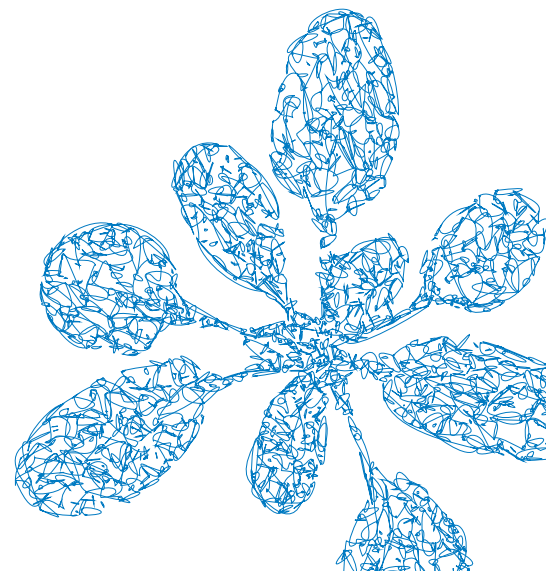Temperature-Dependent Flowering of

Arabidopsis thaliana is Modulated by

Natural Variation of *FLOWERING LOCUS M*

Ulrich Lutz

2016

Fakultät Wissenschaftszentrum Weihenstephan
für Ernährung, Landnutzung und Umwelt
Lehrstuhl für Systembiologie der Pflanzen

Temperature-dependent flowering of *Arabidopsis thaliana* is modulated

by natural variation of *FLOWERING LOCUS M*

Ulrich Lutz

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Kay Schneitz
Prüfer der Dissertation: 1. Prof. Dr. Claus Schwechheimer
2. Prof. Dr. Detlef Weigel
3. Prof. Dr. Martin Parniske

Die Dissertation wurde am 15.11.2016 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 08.02.2017 angenommen.

Tᴀʙʟᴇ ᴏꜰ Cᴏɴᴛᴇɴᴛs

ABSTRACT

Plants integrate seasonal cues such as temperature and day length to optimally adjust their flowering time to the environment. Compared to the control of flowering before and after winter by the vernalization pathway, mechanisms that delay or promote flowering during a transient cool or warm period, especially during spring, are less well understood. However, due to global warming, understanding this ambient temperature pathway has gained increasing importance. In *Arabidopsis thaliana* (*A. thaliana*), *FLOWERING LOCUS M* (*FLM*) is a critical flowering regulator of the ambient temperature pathway. *FLM* is alternatively spliced in a temperature-dependent manner and the two predominant splice variants, *FLM-ß* and *FLM-δ*, can repress and activate the induction of flowering in the genetic background of the *A. thaliana* reference accession Columbia-0. The relevance of this regulatory mechanism for the environmental adaptation across the entire range of the species is, however, unknown. Further, whether natural genetic variation of the *FLM* locus plays a role in flowering time regulation across *A. thaliana* accessions remained to be determined. The present thesis describes the identification of a large structural variation in the first intron of *FLM* as causative for low *FLM* transcript levels and accelerated flowering in *A. thaliana* accessions from Scotland, Sweden, and Germany. The effect of this expression-modifying allele is especially prominent at cool (15°C) temperatures. Further, evidence for a potential adaptive role of this structural variation is presented, which is specifically linked to changes in the abundance of *FLM-ß*.

In the following project, phylogenetic footprinting was applied and regions in the promoter and intron 1 were identified to be essential for *FLM* expression and conserved among *Brassicaceae*. In a parallel approach, sequence information of more than 800 *A. thaliana* accessions was combined with expression analysis and association analysis and a small polymorphic region in the *FLM* promoter and a nucleotide triplet in intron six that control differential *FLM* abundance in a haplotype-specific manner was precisely identified. *FLM-ß* was found to be the major physiologically relevant *FLM* splice variant and may therefore be an important modulator to fine-tune flowering among a broad vegetative period.

Climate change leads to increased ambient temperature, which in turn shifts flowering time of crops and negative effects on yield are expected. Therefore, there is an increasing need to buffer these negative effects. The results presented in this thesis may help in the future to readjust flowering time and to develop climate change-resilient crop cultivars.

Zusammenfassung

Pflanzen regulieren den Blühzeitpunkt, indem sie die saisonalen Unterschiede vieler Einflussfaktoren wie der Temperatur und der Tageslänge wahrnehmen. Im Vergleich zur Kontrolle des Blühzeitpunkts vor und nach dem Winter durch den Signalweg der Vernalisation, ist weniger über die Mechanismen bekannt, die den Blühzeitpunkt während einer kurzen Phase mit kühlen oder warmen Temperaturen, vor allem im Frühling, kontrollieren. Aufgrund der globalen Erderwärmung hat dieses Forschungsfeld jedoch an Beachtung gewonnen. Der Transkriptionsfaktor *FLOWERING LOCUS M* (*FLM*) von *Arabidopsis thaliana* (*A. thaliana*) ist eine wichtige Komponente zur Wahrnehmung solcher leichten Unterschiede der Umgebungstemperatur. *FLM* wird über temperaturabhängiges, alternatives Splicing reguliert. Im genetischen Hintergrund der Referenzakzession Columbia-0 werden hauptsächlich die zwei Isoformen *FLM-ß* und *FLM-δ* gebildet, die die Blühinduktion reprimieren beziehungsweise aktivieren. Es ist allerdings nicht bekannt, ob dieser regulatorische Mechanismus für die ökologische Anpassung der gesamten Spezies relevant ist. Außerdem wurde bisher noch nicht untersucht, ob die natürliche genetische Variation des Genlokus *FLM* an der Regulation des Blühzeitpunkts von *A. thaliana*-Akzessionen beteiligt ist. Die vorliegende Arbeit beschreibt die Identifizierung einer großen, strukturellen Variation im ersten Intron von *FLM,* die zu niedrigen Transkriptmengen von *FLM* und zu einem frühen Blühzeitpunkt von *A. thaliana*-Akzessionen aus Schottland, Schweden und Deutschland führt. Der Effekt dieses Alleles, das die Expression von *FLM* modifiziert, ist bei kühleren Temperaturen (15°C) besonders deutlich ausgeprägt. Darüber hinaus werden Hinweise beschrieben, die auf eine potentielle Rolle dieser strukturellen Variation bei der evolutionären Anpassung des Blühzeitpunkts hindeuten. Dieser Einfluss ist durch den spezifischen Effekt der Insertion auf die Abundanz der Transkriptisoform *FLM-ß* zu erklären.

Mithilfe einer phylogenetische Vergleichsanalyse wurden im Folgeprojekt zwei Regionen im Promoter und im ersten Intron von *FLM* identifiziert, die essentiell für dessen Expression sind und innerhalb der Familie der *Brassicaceae* konserviert vorliegen. In einem parallel durchgeführten Ansatz wurden genomische Sequenzeninformationen von mehr als 800 Akzessionen von *A. thaliana* mit einer Expressions- und einer Assoziationsanalyse kombiniert. Dadurch konnte eine kleine, polymorphe Region im Promoter und ein Basentriplett im sechsten Intron von *FLM* identifiziert werden, die die Abundanz von *FLM* in Abhängigkeit ihres Haplotyps regulieren. Weitere Untersuchungen ergaben, dass die Transkriptisoform *FLM-ß* von ausschließlicher physiologischer Bedeutung ist und es wurde geschlussfolgert, dass diese Isoform eine wichtige Komponente zur präzisen Regulierung des Blühzeitpunkts über einen langen vegetativen Zeitraum ist.

Die Erhöhung der Umgebungstemperatur führt zu Änderungen des Blühzeitpunkts von Kulturpflanzen, was negative Auswirkungen auf Ernteerträge erwarten lässt. Die Ergebnisse dieser Thesis könnten in Zukunft helfen, den Blühzeitpunkt anzugleichen und Kulturpflanzen zu entwickeln, die an die Folgen des Klimawandels angepasst sind.

List of Abbreviations

3'-RACE	rapid amplification of 3' cDNA-ends
AGL	AGAMOUS LIKE
AS	alternative splicing
bHLH	basic helix-loop-helix
bp	base pair
CArG	MADS-box transcription factor consensus binding motif 5'-CC[A/T]$_6$GG-3'
cDNA	copy DNA
ChIP	Chromatin Immunoprecipitation
Chr	chromosome
CO	CONSTANS
CRY	CRYPTOCHROME
DHS	DNaseI hypersensitive site
DTB	days to bolting
E1I1p	exon1 intron1 partial
FKF1	FLAVIN-BINDING KELCH REPEAT F-BOX PROTEIN
FLC	FLOWERING LOCUS C
FLM	FLOWERING LOCUS M
FRI	FRIGIDA
FT	FLOWERING LOCUS T
FT15	flowering time at 15°C and long day photoperiod
GI	GIGANTEA
GRP7	GLYCINE-RICH RNA BINDING PROTEIN 7
H3K27me3	trimethylated Lysin 27 of histone 3
H7kb	haplotype group 7 kb *FLM* locus
HI1	haplotype group intron 1 of *FLM*
HI2-6	haplotype group introns 2 to 6 of *FLM*
Hyg	hygromycin
Indel	insertion/deletion
Kan	kanamycin
kb	kilobase
LD	linkage disequilibrium
LINE	long interspersed nuclear element
LOESS	locally weighted scatterplot smoothing
MAF	MADS AFFECTING FLOWERING
maf	minor allele frequency
miR172	micro RNA172
NGS	next generation sequencing
NMD	nonsense mediated decay
ORF	open reading frame
PHD-PRC	PHD protein domain-polycomb repressive complex 2
PHY	PHYTOCHROME
PIF4	PHYTOCHROME INTERACTING FACTOR 4
PoolSeq	pooled sequencing
Pre-mRNA	premature messenger RNA
PTC	premature termination codon
qPCR	quantitative polymerase chain reaction
QTL	quantitative trait locus
RLN	rosette leaf number
RNA-seq	RNA sequencing
RT-PCR	reverse transcriptase polymerase chain reaction
SNP	single nucleotide polymorphism
snRNP/hnRNP	small/heterogeneous nuclear ribonucleoprotein partical
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
Spec	spectinomycin
sqPCR	semi-quantitative polymerase chain reaction
SVP	SHORT VEGETATIVE PHASE
TAIR	The Arabidopsis Information Resource
T-DNA	transfer DNA
TE	transposable element
TF	transcription factor
UTR	untranslated region
VAL1	VIVAPAROUS1/ABI3-LIKE1

List of Figures

Lɪsᴛ ᴏғ Tᴀʙʟᴇs

1.   Introduction

The correct biological timing of flowering is one of the most important factors that control the reproductive success of plants. In order to produce viable seeds and to ensure biological fitness, plants enter the regenerative phase when optimal environmental conditions such as light duration and ambient temperature are favorable. Understanding flowering time regulation of crops is crucial, since the precisely timed transition from the vegetative to the generative phase is the basis for crop production whenever seeds are harvested. Herein, the knowledge of flowering time genes of the model species *Arabidopsis thaliana* (*A. thaliana*) is an important resource, which helps to identify and characterize orthologous flowering time regulators of cultivated crop species (Trevaskis et al., 2007; Distelfeld et al., 2009; Blümel et al., 2015). Since the pioneering study of multiple flowering time mutants of *A. thaliana* in 1998 by Koornneef et al., extensive progress has been made and more than 300 genes with a role in flowering time control have been identified so far (Koornneef et al., 1998; Bouché et al., 2016). Despite the usefulness of artificially generated knock-out mutants, species-wide genetic variation of *A. thaliana* provides a vast resource of naturally occurring expression-modifying alleles. These alleles contribute to the molecular and evolutionary basis of natural variation of *A. thaliana* flowering whose elucidation became a major task of plant biologists.

1.1.   Natural variation of *Arabidopsis thaliana*

The native natural habitat of the model plant *A. thaliana* is within Europe and Central Asia but also has a young demographic history in North America. *A. thaliana* has adapted to a wide range of different climatic environments within the Northern hemisphere (Sharbel et al., 2000; Hoffmann, 2002; Schmid et al., 2006; Weigel, 2012). *A. thaliana* frequently grows in dry and disturbed areas and typically at the margins of agricultural areas and on farmed grassland. Since it is mainly self-fertilizing, genotypes of wild *A. thaliana* strains (accessions) are fixed and can be repetitively used to perform several phenotyping experiments. Since it can be outcrossed, crossings of *A. thaliana* accessions can be generated for genetic mapping analysis and natural variation can be used as a resource for genetic variation. Natural genetic variation can be investigated to identify the genetic basis of traits with complex structure. For functional studies, identifying and utilizing natural variation is advantageous, since artificial knock-out mutants often are lethal or phenotypic effects are masked due to genetic redundancy. Furthermore, natural genetic variation can be exploited from the evolutionary point of view, to investigate how ecologically relevant traits evolve in natural environments.

Studies, which exploit natural variation to identify the molecular basis of the complex trait of flowering time by linking genotypes to phenotypes have identified many new regulators as well as new natural alleles of already known regulators of flowering time. Herein, recombinant inbred line collections were used frequently (Koornneef et al., 2004; Alonso-Blanco et al., 2009; Weigel, 2012). These collections are sets of lines that all derive from a single cross of two parental *A. thaliana* accessions and high homogeneity is obtained by recurrent inbreeding for several generations (> F7). Hence, these collections can be repetitively used to perform phenotyping experiments while genotyping is only necessary once. Despite the fact, that such genetic material is of great value for

the community to identify quantitative trait loci (QTL) underlying specific traits, it is also very time consuming and expensive to generate such resources. Quickly generated F2 populations as well as introgression lines can be used to identify one or a few loci with a major effect on a trait that has a simple genetic architecture (Balasubramanian et al., 2006; Mendez-Vigo et al., 2013).

With the fast progress in development and application of next generation sequencing (NGS) technologies, the cost of sequencing for one *A. thaliana* genome has dropped in the last years (Goodwin et al., 2016). Through the generation of genome wide polymorphism data from large numbers of accessions new possibilities arose to determine the functional molecular basis of phenotypic variation. More precise information about heterogeneous alleles that are involved in adaptation can be gained. Further, ecological and evolutionary studies can be performed to learn more about *A. thaliana* adaptation, e.g. to different climates. With the aim to generate an extensive resource of genomic information, the 1001 *A. thaliana* Genomes Project has recently released a well curated set of 1,135 genome sequences (The 1001 Genomes Consortium, 2016). However, before that release, several reports described datasets of smaller subsets of accessions (e.g., Ossowski et al., 2008; Cao et al., 2011; Long et al., 2013). A high density variant map of one SNP per 10 bp was generated. On average 84% of the TAIR10 Col-0 reference genome was covered, showing that, in some instances , a lot of large structural variation is present in the genomes. Approaches that apply a combinatorial strategy of *de novo* assembly, guided by the Col-0 reference sequence, could successfully add information about such large structural variation (Schneeberger et al., 2011; Zapata et al., 2016). Previous studies that used sets of SNP markers and smaller subsets of *A. thaliana* accessions found that high genetic divergence between two populations correlated with geographic distance. Such isolation by distance effects are due to long-term geographical isolation while gene flow between populations is very low. An image emerged that accessions from Central and Northern Europe represent admixed populations with highly reshuffled genomes likely resulting from various recombination events. Such genetic admixture reflects *A. thaliana* recolonization of Central and Northern Europe from glacial refugia in Asia and the Iberian Peninsula (Sharbel et al., 2000; Nordborg et al., 2005; Schmid et al., 2006).

1.2.   The MADS-box family of transcription factors in plants

MADS-box transcription factors form a large gene family in plants and are important regulators of plant development and especially flowering time (Hemming and Trevaskis, 2011; Ng and Yanofsky, 2001). They are named after the genes *MINICHROMOSOME MAINTERNANCE 1* from *Saccharomyces cerevisiae*, *AGAMOUS* from *Arabidopsis thaliana*, *DEFICIENS* from *Antirrhinum maju*s and *SERUM RESPONSE FACTOR* from *Homo sapiens* (Shore and Sharrocks, 1995). Their common feature is the conserved DNA binding domain, the MADS-box. They can be found in animals, fungi, and plants, whereby the family expanded especially in plants where over 100 genes are known to date. Those can be divided into the classes of type I and type II. Since all MADS-box genes described in the present thesis belong to class II, only this class will be introduced (Gramzow et al., 2010). The type II MADS-box transcription factors of plants have seven exons on average and all possess a characteristic protein structure: the DNA-binding MADS-domain (M), the Intervening-domain (I), a Keratin-like domain (K), and a C-terminal domain (C) (De Bodt et al., 2003). Subclasses of the type II genes are the MIKC^c and MIKC* classes, whereby this classification is based on the number of exons

of the I domain and structural differences of the K domain (Henschel et al., 2002). MIKC^c proteins have a short, and MIKC* have a long I domain and this domain influences the specificity of dimer formation, together with the MADS domain (Henschel et al., 2002; Zobell et al., 2010). The MADS domain is the DNA binding domain, which is usually located in the first, largest exon (Henschel et al., 2002). In order to possess DNA binding- and transcriptional activity, MADS-box proteins engage in homo- or heterodimeric complexes, while mainly interacting with another MADS-box protein. The MIKC^c-type protein FLC mostly engages in higher-order multimeric complexes of sizes larger than 100 kDa. This is much larger than dimeric complexes of, for example, FLC with SVP or members of the FLC-clade, the MAFs, which have a size of 40-50 kDa. Therefore, it is nearby that these proteins predominantly function as multimeric but not dimeric complexes (Helliwell et al., 2006). Complexes of MADS-box proteins mostly bind to the CArG-consensus DNA motif 5'-CC[A/T]$_6$GG-3' or slight variations thereof. However, binding is not restricted to CArG motifs and some rather rare binding events of MADS-box transcription factors to target motifs with long A-motifs were observed (Matys et al., 2006; Mathelier et al., 2014; Pajoro et al., 2014; Chow et al., 2016).

1.3.   Flowering time control of *A. thaliana*

Plants are sessile organisms that have adapted to their ecological habitats to optimize flowering time in order to guarantee reproductive success and survival. *A. thaliana* integrates a large number of exogenous and endogenous signals to perfectly adjust the transition from the vegetative to the generative phase through several pathways: the photoperiod pathway, vernalization pathway, ambient temperature pathway, sugar signaling pathway, gibberellin pathway, age pathway, and autonomous pathway (Amasino, 2010; Amasino and Michaels, 2010; Srikanth and Schmid, 2011; Posé et al., 2012). The ladder three integrate endogenous signals while the former respond to external cues. Flowering time of *A. thaliana* is well studied and more than 300 genes are known to be involved (Blümel et al., 2015; Bouché et al., 2016). Since light and temperature are the most important external cues for flowering time regulation and are relevant to the present thesis, this introduction focuses on the photoperiod and the temperature (vernalization and ambient temperature) pathways **(FIG. 1)**.
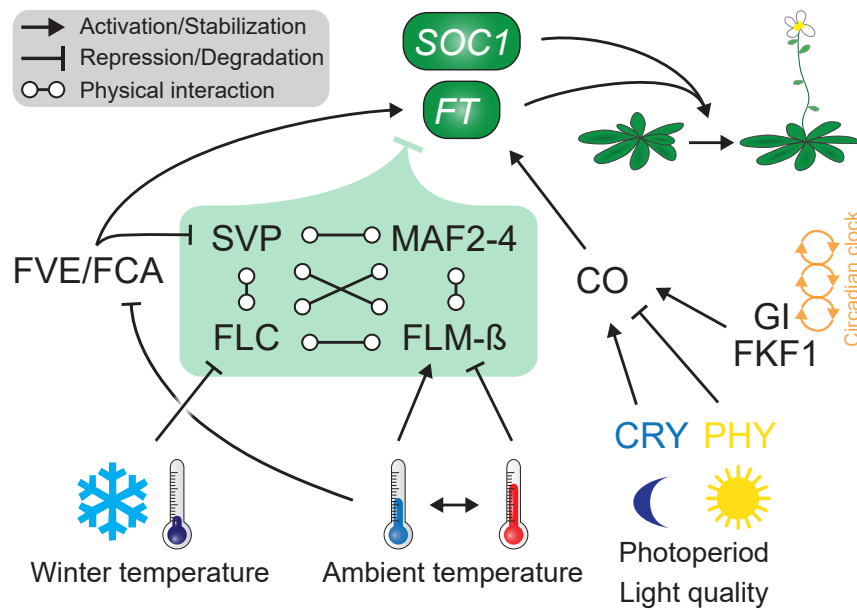
FIGURE 1: Simplified scheme of flowering time regulation of *A. thaliana* through the photoperiod, vernalization, and ambient temperature pathways. The circadian clock is composed by three interlocked loops, which are active at different times of a day. The GI/FKF1 complex induces *CO* expression (Imaizumi et al., 2005; Sawa et al., 2007; Fornara et al., 2009). Under long days, CO protein is under the control of signals originating from phytochrome (PHY) and cryptochrome (CRY) photoreceptors (Valverde et al., 2004). Once CO protein is stabilized, it induces *FT* expression in the leaf phloem. FVE and FCA, members of the autonomous pathway, are involved in thermosensitive flowering at 23°C and 16°C (Blazquez et al., 2003; Jeon and Kim, 2011). *FVE* levels are increased in response to higher temperature but no temperature-dependent regulation of *FCA* was shown until now (Jung et al., 2012). A long period of cold temperature (vernalization) leads to epigenetic silencing of *FLC*. In unvernalized or summer-annual plants, FLC directly represses the expression of floral activator genes like *FT* and *SOC1*, in concert with SVP (Michaels and Amasino, 2001; Li et al., 2008; Song et al., 2012). FLC, FLM, and MAF2-4 interact with each other and with SVP and protein complexes thereof delay flowering by repression of, e.g., *FT* and *SOC1* (indicated with the green box) (Li et al., 2008; Gu et al., 2013). The flowering-repressive isoform of FLM, FLM-ß, may represent a core component in a protein complex together with SVP to integrate changes of ambient temperature (Gu et al., 2013; Posé et al., 2013).

1.3.1. The photoperiod pathway

Depending on the flowering regulation in response to day length (photoperiod), plants can be classified as long day (long days promote and short days delay flowering), short day (short days promote and long days delay flowering) or day-neutral plants (not influenced by the length of light period). The plant model species *A. thaliana* is classified as a long day plant. Internal and external signals are combined to perfectly render flowering time dependent on the photoperiod. The circadian clock is an internal signal that cycles in a diurnal pattern and is reset at the end of the day according to the day length (Putterill et al., 2004; Imaizumi and Kay, 2006; Kobayashi and Weigel, 2007). The circadian clock in *A. thaliana* is built up of three main components: an early or morning loop, a central loop and a late or evening loop, which regulate each other by positive or negative feedback regulation (Dunlap, 1999; Harmer et al., 2001). Important integrators of the circadian clock to flowering time regulation are *GIGANTEA* (*GI*) and *FLAVIN-BINDING KELCH REPEAT F-BOX* (*FKF1*) (Fowler et al., 1999; Imaizumi et al., 2003). The loss of *GIGANTEA* expression disrupts the circadian rhythm, down-regulates *CONSTANS* (*CO*), a major photoperiod output gene, and delays flowering (Putterill et al., 1995; Suarez-Lopez et al., 2001). The *CO* transcript shows a peak at the end of the light period in long days, but during the night (after dusk) in short days. Light is necessary to stabilize the CO protein and CO-mediated flowering is prevented in short days (Suarez-

Lopez et al., 2001; Mockler et al., 2003; Valverde et al., 2004). When plants are transferred from long day conditions to continuous light mutants of *CO* flower late, whereas overexpression of *CO* accelerates flowering under long day but not under short day photoperiod (Putterill et al., 1995). In the final step of the photoperiod pathway, CO induces the expression of *FLOWERING LOCUS T* (*FT*) under long days. To do so, the CO protein must be activated and stabilized by photoreceptors of the PHYTOCHROME (PHY) and CRYPTOCHROME (CRY) families (Yanovsky and Kay, 2002). These photoreceptors are day-length sensors that regulate flowering by external light stimuli. Under long day conditions, CO protein is unstable at the beginning of the day but stabilizes at the end of the light period, few hours before dusk. In this stabilized state, CO induces the expression of *FT*, a small transcription cofactor (Kardailsky et al., 1999; Kobayashi et al., 1999). The loss of *FT* results in late flowering, while overexpression of *FT* anticipates flowering, even in the *CO* mutant (Kardailsky et al., 1999). This indicates that *FT* acts independently of the photoperiod pathway as a major floral integrator (Kardailsky et al., 1999; Kobayashi et al., 1999). Although *FT* expression is regulated by different inputs, CO is the main regulator in the photoperiodic pathway (Kobayashi and Weigel, 2007; Imaizumi, 2010). After reaching a specific threshold, FT protein, also known as 'florigen', moves to the shoot meristem to act as a long distance signal to induce flowering (Corbesier et al., 2007; Kobayashi and Weigel, 2007; Mathieu et al., 2007; Liu et al., 2012).

