-1	A	G	T		T	C	T	G	C		T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Vag-2						C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T	G	C	C	T	G	C		A	C		G	G	A	G	A		C	C		G	G	A
Vag-3						C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T	G	C	C	T	G	C		A	C		G	G	A	G	A		C	C		G	G	A
Lom-1	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Lom-2	A				T	C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Sk-1	A				T	C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C		G	G	A
Sk-2	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Lom-4	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Vgn-1	A	G	T	T	T	C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C	G	G	G	A	G	A		C	C		G	G	A
Had-1	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Had-2	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Had-3	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Lod-1	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Lod-2	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Tje-1	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Lod-3	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Lod-4	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Vgn-2	A		T		T	C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Sog-1	A	G	T		T	C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T	G	C	C	T	G	C		A	C		G	G	A	G	A		C	C		G	G	A
Ors-1	A	G	T	T	T	C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C	G	G	G	A	G	A		C	C		G	G	A
Veg-1	A				T	C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C		G	G	A
Eid-1	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C		G	G	A
Kon-1	A		T		T	C	T	G		A	T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Kon-2	A		T		T	C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C	T	G	G	A
Kon-3	A		T		T	C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T	G	C	C	T	G	C		A	C		G	G	A	G	A		C	C		G	G	A
Kvi-1	A	G	T	T	T	C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C	G	G	G	A	G	A		C	C		G	G	A
Kvi-2	A	G	T	T	T	C	T	G			T	A	G	C	T	C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C	G	G	G	A	G	A		C	C		G	G	A
Vin-1	A	G	T		T	C	T	G								C	C	T	G	C	T	C	C	A	A	T		C	C	T	G	C		A	C		G	G	A	G	A		C	C		G	G	A
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*		*	*	*	*	*	*	*	*		*	*	*	*	*	*

Table S8. Polymorphisms table of *PHYC* in *Arabidopsis thaliana.* Numbers indicate positions relatively to start codon ATG (3352 bp in total). NAC - no amino acid change. Asterisk (*) represents polymorphisms reported by others.

PAPER III

Genetic diversity and population structure in *Arabidopsis thaliana* from Northern Europe

Anna M. Lewandowska-Sabat, Siri Fjellheim, Odd Arne Rognli

Department of Plant and Environmental Sciences, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway

Corresponding author
E-mail: odd-arne.rognli@umb.no
Telephone: +47 64 96 55 78
Fax: +47 64 96 59 01

Abstract

Genetic diversity and population structure in *Arabidopsis thaliana* populations from Norway were studied in order to understand the demographic history of populations adapted at the northernmost boundary of the species. Moreover, in order to elucidate possible colonization routes Norwegian populations were compared to a worldwide sample of *A. thaliana*. We genotyped 282 individuals from 31 local populations using 149 single nucleotide polymorphism (SNP) markers. A high level of population subdivision (F_{ST}=0.85 ± 0.007) was found indicating that *A. thaliana* is highly structured on a regional scale. Significant relationships between genetic and geographic distances were found indicating isolation-by-distance mode of evolution. The genetic diversity was much lower in populations from the North (65-68°N) than populations from the South (59-62°N) consistent with a northward expansion pattern. Moreover, a higher level of linkage disequlibrium (LD) was found within the Northern populations. A Neighbor-Joining (NJ) tree showed that populations from Northern Norway form a separate cluster while the remaining populations are distributed over a few minor clusters. There seems to have been minimal gene flow between populations of different regions, specifically between the geographically distant Northern and Southern populations. Our data suggest that Northern populations represent a homogenous group that may have been established from a few founders during northward expansion, while populations from central part of Norway constitute an admixed group established by founders of different origins, most probably a result of human mediated gene flow. Moreover, Norwegian populations appeared as homogenous and isolated as compared to a worldwide sample of *A. thaliana*, but they stil group together with Swedish populations, which may indicate common colonization histories of these populations of *A. thaliana*.

Key words: *Arabidopsis thaliana*, gene flow, isolation by distance, linkage disequilibrium, population structure.

Introduction

Scandinavia constitutes the northern distribution limit of many European plant species. At the margin severe environmental conditions makes reproduction and survival challenging tasks. Genetic drift, gene flow and natural selection determine genetic variation and structure of populations, and genetic characterization is a key factor in understanding historical events, population dynamics and footprints of selection. During geographical range expansion genetic variation in populations is often reduced at the margin of the the species' distribution range due to bottlenecks (Petit et al. 2002; Eckert et al. 2008). In populations that experience size fluctuations and become progressively smaller towards the margin genetic variation is also reduced through genetic drift (Vucetich and Waite 2003). Moreover, depletion of genetic variation in marginal populations may decrease the ability of a species to respond to selection (Hewitt 2000). Populations from regions with steep environmental gradients may experience instability in population size, increased drift and genetic divergence (Pamilo and Savolainen 1999), and this might result in decreased variability of populations from extreme environment.

In recent years, increasing interest in elucidating the genetic structure of the model plant *Arabidopsis thaliana* have spurred several studies on the geographical pattern of genetic diversity, both at the regional and global scale. This pattern is influenced both by historical and recent changes in population size and gene flow (Vucetich and Waite 2003). *A. thaliana* is nearly completely self-fertilized and therefore its potential long-distance gene flow is due to seed and not pollen dispersal (Abbott and Gomes 1989). *A. thaliana* is commonly found in human-disturbed areas and its seeds are viewed as being easily transported by human activity. Much of the dispersal in *A. thaliana* is most likely unintentional, since this plant is of no agricultural interest (Beck et al. 2008). In earlier studies of genetic variation in *A. thaliana*, no association between genetic and geographic distance (isolation by distance, IBD) has been found in global samples of *A. thaliana* (King et al. 1993; Todokoro et al. 1995; Bergelson et al. 1998; Miyashita et al. 1999). The lack of IBD was attributed to a recent expansion of the species associated with spread of agriculture. However, later studies have demonstrated IBD and significant population structure among natural populations of *A. thaliana* on a wide geographic scale (Sharbel et al. 2000; Nordborg et al. 2005; Ostrowski et al. 2006; Schmid et al. 2006; Beck et al. 2008). A lower level of IBD as a result of human activity was detected among native European populations compared to populations from more undisturbed areas in

Asia (Beck et al. 2008). Therefore, genetic structure resulting from historical expansions of plant species could be masked by long distance human-mediated dispersal.

It has been suggested that *A. thaliana* re-colonized Scandinavia after the last glaciations (~10,000 ya) via southern and northeastern routes (Sharbel et al. 2000; François et al. 2008). Consequently, the level of genetic variation in Northern populations should be lower due to stepwise postglacial migration either from South or Northeast. Thus, significant isolation by distance should be present among these populations indicating genetic structure. However, the few studies of population structure of *A. thaliana* from its peripheral Northern distribution range that have been conducted include only a few populations (Kuittinen et al. 1997; Stenøien et al. 2005; Lundemo et al. 2009) and none from its ultimate Northern distribution boundary. Decline in genetic variation at microsatellite loci was found among Northern haplotypes of *A. thaliana* (Kuittinen et al. 1997) and only studies of Lundemo et al. (2009) on *A. thaliana* and Kärkkäinen et al. (2004) on populations of the outbreeding relative *Arabidopsis lyrata* from Scandinavia have shown significant IBD.

To elucidate the evolutionary history of *A. thaliana* at its northernmost distribution range we have studied genetic structure of local Norwegian populations. In Norway *A. thaliana* grow in sparse, extremely fragmented habitats throughout the central and Northern part of the country and in continuous habitats along the Southern coast (Naturhistoriska riksmuseet 1998, http://linnaeus.nrm.se/flora/di/brassica/arabs/arabtha.html). Habitat fragmentation should in principle lead to reduced gene flow and lower genetic diversity within populations relative to areas with less fragmentation and greater population density.

Our sampling strategy focused on individuals representing native peripheral populations originating from isolated sites with no human influence. Most of these populations were found on sites with relatively little human impact, i.e. isolated rock shelves within a hill and coastal rocks, far from pastures and agricultural landscapes. Most likely, they represent populations stemming from post-Pleistocene colonization. However, since *A. thaliana* is spreading as a weed, it therefore often occur as human commensal, growing along roadsides, railways, river banks, pastures and cultivated landscapes. Thus, in spite of the fact that our sampling strategy was directed towards isolated sites, we have probably not been able to entirely avoid populations that are a result of recent introductions. A collection of 282 individuals sampled from 31 populations covering latitudinal and altitudinal gradients in Norway, was genotyped using genomewide single nucleotide polymorphisms (SNPs) at 149 loci. The objectives of this study were to i) describe genetic structure and degree of IBD in Norwegian populations, ii) study the genetic relationship between Norwegian populations

and a worldwide sample of *A. thaliana* populations, and iii) analyze the genetic diversity in order to increase our knowledge about the demographic and evolutionary processes acting on populations at the northernmost margin of their distribution.

Materials and Methods

Study species and plant material

Arabidopsis thaliana (L.) Heyhn is an annual self-fertilizing weedy plant, which is distributed worldwide, tolerates little competition, and is adapted to dry habitats. Populations growing in Norway (up to 68°N) are at the northernmost distribution edge of the species. In this study, 31 populations of *A. thaliana* collected across latitudinal (59-68°N) and altitudinal gradients (2-850 m a.s.l.) in Norway (Table 1) were used.