1.3.2. The vernalization pathway

Wild strains of *A. thaliana* can be classified into summer-annual and winter-annual types. Summer-annual accessions accomplish one or even more generative phases in a single season and are not dependent on a previous winter period. Flowering of winter-annual accessions depends on a long (several weeks) cold (around 4°C) period. They germinate during autumn and the young plants overwinter in the vegetative phase as rosettes and continue growth until initiation of the generative phase in spring (Shindo 2007).

In winter-annual accessions of *A. thaliana*, the well-studied vernalization pathway prevents premature flowering in many plant species before the long cold winter period and serves as a protective mechanism to prevent the direct exposure of sensitive floral meristems to lethal cold temperatures (Trevaskis et al., 2007; Distelfeld et al., 2009; Song et al., 2012; Hepworth and Dean, 2015). In winter-annual accessions, flowering without prior vernalization is strongly delayed by the MADS-box transcription factor *FLOWERING LOCUS C* (*FLC*), which directly represses the expression of floral activator genes like *FT* and *SUPPRESSOR OF OVEREXPRESSION OF CO1* (*SOC1*) (Michaels and Amasino, 1999; Johanson, 2000; Michaels and Amasino, 2001; Lee et al., 2007; Li et al., 2008; Song et al., 2012). Vernalization is a multi-step process that consists of the activation of *FLC*, the reprogramming of *FLC* chromatin during winter and the stable maintenance of the repressed state during development until completion of a generation (Song et al., 2013). The activation of *FLC* requires many chromatin modifiers and their function is enhanced by the coiled-coil domain protein FRIGIDA (FRI), which supports the recruitment of the chromatin modifiers to the promoter region of *FLC* (Johanson, 2000; Geraldo et al., 2009; Choi et al., 2011; Crevillen and Dean, 2011). When temperature drops during winter, *FLC* is gradually silenced through several regulatory levels, depending on the length of the cold period. The early *FLC* suppression is mediated by upregulation of *FLC* antisense transcript and a prolonged period of cold temperature leads to accumulation of the

epigenetic transcriptional repressive mark H3K27me3 (trimethylation of Lys27 of histone H3) around the transcription initiation site of *FLC*, the so-called nucleation region. This step is mediated by the POLYCOMB REPRESSOR COMPLEX 2 (PRC2)-like complexes (Finnegan and Dennis, 2007; De Lucia et al., 2008; Angel et al., 2011). When temperature increases during spring, *FLC* remains efficiently silenced until the finalization of development, as H3K27me3 marks quickly spread over the whole *FLC* locus (Finnegan and Dennis, 2007; De Lucia et al., 2008; Angel et al., 2011). Importantly, this epigenetic repression of *FLC* is reset during meiosis from both the paternally and the maternally derived silenced *FLC*. Hence, it is guaranteed that the progeny is again vernalization sensitive and does not flower before spring (Sheldon et al., 2008; Choi et al., 2009; Song et al., 2012). However, the molecular mechanism underlying this epigenetic resetting is not known yet.

1.3.3.   The ambient temperature pathway

The ambient temperature pathway of *A. thaliana* controls flowering time in a physiological temperature range of 12°C to 27°C and integrates short term temperature fluctuations during the vegetative growth period (Balasubramanian and Weigel, 2006; Kumar and Wigge, 2010; Wigge, 2013). Compared to the well studied photoperiod and vernalization pathways, the underlying molecular and genetic basis of the ambient temperature pathway is just beginning to be understood (Verhage et al., 2014; Capovilla et al., 2015). A rise of temperature by only a few centigrade (°C) due to global warming threatens agricultural production systems (Wheeler and von Braun, 2013; Pachauri and Meyer, 2014; Moore and Lobell, 2015). Therefore, understanding the ambient temperature pathway has recently gained increasing attention.

After the experience of a long cold winter period, the vernalization pathway has only a minor residual role in temperature-dependent flowering of *A. thaliana* (Blazquez et al., 2003; Gu et al., 2013; Lee et al., 2013). Following winter and during spring, cooler temperature delays and warmer temperature promotes flowering of *A. thaliana* **(FIG. 2)**. E.g., a temperature decrease from 23°C to 16°C delays flowering of long day grown *A. thaliana* plants by twelve days and by the formation of six leaves (Lempe et al., 2005). However, there are many examples of other plant species showing that this pattern is not conserved and that cooler temperatures can promote flowering in comparison to warmer conditions (Anderson et al., 2011; Nakano et al., 2013).
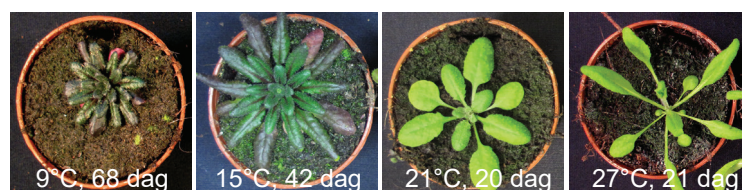
FIGURE 2: Pictures of *A. thaliana* plants grown in 15°C and 21°C temperature. The ambient growth temperature and the age of the plants (days after germination, dag) is indicated. Plants at 15°C, 21°C, and 27°C were bolting when the photographs were taken. No bolting was yet observed of the plant grown at 9°C.

One of the first implications on the genetic control of flowering time upon changes of the ambient temperature arose from the observation that disruption of the autonomous pathway genes *FVE*, a member of a chromatin remodeling complex, and *FCA*, a factor involved in mRNA polyadenylation, an impaired temperature-dependent induction of flowering (Blazquez et al., 2003).

FCA expression and protein levels increase in response to higher temperature but no temperature-dependent regulation of *FVE* was shown so far (Jung et al., 2012). FCA and FVE participate in the regulation of flowering through activation of *FT* and repression of the MADS-box transcription factor *SHORT VEGETATIVE PHASE* (*SVP*) (Lee et al., 2007).

Disruption of *SVP* leads to very early flowering and the sensitivity of flowering in response to changing temperature in a wide range of temperatures (5° to 27°C) is strongly diminished. Therefore, *SVP* is a potent repressor of flowering and represents a central component of the ambient temperature pathway (Hartmann et al., 2000; Lee et al., 2007; Lee et al., 2013). SVP represses the activation of floral activator genes like *FT* and *SUPPRESSOR OF OVEREXPRESSION OF CO1* (*SOC1*) by direct binding to their regulatory chromatin regions (Li et al., 2008). *SVP* is mainly expressed in leaf veins and mesophyll cells but *SVP* expression is only slightly regulated in response to temperature changes (Hartmann et al., 2000; Lee et al., 2007; Liu et al., 2007). However, protein levels do not follow these static transcript levels and a strong decrease in SVP protein levels was observed in response to increasing temperature. Facilitated through SVP protein degradation via the 26S proteasome, this posttranslational regulation may be an important mechanism of SVP-dependent regulation of flowering time (Lee et al., 2007; Lee et al., 2013).

At ambient temperatures of 12°C to 23°C, the MADS-box transcription factors of the so-called *FLC*-clade *FLOWERING LOCUS M* (*FLM*)*,* also known as *MAF1* and *AGL27*, and *MADS AFFECTING FLOWERING* (*MAF*) *2–4*, also known as *AGL31*, *AGL70*, and *AGL69*, are integrators of thermal changes. *FLM* is closely related to *MAF3*, and *MAF2-4* and are arranged in a gene tandem together with *MAF5*, which has an unclear role in flowering time regulation (Ratcliffe et al., 2001; Martinez-Castilla and Alvarez-Buylla, 2003; Balasubramanian et al., 2006; Lee et al., 2007; Gu et al., 2013; Posé et al., 2013). A loss-of-function mutant of *FLM* displayed similar but weaker phenotypes than *svp* and it was suggested that *FLM* is another central repressor of the ambient temperature pathway acting upstream and in dependency of *SVP* (Balasubramanian et al., 2006; Lee et al., 2007; Lee et al., 2013; Posé et al., 2013). Loss-of-function mutants of *MAF2*, *MAF3*, and *MAF4* showed earlier flowering and reduced thermosensitivity but to a much lesser extend than mutants of *FLM*. *MAF2* acts in parallel with *FLM* and a *flm maf2* double mutant largely resembles the early flowering and temperature-insensitivity of *svp* (Gu et al., 2013; Airoldi et al., 2015). Further, *MAF2* is regulated by vernalization and thereby prevents premature response of vernalization due to short periods of warm ambient temperatures during an otherwise cold period (Ratcliffe et al., 2003; Rosloski et al., 2013; Airoldi et al., 2015). As shown for the *FLM* mutant, phenotypes of *maf2-3* mutants were also most pronounced at cooler temperatures (Ratcliffe et al., 2003; Scortecci et al., 2003; Gu et al., 2013; Airoldi et al., 2015).

Studies on protein interaction have shown that the members of the *FLC*-clade interact with each other. MAF2 and MAF3 dimerize with FLM, and MAF4 with FLM and FLC (Li et al., 2008; Gu et al., 2013). FLC, FLM, MAF2 and MAF4 are able to interact with SVP, likely to form higher-order protein complexes, which then repress floral activator genes like *FT* and *SOC1* (Li et al., 2008; Gu et al., 2013). These FLC-clade protein complexes are predicted to have an overlapping function, because they genetically act in a partial redundant manner (Gu et al., 2013). Their ability to interact with each other and with SVP, and their temperature-dependent regulation may serve as an important mechanism to dynamically adjust flowering in response to changing ambient temperatures. SVP and FLM may thereby be the most important core components **(Fig. 1)**.

1.4.   Constitutive and alternative splicing in plants

Alternative splicing (AS) is an important regulatory level of gene regulation in *A. thaliana* and several genes, which are involved in flowering time regulation. The genes of plants have a multiexon structure and an average intron size of around 160 bp, what is considerably smaller than the introns of animal genes (Sakharkar et al., 2004; Marquez et al., 2012). Besides several mechanisms of posttranscriptional gene regulation such as, the regulation of mRNA stability, translation and posttranslational modifications, the removal of the introns and the parsing of the exon boundaries of unprocessed precursor messenger RNA (pre-mRNA) are important levels of gene regulation. Splicing is guided by four loosely conserved core sequences: The 5' and the 3' splice site (ss) at exon-intron-junctions, the branch point, and an polypyrimidine sequence, whereas the latter two are located just few bases upstream of the 3' splice site. The human spliceosome was shown to be a macromolecular complex that consists of several small nuclear RNAs (snRNAs), small ribonucleopreotein particles (snRNPs) and more than 150 additional proteins (Jurica and Moore, 2003; Will and Luhrmann, 2011). The composition of plant spliceosomes has not been characterized so far. However, analysis of homology between the human and *A. thaliana* genomes showed that the *A. thaliana* genome may code for a much higher number (>200) of splicing factors than were identified in humans (Wang and Brendel, 2006; Koncz et al., 2012). Such higher complexity may reflect the increased redundancy between splicing factors in plants. This is supported by the observation, that disruption of several snRNP proteins was not lethal but led to diverse defects in developmental processes (Staiger and Brown, 2013).

Alternative splicing (AS) is a level of the cotranscriptional splicing process, which tremendously increases the coding potential of the *A. thaliana* genome. AS acts through the alternative assembly of exons of intron-containing open reading frames (ORF) and several types of AS events have been described (Reddy, 2007). Among these different types, intron retention and exon skipping make up around 50% of AS events. Further, differential 5' or 3' splice site choice that leads to expanded or reduced exon length account for around 23% of *A. thaliana* AS events (Reddy, 2007; Kalyna et al., 2012; Marquez et al., 2012) **(Fig. 3)**.
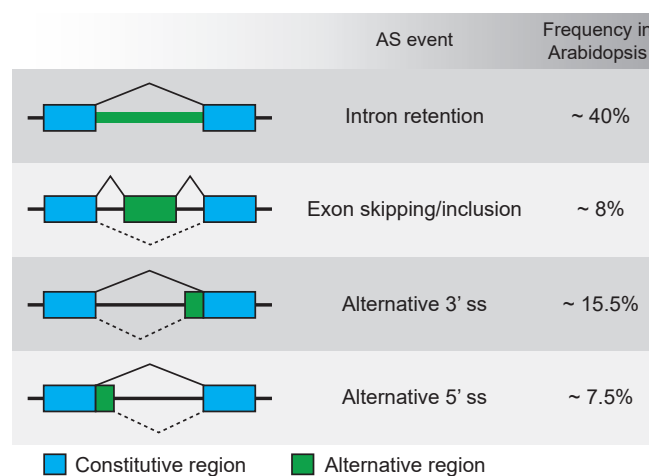
FIGURE 3: Frequent types of alternative splicing events in *A. thaliana*. Proportion of four most frequent types of AS events in *A. thaliana* as reported by Barta et al., 2012, and Marquez et al., 2012. (Modified from Reddy et al. 2013).

Not every alternatively spliced transcript is translated into functional protein. An important downstream quality control mechanism of AS is the degradation of aberrant spliced transcript through the nonsense-mediated decay (NMD) pathway (Kervestin and Jacobson, 2012). Some examples of typical features of NMD sensitive transcripts are a premature termination codon (PTC) or long 3' untranslated regions (3'-UTR) (Isken and Maquat, 2007). In *A. thaliana*, most NMD sensitive transcripts contain one of these frequent sequence features and it was estimated that around 13% of intron-containing *A. thaliana* genes are regulated through AS coupled to NMD (Kalyna et al., 2012).

1.5.   The role of alternative splicing for temperature controlled flowering

Alternative splicing has important effects on *A. thaliana* flowering time control in response to changing ambient temperature. Four variants of *FLM* (α, β, γ, δ) were initially identified, whereby the most abundant ones are *FLM-ß* and *FLM-δ*, at least in the reference accession Col-0 (Scortecci et al., 2001; Lee et al., 2013; Posé et al., 2013) **(Fig. 4)**.
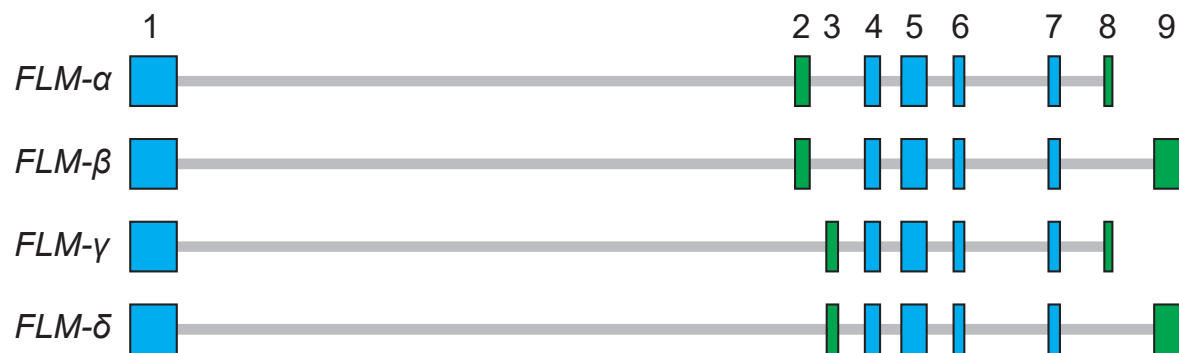
FIGURE 4: Graphic representation of the *FLM-α*, *FLM-ß*, *FLM-γ*, and *FLM-δ* isoforms. Constitutively used exons and alternatively used exons are indicated as blue and green boxes, respectively. Introns are indicated as gray line. Exon numbers are indicated at the top.

Within the ambient temperature range, *FLM* is subject to temperature-dependent alternative splicing and *FLM-ß* and *FLM-δ* are generated through the incorporation of either a mutually exclusive exon 2 (*FLM-ß*) or 3 (*FLM-δ*). When either *FLM-ß* or *FLM-δ* were overexpressed under the control of the Cauliflower mosaic virus (CaMV) 35S promoter, *FLM-ß* overexpression led to delayed flowering and *FLM-δ* overexpression led to accelerated flowering. Based on these observations and by further chromatin-immunoprecipitation (ChIP) and yeast-2-hybrid (Y2H) experiments, a model was proposed according to which *FLM-ß* and *FLM-δ* have antagonistic functions in flowering time control (Posé et al. 2013). Both FLM isoforms, FLM-ß and FLM-δ, physically interact with SVP. However, the interesting downstream effects of these complexes are antagonistic: the FLM-ß-SVP dimer represses the expression of floral activator genes like *SOC1*, *TEMPRANILLO2* (*TEM2*), *SEPALLATA3* (*SEP3*), and *FT* by direct chromatin association. FLM-δ also interacts with SVP but this complex is unable to bind DNA but competes with FLM-ß. Therefore, FLM-δ was proposed to act as a dominant-negative FLM variant to promote flowering in response to increased ambient temperature (Gu et al., 2013; Lee et al., 2013; Posé et al., 2013). Further, FLM-ß was shown to possess DNA-binding abilities only in presence of SVP, indicating that FLM-ß-FLM-ß dimeric complexes were not sufficient to repress floral activator genes and that the presence of SVP is necessary for FLM-ß

function **(Fig. 5)**. However, recent studies put this model in a new perspective. A model was proposed by which *FLM-ß* abundance is regulated through alternative splicing and a large number of splice variants, including *FLM-δ,* are produced at the expense of *FLM-ß* but those are likely subject to degradation via the NMD pathway. Only *FLM-ß* may be translated and interact with SVP and members of the *FLC*-clade to participate in flowering repression in response to decreased ambient temperature (Sureshkumar et al., 2016).
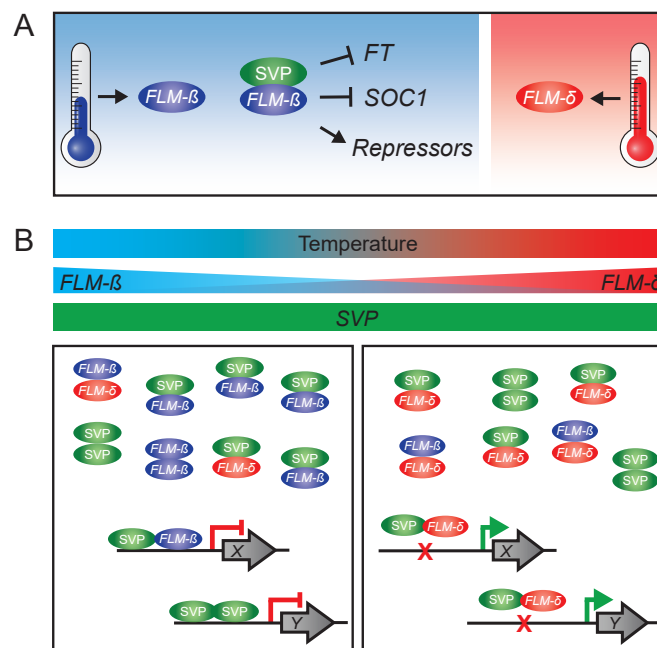
FIGURE 5: Model of FLM-ß function in complex with SVP. A. *FLM* splicing is temperature dependent and FLM-ß-SVP heterodimers are able to directly repress the expression of floral activators and to activate the expression of floral repressors. **B**. Model of FLM-SVP complex function in dependency with temperature. *FLM-ß* and *FLM-δ* transcript levels decrease and increase with increasing temperature, respectively. *SVP* transcript levels are not regulated by temperature. At low temperatures FLM-ß-SVP heterodimeric and SVP-SVP homodimeric complexes are abundant, which both regulate the expression of flowering related genes. Under warmer temperatures, the *FLM-δ* isoform becomes more abundant, which is also able to form heterodimeric complexes with SVP. However, these complexes are impaired in DNA-binding ability, which is why it acts as a negative dominant protein complex. (Modified from Posé et al., 2013).

MAF2, which is another regulator of ambient temperature controlled flowering and a member of the *FLC*-clade of MADS-box transcription factors, is also regulated by AS. Similar to *FLM*, levels of two major *MAF2* isoforms var1 and var2 increased and decreased, respectively, when plants were transferred to cold temperature. However, only *MAF2* var1 codes for a functional protein, and similar to *FLM-ß*, this isoform acts as flowering repressor. The non-functional splice variant var2 contains a PTC and is mainly produced at higher temperatures (Ratcliffe et al., 2003; Gu et al., 2013; Airoldi et al., 2015).

As mentioned earlier, the *A. thaliana* genome codes for a high number of splicing factors. Amongst others, serine/arginine proteins (SR), heterogeneous nuclear ribonucleoproteins (hnRNPs), and small nuclear ribonucleoprotein particles (snRNPs) are known to be important splicing factors in *A. thaliana* and some were identified to be upregulated when plants were transferred from 16°C to 23°C (Balasubramanian et al., 2006). However, it remains to be investigated, which splicing factors are involved in AS regulation of flowering time regulators like *FLM* and *MAF2*.

1.6.   Objectives

Increasing global temperature threatens the productivity of agricultural systems. Therefore, research on the ambient temperature pathway, which regulates flowering time in response to changing ambient temperature, came into focus. It is largely unknown, whether natural genetic variation of genes that act in the ambient temperature pathway contributes to regulation and adaptation of flowering time of *A.thaliana*.

Hence, one aim of the present thesis was to investigate natural variation of flowering time of *A. thaliana* accessions in the ambient temperature range and to elucidate its molecular basis by genetic mapping. Thus, a novel natural allele of the MADS-box transcription factor *FLOWERING LOCUS M* (*FLM*) was identified and its effect on *FLM* expression and flowering time regulation was characterized. This work was published in PLOS Genetics (see Appendix).

During this project, the investigation of publicly available genome sequences of *A. thaliana* accessions uncovered that *FLM* harbors substantial nucleotide diversity on a species-wide level, especially in the non-coding region. Regulatory regions of *FLM* were not yet described and it was not known how *FLM* expression is controlled. Therefore, a follow-up project aimed at identifying such regulatory regions. Subsequently, genetic variation at these regulatory regions and its consequence on *FLM* expression should be investigated. The findings of this study provide first examples of *cis*-regulatory regions of *FLM.*

2.   Results

2.1.   Modulation of temperature-dependent flowering by natural variation of *FLM*

Natural variation of flowering in response to changes in the ambient growth temperature was observed among *A. thaliana* accessions (Lempe et al., 2005; Werner et al., 2005; Balasubramanian et al., 2006; Lee et al., 2013). However, at the beginning of the present thesis, the underlying genetic basis of this phenotypic diversity was largely unknown. This motivated the initiation of the project to uncover as of yet unidentified alleles of genes, which control thermoresponsive flowering. Hence, the first part of the present thesis aimed to genetically map and characterize natural alleles of *A. thaliana* accessions and to elucidate their potential role for adaptation of flowering time during *A. thaliana* evolution.

2.1.1.   Killean-0 is an early flowering *A. thaliana* accession

To understand the variation in flowering time in response to temperature, flowering time of a collection of *A. thaliana* accessions was determined at 15°C and 21°C. Thus, the accession Killean-0 (Kil-0) was identified as an early flowering accession at both temperatures. When compared to the reference accession Col-0, Kil-0 also flowered earlier under different photoperiods (long day and short day light regimes) **(Fig. 6A - C)**. Flowering is promoted by the experience of long periods of cold during the winter (vernalization) and the vernalization pathway could, therefore, contribute to this early flowering behavior of Kil-0. However, when the effect of a six week cold-treatment (4°C) on flowering in Kil-0 and Col-0 was compared, only a minor effect on flowering in both accessions between non-vernalized and vernalized plants was detected **(Fig. 6D)**. Furthermore, when testing the expression of *FLC* (*FLOWERING LOCUS C*), the major vernalization-responsive gene in *A. thaliana*, its abundance was as strongly reduced in Kil-0 as in the vernalization-independent Col-0 **(Fig. 6E)**. Taken together, this suggested that the early flowering of Kil-0 at 15°C was regulated by the ambient temperature pathway and independently from the vernalization pathway. This latter conclusion was also in line with previous reports that had identified Kil-0, like the Col-0 reference, as a vernalization-independent summer annual accession (Lempe et al., 2005; Werner et al., 2005a).
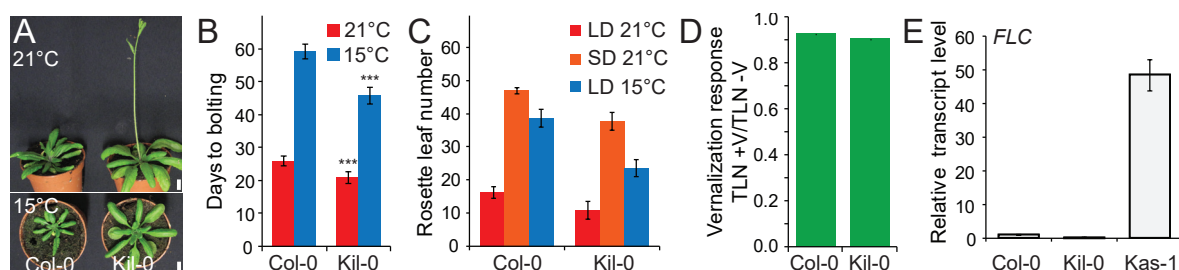
Figure 6: Kil-0 flowered earlier than Col-0 but independent of vernalization. A. Representative photographs of Col-0 and Kil-0 under 21°C and 15°C and continuous light. **B.** Quantitative flowering time analysis of Col-0 and Kil-0 under 21°C and 15°C and continuous light. Measured in days to bolting (DTB). **C.** Quantitative flowering time analysis of Col-0 and Kil-0 under different growth conditions. Measured in rosette leaf number (RLN). LD, long days; SD, short days; CC, continuous light. Bars represent s.d. of biological replicates (C. and D.). **D.** Vernalization response of Col-0 and Kil-0. Total leaf number of vernalized plants (8 weeks at 6°C) was compared to non-vernalized plants. **E.** qRT-PCR analysis of *FLC* of twelve days old non-vernalized plants grown under 21°C long day photoperiod. Kas-1 was integrated as representative winter-annual accession with high *FLC* levels. Student's t-tests were performed in comparison to the wild type: ***, p < 0.001.

2.1.2. *FLM* is the causative locus for early flowering in Kil-0

When grown at 15°C, Kil-0 (39.9 days to bolting, DTB) bolted about two weeks earlier than Col-0 (52.4 DTB) **(Fɪɢ. 6B)**. This prominent early flowering behavior of Kil-0 at 15°C (hitherto FT15) allowed to reliably distinguish between Kil-0 and Col-0. F1 hybrids of the two accessions indicated that early flowering of Kil-0 approximated the inheritance of a recessive trait **(Fɪɢ. 7A)**. The subsequent analysis of segregating F2 plants showed that the Kil-0 flowering phenotype segregated in an approximately monogenic manner with an average value for the late flowering segregants of 54.7 DTB compared to 59.3 DTB in Col-0 or 45.9 DTB of Kil-0 **(Fɪɢ. 7B ᴀɴᴅ C)**. Thus, FT15 in Kil-0 was largely controlled by one major recessive locus.
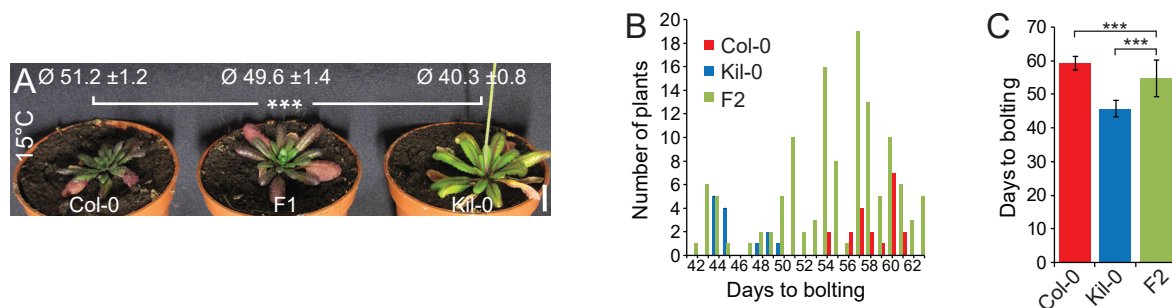
FIGURE 7: Early flowering of the accession Kil-0 was likely controlled by one major locus. A. Representative photographs of Col-0 x Kil-0 F1 hybrids under 15°C continuous light. Average values ± s.d. are shown in the picture. **B.** Distribution of flowering time in the F2 segregating population obtained from crossing Col-0 with Kil-0. Presented data are distribution of Col-0 (n = 20), Kil-0 (n = 15), and F2 (Col-0 x Kil-0; n = 124) plants. **C**. Average values and s.d. of flowering time of F2 segregating population shown in B. Student's t-tests were performed as indicated in the figure: ***, p < 0.001.

Using the extreme phenotypic borders of a Col-0 x Kil-0 F2 population the FT15 trait was rough-mapped to a 968 kb genomic region on the lower arm of chromosome 1 **(Fɪɢ. 8A ᴀɴᴅ B)**. First, F2 plants with a recombination in the interval of interest were selected by screening the genotypes with two flanking markers. After selfing of these plants, plants of the F3 generation were phenotyped for FT15 that still showed a recombination in this interval and, in this manner, identified 49 and 41 recombinant F3 plants with an early (Kil-0) or late (Col-0) FT15 flowering time, respectively **(Fɪɢ. 7C ᴀɴᴅ 8D)**. The interval of interest was further narrowed down to a 151 kb region between 28.9 and 29.1 Mb on Col-0 chromosome 1 with additional markers **(Fɪɢ. 8E)**. Then, 15 early and nine late flowering F3 recombinants were selected and pooled to obtain genomic sequences of these pools by next generation sequencing (PoolSeq). The sequence reads were aligned to the Col-0 reference genome. In parallel, using next generation sequencing a Kil-0 genome sequence was obtained and 309 high confidence single nucleotide polymorphisms (SNP) that allowed distinguishing between Kil-0 and Col-0 were detected. SNP allele frequencies (*f*) within the 151 kb mapping interval were determined. To balance out local fluctuations, locally weighted scatterplot smoothing (LOESS) was applied and the difference (*Δf*) in SNP allele frequency between the two pools of each LOESS-smoothed *f*-value was calculated for each SNP. Using a *Δf* threshold > 25%, a peak region of 31.3 kb was determined as the final mapping interval in Kil-0 **(Fɪɢ. 8E)**.
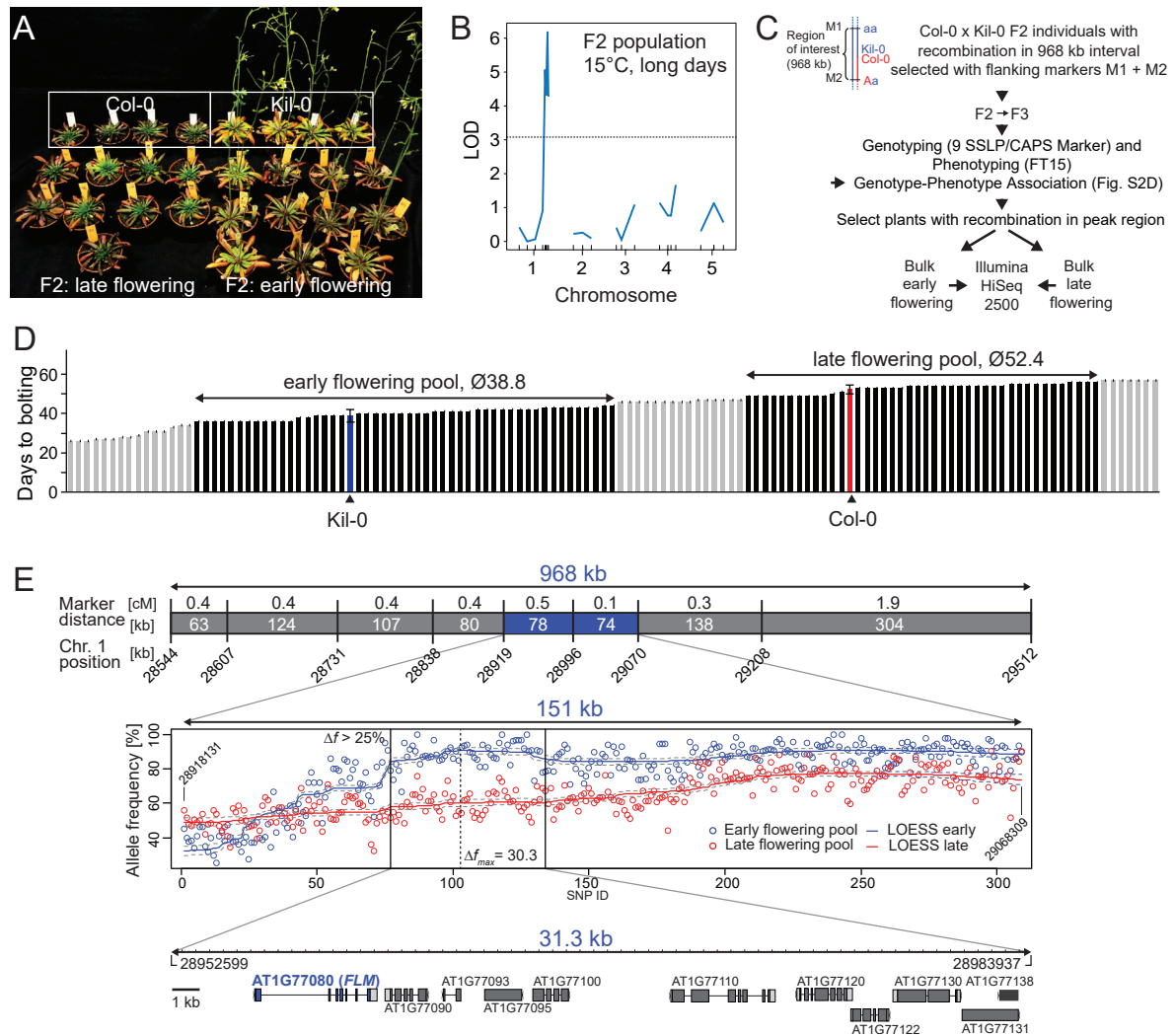
FIGURE 8: The underlying locus of Kil-0 early flowering was identified by genetic mapping. A. Representative photographs of Col x Kil-0 F2 plants selected from the extreme phenotypic borders (see also Fig. 7B). **B**. Rough mapping using ten early and ten late flowering F2 plants as representatively shown in A. The 5% LOD significance threshold as calculated from 1000 permutations is indicated with the dotted horizontal line. Markers were selected from Pacurar et al. (2012) or generated as described in MATERIAL AND METHODS and are provided as TABLE 3. **C**. Schematic representation of the mapping procedure. A previously identified genetic interval of 968 kb was flanked by two markers (indicated as M1 and M2) and recombinants were identified by screening of 1049 F2 plants. Hence, 43 F2 plants with M1$^{Kil-0/Col-0}$ and M2$^{Kil-0/Kil-0}$ genotypes (and vice versa) were selected. Since the early flowering phenotype of Kil-0 was not fully penetrant and phenotypical segregation not exactly in a 3:1 ratio, F2 individuals were subsequently selfed to obtain 43 F3 recombinant lines. Four plants per F3 familiy (in total 172 plants) were individually genotyped and phenotyped at 15°C and continuous light. F3 plants without recombination were discarded. A genotype-phenotype association of the remaining set of was calculated and a peak region of 151 kb was identified. F3 plants with a recombination in the peak region were bulked to early (Kil-0-like, 15 plants) and late (Col-0-like, nine plants) flowering bulks, respectively and sequenced with a Illumina HiSeq 2500 (Illumina, San Diego, USA) to obtain high-quality SNP data. **D**. Quantitative flowering time analysis of 127 F3 recombinant plants from 43 families in DTB grown under 15°C continuous light. Each bar represents a single plant. The parental lines Col-0 and Kil-0 are indicated as blue and red bar, respectively, with s.d. shown for 15 replicates. The horizontal arrows depict the value range of the early and the late flowering pool, as indicated as dark grey bars. The average values are indicated on the arrows. Plants that were not included in the analysis due to phenotypic values distinct from the value range of the parental lines are indicated by light grey bars. **E**. Results of the mapping procedure described in A. The 968 kb interval with flanking markers M1 and M2 and seven internally located markers is shown as vertical lines with the genetic (cM) and physical (kb) marker distance (calculated from the dataset of 1,049 F2 individuals described in A. and the physical position on chromosome 1 (kb). The 151 kb peak region is shown in blue. The middle panel shows the respective allele frequency of the 309 SNPs of either the early (Kil-0-like, in blue) or late (Col-0-like, in red) flowering pool. Red and blue lines show the respective LOESS-smoothed allele frequency values with the 95% confidence interval as grey dotted line. Vertical black lines show the final mapping interval of 31.3 kb with $\Delta f > 25\%$. The SNP with the highest difference between early and late pool (Δf_{max}) is marked with a dotted black line. Start and end positions on chromosome one are indicated as numbers in boxes (bp). The lower panel shows a detailed illustration of the eleven annotated gene models in the 31.3 kb interval with *FLM* marked in blue.

In parallel, a gene expression dataset was generated by RNA sequencing (RNA-seq) from 10 days old Kil-0 and Col-0 plants grown at 21°C. The analysis of the RNAseq data was performed in collaboration with Matthias Pfeifer (Helmholtz Zentrum, München, Germany). Transcripts were detected for seven of the eleven annotated genes within the 31.3 kb region of interest. Only the MADS-box transcription factor *FLM*, which has a role in temperature-dependent flowering control, showed a strong and significant downregulation when Kil-0 was compared to Col-0 (Scortecci et al., 2001; Ratcliffe et al., 2001). *FLM* was also the most strongly downregulated gene between Kil-0 and Col-0 when we specifically analyzed the expression of 267 genes with an implied role in flowering time regulation **(Fig. 9A)**. This data thus suggested that *FLM* may be the causative gene for early flowering in Kil-0 (*FLM^{Kil-0}*).

2.1.3. *FLM^{Kil-0}* harbors a large structural variation

To uncover a putative causative polymorphism at the *FLM^{Kil-0}* locus, the Kil-0 genome sequence was mapped to the Col-0 reference. When comparing *FLM^{Kil-0}* and *FLM^{Col-0}*, only a few polymorphisms were found in the *FLM* introns and the promoter region but no polymorphisms in the *FLM* coding sequence. Interestingly, however, theaverage 25-fold paired-end read coverage of the genome resequencing was reduced to only threefold at the beginning of the first *FLM* intron **(Fig. 9B)**. Since a structural polymorphism at this site could potentially cause such a strong reduction in read coverage, the Kil-0 genomic sequencing reads were *de novo* assembled and an insertion in *FLM^{Kil-0}* that was flanked by two contigs corresponding to two halves of the *FLM^{Col-0}* locus at the predicted insertion site was identified **(Fig. 9B)**. PCR analysis then identified a 5.7 kb insertion in the first intron of *FLM^{Kil-0}* **(Fig. 9B)**. In cooperation with Heidrun Gundlach (Helmholtz Zentrum, München, Germany), it was found that the inserted 5.7 kb insertion was remotely related (68% identity, 86% coverage) to the *A. thaliana* LINE element ATLINE1_8 (hereafter called LINE) that belonged to the class of non-LTR (non-LONG TERMINAL REPEAT) LINE (LONG INTERSPERSED ELEMENT) elements **(Fig. 9C)**. This LINE contained six, mostly fully covered, typical retrotransposon domains as well as a perfect copy of the second exon of a *RIBONUCLEASE H-LIKE* gene corresponding to AT1G04625 on chromosome 1 of the Col-0 reference genome **(Fig. 9C)**. This structural polymorphism may be causative for the *FLM*-dependent early flowering phenotype of Kil-0.
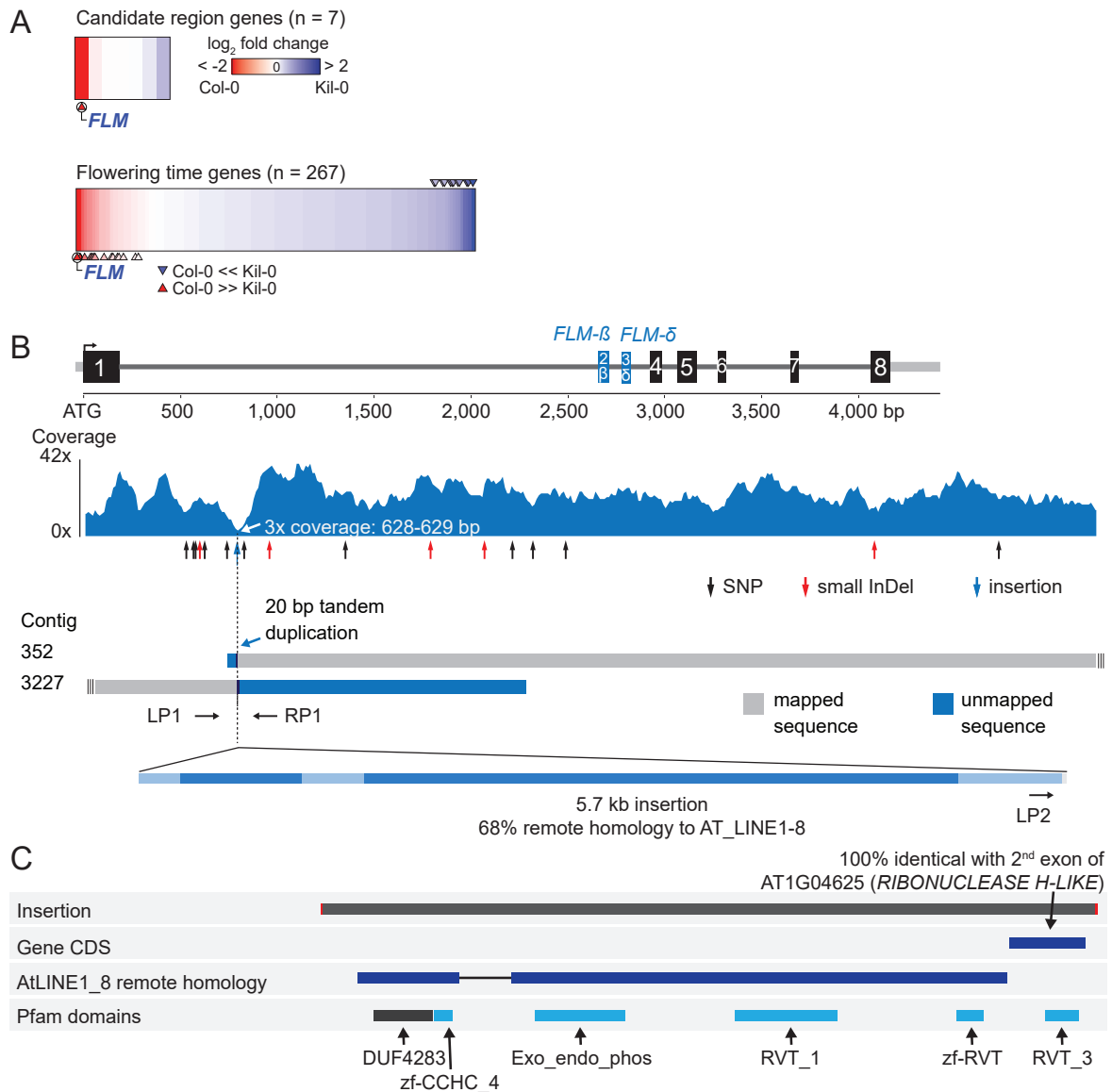
FIGURE 9: A large structural variation in the first intron is the causative polymorphism for FLM^{Kil-0} accelerated flowering.
A. Heatmap of differentially expressed genes in Kil-0 versus Col-0 in 21°C as analyzed by RNA-seq. The upper bar shows seven expressed genes out of eleven annotated gene models in the 31.3 kb mapping interval that is depicted in B. The lower bar shows expression values of 267 genes with an implied function in flowering time regulation. Fold changes are log2 transformed whereby blue shows upregulation and red downregulation in Kil-0 versus Col-0, respectively. Significantly different (P < 0.01) up- and downregulated genes are depicted with blue and red arrows above and below the heatmap panels, as also shown by the legend. **B**. Gene model of the FLM^{Kil-0} locus and position of the insertion. Shown are the Col-0 reference gene models of FLM-$ß$ and FLM-$δ$. Black boxes indicate exons, light grey lines show UTRs and dark grey lines indicate introns. TAIR10 FLM isoform coding annotations are shown above. Numbering is based on the A of the ATG start codon as +1. The read mapping track is shown below. The threefold covered region is depicted with an arrow. Below the *de novo* read assembly of Kil-0 sequencing reads is shown. The two contigs (contig 352 and contig 3227) were reassembled to FLM^{Col-0}. The light grey bars indicate the mapped sequence. The sequence that could not be mapped to the reference sequence is indicated in blue. Note, that the two contigs did not assemble each other indicating that the insertion sequence could not be completely recovered from the *de novo* analysis. Primers used for screening of accessions as mentioned in the text are indicated as LP1, LP2, and RP1. The inserted sequence in intron 1 of FLM^{Kil-0} is described at the bottom. The light grey bar indicated the complete 5.7 kb insertion. The dark blue bar indicates the partition of the insertion that shows 68% similarity to a LINE retrotransposon sequence. **C**. Detailed description of the individual domains and the respective homologies of the inserted sequence in intron 1 of FLM^{Kil-0}.

2.1.4. FLM^{Kil-0} is a functional temperature-sensitive *FLM* allele

The loss of *FLM* function was previously reported to result in early flowering in Col-0 when grown at 23°C (Scortecci et al., 2001; Balasubramanian et al., 2006). To be able to compare the effects of the FLM^{Kil-0} candidate locus with those of the *flm-3* loss-of-function mutant and the reference *FLM* allele from Col-0 (FLM^{Col-0}) in a homogenous genetic background, an introgression line $Col^{FLM-Kil}$ was established by six backcrosses (BC6) of FLM^{Kil-0} into Col-0. Subsequent flowering time analyses revealed a delay in flowering at 15°C and 21°C in comparison to *flm-3* but a significantly faster flowering of $Col^{FLM-Kil}$ in comparison to Col-0 **(FIG. 10A AND B)**. Conversely, to understand the effect of the *flm-3* loss-of-function on flowering in Kil-0 an introgression of *flm-3* into Kil-0 until BC4 (Kil-0^{flm-3}) was also established. In line with the conclusions from the $Col^{FLM-Kil}$ analysis, Kil-0^{flm-3} flowered earlier than Kil-0, again indicating that the effects of FLM^{Kil-0} were intermediate between those of the functional Col-0 allele and the *flm-3* loss-of-function allele **(FIG. 10A AND B)**. Interestingly, the phenotypic differences between Col-0 and $Col^{FLM-Kil}$ or Col-0 and *flm-3* allele were more pronounced at 15°C than at 21°C **(FIG. 10B)**. This observation was in agreement with the flowering time analysis of *flm-3* in a range of growth temperatures, which revealed that *FLM* contributed particularly prominently to flowering time control at 15°C **(FIG. 10C)**. In summary, the *FLM* function was partially compromised in the FLM^{Kil-0} allele and *FLM* strongly contributed to flowering at 15°C in Col-0 and Kil-0. The strong effect of *FLM* on flowering at 15°C was interesting since it had previously been proposed that this growth temperature much better resembles natural growth conditions of *A. thaliana* than the commonly used 21°C (Hoffmann, 2002; Weigel, 2012).
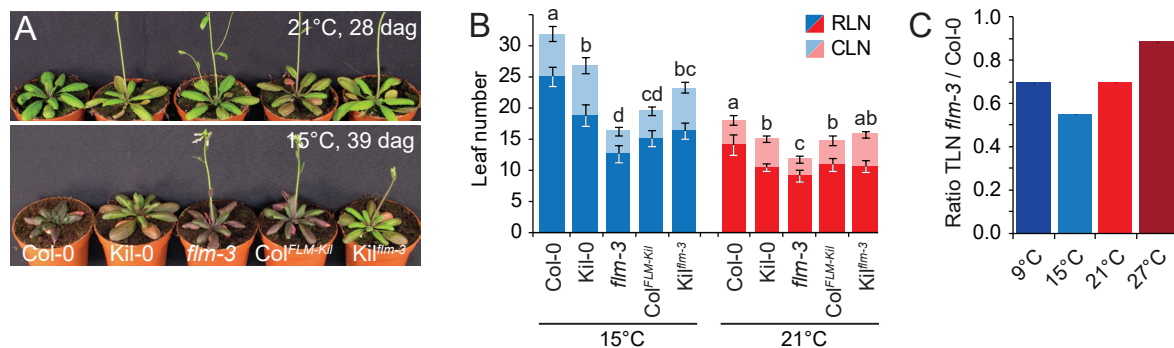
FIGURE 10: The effects of the alleles FLM^{Kil-0} and *flm-3* depend on the genetic background. A. Representative photographs of $Col^{FLM-Kil}$ and Kil^{flm-3} backcross lines in comparison to Col-0, Kil-0, and *flm-3* under 15°C and 21°C temperature and long day photoperiod. Photographs were taken at 28 and 39 days after germination (dag) as indicated. **B.** Average and s.d. of flowering time analysis. RLN, rosette leaf number; CLN, cauline leaf number. Similar letters indicate no significant difference of total leaf number (Tukey HSD, p < 0.05). **C.** Ratio of total leaf number (TLN) of *flm-3* compared to Col-0 grown at long day photoperiod and the indicated temperature.