Genotyping

Original seeds sampled from the field or multiplied one generation in a growth chamber were used to establish plants. Leaf materials were harvested from individual plants grown under a 16 h light/8 h dark day/night regime at 23°C in a growth chamber for 2.5 weeks. Genomic DNA was extracted from leaf tissue using DNeasy 96 Plant Kit (Qiagen GmbH, Hilden, Germany) following the protocol of the manufacturer and 0.4µg of DNA (10ng/µl) was used for genotyping. Four to ten individuals from each of 31 populations (n=282; for number of individuals genotyped from each population see Table 1) were genotyped using 149 single nucleotide polymorphisms (SNPs) markers distributed evenly across the entire genome. Genotyping was performed using the Sequenom MassArray system at Sequenom (San Diego, CA, USA) and at the University of Chicago DNA sequencing facility (Chicago, IL, USA). The primer sequences as well as physical distances of SNP markers are listed on the web (http://borevitzlab.uchicago.edu/resources/molecular-resources/snp-markers).

Statistical analyses

We calculated the proportion of polymorphic loci (PPL) as the number of polymorphic loci per total number of loci, observed heterozygosity H_O, expected heterozygosity H_E and average gene diversity over loci in each population. Student's t-test was used to compare H_O,

H_E, gene diversity and PPL between Northern and Southern populations. Simple correlation analyses between PPL, H_O, H_E, gene diversity and latitude of population origin were performed. Population differentiation and differentiation between populations were estimated based on pairwise comparisons using F_{ST} (Weir and Cockerham 1984). Analysis of molecular variance (AMOVA) was performed to asses between- and within population variation. Estimation of the different molecular diversity indices, population genetic parameters and analysis of molecular variation AMOVA was performed using Arlequin version 3.1 (Excoffier et al. 2005). Linkage Disequilibrium (LD) for each pair of loci was calculated separately for each chromosome and genomewide using the correlation coefficient r^2 (Hill and Weir 1994) and rxc software (Zaykin et al. 2008, http://www.niehs.nih.gov/research/atniehs/labs/bb/staff/zaykin/rxc.cfm). LD decay along chromosomes and genomewide was tested by regressing pairwise measures of r^2 (monomorphic loci within population groups were not included in the analysis) against genetic distances between loci (in base pairs). The level of LD was calculated as the proportion of pairwise loci comparisons with r^2 higher than 0.7. LD was calculated across all populations as well as within population groups (Table 1).

An estimator for η ($\eta = N_e u$) was calculated from the formula $\eta = 0.75H / (3 - 4H)$ according to Lynch and Conery (2003). N_e is the effective population size, u is the substitutional mutation rate per nucleotide per generation and H is the average number of substitutions between silent sites in two randomly sampled allelic sequences. H was estimated as the mean number of pairwise differences θ_{Pi} (Tajima 1983) over loci for each population. η was compared to elucidate differences in N_e between populations. We calculated $N_e u$ instead of N_e due to the fact that mutation rates for SNPs may be higher than suggested (Wright et al. 2002).

We used STRUCTURE 2.2.3 (Pritchard et al. 2000; Falush et al. 2003; through the BIOPORTAL http://www.bioportal.uio.no/) to assign individuals to a number of putative clusters (K). Structure was run for 1 000 000 iterations after a burn-in length of 100 000 iterations for deduction of each of K populations. Five runs were considered for each K, with K ranging from 1 to 15. The admixture model was run with allele frequencies uncorrelated using all populations as well as omitting populations from Central Norway. The most likely number of clusters was detected calculating an *ad hoc* statistic ΔK based on change in the log probability of data between successive K clusters (Evanno et al. 2005) using STRUCTURE-SUM (Ehrich D 2006) in R 2.9.2 (R Development Core Team 2009). Similarity coefficients among each pair of structure runs for each K (Rosenberg et al. 2002) were calculated using

the same method. Overall proportion of membership of the sample in each of the 5 or 3 clusters was averaged over 5 runs and used to present genetic composition of each individual.

Neighbor-Joining trees for Norwegian and worldwide populations were constructed in TREEFINDER (Jobb 2004) with the Nei's distance matrix (D_A, Nei et al. 1983) calculated using Populations 1.2.30 (http://bioinformatics.org/~tryphon/populations/#ancre_distances) as a measure of relative genetic distances between populations. SNP genotype data of worldwide accessions were kindly provided by Alexander Platt (University of Southern California, Center of Excellence in Genome Science, Office 413J 1050 Childs Way, Los Angeles, CA 90089).

Isolation by distance (IBD) was analyzed for all 31 Norwegian populations as well as for Southern and Northern (68-65°N, Table 1) populations separately. Genetic distances were estimated as shared allele distance, D_{SA} (Chakraborty and Jin 1993) and compared to geographical distances based on population coordinates using a Mantel test (Mantel 1967).

Results

Variation among and within populations

We removed 11 loci that failed to give any results and used 138 loci in further analyses. AMOVA analyses show that the highest level of variation was found between populations within geographical subregions (regions defined in Table 1) which accounted for 55.6%. Variation between geographical subregions accounted for 30.6% while variation within populations was estimated to be 11.7%. H_O, H_E, gene diversity and PPL were lower in Northern than in Southern populations (Table 2). Average H_O and H_E (± SE) across populations were 0.009 (± 0.001) and 0.051 (± 0.003), respectively. PPL in populations ranged from 0 to 0.421 with a mean of 0.13 (± 0.03). Gene diversity in populations ranged from 0-0.141 with an average of 0.041 (± 0.01). Summary of molecular diversity indices for each population is presented in Table 2. Correlations between H_O, H_E, PPL, gene diversity and latitude were significant (r=-0.46, -0.58, -0.55, -0.58, respectively; $p \leq 0.01$; see Fig. 1 for gene diversity and latitude).

Average pairwise F_{ST} (mean ± SE) across populations was 0.85 ± 0.007, ranging from 0.31 to 1.00 between all pairs of populations. A positive and significant correlation was found between shared allele distance (D_{SA}) and geographical distance for all populations (r=0.69, p=0.001; Fig. 2) which indicate IBD. Positive and significant correlations were also found

when the Southern and Northern (68-65°N) groups of populations were tested separately (r=0.60, 0.66; p=0.001; respectively, Fig. S1) indicating IBD within these two spatially separated regions.

Except for population Vgn-1, populations from Northern Norway appeared as two homogenous groups (see Fig. 3 and Fig. 4) based on the STRUCTURE analyses. Pairwise F_{ST} between populations within these groups were non significant indicating genetic similarity of populations. For all pairwise F_{ST} see Table S1.

Significant associations between LD (pairwise r^2) and genetic distance were found for chromosome 1 and genomewide across all populations. The highest level of LD was found in Northern populations and the lowest in Central Norwegian populations. Genomewide LD decay was observed across all populations and in Central Norwegian populations. The level of LD decreased with number of populations, except for Northern populations that were outside the overall pattern of a distribution (Table 3, Fig. S2).

$N_e u$ within populations was highest for populations from Central, Southeastern Norway and from the Southwestern Coast and lowest for populations from Northern Norway and the Northwestern Coast (Fig. 5).

Genetic structure

The number of clusters was detected by STRUCTURE applying a Bayesian algorithm at K=5, however the similarity coefficient for runs at K=5 was non significant (Fig. S3a and b). Therefore, structure analyses were repeated after omitting populations from Central Norway that seem to represent admixed individuals originating from several founders. Consequently, the number of clusters was detected at K=3 and the similarity coefficient was significant (Fig. S3c and d). Population genetic structure at ΔK=5 and ΔK=3 are presented in Fig. 3.

Nei's distance neighbor-joining tree of 31 Norwegian populations clustered the sample into 4 main groups generally reflecting geographic origin of the populations. A dendrogram showing the relationship among the populations is presented in Fig. 4.

Nei's distance neighbor-joining tree of a worldwide sample of *A. thaliana* populations clustered the populations into several minor groups (Fig. 6, Table S2 for names and origins of populations). The Norwegian populations and most of the Swedish as well as the Irish populations, plus two British and one Danish population clustered together in one clade. Some Asian, Central European and Spanish populations also grouped with this clade. One Norwegian population (not sampled by the present authors), several Swedish and Finnish

populations grouped in separate clades. Most of the populations from USA, Canada, UK and Northern and Western Europe were grouped together as well as did the Asian populations.

Discussion

Norwegian populations of A. thaliana *show complex genetic structuring*

Arabidopsis thaliana is often found in human-disturbed areas within its native world-wide distribution range (Beck et al. 2008). In Norway, its distribution in highly fragmented habitats results in some populations being extremely isolated with limited gene flow and little impact by human activity (Naturhistoriska riksmuseet 1998; Fremstad 1997). These populations are believed to be a result of postglacial colonization. In the present study we targeted sampling of populations from isolated sites with possibly no major impact by human activity. This sampling strategy was expected to reveal genetic structure among Norwegian populations of *A. thaliana* resulting from geographical range expansion (Beck et al. 2008). However, we found a cryptic population structure (Fig. 3) which likely is a result of both natural colonization and human-mediated seed dispersal. This human-mediated gene flow may cause departure from a historical stepping-stone model of genetic structure and isolation by distance (Beck et al. 2008). Particularly, this is the case in the Central Norwegian populations which appear as the most heterogeneous group and most likely consists of a mixture of ancient and newly introduced populations. The STRUCTURE analysis infers that the Central populations represent a mixture of genotypes from all geographic regions (Fig. 3) and they do not form one consistent group in the cluster analysis (Fig. 4). Moreover, the LD decay in the Central populations is significant (Table 3, Fig. S2), which indicate that recent admixture events and recombination events reducing LD may be related to their location along the main traditional trade routes, but also present-day gene flow. In addition, the largest effective population sizes were found within Central Norwegian populations (Fig. 5) indicating that in these populations more individuals contribute gametes to the next generation and that they probably had more founders than Northern populations. Since historical times the geographic region harboring the Central populations has had important trade routes crossing and is an active agricultural region. Thus, human mediated transport of *A. thaliana* seeds to this region is highly likely.

An opposite pattern was found in Northern Norwegian populations, for which the STRUCTURE analysis revealed that they appear remote and isolated from the remaining

populations (Fig. 3). Moreover, the highest level of LD was found among these populations (Table 3, Fig. S2) which confirms that Northern populations are a homogenous group that has experienced less recombination with immigrant genotypes than Central populations. High LD in these populations is probably due to population bottlenecks followed by inbreeding, limited gene flow and lower effective recombination rate. Moreover, this group of populations as well as Northwestern Coastal populations has the lowest effective population size (Fig. 5) which indicates that very few individuals contribute gametes to the next generation in the North. In these small and isolated populations genetic drift will have a large effect and differentiate them genetically from larger populations further South in Norway. The present results suggest that there has been enough time for genetic drift and mutations to differentiate populations from the North from the remaining Norwegian populations while gene flow associated with human activity in Central Norway has resulted in haplotype sharing and cryptic genetic structure among these populations. Moreover, Northern populations originate from undisturbed regions where the chances for spreading by agricultural activity are limited.

Furthermore, as revealed by STRUCTURE analyses Northern populations, except for population Vgn-1, are comprised of two groups (Fig. 3). These Northern populations may have originated from two different founders or diverged recently since they still cluster together but separately from the remaining Norwegian populations (Fig. 4). Genetic drift may have differentiated them from each other since they are situated on two different islands.

Southern and Southeastern populations group together with Coastal populations and this is particularly evident in the K=3 diagram, and at K=5 they share some alleles with populations from Central Norway. Coastal populations represent distinct fractions which are admixtures of Central genotypes. This is also apparent in the NJ-tree where these populations, except for Vin-1, form separate groups (Fig. 4). One Northern population (Vgn-1) is genetically similar to the Coastal group (Fig. 3) which may indicate that they have similar colonization history.

A clear pattern of a North-South differentiation emerges from the STRUCTURE analyses. This pattern may be due to different colonization routes in the South and the North (Sharbel et al. 2000; François et al. 2008) but also due to genetic drift in the Northern populations with small N_e.

The high level of LD was found in the Southern and Southeastern population groups, however this could be overestimated due to the low number of populations in these groups (Table 3, Fig. S2). The observed differences in levels of LD between Northern and Central

Norwegian populations are thus not an effect of different sample size (Fig. S2) but a result of differences in gene flow within these two regions. Across all populations LD decays rapidly, which is consistent with other studies on global population samples (Nordborg et al. 2002 and 2005; Kim et al. 2007), and is due to the fact that global population samples contain all historical polymorphisms and recombinations gathered over thousands of generations (Kim et al. 2007).

AMOVA analyses indicated high levels of genetic variability within and between regions, relatively higher than in other studies (Nordborg et al. 2005), but comparable to other studies of Norwegian *A. thaliana* populations (Stenøien et al. 2005; Lundemo et al. 2009). Moreover, strong differentiation as measured by F_{ST} even between populations located closely was also observed. This is an additional indication of strong genetic drift within populations and limited gene flow between populations within regions.

Norwegian populations are homogenous and isolated compared with a worldwide sample of A. thaliana

Even though there is strong differentiation between Norwegian populations of *A. thaliana* (Fig. 3, Fig. 4), they appear isolated in the analysis encompassing the worldwide population sample (Fig. 6). The study of genetic variation in *Arabis petraea* (Jonsell et al. 1996) showed that Swedish populations were genetically closer to Norwegian and distinct from Russian populations indicating different colonization routes for Scandinavian and Russian populations. This may also be the case for Scandinavian *A. thaliana* populations (Fig. 6, Table S2). Based on the relationship between the worldwide sample of populations and the Norwegian populations (Fig. 6), we might conclude that *A. thaliana* have been introduced to Norway from Sweden since most of Swedish and Norwegian populations cluster together. Finnish and Russian populations do not group with the remaining Scandinavian populations which might suggest different colonization patterns for these populations and excludes a possible immigration route of Norwegian populations from the Northeast. Some Coastal populations may have been introduced through main trade routes by sea, possibly dating back to the Norwegian Saga era as shown for *Festuca pratensis* (Fjellheim and Rognli 2005). Particularly, the coastal populations Veg-1 and Veg-2 (65°N) and the Northern population Vgn-1 share alleles with populations from the Southwestern Coast (Fig. 3) and cluster together with British and Irish populations (Fig. 6, Table S2). Populations from separate refugia will differ in only low frequency alleles due to mutation and drift, while high

frequency alleles should be common over a broad global scale and therefore genealogical relationships among populations are still present. On the other hand, assignment tests may not always be fully correct in inferring the historical origin of population since genetic drift or human-mediated colonization could remove genetic effects of ancient demographic postglacial colonization events.

Genetic variation is depleted in populations from the most peripheral geographic range of A. thaliana

The post-Pleistocene colonization process reduces genetic diversity in recently colonized peripheral part of the species range relative to the central refugia (Comes and Kadereit 1998). This depletion of genetic variation in marginal populations is due to relatively few founder individuals and population bottlenecks (Hewitt 2000). So far this prediction has been confirmed for many species at the margin of their distribution (Eckert et al. 2008), but except for a few studies (Todokoro et al. 1995; Kuittinen et al. 1997), no depletion of genetic variation towards the range margin of *A. thaliana* has been found (Stenøien et al. 2005; Lundemo et al. 2009). Our findings suggest that the variation in Northern populations might have been lost during the northward expansion of the species due to bottlenecks and few founder individuals (Fig. 1, Table 2). As a result, low estimated effective population size (Fig. 5) and limited immigration of unrelated genotypes due to extreme habitat isolation result in low genetic diversity and strong genetic differentiation from Southern populations. The opposite scenario is found in Central populations, where high effective population size (Fig. 5) and low LD (Fig. S2) seem to be a result of admixture with migrant genotypes. The extremely low levels of genetic diversity found in Northern populations in our study compared to other studies by Stenøien et al. (2005) and Lundemo et al. (2009) may be a result of sampling differences. The Northern populations in our study represent extreme, isolated habitats, situated far from grasslands and agricultural sites and they likely represent locally adapted populations. The lower diversity in Northern populations may also be due to an island effect as both the Northern and Northwestern Coast population groups are located on islands which restrict gene flow and migration.

The significant relationship between geographic and genetic distances found among the Norwegian populations (Fig. 2) as well as within the Southern and Northern populations separately (Fig. S1) indicate that isolation by distance is not due to extreme dissimilarity between these two geographically remote population groups (Fig. 2) but due to a stepwise

mode of postglacial colonization from South to North. Isolation by distance is also generally reflected in the dendrogram (Fig. 4), and suggests that the original biogeographic pattern of *A. thaliana* distribution is still visible and has not been completely masked by human activity.

Our study illuminate the difficulties created by cryptic genetic structures and demonstrate that working with a species that comprise of mixture of populations resulting from ancient colonization and antropochorous spreading may be challenging. As demographic factors contribute to a genome-wide departure from a neutral model of sequence polymorphism, very often significant associations between phenotypic and genotypic variation are product of false positive associations caused by population structure (Caicedo et al. 2004; Helgason et al. 2005; Zhao et al. 2008). Elucidating whether genetic variation is geographically structured at regional scales will allow us to interpret polymorphism data correctly. In view of this, our results might have major implications for developing sampling strategies and designing *A. thaliana* studies with special focus on local adaptation.

Acknowledgements

We would like to thank Yan Li and Justin Borevitz for genotyping our populations and Alexander Platt (University of Southern California, Center of Excellence in Genome Science) for providing the genotype data of the worldwide population sample. We are also indebted to Sverre Lundemo and Simen Rød Sandve for helpful remarks on the statistical analyses. This study was supported by a grant from the Research Council of Norway: The Norwegian Arabidopsis Research Centre, project no. 151991/150, and a PhD scholarship to Anna Lewandowska-Sabat from the Norwegian University of Life Sciences.

References

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

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

Bergelson J, Stahl E, Dudek S, Kreitman M (1998) Genetic variation within and among populations of *Arabidopsis thaliana*. Genetics 148: 1311-1323.

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. PNAS 101: 15670-15675.

Chakraborty R, Jin L (1993) Determination of relatedness between individuals using DNA fingerprinting. Human Biology 65: 875-895.

Comes HP, Kadereit JW (1998) The effect of Quaternary climatic changes on plant distribution and evolution. Trends in Plant Science 3: 432-438.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Molecular Ecology 17: 1170-1188.

Ehrich D (2006) AFLPdat: a collection of R functions for convenient handling of AFLP data. Molecular Ecology Notes 6: 603-604.

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

Excoffier L, Laval G, Schneider S (2005) Arlequin version 3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47-50.

Falush D, Stephens M, Pritchard J (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Fjellheim S, Rognli O (2005) Molecular diversity of local Norwegian meadow fescue (*Festuca pratensis* Huds.) populations and Nordic cultivars: consequences for management and utilisation. Theoretical and Applied Genetics 111: 640-650.