2.1.5. Gene expression and alternative splicing of *FLM* are differentially regulated in Kil-0

Since the FLM^{Kil-0} allele had weaker effects on flowering time than the *flm-3* loss-of-function allele and since FLM^{Kil-0} did not reveal any major polymorphisms in the *FLM* coding region, it might be possible that the temperature-sensitive flowering phenotype of Kil-0 may be caused by changes in *FLM* expression or *FLM* alternative splicing. To examine the effect of temperature on *FLM* transcription and splicing in Kil-0, Col-0 and Kil-0 plants were grown for seven days at 21°C and then transferred for three days to 9°C, 15°C, 21°C, and 27°C, respectively. In agreement with previously

published data, we observed a decrease of *FLM-ß* and an increase of *FLM-δ*, respectively, in response to warmer temperatures (**Fig. 11A and B**) (Lee et al., 2013; Posé et al., 2013). Importantly, similar temperature-dependent changes in the abundance of the *FLM* isoforms were maintained in Kil-0 and Col*FLM-Kil*. However, whereas the increases in *FLM-ß* and *FLM-δ* abundance were maintained in relative terms within the different genotypes, the overall transcript abundance of *FLM-ß* and *FLM-δ* was decreased sixfold and 27-fold between Kil-0 or Col*FLM-Kil* and Col-0 at 21°C. To understand the effects of this differential regulation, the expression levels of *SVP*, *FT*, and *SOC1* were determined. Whereas *SVP* expression was comparable between the different genotypes, *FT* and *SOC1* were expressed more strongly in the early flowering Kil-0 or Col*FLM-Kil* than in Col-0 (**Fig. 11C - G**). Thus, the reduced expression of *FLM* was likely the cause for the early flowering of Kil-0 and Col*FLM-Kil* and the temperature-dependent differential accumulation of *FLM-ß* and *FLM-δ* might be the basis of the temperature-sensitive flowering time in Kil-0. Since reduced expression of *FLM-ß* had a stronger effect on flowering time than *FLM-δ*, the *FLM-ß* levels seemed to provide the dominant cue for flowering time regulation in Kil-0.
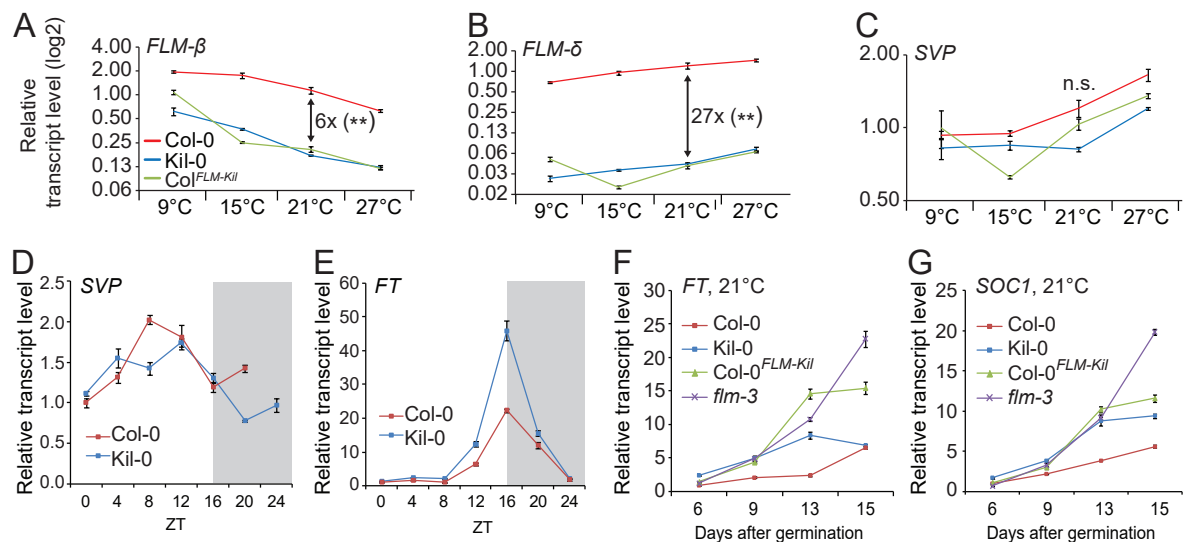
FIGURE 11: Kil-0 and Col*FLM-Kil* show strong downregulation of *FLM-ß* and *FLM-δ*. A. - C. qRT-PCR analysis of *FLM-ß*, *FLM-δ*, and *SVP* expression in ten days old Col-0, Kil-0 and Col*FLM-Kil* plants. Plants were grown for seven days at 21°C and then transferred to the indicated temperature and grown for further three days. Ratio of fold change values of 21°C versus 15°C are indicated in the graph. Student's t-tests were performed as indicated in the figure: **, p < 0.01. **D.** and **E.** qRT-PCR analysis of *SVP* and *FT* during a long day photoperiod and 21°C of Col-0 and Kil-0. Dark period is indicated by grey area. Samples were harvested every four hours. **F.** and **G.** qRT-PCR analysis of *FT* and *SOC1*. Col-0, Kil-0, *flm-3* and Col*FLM-Kil* were grown at 21°C long day photoperiod. Samples were taken when plants were 6, 9, 13, and 15 days old, respectively, at ZT16 (start of dark). Error bars indicate s.d. of three biological replicates.

2.1.6. The LINE insertion affects *FLM* splice isoform abundance

The large insertion in the first intron of *FLM* could be the causative polymorphism for the low *FLM* abundance or the reduction in the *FLM-ß* isoform in Kil-0. To test this hypothesis, Kil-0 plants were transformed with a genomic fragment of FLM^{Col-0} including 2 kb promoter and 0.5 kb 3'-UTR sequence as well as a variant of FLM^{Kil-0} with a deletion of the LINE insertion ($FLM^{Kil-0\Delta LINE}$). In support of the hypothesis, transgenic expression of $FLM^{Kil-0\Delta LINE}$ and FLM^{Col-0} resulted in delayed flowering when expressed from the transgenes in Kil-0 (**Fig. 12A**). Conversely, we also tested whether the 5.7 kb insertion had a comparable effect on *FLM* transcript abundance and introduced a transgene of FLM^{Col-0} carrying the 5.7 kb LINE insertion in the respective locus into the *flm-3* loss-of-function mutant. Indeed, this insertion correlated with a strong downregulation of *FLM*, which in turn retained the alternative splicing behavior of *FLM* and thus mimicked the expression and splicing pattern observed for FLM^{Kil-0} (**Fig. 12B**). Thus, it was very likely that the LINE insertion in the first intron of *FLM* was the causative polymorphism for reduced *FLM* transcription, *FLM* isoform accumulation and accelerated flowering in Kil-0.
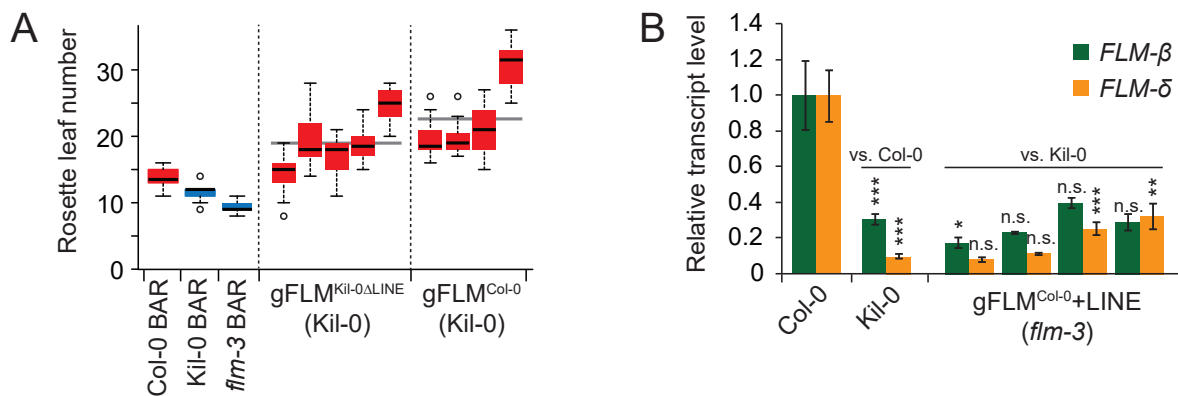
FIGURE 12: Transgenic complementation identifies the large intron insertion as the causative polymorphism. A. Quantitative flowering time analysis (RLN) of four to five independent T2 lines being either hemi- or homozygous for the indicated construct (Kil-0 background). 14 - 20 plants were grown in 21°C and long day photoperiod. As control, BASTA-resistant Col-0, Kil-0 and *flm-3* lines were used. **B**. qRT-PCR analysis of *FLM-ß* and *FLM-δ* expression of four independent homozygous T3 lines of the indicated construct (*flm-3* background). Plants were grown for ten days under 21°C long day photoperiod. Error bars indicate s.d. of three biological replicates. Student's t-tests were performed as indicated in the figure: *, $p < 0.05$; **, $p \leq 0.01$; ***, $p < 0.001$; n.s., not significant.

To gain a detailed understanding of the molecular effect of the LINE insertion on *FLM* transcription and splicing, the exon usage of Col-0 and Kil-0 was compared using the before mentioned RNA-seq dataset obtained from plant material grown at 21°C (RNA-seq data analysis was done in cooperation with Mathias Pfeiffer, Helmholtz Zentrum, München, Germany). In Col-0, there was a high presence of all exons with a strong differential use of the alternative exon 2 or 3 that define the *FLM-ß* and *FLM-δ* isoforms, respectively. Furthermore, the *FLM-ß*-specific exon 2 was more abundant at 21°C than the defining *FLM-δ* exon 3. In Kil-0, generally fewer reads were detected for all exons including exon 1 suggesting that the LINE insertion between exon 1 and exon 2 may control overall *FLM* transcript abundance (**Fig. 13A**). The read-coverage of the subsequent exons including that of the alternatively used *FLM-ß* exon 2 and *FLM-δ* exon 3 was more strongly decreased in Kil-0 when compared to exon 1 suggesting that the LINE element may also negatively control *FLM* splicing efficiency (**Fig. 13B**). This observation was verified by qRT-PCR (**Fig. 13C**). Since the reduced exon reads in Kil-0 could be the consequence of the

increased size of intron 1 in Kil-0 and possibly inefficient splicing, the RNA-seq reads from Col-0 and Kil-0 were remapped against the FLM^{Col-0} genomic locus and read mapping was specifically examined over the exon-intron junctions. Whereas read coverage dropped sharply at the exon 1/intron 1 junction in Col-0, reads covering the beginning of intron 1 could be readily retrieved in Kil-0 **(Fig. 13D)**.
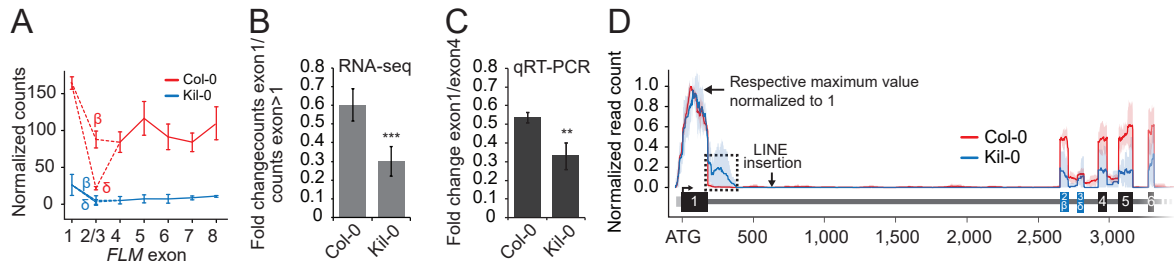
FIGURE 13: The large structural variation in the first intron affects *FLM* splicing efficiency. A. Exon-usage analysis along the *FLM* coding sequence. **B**. Ratios between the read count of exon 1 versus exons 2 - 7 as shown in A. **C**. Reproduction of the exon-usage analysis by qRT-PCR using primers located in the *FLM* exon 1 and exon 4. Student's t-tests was performed as indicated in the figure: **, p ≤ 0.01. **D**. Gene expression determined by RNA-seq in the *FLM* gene locus (including UTRs, and exons and introns of *FLM-ß* and *FLM-δ*) with a per nucleotide resolution. The number of mapped RNA-seq reads were normalized to range from 0 (no expression) to 1 (maximum expression). Lines indicate the average expression level and light blue and light red areas indicate the 5% and 95% confidence intervals that were determined across the biological replicates of one genotype. Parts of the *FLM* locus are shown with light grey as the 5'-UTR, black box the exon 1 and dark grey line parts of the intron 1. The position of the LINE insertion in *FLM*[Kil-0] is indicated with an arrow. Numbering is based on the A of the ATG start codon as +1.

Then, qRT-PCR with primers spanning only exon 1 or exon 1 and parts of intron 1 were used to estimate the amount of transcripts containing intron 1 sequences **(Fig. 14A)**. The existence of intron-containing transcripts suggested that the LINE insertion resulted in inaccurate splicing at the 5'-splice site of Kil-0 intron 1 **(Fig. 14B)**. This inaccuracy may impact the abundance of the *FLM-ß* and *FLM-δ* full-length transcripts but, importantly, not the existence of the respective full length mRNAs as confirmed by semi-quantitative RT-PCR **(Fig. 14C)**. Apart from the inaccuracy of splicing, also a premature termination of transcription could be a possible reason for intron-containing transcripts. Hence, to approximate the rate of *FLM* transcription, the level of nascent *FLM* pre-mRNA was determined using primers in intron 1 upstream of the insertion, and the introns 2 and 6. No significant differences in pre-mRNA levels were detected when amplifying intron 1 **(Fig. 14D)**. However, the relative abundance of sequences of the introns 2 and 6, which are located downstream from the insertion were strongly reduced in Kil-0, indicative for a partial premature termination of *FLM* transcription **(Fig. 14D)**.
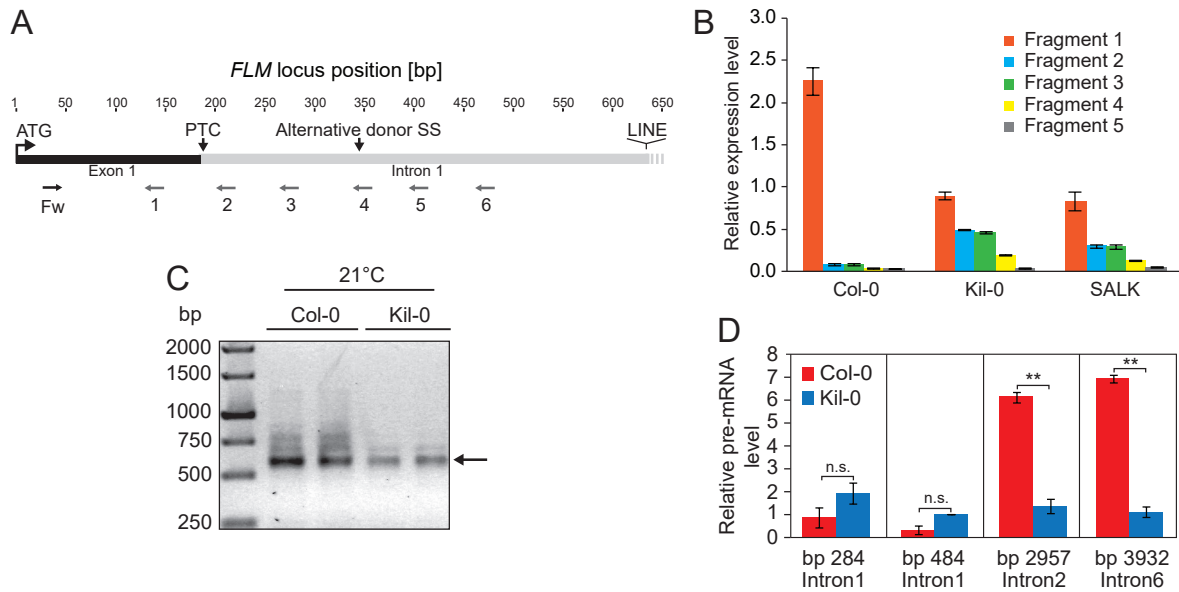
FIGURE 14: Not only aberrant splicing but also premature termination of transcription may reduce FLM levels of FLM^Kil-0. A. Schematic representation of the binding position of the primers used for the qRT-PCR quantification of E1I1p intron 1 sequence-containing transcripts and the 3'-RACE-PCR as indicated by arrows below the gene model; PTC, premature termination codon. **B.** qRT-PCR analyses of intron 1 sequence-containing transcripts, which were amplified using the reverse primers 1 - 6 as indicated in FIG. 14B. Error bars indicate s.d. of three biological replicates. Note that the comparison of the expression values between the fragments are approximations since the primer efficiencies may not be exactly identical. **C.** Semi-quantitative PCR using cDNA samples of Col-0 and Kil-0. **D.** qRT-PCR quantification of unspliced pre-mRNA from isolated nuclei. Primers are located introns 1, 2, and 6 of FLM and introns 2 and 3 of the reference gene ACT8. The position of the respective reverse primer relative to the ATG start codon is indicated. Error bars indicate s.e. of two biological replicates. Student's t-tests were performed as indicated: **, $p < 0.01$; n.s., not significant.

Transposon insertions can lead to alternative polyadenylation and it was examined whether the corresponding intron 1 sequence-containing transcripts were polyadenylated by and performing 3'-RACE (rapid amplification of cDNA 3'-ends) experiments (Wu et al., 2011; Duc et al., 2013; Tsuchiya and Eulgem, 2013) **(FIG. 14B)**. The 3'-RACE-PCR yielded two fragments, one corresponding in size to *FLM-ß* or *FLM-δ* transcripts and one smaller fragment, which was much more abundant in Kil-0 than in Col-0 **(FIG. 15A)**. These low molecular PCR products were isolated, cloned, and sequenced and six different short polyadenylated transcripts containing intron 1 sequences were determined **(FIG. 15B)**.
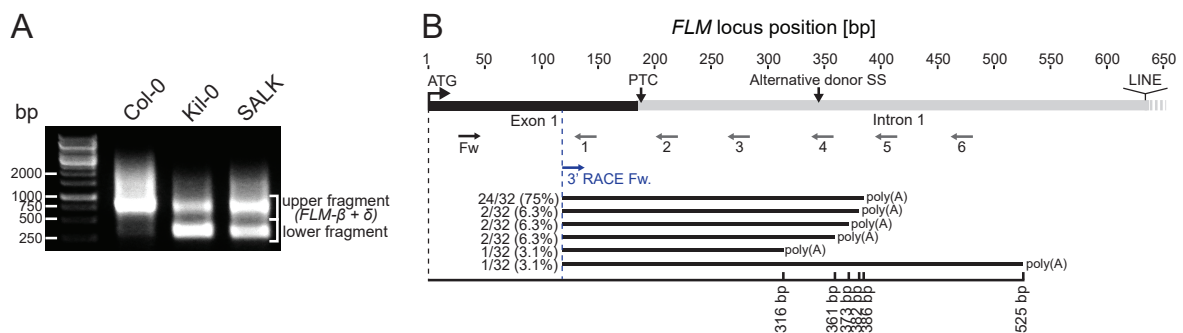
FIGURE 15: Elucidation of the sequence identity of short E1I1p-transcripts. A. Image of an agarose gel with the analysis of 3'-RACE-PCR products. The upper and lower fragments were isolated for sequencing. **B.** A schematic representation of the identified intron 1 sequence-containing E1I1p polyadenylated transcripts. Numbers on the left indicate the frequency of each transcript among the 32 individually sequenced clones. The black line indicates the length of each transcript respective to the ATG start codon.

In conclusion, the early flowering of Kil-0 correlated with the presence of a LINE insertion in *FLM* intron 1. Further, the LINE insertion did not affect *de novo FLM* transcription initiation but partially impaired the formation of full length transcripts, possibly as a result of premature transcription termination and aberrant splicing. Many other mechanisms that were not directly examined may also explain the reduction of *FLM* transcript in Kil-0.

2.1.7. The geographic distribution of the *FLM*$^{Kil-0}$ allele is consistent with a recent adaptive selective sweep

To understand to what extent the *FLM*$^{Kil-0}$ structural polymorphism was present in other *A. thaliana* natural variants, a highly genetically diverse set of 415 *A. thaliana* accessions composed of accessions represented in a previously published Hapmap population and other selected laboratory accessions was screened (Li et al., 2010) **(Fig. 16A)**. To this end, PCRs on genomic DNA from these accessions were performed with a diagnostic primer set that allowed distinguishing between the *FLM*$^{Kil-0}$ and the *FLM*$^{Col-0}$ alleles. *FLM* genomic fragments were amplified from all accessions and in this way a further nine accessions with an insertion of identical size and insertion position as detected in Kil-0 were identified. Subsequent DNA sequence analysis of the nine *FLM*$^{Kil-0}$-like accessions confirmed that all accessions carried the insertion in the same position as Kil-0. Therefore, these ten accessions are refered to as *FLM*^{LINE} accessions **(Table 1 and Fig. 16A - C)**.

TABLE1: List of accessions carrying the *FLM*$^{Kil-0}$ allele. *FRI*, *FRI GIDA* allelic status as retrieved from Werner et al. (2005) and Lempe et al. (2005) or our own analysis using primers described in Johanson et al. (2000). Vernalization response was calculated from the ratio of total leaf number of vernalized (eight weeks at 6°C) compared to unvernalized plants. Geographic data was retrieved from www.1001genomes.org.

Name	Stock	Latitude	Longitude	*FRI* allele	Vernalization response	Country	Location
Go-0	CS6721	51.5	9.9	Col-0	1.0	Ger	Göttingen
Hl-0	CS6737	52.1	9.4	Col-0	1.0	GER	Holtensen
Hn-0	CS6739	51.3	8.3	Col-0	0.8	GER	Hennetalsperre
Kil-0	CS6754	55.6	-5.7	Col-0	0.9	UK	Killean
Me-0	CS1364	51.9	10.1	Col-0	1.1	GER	Mechtshausen
Sp-0	CS6862	52.5	13.2	-	0.4	GER	Berlin-Spandau
Ste-0	CS6864	52.6	11.9	-	0.5	GER	Stendal
Ull2-3	CS22587	56.1	14.0	-	1.0	SWE	near Kristianstad
Wt-3	CS6894	52.3	9.3	Col-0	1.0	GER	Wietze
Unknown	-	-	-	Col-0	1.0	-	-
Col-0	CS28168	38.3	-92.3	Col-0	0.9	USA	Columbia

When the geographic distribution of the *FLM^{LINE}* accessions was analyzed based on geographic information of the collection points, a distribution over Scotland (Kil-0), Sweden (Ull2-3) and Germany (eight accessions) with a central distribution in Germany were found **(Fig. 16D)**. In cooperation with Jörg Hagmann and Congmao Wang (Max Planck Institute for Developmental Biology, Tübingen, Germany), genome sequences of 1128 accessions from the 1001 *A. thaliana* Genomes Project was analyzed using read coverage over the insertion site of the *FLM^{LINE}*-locus and its 5'- and 3'-flanking sequences **(Fig. 16E)**. This analysis, independently, identified the accession El-0 (Germany) as one further *FLM^{LINE}* accession, which was not included in the further analysis.

Sequence analyses of the LINE insertion of the ten *FLM^{LINE}* accessions revealed a very high homology between them **(Fig. 16F)**. When considering a spontaneous mutation rate of 6 to 7×10^{-9} per site per generation, around one to three seed generations per year, and the absence of any selective pressure on the inserted sequence, the common ancestor was calculated to have originated 8,000 to 30,000 years ago (Schultz et al., 1999; Ossowski et al., 2010).