François O, Blum MGB, Jakobsson M, Rosenberg NA (2008) Demographic history of European populations of *Arabidopsis thaliana*. PLOS Genetics 4: e100075.

Fremstad E (1997) Vegetation types in Norway (Vegetasjonstyper i Norge). Norsk Institutt for Naturforskning Temahefte 12: 1-279.

Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J, Stefansson K (2005) An Icelandic example of the impact of population structure on association studies. Nature Genetics 37: 90-95.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. Nature 405: 907-913.

Hill WG, Weir BS (1994) Maximum-likelihood estimation of gene location by linkage disequilibrium. American Journal of Human Genetics 54: 705-714.

Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. BMC Evolutionary Biology 4: 18.

Jonsell B, Kustis K, Nordal I (1996) Genetic variation in *Arabis petraea*, a disjunct species in northern Europe. Ecography 18: 321-332.

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM et al. (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics 39: 1151-1155.

King G, Niehnuis J, Hussey C (1993) Genetic similarity among ecotypes of *Arabidopsis thaliana* estimated by analysis of restriction fragment length polymorphisms. Theoretical and Applied Genetics 86: 1028-1032.

Kuittinen H, Mattila A, Savolainen O (1997) Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. Heredity 79: 144-152.

Kärkkäinen K, Løe G, Agren J (2004) Population structure in *Arabidopsis lyrata*: evidence for divergent selection on trichome production. Evolution 58: 2831-6.

Lundemo S, Falahati-Anbaran M, Stenøien HK (2009) Seed banks cause elevated generation times and effective population sizes of *Arabidopsis thaliana* in northern Europe. Molecular Ecology 18: 2798-811.

Lynch M, Conery J (2003) The origins of genome complexity. Science 302: 1401-1404.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research 27: 209-220.

Miyashita NT, Kawabe A, Innan H (1999) DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. Genetics 152: 1723-1731.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. Journal of Molecular Evolution 19: 153-170.

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J et al. (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics 30: 190-193.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C et al. (2005) The pattern of polymorphism in *Arabidopsis thaliana*. PLOS Biology 3: e196.

Ostrowski MF, David J, Santoni S, Mckhann H, Reboud X et al. (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.

Pamilo P, Savolainen O (1999) Post-glacial colonization, drift, local selection and conservation value of populations: a northern perspective. Hereditas 130: 229-238.

Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R et al. (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecology and Management 156: 49-74.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK (2002) Genetic structure of human populations. Science 298: 2381-2385.

Schmid KJ, Torjek O, Meyer R, Schmuths H, Hoffmann MH et al. (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.

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

Stenøien HK, Fenster CB, Tonteri A, Savolainen O (2005) Genetic variability in natural populations of *Arabidopsis thaliana* in northern Europe. Molecular Ecology 14: 137-148.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437-460.

Todokoro S, Terauchi R, Kawano S (1995) Microsatellite polymorphisms in natural populations of *Arabidopsis thaliana* in Japan. Japanese Journal of Genetics 70: 543-554.

Vucetich JA, Waite TA (2003) Spatial patterns of demography and genetic processes across the species' range: null hypotheses for landscape conservation genetics. Conservation Genetics 4: 639-645.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Wright SI, Lauga B, Charlesworth D (2002) Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. Molecular Biology and Evolution 19: 1407-1420.

Zaykin, DV, Pudovkin AI, Weir BS (2008) Correlation-based inference for linkage disequilibrium with multiple alleles. Genetics 180: 533-545.

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C et al. (2007) An *Arabidopsis* example of association mapping in structured samples. PLOS Genetics 3: e4.

Figure legends

Figure 1. Relationship between average gene diversity over loci and latitude of origin of populations (r=0.58, p=0.001).

Figure 2. Relationship between pairwise genetic (shared allele distance – D_{SA}) and geographic distance (km) among 31 populations of *Arabidopsis thaliana* tested using Mantel test (r=0.69, p=0.001).

Figure 3. Genetic structure and geographic origin of Norwegian *Arabidopsis thaliana* populations. Each population is represented by one chart, showing the relative proportion of membership to the different clusters for K=5 (n=31) and K=3 (n=21) as deduced from STRUCTURE analyses.

Figure 4. Distance neighbor-joining tree for 31 Norwegian populations of *Arabidopsis thaliana*, Col-0 and L*er*. Colors indicate the 5 different geographical and genetic groups of populations: green – Southeastern and Southern Norway, red – West Coastal Norway; blue – Northern Norway and orange – Central Norway.

Figure 5. $N_e u$ within populations and within groups of populations of *Arabidopsis thaliana* from 6 geographical regions in Norway. SW – Southwestern Coast, NW – Northwestern Coast.

Figure 6. Distance neighbor-joining tree for 31 Norwegian populations of *Arabidopsis thaliana* and a worldwide sample of populations. Colors indicate geographical regions or country of origin of populations. For population names see supplementary Table S2. SNP genotype data of the worldwide population sample were kindly provided by Alexander Platt.

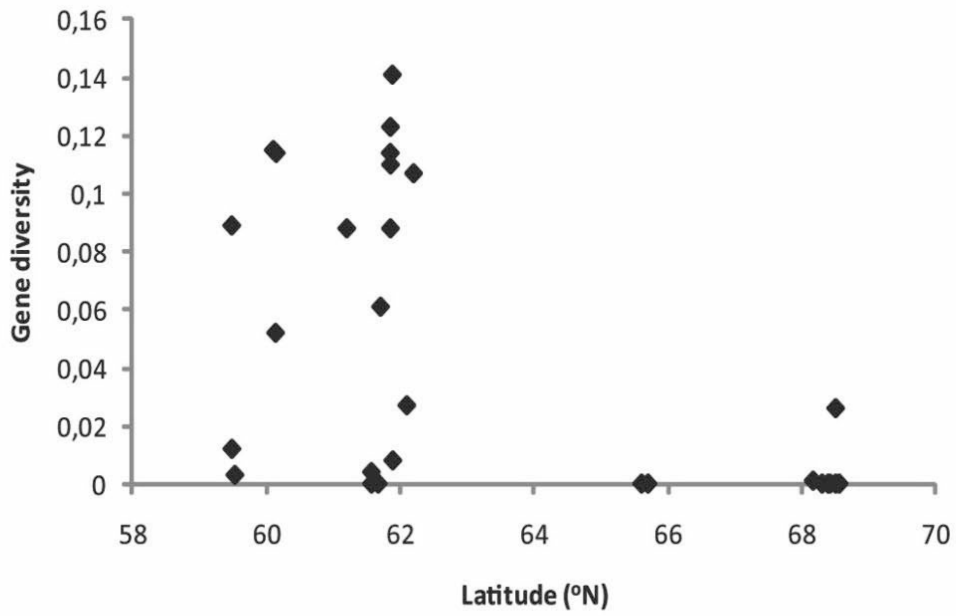


Figure 1

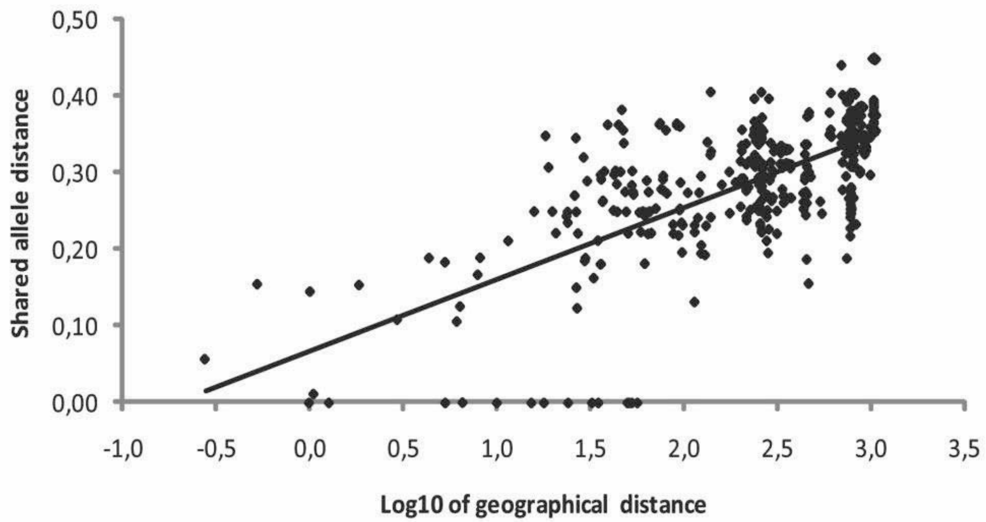


Figure 2

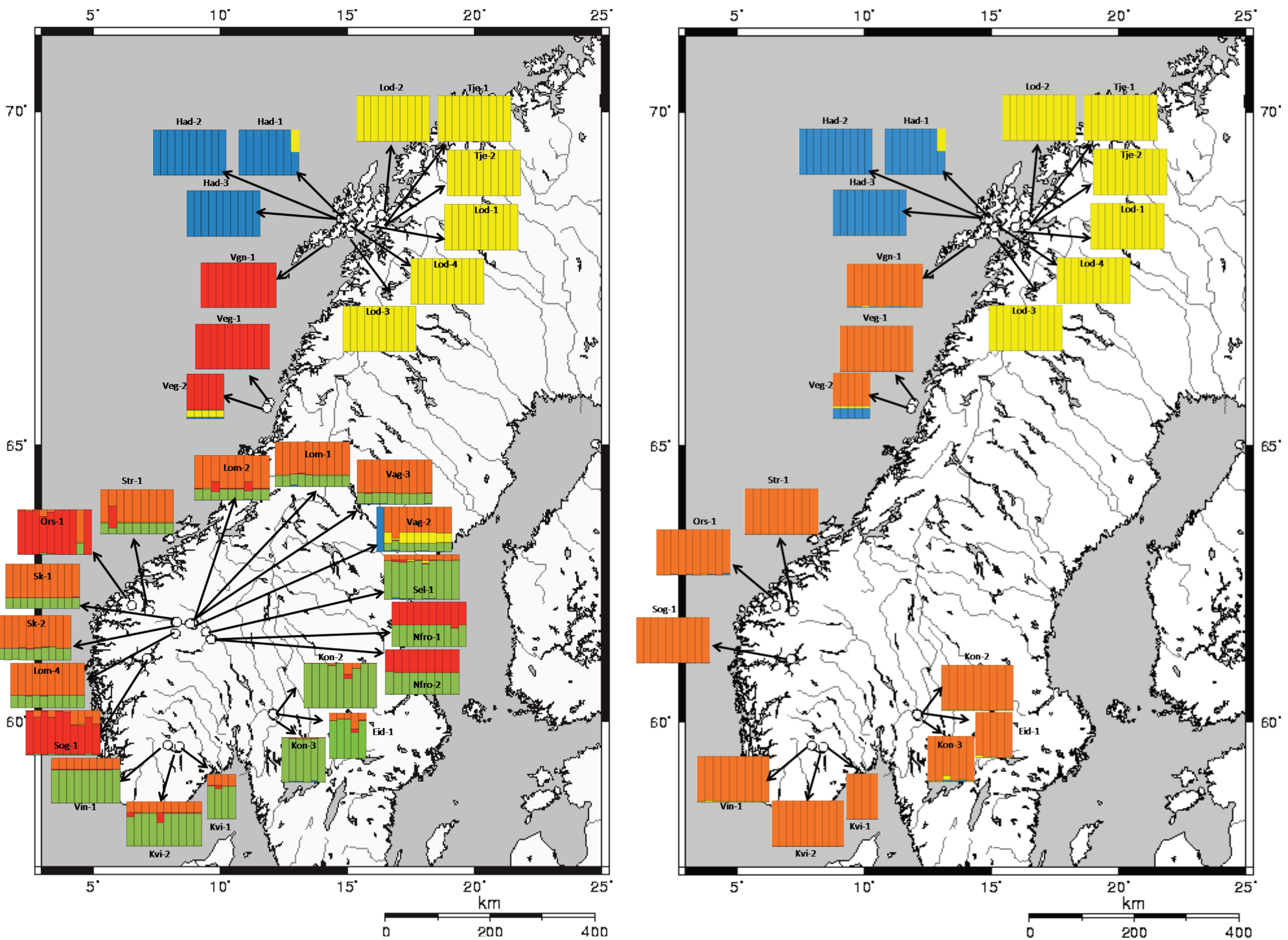


Figure 3

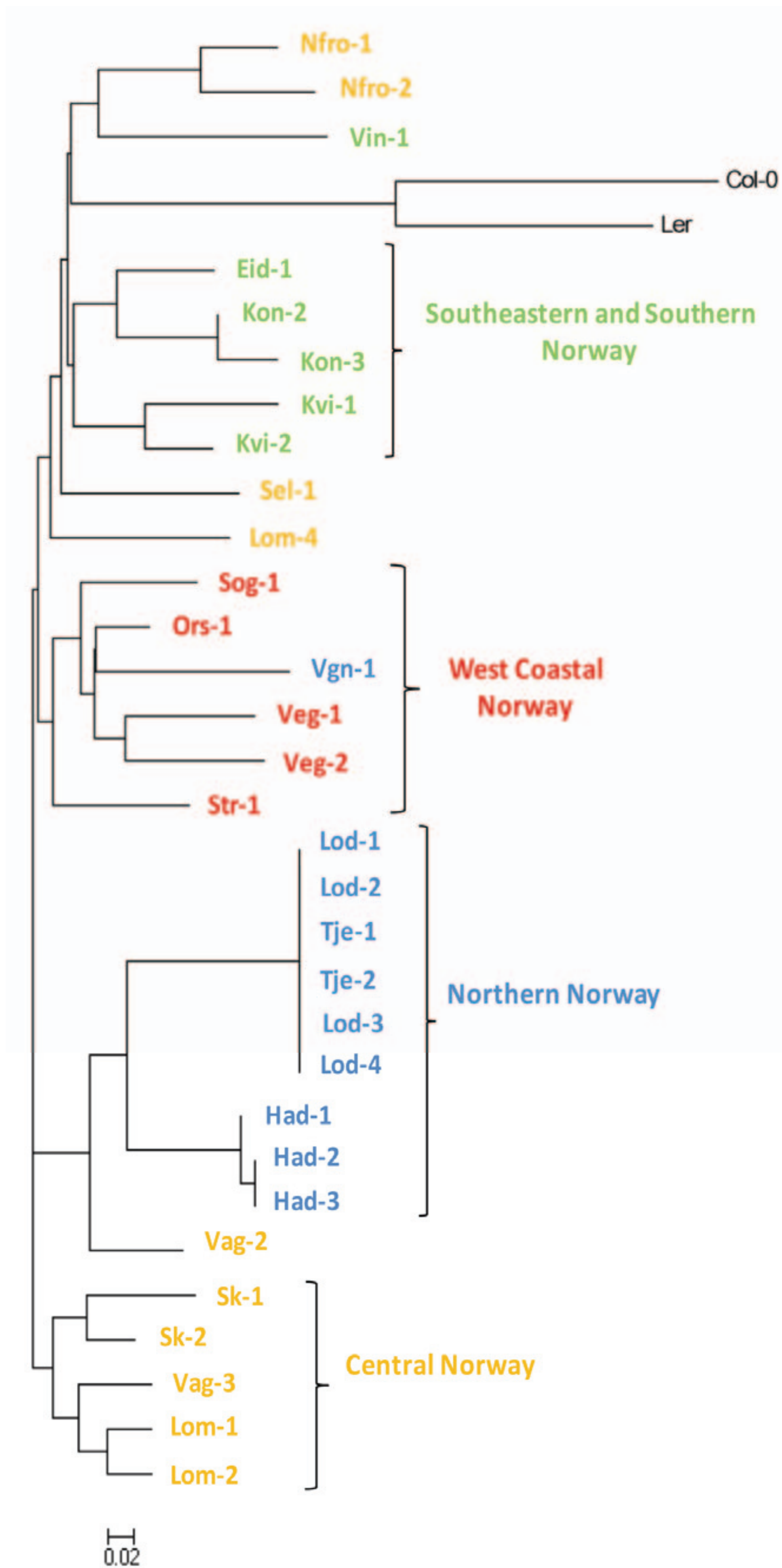


Figure 4

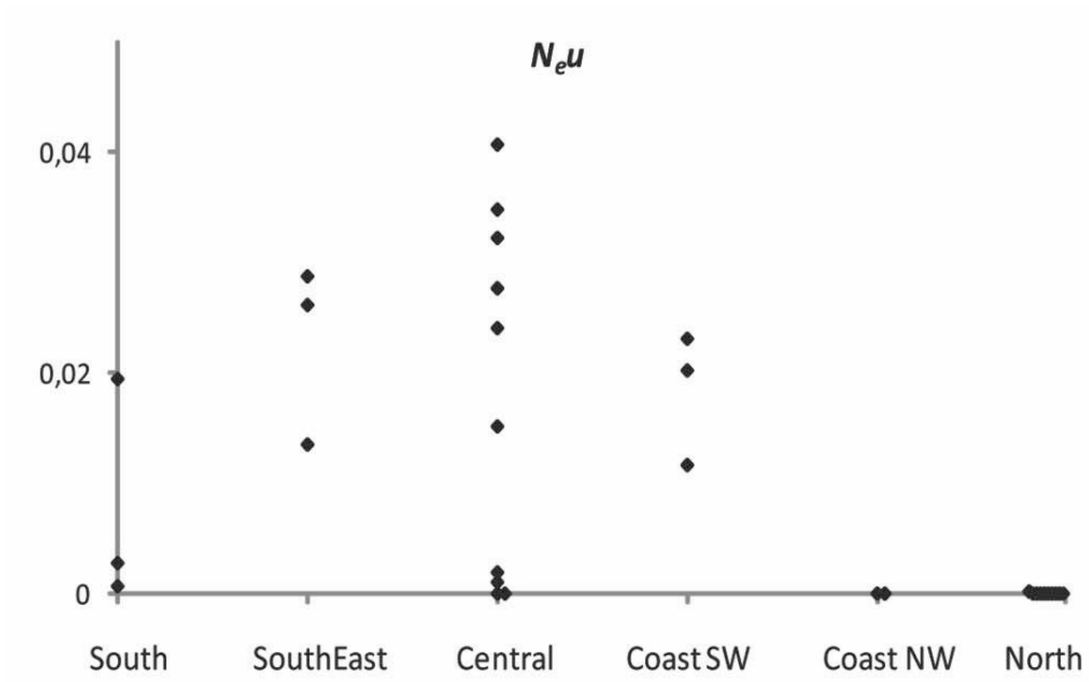


Figure 5

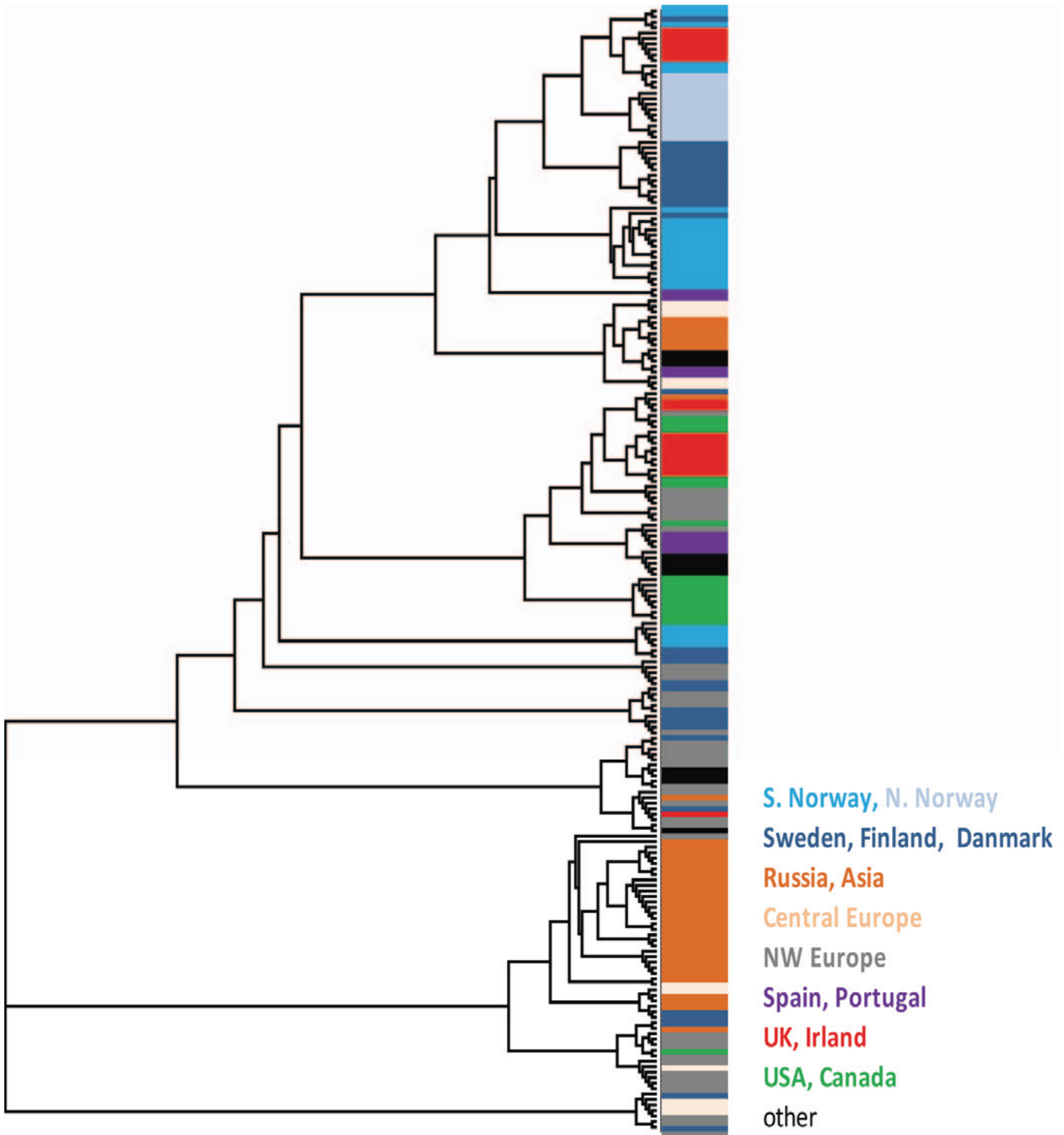


Figure 6

Table 1. Populations of *Arabidopsis thaliana* from Norway used in the study. Ind. = numbers of individuals genotyped per population.

Region	Population name	# of ind.	Latitude (°N)	Longitude (°E)	Altitude (m a.s.l)	Site of origin
Northern Norway	Vgn-1	10	68°10'11"	14°13'15"	3	Vågan
	Had-1	8	68°30'34"	14°43'57"	50	Hadsel
	Had-2	10	68°30'42"	14°45'44"	60	Hadsel
	Had-3	10	68°30'27"	14°53'32"	60	Hadsel
	Lod-1	10	68°23'51"	15°56'6"	100	Lødingen
	Lod-2	10	68°33'56"	16°19'7"	100	Lødingen
	Tje-1	9	68°25'29"	16°22'9"	110	Tjeldsund
	Tje-2	10	68°25'29"	16°22'9"	80	Tjeldsund
	Lod-3	10	68°18'17"	15°06'51"	82	Lødingen
	Lod-4	10	68°23'59"	15°08'23"	80	Lødingen
West Coast	Veg-1	10	65°42'24"	11°56'4"	3	Vega
	Veg-2	5	65°36'45"	11°50'18"	5	Vega
	Str-1	9	62°6'12"	7°12'11"	10	Stranda
	Ors-1	10	62°12'19"	6°32'1"	2	Ørsta
	Sog-1	10	61°12'33"	7°07'39"	8	Sogndal
Central Norway	Nfro-1	10	61°34'38"	9°39'42"	735	Nord-Fron
	Nfro-2	9	61°34'44"	9°38'28"	816	Nord-Fron
	Sel-1	10	61°42'58"	9°26'26"	687	Sel
	Vag-2	10	61°51'47"	8°58'43"	640	Våga
	Vag-3	9	61°51'28"	8°56'35"	400	Våga
	Lom-1	10	61°51'42"	8°49'30"	525	Lom
	Lom-2	10	61°51'49"	8°49'2"	700	Lom
	Sk-1	10	61°53'56"	8°15'14"	604	Skjåk
	Sk-2	10	61°53'18"	8°18'43"	598	Skjåk
	Lom-4	10	61°40'52"	8°13'51"	850	Lom
Southeastern Norway	Eid-1	5	60°6'44"	12°7'28"	200	Eidskog
	Kon-2	9	60°9'10'	12°5'6"	350	Kongsvinger