A subsequent phylogenetic analysis using a *FLM* genomic fragment from the ten *FLM^{LINE}* accessions together with further *FLM* genomic sequences from Col-0 and 88 randomly chosen accessions from the 1001 *A. thaliana* Genomes Project indicated that all accessions clustered in one clade of the phylogenetic tree and are thus strongly related **(Fig. 16G and H)**. To test whether these accessions were also strongly related when analyzed at a genome wide scale, published haplotype data, which included all ten *FLM^{LINE}* accessions, was investigated (Atwell et al., 2010; http://arabidopsis.usc.edu). This analysis revealed that the ten *FLM^{LINE}*-accessions belong to genetically differentiated clades indicating that they represented truly independent accessions.
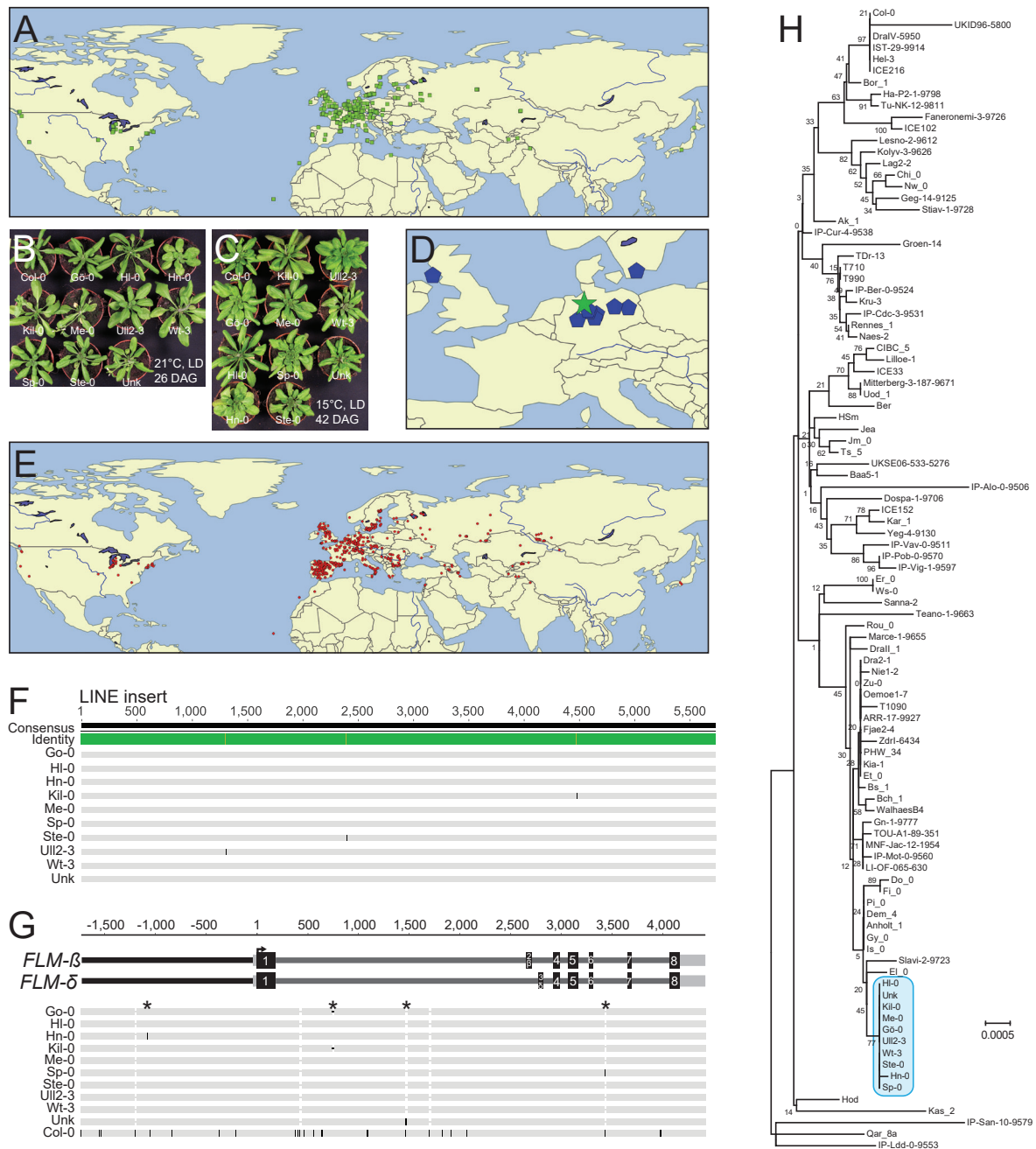
FIGURE 16: The *FLM^{Kil-0}* allele can be identified across accessions. A. Geographic distribution of the accessions used for the PCR-based screens for *FLM^{LINE}* accessions and further large structural variations in intron 1 of *FLM*. **B.** Geographical distribution of the *FLM^{LINE}* accessions. The centroid is indicated as light grey star. Note that the accession with ambiguous name was excluded due to missing geographic information. **C.** and **D**. Representative photographs of the *FLM^{LINE}* accessions grown in 21°C and 15°C and under long day photoperiod. **E**. Geographic distribution of the accessions used for the analysis of unmapped reads. **F**. Multiple alignment of the 5.7 kb inserted sequence of the *FLM^{LINE}* accessions. Black marks lines indicate SNPs. **G**. Multiple alignment of the *FLM* locus and 1.8 kb upstream sequence of the *FLM^{LINE}* accessions. Note that the LINE insertion at +631 was not considered in this analysis. Polymorphic sites between the *FLM^{LINE}* accessions are indicated with an asterisk. Col-0 sequence was used as reference. Numbering is based on the A of the ATG start codon as +1. **H**. Neighbor joining tree showing the genetic relationships among *FLM* sequences. Sequences of ten *FLM^{LINE}* accessions were combined with additional 88 sequences randomly chosen from the 1001 *A. thaliana* Genomes database. The *FLM^{LINE}* -clade is highlighted with a blue box.

All *FLM^LINE* accessions showed strong downregulation of *FLM-ß* and *FLM-δ* transcript abundance, albeit to different extents **(FIG. 17A)**. Since the low *FLM-ß* and *FLM-δ* levels in Kil-0 are the cause for early flowering we examined flowering time of the *FLM^LINE* accessions. A vernalization-requirement may interfere with such an analysis, therefore we first examined the vernalization requirement of these ten accessions. This revealed that two relevant accessions carried functional *FRI* and *FLC* alleles and were vernalization-dependent. Therefore, these were excluded from the further analysis. When summarized, the eight vernalization-independent *FLM^LINE* accessions flowered earlier than Col-0, which was particularly prominent at 15°C. Thus, in the absence of a vernalization-requirement, the LINE insertion polymorphism might contribute to earlier flowering at 15°C **(FIG. 17B AND C)**.

In conclusion, the *FLM^Kil-0* allele may have arisen recently and spread geographically. In the respective accessions, the *FLM^Kil-0* allele may contribute to their early flowering in a background-dependent and temperature-dependent manner.
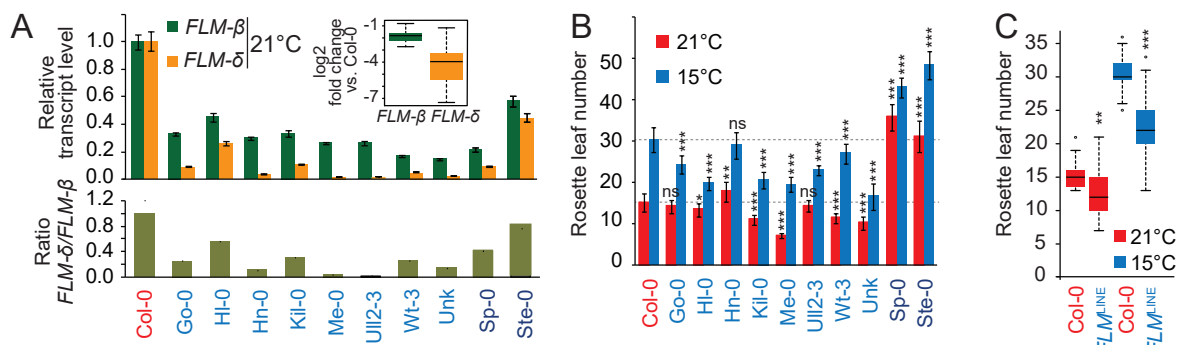
FIGURE 17: The *FLM^Kil-0* allele accelerates flowering depending on the genetic background. A. qRT-PCR analysis of *FLM-ß* and *FLM-δ* expression of all *FLM^LINE* accessions grown under 21°C long day photoperiod. Fold changes are normalized with Col-0 as control. Insert graph shows the summarized log2 transformed fold changes of *FLM^LINE* accessions only (Col-0 excluded). Below the ratios of *FLM-δ* versus *FLM-ß* fold change values are shown with Col-0 set to 1. Vernalization-sensitive accessions are indicated in dark blue. Bars indicate s.e. of three measurements. **B.** Quantitative flowering time analysis of *FLM^LINE* accessions grown under 21°C or 15°C long day photoperiod. Control lines indicating the Col-0 values are dashed. **C**. Averaged RLN values of all eight summer-annual *FLM^LINE* accessions. Student's t-tests were performed in comparison to Col-0 unless indicated otherwise in the figure: *, $p < 0.05$; **, $p \leq 0.01$; ***, $p < 0.001$; n.s., not significant.

2.1.8. The first intron carries important regions for isoform-specific *FLM* regulation

Specifically within the family of MADS-box transcription factor genes, various cases are known where structural polymorphisms located in the first intron enhance or repress gene expression (Hong et al., 2003; Distelfeld et al., 2009; Schauer et al., 2009; Yoo et al., 2011; Coustham et al., 2012). To test if *FLM* intronic sequences contribute to *FLM* expression, *flm-3* mutants were transformed with constructs for the expression of the *FLM-ß* or *FLM-δ* coding sequences under control of a 2.1 kb *FLM* promoter fragment. These experiments were performed by David Posé (Max Planck Institute for Developmental Biology, Tübingen, Germany) **(FIG. 18A)**. Importantly, none of the resulting *FLM-ß* or *FLM-δ* T1 transformants expressed significant levels of the respective transgene or showed a suppression of the *flm-3* early flowering phenotype **(FIG. 18B AND C)**. *FLM* is expressed from the corresponding genomic constructs containing all introns. Further, *FLM* is expressed from the above-described construct when only intron 1 is included (Posé et al., 2013). Hence, intron 1 may be essential for *FLM* expression.
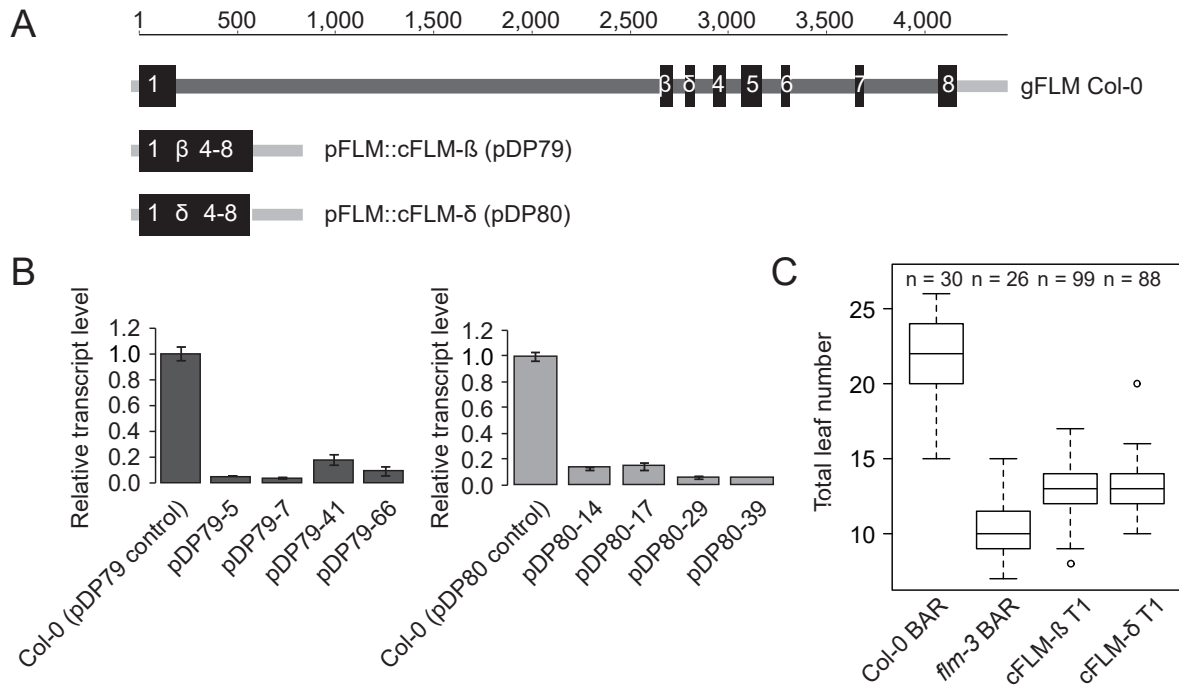
FIGURE 18: The first intron is essential for *FLM* basal expression. A. Constructs used by David Posé and Markus Schmid (Max Planck Institute for Developmental Biology, Tübingen, Germany) to determine a putative regulatory role of *FLM* intronic sequence. Numbering is based on the A of the ATG start codon as +1. **B**. qRT-PCR analysis of *FLM-ß* and *FLM-δ* of each four independent T2 lines transformed with the constructs described in A. Bars indicate s.e. of two biological replicates. **C.** Quantitative flowering time analysis of independent T1 transformants with the constructs described in A. BASTA-resistant Col-0 and *flm-3* lines were used as controls. Number of plants used is indicated in the top of the graph. These experiments were performed by David Posé (Max Planck Institute for Developmental Biology, Tübingen, Germany).

To address whether sequence identity, sequence length or the specific insertion site within intron 1 conferred the effect of the LINE element on *FLM* expression and splicing, the effects of T-DNA or *DS* (*DISSOCIATOR*) transposon element insertions in *FLM* intron 1 were examined (Alonso et al., 2003; Kuromori et al., 2004; Kleinboelting et al., 2012) **(Fɪɢ. 19A)**. These intron 1 insertions were of similar size (4.5 kb to 5.3 kb) to the 5.7 kb LINE insertion and present in the Col-0 and No-0 (Nossen-0) accessions: Salk_068360 (Salk, Col-0), RATM13-4593-1 (RIKEN, No-0) and GK_487H01 (GABI, Col-0) **(Fɪɢ. 19A)**. Additionally, the accession Co-1 was identified, which carries an around 1.5 kb insertion in *FLM* intron 1 **(Fɪɢ. 19A)**. When we determined the effects of these structural variants on *FLM* transcript accumulation and alternative splicing in plants grown at 15°C and 21°C, the insertion was found to reduce *FLM* transcript abundance when compared to the controls **(Fɪɢ. 19B)**. Furthermore, *FLM-δ* was more strongly reduced than *FLM-ß.* Hence, increases in intron length, regardless of the molecular identities of the insertions, may result in decreased *FLM-ß* and *FLM-δ* expression and changes in the ratio between the isoforms. At the same time, the temperature-sensitive regulation of the *FLM* isoforms was maintained in all lines.
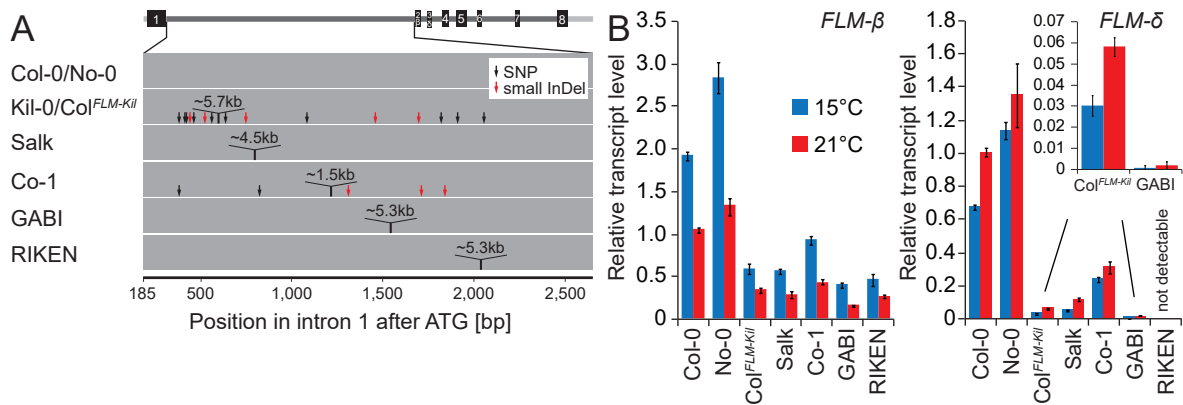
FIGURE 19: The effect of large insertions in the first intron of *FLM* on transcript levels depends on their size and position.
A. Schematic representation of the insertion lines used. The *FLM-ß* genomic gene model is indicated with black boxes as exons and grey lines as introns. The intron 1 is depicted below as grey bar. The position of the insertion is indicated with black arrows with the respective approximate size of the insertion. Numbering is based on the A of the ATG start codon as +1. **B**. qRT-PCR analysis of *FLM-ß* and *FLM-δ* expression of the lines shown in A. Plants were grown for seven days at 21°C and then shifted to 21°C or 15°C long day photoperiod. The insert graph shows the *FLM-δ* values of Kil-0 and GABI lines for clearer visibility. Bars indicate s.e. of three measurements.

To examine the phenotypic consequences of the observed transcriptional changes of the *FLM* insertion lines, their flowering was evaluated at 15°C and 21°C. Regardless of the differences in expression of the *FLM* isoforms, all lines flowered earlier than the respective wild type, and this effect was particularly prominent at 15°C **(Fig. 20A)**. Importantly, no pleiotropic effects on plant growth or plant height were noticed for the tested alleles suggesting that *FLM* acts specifically on flowering time regulation **(Fig. 20B)**. Next, it was evaluated to what extent changes in *FLM-ß* or *FLM-δ* abundance could explain the observed difference in flowering time. To this end, flowering time was correlated to *FLM* expression **(Fig. 20C)**. Expression of *FLM-ß* correlated much better with flowering time than *FLM-δ* and this effect was particularly pronounced at 15°C **(Fig. 20C)**. Thus, *FLM-ß* expression levels alone, rather than the ratio between *FLM-ß* and *FLM-δ,* have a prominent effect on flowering time regulation in *A. thaliana*.
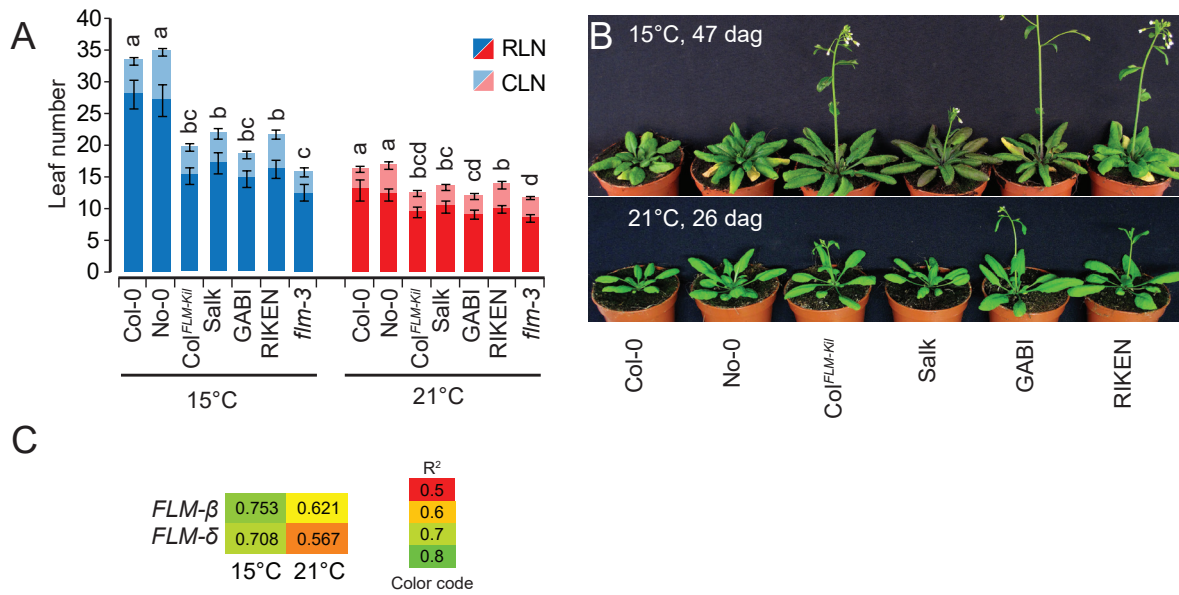
FIGURE 20: Reduction of *FLM-ß* explains acceleration of flowering. A. Quantitative flowering time analysis of the insertion lines shown in Fig. 18A with Col-0 and No-0 as wild type controls. Similar letters indicate no significant difference of total leaf number (Tukey HSD, $p < 0.05$). **B.** Representative photographs of lines described in Fig. 19A under 15°C and 21°C temperature and long day photoperiod. Photographs were taken at 47 and 26 days after germination. **C.** Correlation analysis of the flowering time experiment shown in A. (15°C and 21°C) with the expression values shown in Fig. 19B. Flowering time values of *flm-3* were also included with its *FLM* expression set to 0 (no expression). R^2 values from simple linear regression are shown.

2.2. Haplotypes of *FLM* non-coding regions fine-tune flowering time in ambient temperatures in *A. thaliana*

The first part of the present thesis describes the identification of the *FLM^LINE* allele, which possesses a large structural variation in the first intron of *FLM* and causes low *FLM* transcript levels and accelerated flowering in *A. thaliana* accessions. The effect of this expression-modifying allele was especially prominent at cool (15°C) temperature. Evidence for a potential adaptive role of this structural variation is presented.

During this project, the investigation of publicly available genome sequences of *A. thaliana* accessions uncovered that *FLM* harbors substantial nucleotide diversity on a species-wide level, especially in the non-coding region. As described in the first part of the present thesis, the non-coding, intronic sequence of *FLM* possesses regulatory function. As regulatory regions of *FLM* were not described yet and as it was not known how *FLM* expression is regulated, the idea arose that regulatory regions of *FLM* can be identified by utilizing natural genetic variation as a 'toolbox' of alleles, when combined with datasets of *FLM* expression of natural *A. thaliana* accessions.

2.2.1. *FLM* intronic sequences are required for basal and temperature-sensitive gene expression

An intronless *FLM* expressed from a functional *FLM* promoter fragment failed to express *FLM* to detectable levels and was not able to rescue the flowering time phenotype of a *flm-3* loss-of-function mutant **(Fɪɢ. 18)**. Next, the *FLM* expression in transgenic lines expressing a genomic *FLM* fragment containing all six introns (pFLM::gFLM) was compared to the expression of lines expressing the *FLM-ß* (pFLM::FLM-ß) or *FLM-δ* (pFLM::FLM-δ) splice variants retaining intron 1 but lacking introns 2 to 6 **(Fɪɢ. 21A)**. These constructs and the transgenic lines were generated by David Posé, provided by Markus Schmid (Max Planck Institute for Developmental Biology, Tübingen, Germany) and described in Posé et al. (2013). At the ambient temperatures 15°C, 21°C and 27°C, the presence of intron 1 was sufficient to restore the basal expression of *FLM* but introns 2 - 6 were required for the temperature-sensitive regulation of the splice variants *FLM-ß* and *FLM-δ* **(Fɪɢ. 21B)**. Thus, intron 1 may contain critical information for *FLM* expression and introns 2 - 6 may contribute to temperature-dependent gene expression.
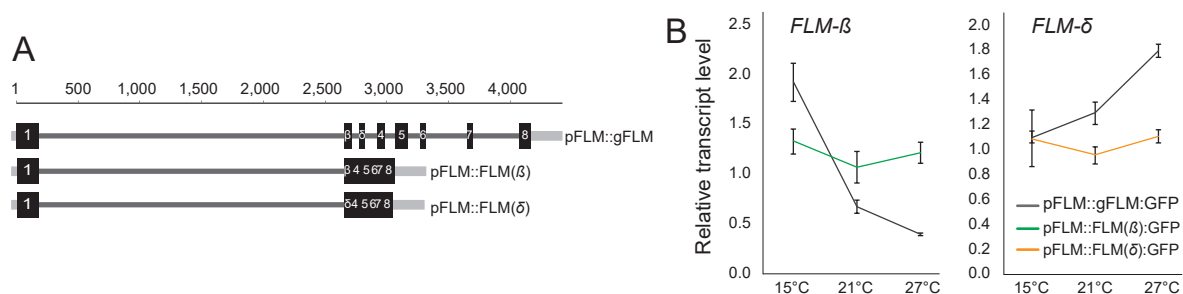
FIGURE 21: The intronic region of *FLM* is important for temperature-sensitive *FLM* regulation. A. Schematic representation of the pFLM::gFLM, pFLM::FLM-ß, and pFLM::FLM-δ constructs. Black boxes indicate exons, grey boxes are 5'- or 3'-untranslated regions and dark-grey lines are introns. **B**. qRT-PCR analysis of *FLM-ß* and *FLM-δ* expression in ten days old plants harboring the transgenes shown in A. Seven days old seedlings grown at 21°C were transferred to the indicated temperature and grown for further three days. Error bars indicate s.d. of three biological replicates. The constructs and the transgenic lines were generated by David Posé and provided by Markus Schmid (Max Planck Institute for Developmental Biology, Tübingen, Germany)

2.2.2. Non-coding variation defines major *FLM* haplotypes in *A. thaliana* accessions

To obtain a better picture of the distribution of nucleotide diversity along the genomic *FLM* locus, natural nucleotide variation was investigated using publicly available genome data from > 800 *A. thaliana* accessions (The 1001 Genomes Consortium, 2016). In agreement with the results from a genome-wide analysis of 80 accessions, the highest diversity was found in the 2 kb *FLM* promoter (π = 0.0036) and intronic sequence (π = 0.0022) whereas the coding sequence was more strongly conserved (π = 0.0007) (Cao et al., 2011) **(Fig. 22A and B)**. *FLM* showed a similar number and distribution of polymorphic sites when compared to 54 other type II MADS box genes of the same phylogenetical clade (Martinez-Castilla and Alvarez-Buylla, 2003) **(Fig. 22B and C)**. In cooperation with Thomas Nussbaumer (Helmholtz Zentrum, München, Germany), a haplotype analysis of the 7 kb *FLM* genomic region using the sequence between bp -2,200 to +4,800 (relative to the A of the ATG start codon) was conducted. A total of 45 promoter and intronic SNPs with a minor allele frequency (MAF) ≥ 5% derived from a set of 776 accessions was integrated **(Fig. 22D)**.
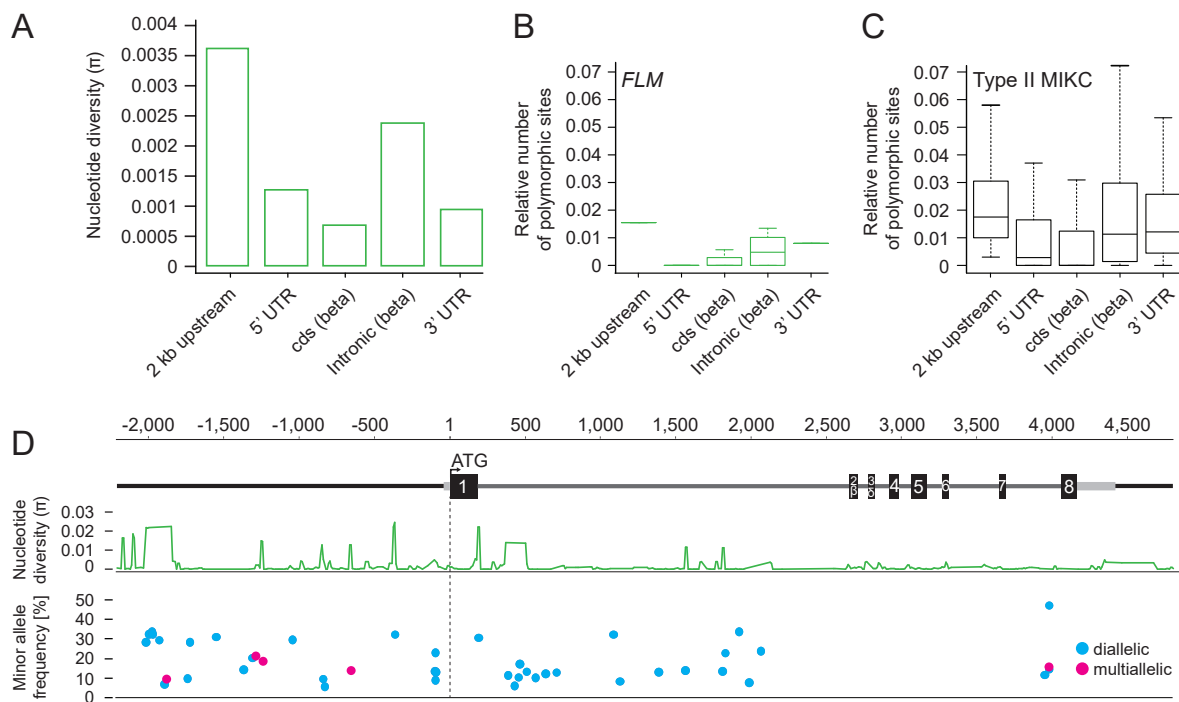
FIGURE 22: The *FLM* locus shows a similar nucleotide diversity when compared to genes of the same phylogenetic clade. A. Nucleotide diversity of elements of the *FLM* locus among 850 accessions. **B**. and **C**. The relative number of polymorphic sites of the *FLM* locus by element type (B) and the combined relative number of polymorphic sites of 54 MADS-box type II genes among 776 accessions (C) (Martinez-Castilla and Alvarez-Buylla, 2003). **D**. Representation of the position and the frequency of 45 SNPs from 776 accessions, which were used for haplotype clustering.