	Kon-3	6	60°8'53"	12°5'6"	300	Kongsvinger
Southern Norway	Kvi-1	4	59°29'38"	8°24'55"	703	Kviteseid
	Kvi-2	10	59°29'25"	8°24'5"	680	Kviteseid
	Vin-1	9	59°32'9"	7°56'58"	524	Vinje

Table 2. Summary of molecular diversity indices. H_O – mean observed heterozygosity; H_E – mean expected heterozygosity; PPL – proportion of polymorphic loci; Gene diversity – averaged over loci. P-values for two-tail *t*-tests comparing Northern and Southern populations.

Region	Population name	H_O	H_E	Gene diversity	PPL
Northern	Vgn-1	0.001	0.001	0.001	0.008
	Had-1	0.022	0.023	0.026	0.197
	Had-2	0.000	0.000	0.000	0.000
	Had-3	0.002	0.003	0.000	0.000
	Lod-1	0.000	0.000	0.000	0.000
	Lod-2	0.000	0.000	0.000	0.000
	Tje-1	0.000	0.000	0.000	0.000
	Tje-2	0.000	0.000	0.000	0.000
	Lod-3	0.000	0.000	0.000	0.000
	Lod-4	0.000	0.000	0.000	0.000
	Veg-1	0.000	0.000	0.000	0.000
	Veg-2	0.000	0.000	0.000	0.000
Southern	Sog-1	0.006	0.141	0.088	0.193
	Str-1	0.024	0.026	0.027	0.233
	Ors-1	0.020	0.161	0.107	0.302
	Nfro-1	0.000	0.004	0.004	0.020
	Nfro-2	0.004	0.004	0.000	0.000
	Sel-1	0.016	0.079	0.061	0.157
	Vag-2	0.000	0.087	0.088	0.388
	Vag-3	0.000	0.118	0.114	0.286
	Lom-1	0.020	0.126	0.123	0.348
	Lom-2	0.010	0.124	0.110	0.273
	Sk-1	0.003	0.010	0.008	0.015
	Sk-2	0.012	0.152	0.141	0.321
	Lom-4	0.001	0.001	0.000	0.000
	Eid-1	0.051	0.113	0.115	0.382
	Kon-2	0.051	0.121	0.114	0.421
	Kon-3	0.009	0.059	0.052	0.113
	Kvi-1	0.011	0.015	0.012	0.047
	Kvi-2	0.021	0.118	0.089	0.303
	Vin-1	0.003	0.003	0.003	0.007
	P-values	0.008	<0.0001	<0.0001	<0.0001

Table 3. Linkage Disequilibrium (LD) level and decay for each chromosome and genomewide across all populations and within population groups of *Arabidopsis thaliana* from Norway.

Populations	**chromosome**	**LD decay (p-value)**	**LD level (%)**	**No of populations**
Total	1	**0.000**	0	31
	2	0.474	0.4	
	3	0.163	0.6	
	4	0.977	0	
	5	0.067	0.2	
	genomewide	**0.000**	0.1	
Northern Norway	1	0.718	10.1	10
	2	0.162	7.4	
	3	0.795	9.4	
	4	0.540	14.4	
	5	0.639	11.3	
	genomewide	0.399	**10.4**	
West Coast	1	0.785	5.9	5
	2	0.529	2.6	
	3	0.286	2.3	
	4	0.016	2.6	
	5	0.861	1.2	
	genomewide	0.707	3.8	
Central Norway	1	0.000	1.4	10
	2	0.856	2.2	
	3	0.029	1.8	
	4	0.216	2.1	
	5	0.768	0.8	
	genomewide	**0.000**	1.4	
Southeastern Norway	1	0.495	11.4	3
	2	0.571	10.4	
	3	0.152	5.9	
	4	0.879	8.4	
	5	0.632	4.4	
	genomewide	0.177	**8.9**	
Southern Norway	1	0.333	13.3	3
	2	0.068	4.8	
	3	0.641	12.9	
	4	0.688	1.6	
	5	0.907	5.4	
	genomewide	0.834	**9.4**	

Supplementary materials

Figure legends

Figure S1. Relationship between pairwise genetic (shared allele distance – D_{SA}) and geographic distance (km) among 19 Southern populations (left panel) and 12 Northern populations (right panel) of *Arabidopsis thaliana* tested with Mantel test (r=0.60, 0.66; p=0.001).

Figure S2. Level of Linkage Disequilibrium (LD, %) plotted against number of populations in different groups. The level of LD across all populations and within each population group was calculated as a proportion of pairwise loci comparisons with r^2 higher than 0.7 for all chromosomes. SEN – Southeastern Norway, SN – Southern Norway, NN – Northern Norway, CN – Central Norway, WeC – West Coast, T – total populations.

Figure S3. Detection of the number of groups K according to Evanno et al. (2005). A) Mean delta K for all populations (n=31); B) Similarity coefficient for different runs for each K (n=31); C) Mean delta K after omitting populations from Central Norway (n=21); D) Similarity coefficient for different runs for each K (n=21). Asterisk represents significant similarity coefficient for K=3. An admixture model with allele frequencies uncorrelated was run in both cases.

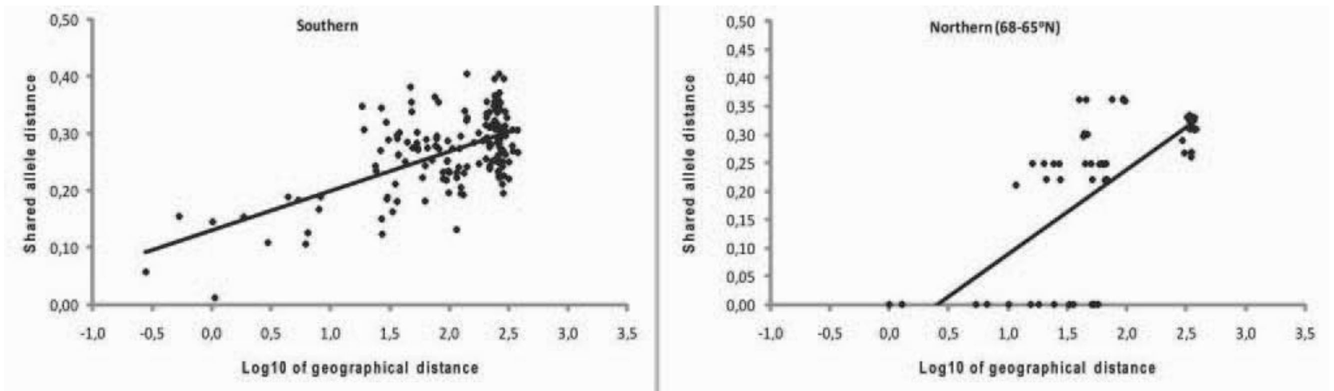


Figure S1

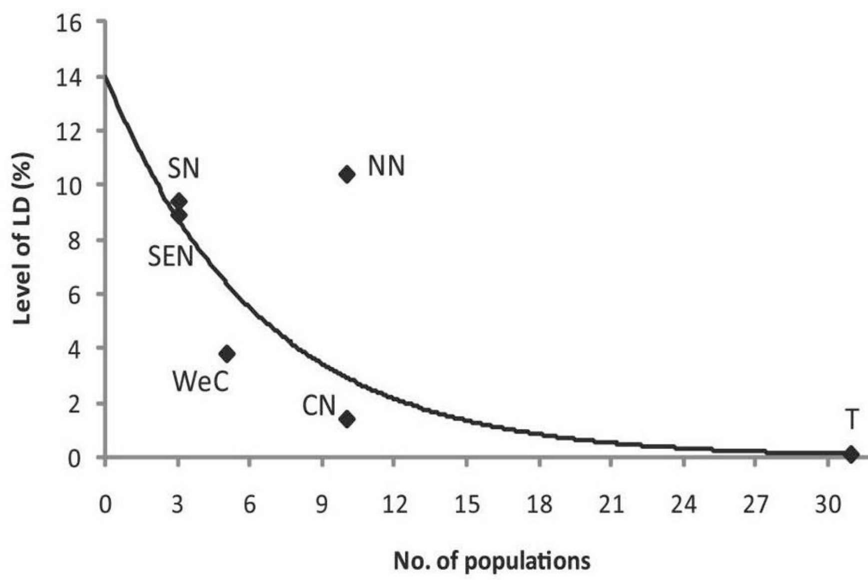


Figure S2

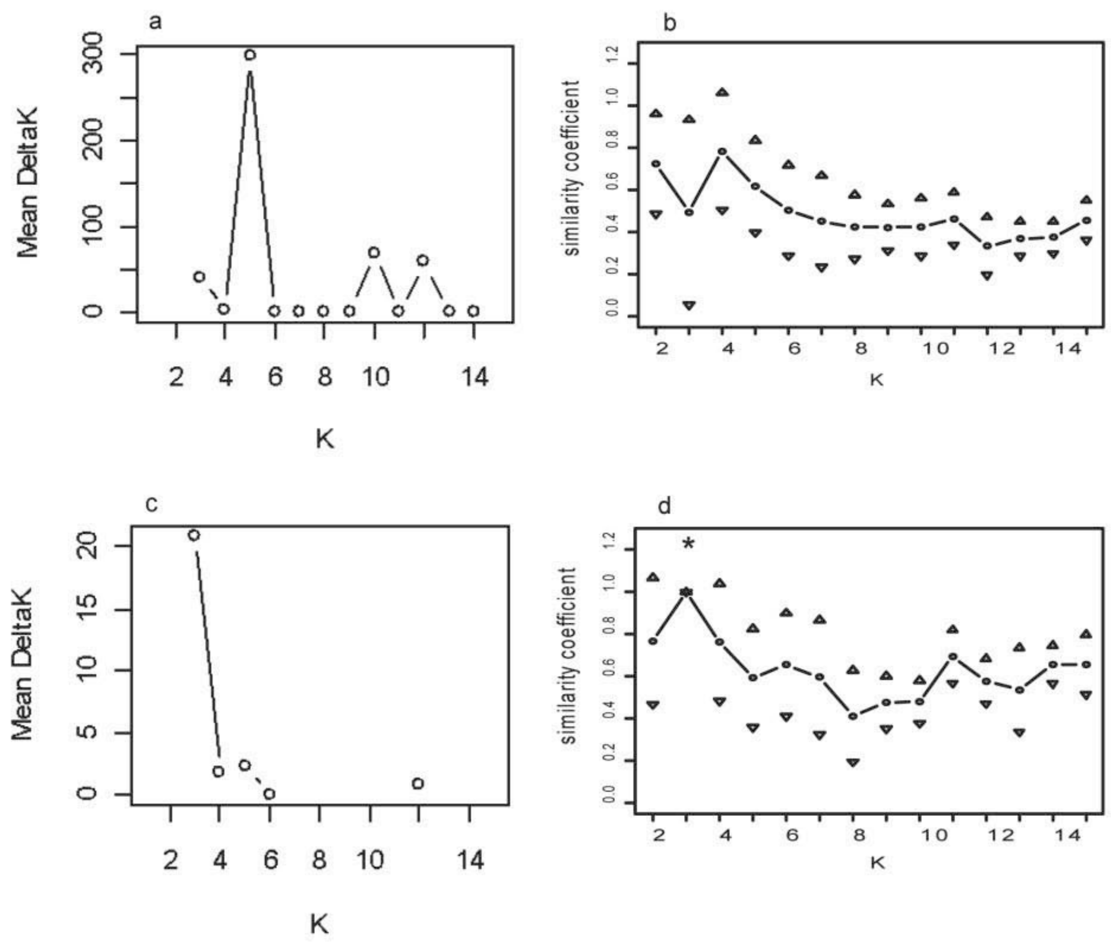


Figure S3

Table S1. Pairwise F_{ST} between 31 Norwegian populations of *Arabidopsis thaliana*. Only significant estimates are listed.