Then, the attention was focused on the ten most frequent haplotypes H1 - H10, which represented 378 (49%) accessions **(FIG. 23A)**. Most of the accessions belonging to these groups showed a broad geographic distribution, however, two haplotypes were rather specific for the Iberian Peninsula (H5) or the coastline of Italy and Greece (H10) **(FIG. 23B)**.
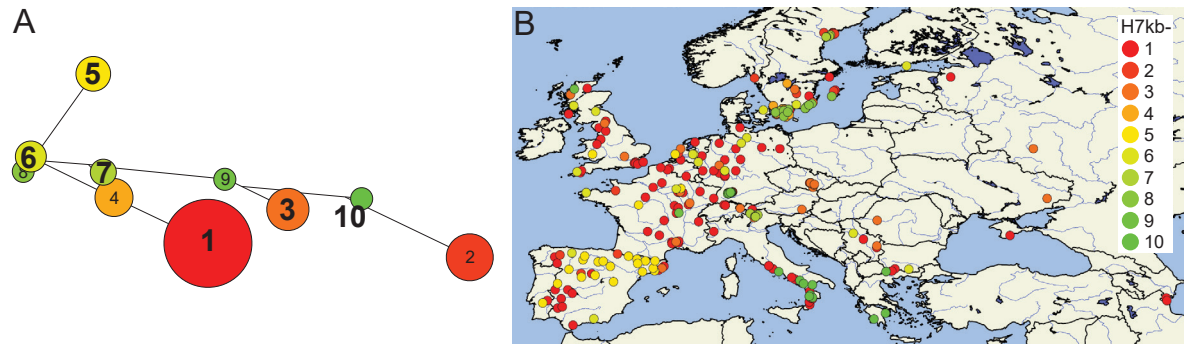
FIGURE 23: Genetic and geographic distribution of the major *FLM* haplotypes. A. Haplotype network analysis of the ten largest haplotype groups of the 7 kb *FLM* genomic region based on 45 SNPs. Circle size illustrates number of accessions in each group. Numbers indicate the corresponding haplotype group numbers. Bolt numbers indicate groups from which representative accessions were chosen. **B.** Geographic distribution of the accessions of the ten largest haplotype groups of the 7 kb *FLM* genomic region.

41 representative accessions were selected from six out of the ten groups (H7kb-1, 3, 5-7, 10). These selections were performed randomly except the only criterion of availability of the seed stock at he European Arabidopsis Stock Centre (NASC). Furthermore, the hypothesis that intron 1 carries essential elements for the basic transcriptional expression of *FLM-ß* and *FLM-δ* whereas the remaining intronic sequence (intron 2 - 6) rather influences the temperature-sensitive regulation and splicing was followed up. To this end, the set of 41 accessions was complemented with further eleven accessions that share the same haplotype when only considered intron 1 (named HI1) but differ in the haplotype of intron 2 to 6 (named HI2-6) **(Fɪɢ. 24A - D)**.
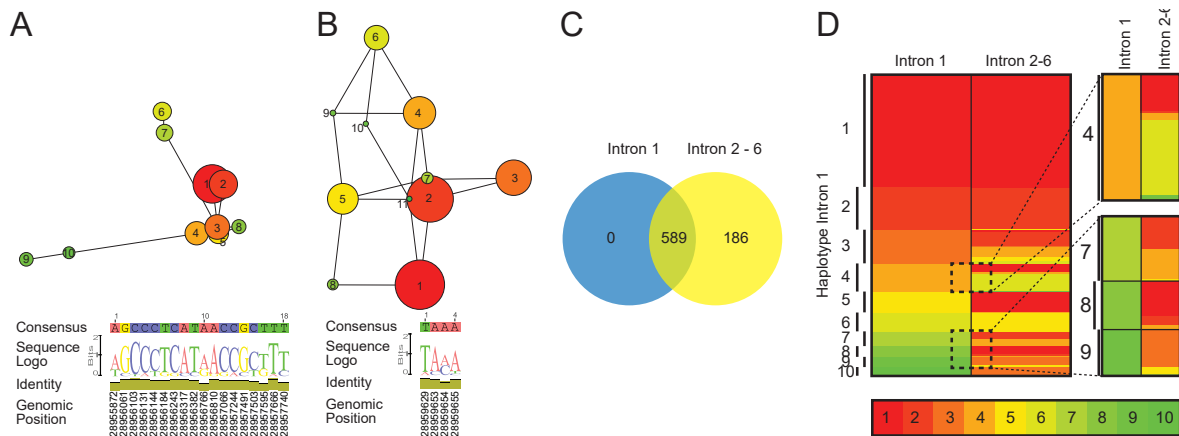
FIGURE 24: Identification of accessions with identical haplotype of intron 1 but different haplotype of introns 2 - 6. A. and **B.** Haplotype network of the ten and eleven largest haplotype groups when SNPs of the intron 1 (A) and the introns 2 - 6 (B) were considered in isolation. Numbering and color according to the group size. The sequence logo below indicates the alleles, the frequency and the genomic position of the included SNPs. **C**. Venn diagram indicating the overlap of accessions that are included in both the ten largest groups of intron 1 and intron 2 - 6 haplotype groups as shown in A and B. **D**. Clustering of the haplotype groups of the 589 accessions that are included in both the ten largest groups of intron 1 and intron 2 - 6 haplotype groups. Accessions were sorted by the haplotype group number of intron 1 and the according haplotype group number of intron 2 - 6 is shown on the right. An enlargement of a subset of accessions is shown on the right. The color coding for haplotype group number is shown at the lower panel of the graph.

The common reference strain Col-0 possessed an H7kb-3 haplotype. The accession Kil-0, which was described in detail in the first part of the present thesis, showed an H7kb-1 haplotype, when the insertion in the first intron was ignored. These two accessions were also included **(Fɪɢ. 25A)**. These selections of accessions were based on a haplotype analysis, which integrated the previously described 45 SNPs. To cover a larger range of diversity, additional low frequent SNPs (MAF > 1%) and indels (≤ 2 bp, MAF > 1%), which are polymorphic among the 54 accessions - apart from the already mentioned 45 SNPs - were included. This information was retrieved from the 1001 Genomes GEBrowser resource (http://signal.salk.edu/atg1001/3.0/gebrowser.php). Hence, a set of 119 polymorphic sites along the 7 kb *FLM* region was obtained **(Fɪɢ. 25B)**. Based on these 119 sites, a phylogenetic tree of the 54 accessions was calculated. When compared to the grouping that considered only 45 SNP sites, it was found that, except for the single accession Alst-1, the integration of the additional SNPs and indels did not change the phylogenic arrangement among the accessions **(Fɪɢ. 25C)**.
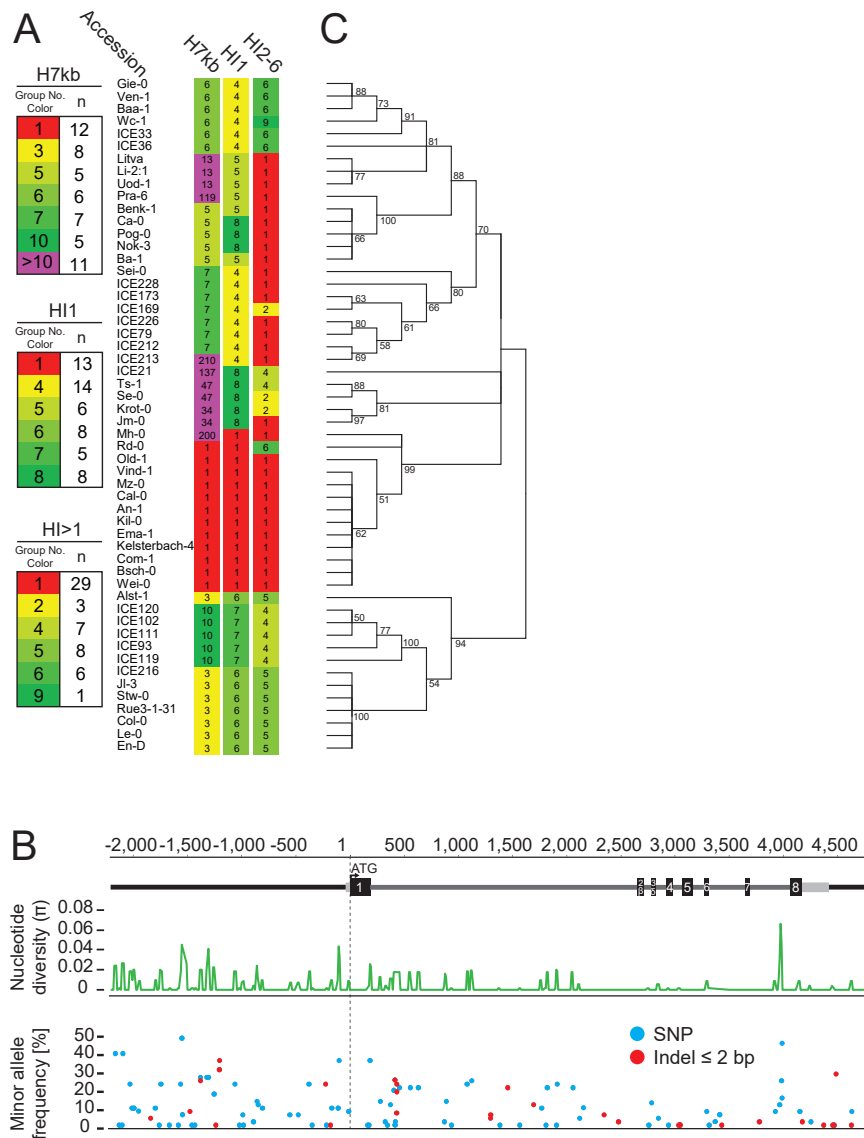
FIGURE 25: The accessions of the *FLM* haplotype set cluster similar when 45 SNPs are compared to 119 polymorphic sites. A. Haplotype group affiliation of the 54 accessions of the *FLM* haplotype set based on 45 SNPs for either the 7 kb *FLM* locus (H7kb), only intron 1 (HI1) or only the introns 2 - 6 (HI2-6). Group number coloring according to group size. **B**. Nucleotide diversity along the *FLM* locus among the 54 accessions of the experimental *FLM* haplotype set. Below the position, the frequency, and the type of the 119 polymorphic sites, which were used for simple single locus association tests, are shown. **C**. Neighbour-Joining tree showing the genetic relationship among accessions of the *FLM* haplotype set based on 119 polymorphic sites (SNPs and indels) along the 7 kb *FLM* genomic locus. Bootstrap values are indicated at each branch.

2.2.3. *FLM* expression may be explained by variation in the non-coding region

The experience of long periods of cold temperature (winter) by the vernalization pathway and the genes *FRI* and *FLC* is a major environmental stimulus to induce flowering (Song et al., 2013). The vernalization requirement differs between *A. thaliana* accessions due to differences in their *FRI* and *FLC* genotype (Lempe et al., 2005; Coustham et al., 2012).

To find polymorphisms regulating *FLM* expression and splicing, a genotype-to-phenotype association analysis should be performed. As the vernalization pathway strongly delayed flowering in 24 of the 54 accessions and consequently suppressed *FLM* effects, not flowering time but *FLM* expression was used as a phenotype for the association analysis **(FIG. 26A)**. To ascertain that *FLM* expression was not affected by *FLC*, *FLM* was examined in non-vernalized wild type Col-0 and *flc-3* mutants as well as in Col-0 carrying a functional vernalization module (*FRI*^Sf-2^*FLC*) (Michaels et al., 1999). Concurrent with previous reports, no influence of *FLC* on *FLM* transcript abundance was detected (Scortecci et al., 2001; Ratcliffe et al., 2001) **(FIG. 26B)**.
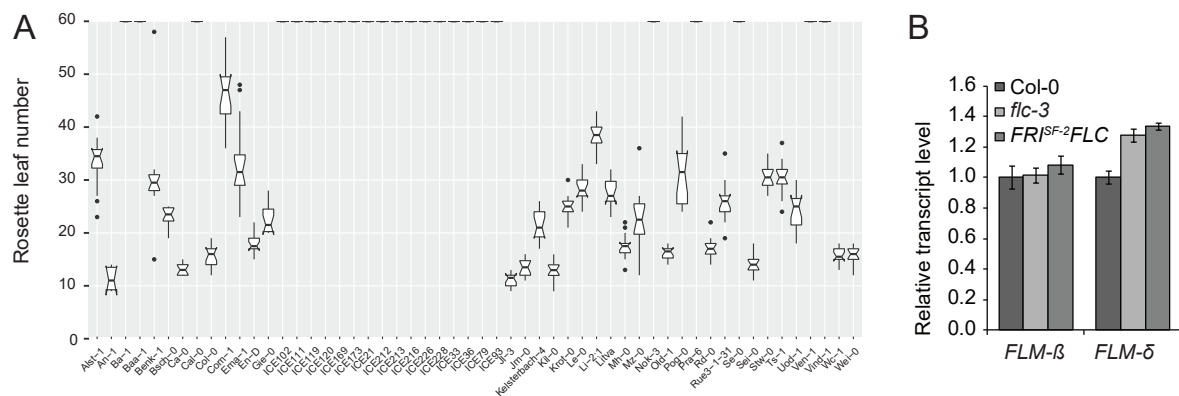
FIGURE 26: Accessions of the *FLM* haplotype set show variation of flowering time. A. Flowering time of the accessions of the *FLM* haplotype set. Accessions that flowered extremely late or even did not flower at all are indicated with 60 rosette leaves. Outliers were determined by 1.5 x interquartile range (IQR). **B.** qRT-PCR analysis of *FLM-ß* and *FLM-δ* expression in twelve days old Col-0, *flc-3* and *FRI*^sf-2^*FLC* plants. Error bars indicate s.d. of three biological replicates.

As described, the expression of the *FLM-ß* and *FLM-δ* splicing variants was used as trait instead of flowering time to investigate putative functional consequences of *FLM* sequence variation. To this end, *FLM* Exon 1 (total *FLM*), *FLM-ß*, and *FLM-δ* expression was determined from plants that had been grown for seven days at 21°C before being transferred for three days to 15°C or 23°C. Additionally, the ratios between *FLM-ß* and *FLM-δ* and between the 15°C and 23°C temperature treatments were calculated **(FIG. 27A - G)**. Concurrent with previous results, strong decreases in *FLM* Exon 1, *FLM-ß* and *FLM-δ* levels of Kil-0 were observed when compared to the Col-0 reference suggesting that this large-scale expression analysis provided biologically meaningful outputs. When evaluated over the entire set, substantial variation in *FLM-ß* and *FLM-δ* levels was observed between the accessions **(FIG. 27H)**. However, the previous observation made in Col-0 and Kil-0 that *FLM-ß* and *FLM-δ* expression decrease and increase, respectively, with increasing temperature was largely maintained **(FIG. 27H)**. At 23°C and 15°C, *FLM* Exon 1 levels strongly correlated with *FLM-ß* but not with *FLM-δ* suggesting that *FLM-ß* is much more abundant than *FLM-δ* in these conditions **(FIG. 27I AND J)**.
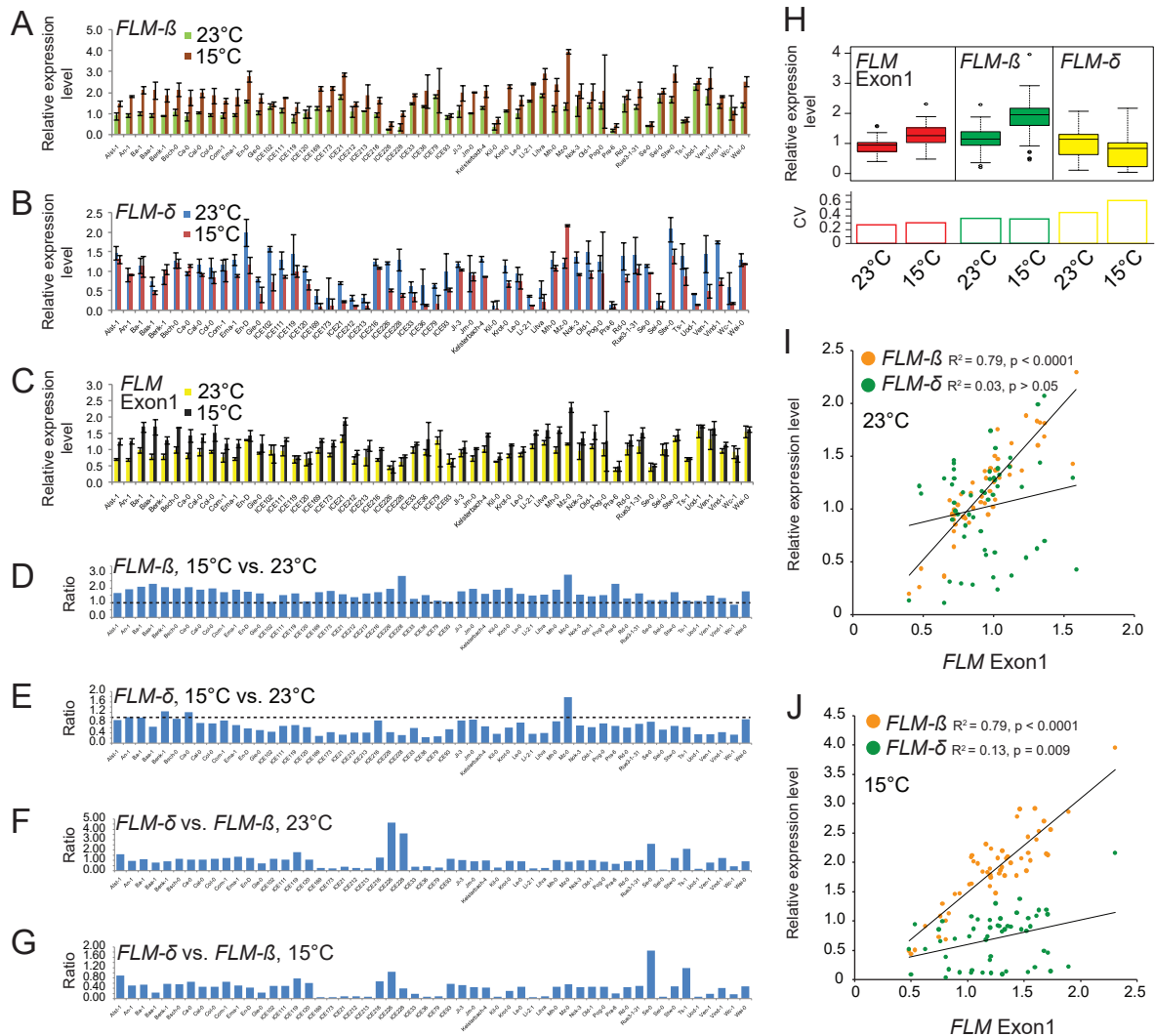
FIGURE 27: Accessions of the *FLM* haplotype set show variation of *FLM* expression. A. - C. qRT-PCR analysis of *FLM-ß* (A), *FLM-δ* (B), and *FLM* Exon 1 (total *FLM*) (C) expression of the 54 accessions of the *FLM* haplotype set. **D. - G.** Relative changes of transcript levels shown in A - C of either temperature (D and E) or isoform (F and G) set relative to each other. **H.** Averaged values of the qRT-PCR analysis of *FLM* Exon1, *FLM-ß* and *FLM-δ* expression (A - C) in the 54 accessions of the *FLM* haplotype set. Outliers were determined based on 1.5 x IQR and are indicated as black dots. CV, coefficient of variation. **I.** and **J.** Correlation of *FLM-ß* and *FLM-δ* transcript levels to the levels of *FLM* Exon 1 at 23°C (H) and 15°C (I) of the 54 accessions of the *FLM* haplotype set.

2.2.4. Highly significant sites in the *FLM* non-coding region influence *FLM* transcript levels

To identify polymorphisms with a potential regulatory function for *FLM* expression, simple single locus association tests using the 119 mainly binary polymorphic sites of the *FLM* haplotype set while the *FLM* expression data was treated as single phenotypes (cooperation with Thomas Nussbaumer, Helmholtz Zentrum, München, Germany). 27 comparisons representing 11 polymorphic sites indicated a significant difference (P < 0.001); seven comparisons representing two sites showed high significance (P < 0.0001) **(FIG. 28A AND B)**.