	Nfro-1	Sk-1	Sk-2	Lom-4	Vgn-1	Had-1	Had-2	Had-3	Lod-1	Lod-2	Nfro-2	Tje-1	Tje-2	Lod-3	Lod-4	Sog-1	Str-1	Ors-1	Veg-1	Veg-2	Eid-1	Kon-2	Kon-3	Kvi-1	Kvi-2	Vin-1	Sel-1	Vag-2	Vag-3	Lom-1
Nfro-1																														
Sk-1	0.98																													
Sk-2	0.78	0.58																												
Lom-4	0.99	0.98	0.75																											
Vgn-1	0.99	0.99	0.80	1.00																										
Had-1	0.96	0.96	0.72	0.97	0.97																									
Had-2	1.00	0.99	0.77	1.00	1.00																									
Had-3	1.00	0.99	0.77	1.00	1.00																									
Lod-1	1.00	0.99	0.81	1.00	1.00	0.96	1.00	1.00																						
Lod-2	1.00	0.99	0.81	1.00	1.00	0.96	1.00	1.00																						
Nfro-2	0.98	0.99	0.78	1.00	1.00	0.97	1.00	1.00	1.00	1.00																				
Tje-1	1.00	0.99	0.80	1.00	1.00	0.95	1.00	1.00			1.00																			
Tje-2	1.00	0.99	0.81	1.00	1.00	0.96	1.00	1.00			1.00																			
Lod-3	1.00	0.99	0.81	1.00	1.00	0.96	1.00	1.00			1.00																			
Lod-4	1.00	0.99	0.81	1.00	1.00	0.96	1.00	1.00			1.00																			
Sog-1	0.77	0.80	0.60	0.80	0.76	0.76	0.81	0.81	0.85	0.85	0.82	0.84	0.85	0.85	0.85															
Str-1	0.96	0.93	0.66	0.95	0.96	0.93	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.72														
Ors-1	0.75	0.75	0.53	0.77	0.70	0.71	0.76	0.76	0.80	0.79	0.78	0.79	0.79	0.80	0.80	0.46	0.65													
Veg-1	1.00	0.99	0.77	1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.78	0.96	0.67												
Veg-2	0.99	0.98	0.73	1.00	1.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.73	0.94	0.63	1.00											
Eid-1	0.88	0.85	0.56	0.89	0.91	0.85	0.90	0.90	0.90	0.90	0.87	0.90	0.90	0.90	0.90	0.68	0.83	0.63	0.90	0.86										
Kon-2	0.83	0.80	0.59	0.84	0.85	0.80	0.85	0.85	0.85	0.85	0.84	0.84	0.85	0.85	0.85	0.69	0.78	0.64	0.86	0.81	0.58									
Kon-3	0.94	0.92	0.66	0.95	0.95	0.90	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.75	0.89	0.71	0.95	0.93	0.69									
Kvi-1	0.98	0.97	0.70	0.99	0.99	0.95	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.74	0.93	0.67	0.99	0.98	0.80	0.74	0.90							
Kvi-2	0.85	0.82	0.63	0.81	0.83	0.82	0.86	0.86	0.86	0.86	0.85	0.85	0.86	0.86	0.86	0.66	0.78	0.60	0.83	0.81	0.67	0.66	0.77	0.63						
Vin-1	1.00	0.99	0.80	1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.79	0.97	0.80	1.00	1.00	0.89	0.85	0.95	0.99	0.82					
Sel-1	0.89	0.87	0.67	0.88	0.90	0.85	0.89	0.89	0.90	0.89	0.89	0.89	0.89	0.90	0.89	0.75	0.85	0.69	0.89	0.86	0.73	0.73	0.82	0.83	0.74	0.90				
Vag-2	0.88	0.86	0.61	0.86	0.87	0.78	0.83	0.83	0.85	0.85	0.89	0.84	0.85	0.85	0.85	0.69	0.80	0.64	0.87	0.82	0.72	0.74	0.81	0.83	0.73	0.87	0.78			
Vag-3	0.82	0.74	0.53	0.81	0.84	0.75	0.80	0.80	0.85	0.85	0.84	0.84	0.85	0.85	0.85	0.63	0.74	0.56	0.81	0.76	0.64	0.65	0.74	0.74	0.67	0.84	0.73	0.58		
Lom-1	0.80	0.73	0.45	0.79	0.83	0.74	0.79	0.79	0.84	0.84	0.80	0.83	0.84	0.84	0.84	0.61	0.72	0.57	0.81	0.74	0.61	0.66	0.73	0.74	0.67	0.83	0.73	0.59	0.44	
Lom-2	0.82	0.74	0.50	0.81	0.83	0.77	0.82	0.82	0.86	0.86	0.83	0.86	0.86	0.86	0.86	0.62	0.75	0.56	0.81	0.76	0.62	0.64	0.71	0.74	0.68	0.85	0.73	0.63	0.50	0.31

Table S2. Names and country of origin of populations used to construct a dendrogram of the worldwide sample of *Arabidopsis thaliana* populations. The populations are listed in the same order (top to bottom) as in the dendrogram.

population name	country	population name	country
Nfro-1	NOR	Kastel-3	UKR
Nfro-2	NOR	Kastel-4	UKR
Lovel-1	DEN	Bak-4	GEO
Vin-1	NOR	Bak-7	GEO
Bur-0	IRL	Bak-5	GEO
Cal-0	UK	Lerik2-3	AZE
Bur-0	IRL	Lerik2-5	AZE
Bur-0	IRL	Lerik2-4	AZE
Bur-0	IRL	Cvi-1	CPV
Mc-0	UK	Cvi-0	CPV
Sog-1	NOR	Cvi-0	CPV
Ors-1	NOR	San Martin-5	ESP
Vgn-1	NOR	Sah-0	ESP
Veg-1	NOR	Por-11	SRB
Veg-2	NOR	Por-1	SRB
Lod-2	NOR	Var2-1	SWE
Lod-1	NOR	UKNW06-018	UK
Tje-1	NOR	Ksk-1	UK
Tje-2	NOR	UKID10	UK
Lod-3	NOR	CAM-57	FRA
Lod-4	NOR	Van-0	CAN
Had-1	NOR	Van-0	CAN
Had-2	NOR	Rmx-A180	USA
Had-3	NOR	Alst-1	UK
TBL 01	SWE	Vind-1	UK
Bil-7	SWE	UKID24	UK
TBL 02	SWE	Laud-1	UK
TBL 03	SWE	UKID93	UK
Bil-5	SWE	Ottb-1	UK
Lov-1	SWE	Edinburgh-1	UK
Eden-1	SWE	Jedb-1	UK
Eden-2	SWE	Kin-0	USA
Sku-28	SWE	Kin-0	USA
Fab-2	SWE	Heem-6	BEL
Fab-4	SWE	Heem-2	BEL
Herr-13	SWE	Heem-8	BEL
Str-1	NOR	Ge-0	SUI
Hen-16-268	SWE	Zu-0	SUI
Vag-2	NOR	Wei-0	SUI
Vag-3	NOR	KNO-1	USA
Lom-1	NOR	VOU-5	FRA
Lom-2	NOR	C24	POR
Sk-1	NOR	C24	POR
Sk-2	NOR	C24	POR
Lom-4	NOR	Co-1	POR
Kvi-1	NOR	Bolsena-6-121	ITA
Kvi-2	NOR	Bolsena-21-134	ITA
Eid-1	NOR	Bolsena-1-111	ITA
Kon-2	NOR	Bolsena-3-115	ITA
Kon-3	NOR	RRS-7	USA
Sel-1	NOR	CSHL-2	USA
Bla-6	ESP	HS-7	USA
Ll-2	ESP	KEN	USA
Kastel-8	UKR	Kno-10	USA

population name	country
Pna-10	USA
Mv-0	USA
Pna-17	USA
Pna-17	USA
Oy-0	NOR
Oy-1	NOR
Oy-0	NOR
Oy-0	NOR
ESA--25	SWE
Aledal-11-63	SWE
Gardby-14-190	SWE
Nok-3	NED
Nok-1	NED
Nok-3	NED
OMO2-1	SWE
Ale-Stenar-41-1	SWE
ANH-8	GER
ANH-13	GER
ANH-22	GER
Spr1-2	SWE
Ull2-3	SWE
Bil-3	SWE
T910	SWE
Col	GER
Rad-17-319	SWE
Bs-5	SUI
Lp2-2	CZE
Bor-1	CZE
Je54	CZE
Dra-1	CZE
Tu-0	ITA
Tu-0	ITA
Tu-0	ITA
Rak-2	CZE
Zdr-1	CZE
Sapporo-0	JPN
Gr3	AUT
Ber	DEN
UKSE06-476	UK
Gr-1	AUT
Gr-1	AUT
Zab-1	CRO
Di-1	FRA
Rak-1	RUS
Leb-1	RUS
Sev	RUS
39753	RUS
Bij	RUS
Neo-3	TJK
Kly-1	RUS
9481A	KAZ
Kz-9	KAZ
Kz-9	KAZ
Kz-10	KAZ

population name	country
9481D	KAZ
Kz-2	KAZ
Kz-13	KAZ
Pan	RUS
Cha-1	RUS
Yeg-2	ARM
Yeg-4	ARM
Yeg-3	ARM
Rsch-0	RUS
Kas-2	IND
Kas-2	IND
Kas-1	IND
Kas-2	IND
Ms-0	RUS
Per-2	RUS
Wa-1	POL
Wa-1	POL
Dog-7	TUR
Dog-6	TUR
Dog-5	TUR
Tamm-27	FIN
Tamm-2	FIN
Tamm-27	FIN
Kyoto	JPN
Pu2-23	CZE
Bor-4	CZE
Bor-4	CZE
ENF	USA
Dra-2	CZE
Zdr-6	CZE
Ler	POL
Dra-0	CZE
Pu2-7	CZE
Pu2-7	CZE
Pu2-8	CZE
Ting-1	SWE
Zloc-5	BUL
Zloc-10	BUL
Zloc-2	BUL
Sav-0	CZE
Jm-0	CZE
Brasarp-11-135	SWE
Jl-1	CZE