The focus was put on the investigation of the potential role of a single base pair deletion (PRO1T/-; bp -215) and a genetically slightly linked SNP (PRO2A/C; bp -93) because they were both located in the proximal part of the *FLM* promoter. Further, three genetically unlinked nucleotides

were investigated because they were positioned as a highly diverse nucleotide triplet in intron 6 (INT6A/C-A/C-A/T/C; bp +3975 - +3977) in an otherwise conserved sequence context **(FIG. 29A AND B)**.
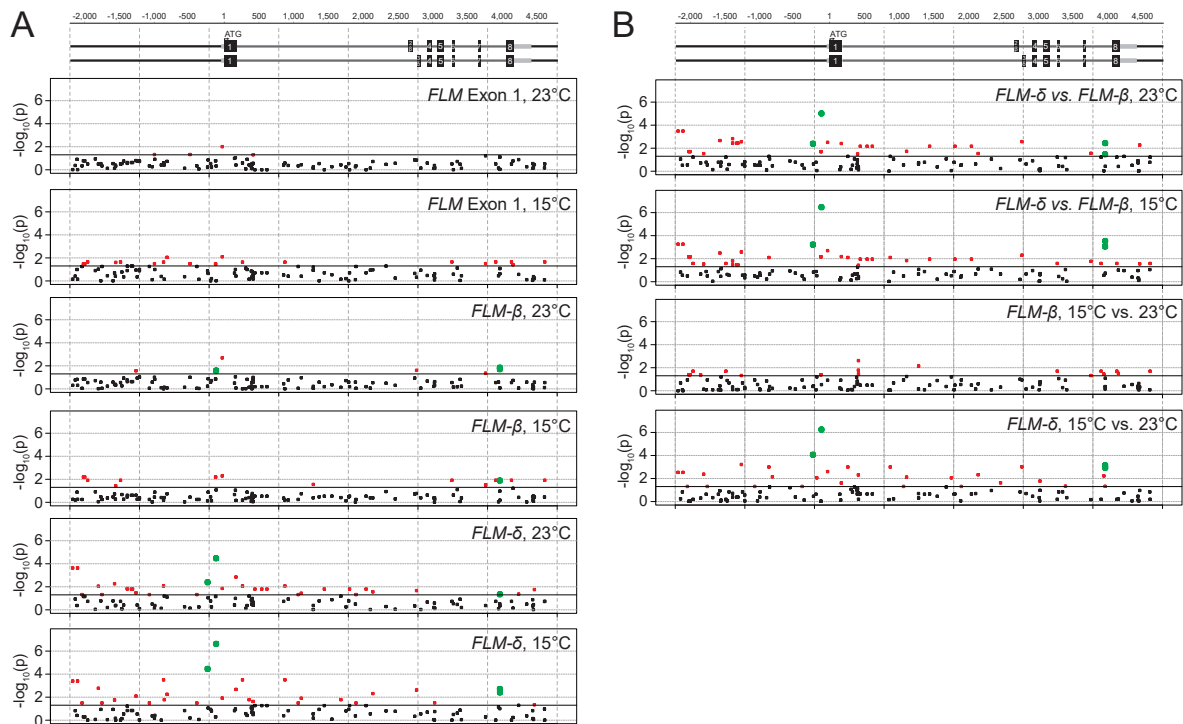
FIGURE 28: Identification of polymorphic sites that significantly associate with *FLM* abundance. A. and **B.** Representation of the –log(10) transformed P-values of the simple single locus association tests of each of the 119 sites that were included in the analysis. The black horizontal line shows the threshold of P = 0.05. The black dots represent polymorphic sites association with P > 0.05, the red dots P < 0.05 and the green dots represent sites with P < 0.05 that were additionally chosen for detailed experimental analysis as indicated in the text. The *FLM* genomic gene model is shown at the top. Black boxes indicate exons, grey boxes are 5′- or 3′-UTRs and dark-grey lines are introns. Each graph represents the analysis of one trait, as indicated in the top right corner of the graph while A shows all comparisons with relative expression values and B shows all comparisons of relative changes of transcript levels of *FLM-ß* and *FLM-δ* in response to change in temperature (15°C versus 23°C) or ratios between the transcript levels (*FLM-δ* versus *FLM-ß*) at 15°C and 23°C.
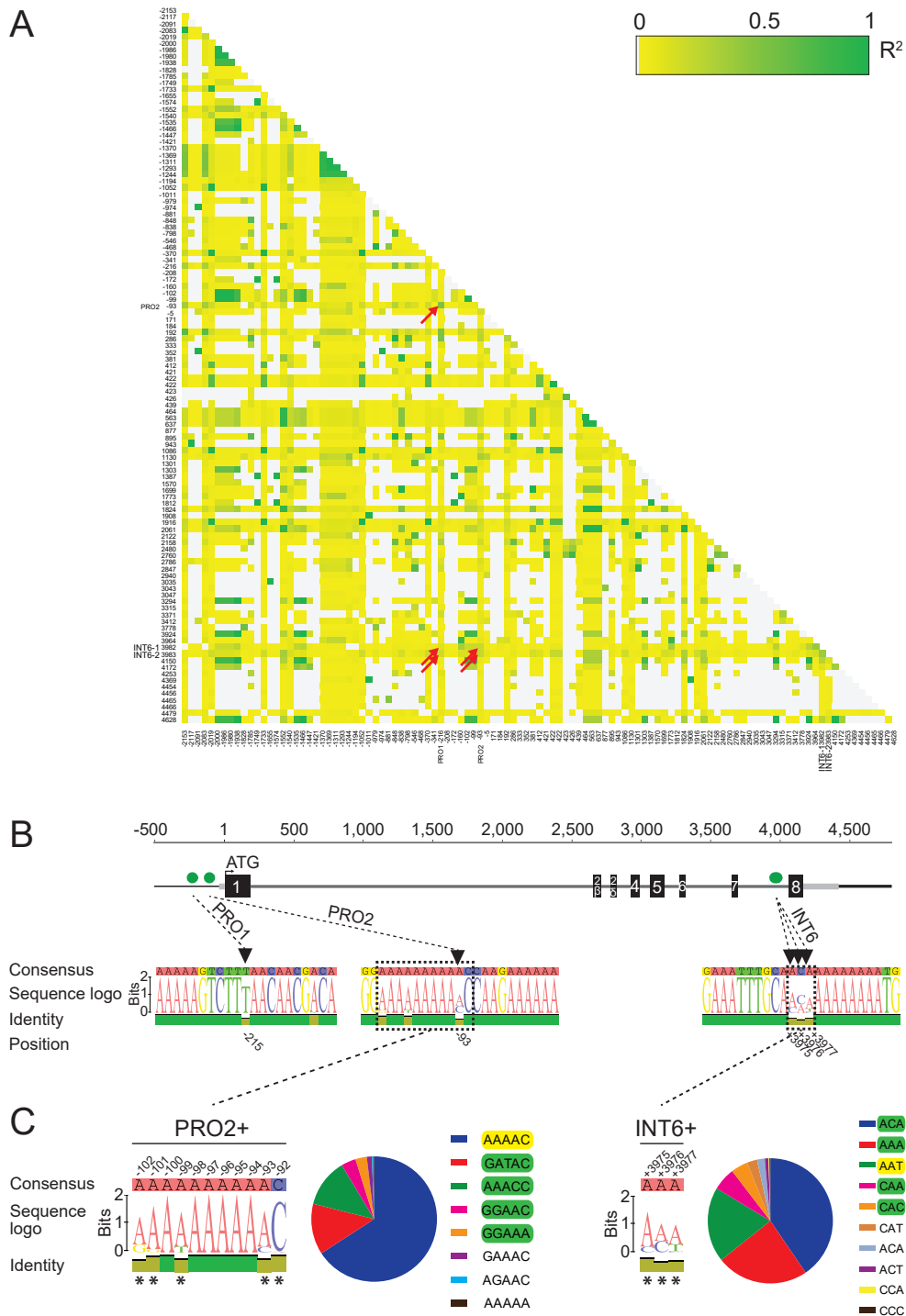
FIGURE 29: The regions PRO2+ and INT6+ show sequence variation. A. Pairwise linkage (R^2) between the sites, which were integrated in the simple single locus association test among the 54 accessions of the *FLM* haplotype set. Only biallelic SNPs are shown (109 of 119). Green indicates $R^2 = 1$ and white $R^2 = 0$ as indicated by the color code. Red arrows point to the sites PRO1, PRO2, and two SNPs of the nucleotide triplet INT6 (bp +3975 and +3976). **B.** Summarizing schematic representation of the sites that showed high association with changes in *FLM* transcript levels. The upper panel shows the *FLM* locus with the green dots indicating the position of the significant sites that were chosen for further investigation. The middle panel shows the sequence variation up- and downstream of these sites among 54 accessions of the *FLM* haplotype set. The position and the names of these sites (PRO1, PRO2, and INT6) are indicated with dotted arrows. The respective allele frequencies are indicated by the sequence logo. **C.** The sequence variation among the sequence set of 840 accessions and the regions PRO2+ and INT6+ are shown. All sites that were polymorphic among the 840 accessions are indicated with asterisks and the respective allele frequencies are indicated by the sequence logo. The pie charts show the frequency of the respective PRO2+ and INT6+ variants. The Col-0 reference variant is indicated in yellow, all variants that were chosen for further investigation are indicated in green.

When the significant effect sizes of their respective major and minor alleles were investigated, the deletion at PRO1 associated with the downregulation of *FLM-δ*. Contrarily, the A to C SNP at PRO2 associated with a downregulation of *FLM-δ* as well as a slight upregulation of *FLM-ß*. Finally, all three nucleotide changes of INT6 largely associated with the downregulation of *FLM-ß* and the upregulation of *FLM-δ* (**FIG. 30**).
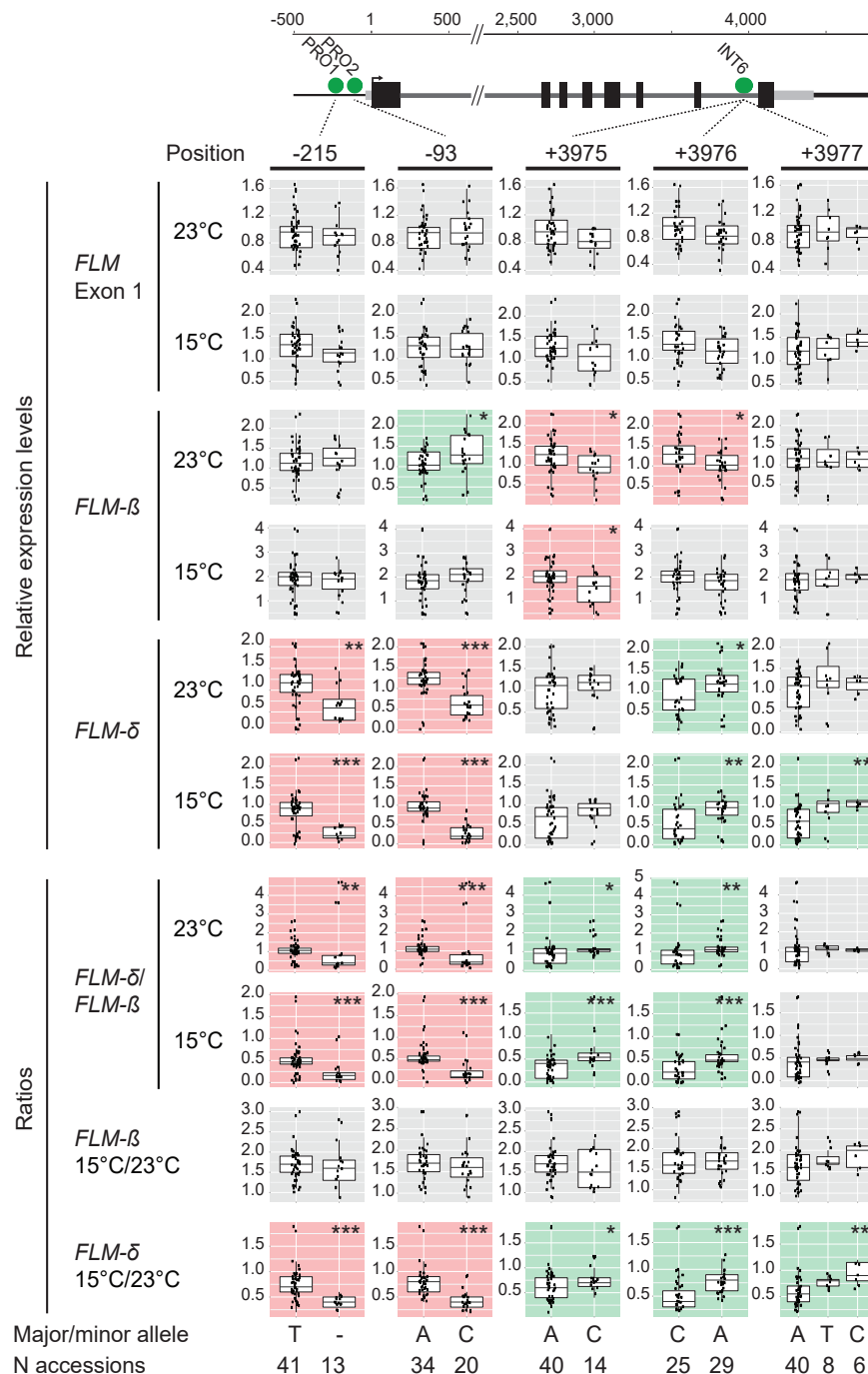
FIGURE 30: The effects of PRO1, PRO2, and INT6 polymorphisms on *FLM* expression. Effects of the PRO1, PRO2 and INT6 alleles on relative expression levels and ratios among the 54 accessions of the *FLM* haplotype set. The background color indicates that the minor allele associates with significant upregulation (green) and downregulation (red), respectively. P-values of the Kruskal-Wallis tests are indicated at the top right of each graph (*, p < 0.05; **, p ≤ 0.01; ***, = p < 0.001). Associations with P > 0.05 are indicated with a grey background. Single values are represented as jittered dots.

2.2.5. Small regions around the significant sites in the promoter and in intron 6 show multiple haplotypes

Before proceeding to verify these associations, sequence variation at the candidate sites in a set of around 840 accessions was reinvestigated in more detail. The region surrounding PRO1 only showed a few additional polymorphic sites that were, however, all lowly frequent (MAF < 1%). Hence, the focus remained on the T deletion. Few bases up- and downstream of PRO2, four additional polymorphic sites were identified with high nucleotide diversity **(FIG. 29C)**. This 10 bp region comprised five polymorphic sites (bp -102 to bp -92) and was named PRO2+. The region around INT6 was highly conserved but the nucleotide triplet region itself showed additional haplotypes and was designated INT6+ **(FIG. 29C)**.

2.2.6. Transgenic experiments verify putative regulatory regions of *FLM*

Next, to understand whether each of the candidate sites is causative for changes in *FLM-ß* and *FLM-δ* transcript levels and their temperature sensitive regulation, a transgenic analysis was performed in a homogeneous genetic background. This analysis was conducted with variants of a Col-0 *FLM* genomic reference construct. Nd-1, an accession with a full deletion of the *FLM* locus and no residual *FLM* transcript was used as genetic background **(FIG. 31A AND B)** (Werner et al., 2005b).
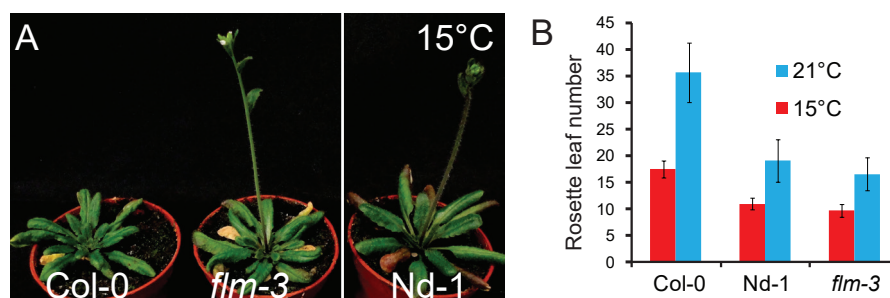
FIGURE 31: Nd-1 is an early flowering *A. thaliana* accession due to a full-length deletion of *FLM*. A. Representative photographs of 42 days-old plants of Col-0, *flm-3*, and Nd-1 grown at 15°C in long day photoperiod. The plant images were spliced together but originate from the same photograph as indicated by a white vertical line. **B.** Quantitative flowering time analysis of Col-0, *flm-3*, and Nd-1 grown at 15°C and 21°C in long day photoperiod.

Based on the *FLM^{Col-0}* reference with the alleles PRO1^T, PRO2+^{AAAAC}, and INT6+^{AAT}, mutations were introduced at PRO1 (PRO1^{ΔT}), PRO2+ (PRO2+^{GATAC/AAACC/GGAAC/GGAAA}) and INT6+ (INT6+^{ACA/AAA/ CAA/CAC}) **(FIG. 32A)**. In addition, transgenic constructs with a 16 bp or 17 bp deletion spanning the regions of either PRO2+ (PRO2+^{Δ16bp}) or INT6+ (INT6+^{Δ17bp}), respectively, were generated **(FIG. 32B)**. Interestingly, PRO2+ contained a PolyA motif and INT6+ was flanked by a PolyA motif ($[A]_{8-10}$). Only four more of these motifs on the sense-strand of the 7 kb locus were found, two of them close to the PRO2+ and INT6+ regions. Since PolyA motifs were found to have functional importance in gene regulation (Suter et al., 2000; Wijnker et al., 2013; O'Malley et al., 2016), the potential causality of these motifs should be investigated by generating a deletion series until a construct that lacked all of the six sense-strand $[A]_{7-11}$ stretches (PolyA_2xΔa, PolyA_2xΔb, PolyA_3xΔ, PolyA_4xΔ, PolyA_5xΔ, PolyA_6xΔ) **(FIG. 32B)**.
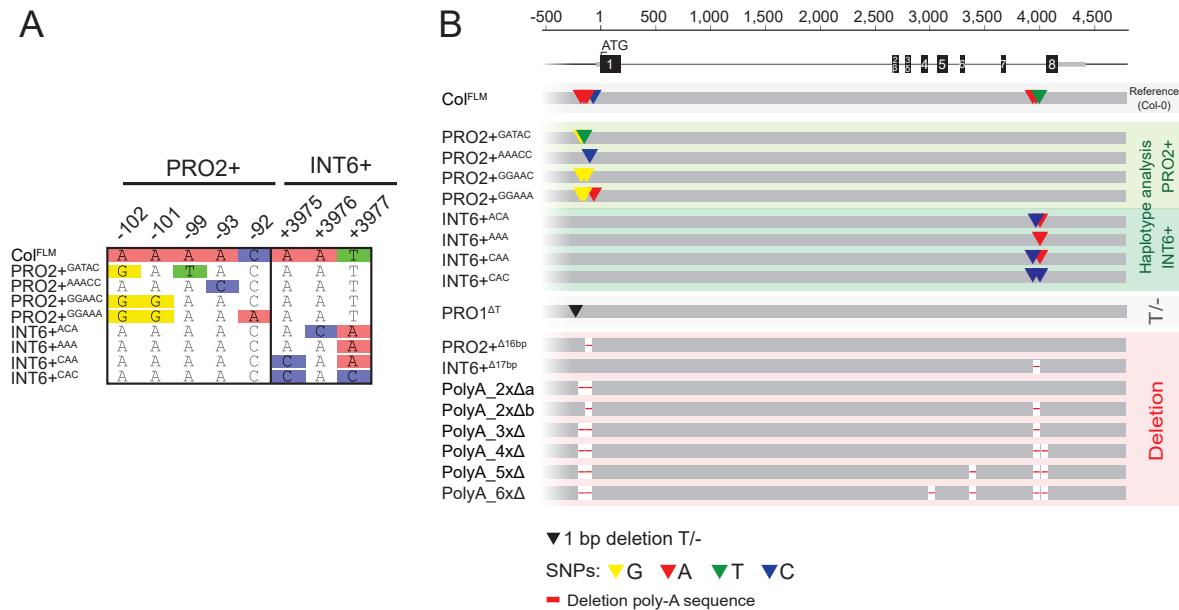
FIGURE 32: Schematic representation of *FLM* variants generated for transgenic experiments. A. and **B.** Alignment of the PRO2+ and INT6+ polymorphisms (A) and schematic representation of the Col^FLM reference construct (pFLM::gFLM) as well as the pFLM::gFLM variants selected for transgenic analysis in the *FLM* deletion accession Nd-1 (B). Bases differing between Col^FLM and the other variants are coloured. Deletions in the deletion constructs of the PRO2+ and INT6+ sites as well as the PolyA motifs are displayed with red lines (not drawn to scale).

Next, the effects of the respective variants on *FLM* expression were analyzed using segregating lines of the T2 generation. To normalize for variability between these transgenic lines, T2 segregating lines were pooled (18 to 45 independent lines per construct, average 30). Plants were grown for seven days at 21°C and then shifted to 15°C and 23°C for another three days to perform qRT-PCR measurements to determine *FLM-ß* and *FLM-δ* abundance. This pooling strategy was validated by demonstrating that the established behavior of *FLM* in the Col-0 and Kil-0 accessions could be faithfully recapitulated when performing equivalent analyses with T2 lines expressing gFLM^Col-0 and gFLM^Col-0 bearing the Kil-0 LINE insertion (gFLM^Col-0+LINE). **(Fig. 33A)**.

Analysis of the PRO1^T mutant variant did not yield differences in *FLM-ß* and *FLM-δ* transcript levels **(Fig. 33D and E)**. Several PRO2+ and INT6+ variants showed increased *FLM-ß* and *FLM-δ* levels (PRO2+^AAACC, PRO2+^GGAAC, INT6+^ACA, INT6+^CAC) or decreased *FLM-ß* and *FLM-δ* levels (PRO2+^GATAC, INT6+^CAA) mostly irrespective of temperature **(Fig. 33B and C)**. Interestingly, the variant INT6+^AAA showed very strong upregulation of *FLM-ß* but no change in *FLM-δ*. The variants PRO2+^Δ16bp and INT6+^Δ17bp did not show significant changes in *FLM-ß* and *FLM-δ* levels. Further, an increasing number of PolyA motif deletions decreased *FLM* levels only slightly **(Fig. 33D and E)**. Nevertheless, the variant with six deletions (PolyA_6xΔ), the highest number generated, showed a strong and significant reduction in *FLM-ß*, but not in *FLM-δ* **(Fig. 33B and C)**. When the thermosensitive regulation of *FLM-ß* and *FLM-δ* was calculated (15°C versus 23°C) and compared to Col^FLM, PRO2+^GGAAA showed a hypersensitive response of *FLM-ß* and a hyposensitive response of *FLM-δ*. Contrarily, INT6+^CAA showed a hyposensitive response to a change in temperature of *FLM-ß* and a hypersensitive response of *FLM-δ*. PolyA_6xΔ showed hyposensitivity of *FLM-ß*, but no changes of *FLM-δ* **(Fig. 33B and C)**.
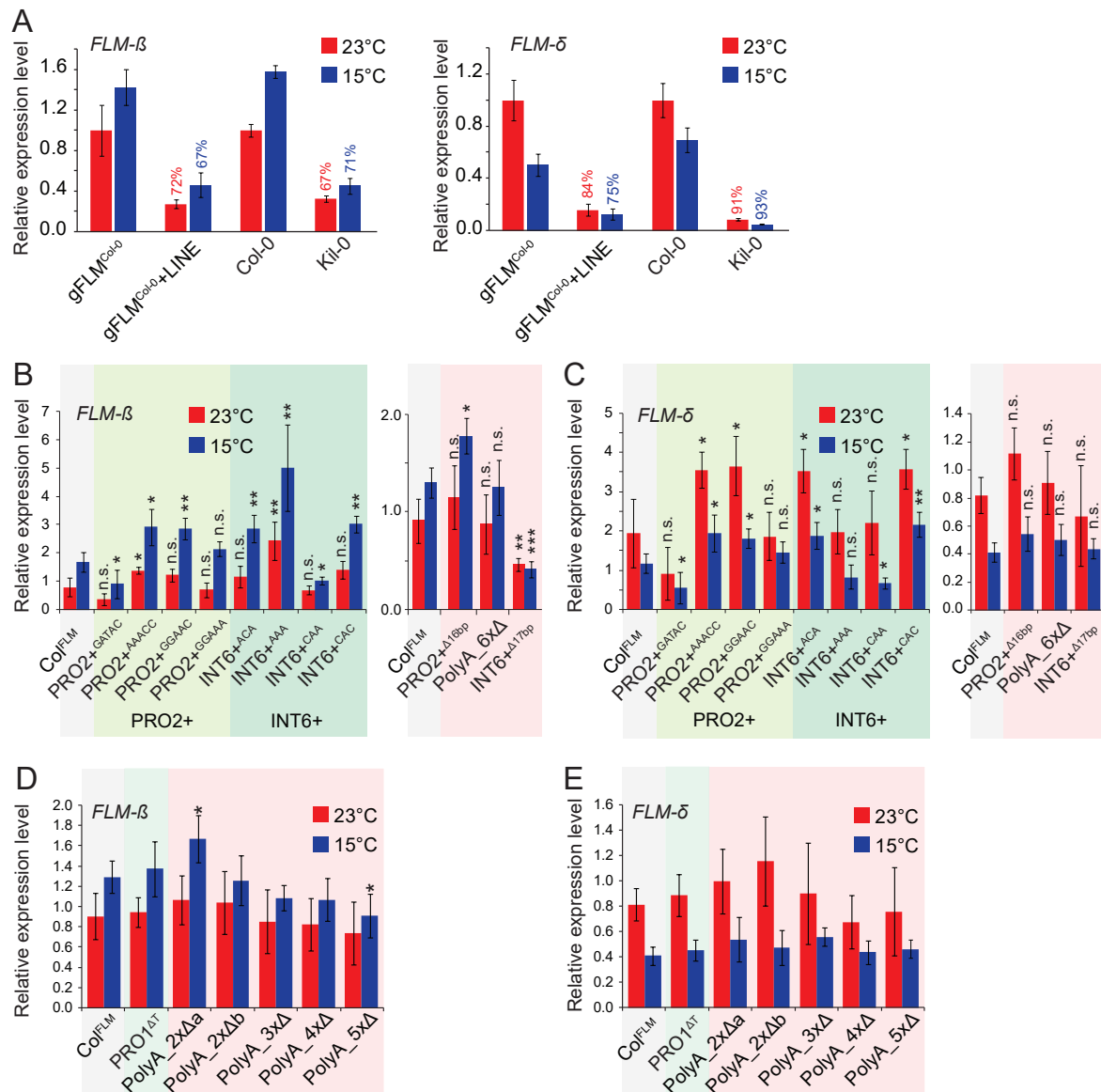
FIGURE 33: The PRO2+, INT6+, and polyA regions are involved in regulation of *FLM* expression and temperature-sensitive regulation of *FLM*. A. qRT-PCR analyses of *FLM-ß* and *FLM-δ* expression at 15°C and 23°C in ten days old *flm-3* mutants complemented with a gFLM^Col-0 or gFLM^Col-0+LINE allele. Pools of 35 - 40 independent T2 transgenic lines were used in comparison to Col-0 and Kil-0 homozygous lines. Shown is the mean and s.d. of three (Col-0, Kil-0) and four replicate pools compromising each eight to ten transgenic lines. Percentages refer to the reduction in *FLM* expression between the gFLM^Col-0+LINE transgenic line in comparison to the gFLM^Col-0 line or between Kil-0 and Col-0, respectively. **B - E.** qRT-PCR analysis of *FLM-ß* (A and C) and *FLM-δ* (B and D) expression in ten days old T2 transformant lines. A pool of 18 to 45 lines was used for each construct. Plants were grown at 21°C for seven days and then grown for further three days at the specified temperature. Error bars indicate s.d. of four replicate pools that compromise each of four to eleven lines. Student's t-tests were performed as indicated: *, p < 0.05; **, p ≤ 0.01; ***, p < 0.001; n.s., not significant.

2.2.7. Phylogenetic footprinting pinpoints essential regions in the promoter and the first intron for *FLM* expression

Phylogenetic footprinting has been successfully used to identify regulatory regions in several MADS-box transcription factor genes (Schauer et al., 2009; Van de Velde et al., 2014; Hong et al., 2003). In cooperation with Manuel Spannagl (Helmholtz Zentrum, München, Germany), putative *FLM* orthologues were selected from five more *Brassicaceae* species. Multiple sequence alignments were calculated together with *A. thaliana FLM* and the homolog gene *MAF3* (Martinez-Castilla and Alvarez-Buylla, 2003). A small region in the promoter and two regions in the first intron with increased sequence conservation were identified **(Fɪɢ. 34A ᴀɴᴅ B)**. To test whether the increased sequence conservation correlated with a regulatory role of these regions, transgenic lines expressing Col-0 variants of *FLM* with a deletion of these three sites were analyzed (PRO(Δ225bp), INT1(Δ373bp), INT1(Δ101bp)) **(Fɪɢ. 34C)**. When *FLM* expression was tested in the same way as described before, it was found that the PRO(Δ225bp) and INT1(Δ373bp) variants did not show substantial *FLM-ß* and *FLM-δ* expression **(Fɪɢ. 34D)**. Therefore, the PRO(Δ225bp) region in the promoter as well as the INT1(Δ373bp) region may be fundamentally important for *FLM* expression. Since the PRO(Δ225bp) region may be essential for *FLM* expression, it was tested whether this region can sufficiently drive *FLM* expression alone by generating a 254 bp minimal-promoter construct PRO(254bp) **(Fɪɢ. 34E)**. This construct also showed strongly reduced *FLM-ß* and *FLM-δ* expression **(Fɪɢ. 34F)**. Hence, the PRO(254bp) region may be essential but not sufficient to drive *FLM* expression and regulation.
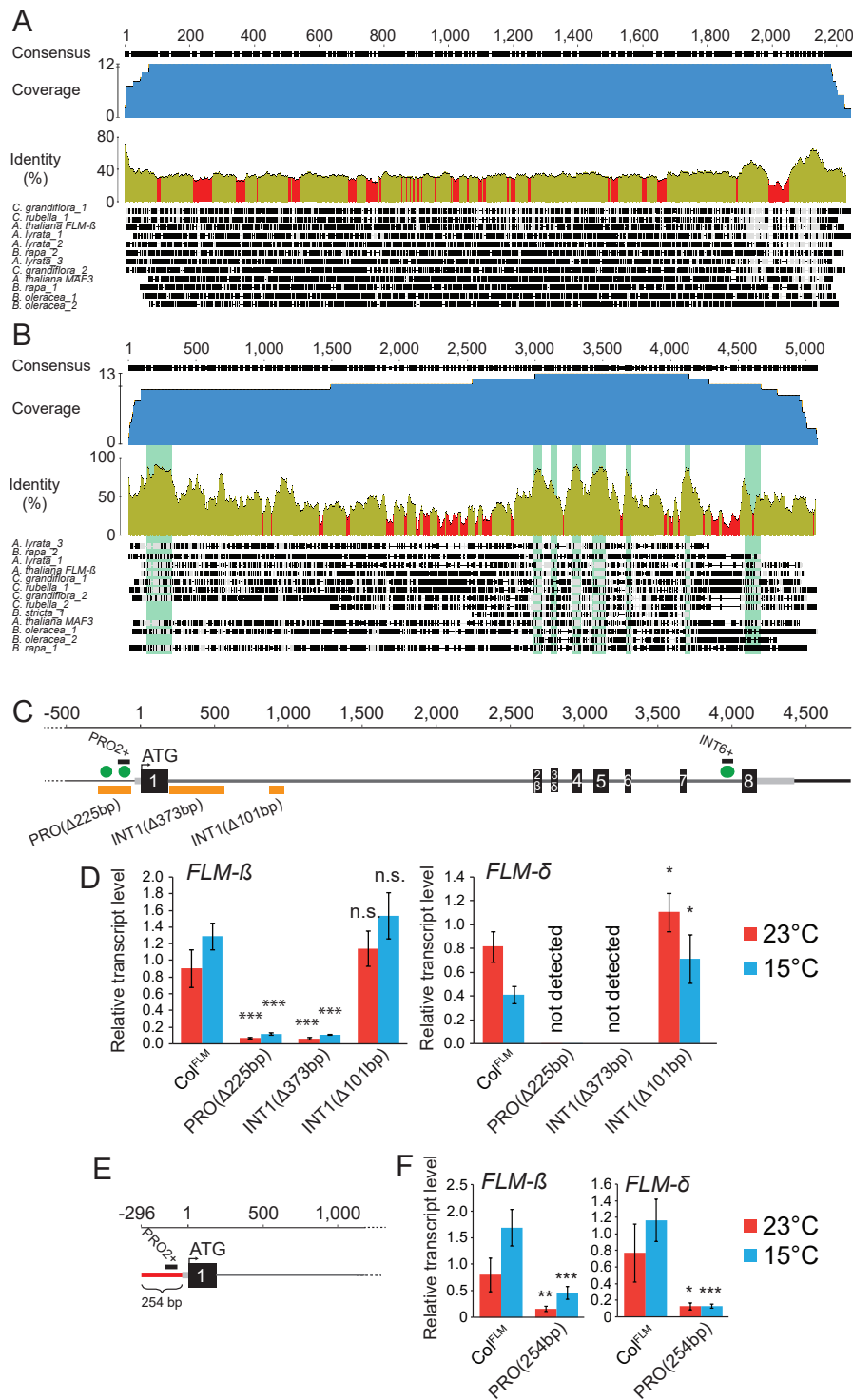
FIGURE 34: Conserved regions among putative *FLM* orthologs are essential for *FLM* expression. A. and **B.** Sequence alignments of promoter (A) and gene sequences (B) of *FLM* and *MAF3* as well as putative *FLM* orthologs from six *Brassicaceae* species. Coverage is shown in blue, identities are based on a sliding window word size = 30 and displayed in ocre (≥ 30%) and red (< 30%). *FLM* exons are shown in green (B). Numbering is indicated according to the position in the alignment. **C.** Schematic illustration of the *FLM* genomic region and transgenic variants used for expression analysis. **D**. qRT-PCR analysis of *FLM-ß* and *FLM-δ* expression in ten days old T2 transformant lines. A pool of 21 to 40 lines was used for each construct. Error bars indicate s.d. of four replicate pools comprising each five to ten lines. **E.** Schematic illustration of the *FLM* minimal promoter construct. The 254 bp upstream promoter region is shown as red line. **F.** qRT-PCR analysis of *FLM-ß* and *FLM-δ* expression in ten days old T2 transformant lines. A pool of 34 and 40 lines was used for each construct. Error bars indicate s.d. of four replicate pools comprising each eight to ten lines. Student's t-tests were performed as indicated: *, $p < 0.05$; **, $p ≤ 0.01$; ***, $p < 0.001$; n.s., not significant.

2.2.8. Changes in the PRO2+ and INT6+ regions modulate flowering time in a linear manner via *FLM-ß*

To correlate *FLM* transcript levels with flowering time, flowering time analysis with eight lines that showed significantly changed *FLM* levels was performed. Since the previous analyses had revealed a particularly prominent role for *FLM* in ecologically more relevant cooler ambient temperatures, this analysis was performed at 15°C. To precisely correlate flowering time with *FLM* transcript levels, additional qRT-PCR expression analyses were performed with this specific set of lines **(Fɪɢ. 35A ᴀɴᴅ B)**. Since this expression analysis was performed with T2 segregating transgenic lines, the same material was used to determine the flowering time phenotype. These T2 lines also include 25% wild type Nd-1 segregants. Therefore, the data including all segregants as well as the data after exclusion of 25% early flowering plants is presented **(Fɪɢ. 35C - E)**. A very strong correlation between *FLM-ß* transcript levels and flowering time was found, while this correlation was lower for *FLM-δ* **(Fɪɢ. 35F ᴀɴᴅ G)**. Furthermore, among these transgenic lines, flowering time responded essentially in a linear manner to changes in *FLM-ß* levels. In the range between 0.1 and 2.5 fold-change, increases by fold-change of 1 resulted in a flowering time delay of 7.9 rosette leaves **(Fɪɢ. 35F)**. Interestingly, compared to the reference, the variant PolyA_6xΔ showed an around threefold downregulation of *FLM-ß* at 15°C but flowered similarly early like the *FLM*-deletion accession Nd-1. This suggested that there may be a critical lower threshold of *FLM-ß* to act as a repressor.

In conclusion, *FLM-ß* may be the major determinant of flowering time at 15°C in this homogenous genetic background. No estimable effects of *FLM-δ* were detected in 23°C and 15°C and it can be supposed that *FLM-δ* has only minor physiological importance for flowering time regulation.
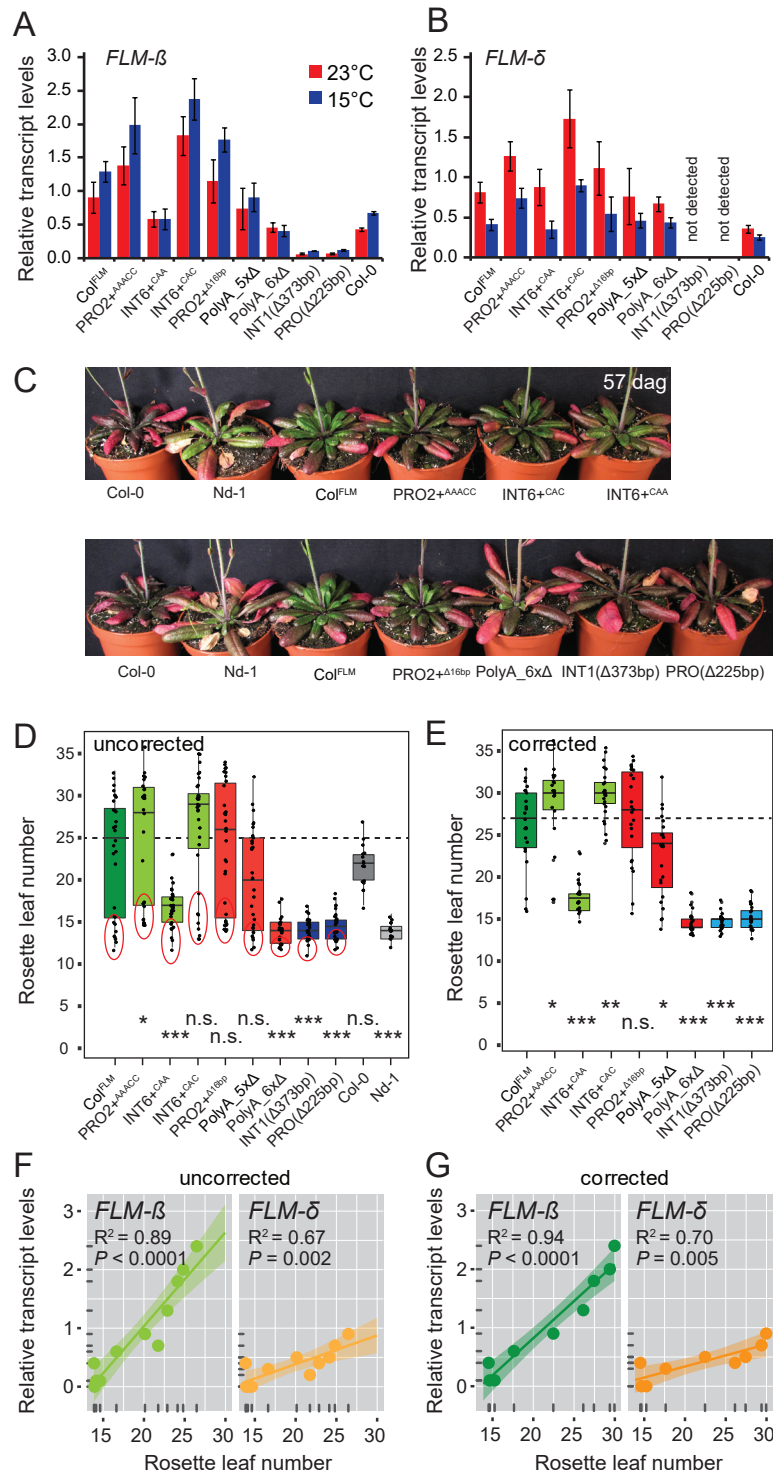
FIGURE 35: *FLM-ß* expression levels of transgenic lines largely explain flowering time. A. and **B**. qRT-PCR analysis of *FLM-ß* (A) and *FLM-δ* (B) expression in ten days old T2 lines and the homozygous line Col-0. A pool of 18 to 40 lines was used for each construct. Error bars indicate s.d. of four replicate pools that compromise each of four to ten lines and three biological replicates for Col-0. **C.** Photographs of representative 57 days old plants expressing Col^{FLM} and PRO2+ and INT6+ variants grown at 15°C in long day photoperiod. The reference wild type lines Col-0, and Nd-1, which was used as genetic background for the transgenic variants, are also shown for comparison. **D.** and **E.** Quantitative flowering time analysis of T2 transformant lines grown at 15°C under long day photoperiod. From each construct three replicate pools with each ten plants were analyzed. The 25% of the plants showing a very early flowering most likely represent the wild type Nd-1 segregants and are indicated with a red circle in D. E represents the data with exclusion of 25% early flowering plants. Single values are represented as jittered dots. Coloring was chosen according to the type of variant. The average value of the reference Col^{FLM} line is indicated as dotted line. Wilcoxon rank tests were performed as indicated: *, p < 0.05; **, p ≤ 0.01; ***, p < 0.001; n.s., not significant. **F.** and **G**. Correlation of *FLM-ß* and *FLM-δ* transcript levels to flowering time with the corrected (F) and the uncorrected (G) flowering time data. The corresponding qRT-PCR analysis of *FLM-ß* and *FLM-δ* expression at 15°C is shown in B. and C. The shaded areas indicate the 95% confidence intervals.

3. Discussion

Ambient temperature during spring is a major cue determining flowering time. Cool temperatures generally delay and warm temperatures promote flowering of *A. thaliana*. The *FLM* locus explains natural flowering time variation in different ambient temperatures but the underlying genetic basis of *FLM*-dependent flowering control remained, however, largely unclear (Werner et al., 2005b; Balasubramanian et al., 2006; el-Lithy et al., 2006; O'Neill et al., 2008; Mendez-Vigo et al., 2010; Salomé et al., 2011).

The present thesis describes the identification of a novel natural intron-insertion allele (*FLM^Kil-0^*) of *FLM*, which affects transcript abundance in an isoform-specific manner and overall accelerates flowering with a most prominent effect in cool temperature (15°C). Further, indications are described that this allele is involved in the adaptation of flowering time of *A. thaliana* (**Fig. 36**). In a follow-up project, genomic data, transcript expression data and phylogenetic footprinting were used to identify regulatory non-coding regions of *FLM*. These regions regulate *FLM* transcript abundance in a haplotype-specific manner and flowering time is highly correlated with the abundance of the splice variant *FLM-ß* (**Fig. 37**).

3.1. Modulation of temperature-dependent flowering by the *FLM^Kil-0^* allele

3.1.1. Whole-genome sequencing of Kil-0 uncovers a non-reference insertion

The genome sequencing reads of Kil-0 were mapped against the Col-0 reference sequence and a strong reduction of read coverage in the first intron of the candidate gene *FLM* of Kil-0 was indicative for an large insertion. By mapping of paired-end reads of a genome resequencing experiment to the Col-0 reference sequence (version TAIR10), this insertion was masked and was only denoted by a small region with very low coverage. *De novo* assembly then uncovered the large insertion event (**Fig. 8**). The invisibility of the insertion by mapping to the Col-0 reference sequence was in agreement with the observation that large parts of the Col-0 genome were not identified in other accessions and *vice versa* (Clark et al., 2007; Zeller et al., 2008; Cao et al., 2011; Schneeberger et al., 2011; Long et al., 2013). Quadrana et al. (2016) applied a targeted approach of *A. thaliana* resequencing data derived from the 1001 *A. thaliana* Genomes Project combined with sequence capture of transposable elements (TE). They elucidated the composition and activity of thousands of recent transposition events, collectively called the "mobilome". Still a lot of likewise large structural variation might wait to be uncovered, e.g. by extensive *de novo* assembly analysis.

3.1.2. The insertion in *FLM^Kil-0^* may act on *FLM* abundance in multiple ways

The inserted sequence in *FLM^Kil-0^* showed remote (68%) homology to a class I LINE retrotransposon family template sequence from *A. thaliana*. The activity of a transposable element (TE) is usually suppressed by epigenetic mechanisms (Lisch, 2009; McCue and Slotkin, 2012). These epigenetic mechanisms are able to negatively affect expression of closely localized genes as it was shown, e.g., for *FWA* (Kinoshita et al., 2004; Lippman et al., 2004). Theoretically, such a silencing mechanism could also explain the decreased *FLM* transcript abundance in Kil-0. However, the latter

hypothesis may not be the molecular explanation for the observed low levels of *FLM* transcript because different types of insertions at similar positions yielded similar results **(FIG. 19)**. A more likely scenario for the underlying basis for *FLM* downregulation than epigenetic silencing may be premature termination of transcription and aberrant splicing due to the enlargement of the first intron. As a consequence, the transcript may be partially truncated as was deduced by mapping RNA-seq reads to the genomic reference sequence (Huang et al., 2012) **(FIG. 13)**. Truncated transcripts that contain premature termination codons (PTC) are likely degradated via the NMD pathway and the *FLM* transcript was shown before to be subjected to the NMD pathway for degradation (Kalyna et al., 2012). Hence, the low abundance of full-length *FLM* transcript in Kil-0 may be due to a splicing defect that leads to truncation of a part of the *FLM* transcript. This aberrant *FLM* RNA transcript may be degradated via the NMD-pathway and is thus not translated to truncated protein.

Interestingly, expression analysis showed that the *FLM* isoforms *FLM-ß* and *FLM-δ* are differentially affected by the insertion **(FIG. 11)**. This effect is very likely independent from the genetic background, since it was also observed in the backcross line Col$^{FLM-Kil}$, in a transgenic complementation line as well as in accessions all carrying the *FLM^{Kil-0}* allele. Again, this specific effect also was observed in lines that carry other types of insertions demonstrating that the insertion had no isoform-specific effect.

3.1.3. The first intron of *FLM* possesses regulatory features

In collaboration with David Posé and Markus Schmid (Max Planck Institute for Developmental Biology, Tübingen, Germany), it was found that *FLM* is not expressed from transgenic constructs that contain *FLM-ß* or *FLM-δ* coding sequence but lack all intronic sequence indicating that the intronic sequence of *FLM* has an enhancing effect on *FLM* expression **(FIG. 18)**. Many examples of regulatory effects of intronic sequences on gene expression (intron mediated enhancement, IME) are known, especially for MIKC-type MADS-box transcription factors (Mascarenhas et al., 1990; Rose, 2008; Schauer et al., 2009; Yoo et al., 2011). E.g., SOC1 directly binds to the intron of *SVP* to impair *SVP* expression (Immink et al., 2012) and FLM was shown to bind to *FT* intron regions to regulate *FT* expression (Gu et al., 2013). Through analysis of several non-allelic insertion lines that harbor insertions at different positions in the first intron position-specific effects were observed **(FIG. 19)**. Hence, it can be concluded that the first intron of *FLM* carries important *cis*-regulatory regions for expression and alternative splicing. In the second part of this thesis an approach is described that used phylogenetic footprinting to identify causative, conserved regions. Interestingly, the 5' region of the first intron was found to be essential for *FLM* regulation and this observation is discussed in due course.

3.1.4. The *FLM^{LINE}* allele may have a role for the adaptation of flowering time

Many natural alleles of genes regulating developmental traits such as flowering time are found only at very low frequencies. Often, these were collected in a geographical narrow region or were only present in unique accessions. Hence, it is likely that these rare alleles have deleterious effects (Alonso-Blanco et al., 2009). For example, only few null-alleles of the potent flowering repressor *FLC* were found, but transposon-insertion alleles that lead to changes in *FLC* expression levels are more frequent (Michaels and Amasino, 1999; Gazzani et al., 2003; Michaels et al., 2003;

Caicedo et al., 2004; Lempe et al., 2005; Quadrana et al., 2016). Therefore, *FLC* null alleles may be deleterious and of disadvantage during adaptation (Koornneef et al., 2004). Here, interesting parallels can be seen to the herein described *FLM^LINE* allele. To date, only a single *FLM* null-allele (*FLM^Nd-1*) was identified in only two accessions (Werner et al., 2005b; Balasubramanian et al., 2006). In comparison to the very low number of accessions that carry the *FLM^Nd-1* full-length deletion allele, the present thesis describes the identification of several types of naturally occurring intron-insertion alleles indicating that these variants may be more frequent, not deleterious, but advantageous for adaptation of flowering time.

Comparison of the LINE insertion sequences showed very high similarity between ten *FLM^LINE* accessions **(Fig. 16)**. One possible explanation for the high conservation of the inserted sequence could be a strong selective pressure. Since no specific function of the inserted sequence was identified, this possibility was however excluded. Further, the age of the insertion was calculated by considering the amount of mutations and an average mutation rate (Ossowski et al., 2010). Thereby it was found that the insertion event may be relatively young. We also found a high sequence homology between the ten *FLM^LINE* accessions when the genomic *FLM* and upstream sequences were compared (without considering the LINE insertion), which again suggested a rather recent insertion event. However, high frequencies of identical alleles can also be found when accessions are collected in spatial proximity that could originate from the same population or from accessions that were collected from distant places but derived from the same population due to human activities (Alonso-Blanco et al., 2009). It is unlikely that the ten *FLM^LINE* accessions were mistakenly taken from the same population because investigations of haplotype data of Li et al. (2010) revealed that they belong to genetically differentiated clades. The rather narrow geographical distribution is likely due to the young demographic history of the *FLM^LINE* allele.

All *FLM^LINE* accessions showed a similar pattern of *FLM* transcript misregulation indicating that the general effect of the insertion is conserved. A similar trend was observed when flowering time of the vernalization-insensitive *FLM^LINE* accessions was compared. Especially at 15°C, *FLM^LINE* accessions tended to flower earlier than Col-0 but high variation of flowering time was observed between them **(Fig. 17)**. As it was shown in detail for a novel allele of *SVP*, the effect of the *FLM^LINE* allele on flowering time may also depend on the genetic background (Mendez-Vigo et al., 2013).

In conclusion, the *FLM^LINE*-allele may crucially contribute to earlier flowering across accessions. The frequency of early flowering-alleles in *A. thaliana* evolution increased through the effects of selection (Toomajian et al., 2006). Thus, *FLM* might be an important gene under natural selection for early flowering as suggested previously (Roux et al., 2006).

3.2. Haplotypes of *FLM* non-coding regions fine-tune flowering time in ambient temperatures in *A. thaliana*

The first project of the present thesis describes a large structural natural variation of *FLM* that promotes early flowering and whose genetic and geographic distribution is consistent with a recent adaptive selective sweep. The second project describes the utilization of genome-wide polymorphism data, transcript expression data and phylogenetic footprinting to identify regulatory regions of the *FLM* locus.

3.2.1. Regulatory regions of *FLM* participate in its thermosensitive regulation

It was found that, besides the *FLM* promoter, intron 1 was essential for *FLM* basal expression and that introns 2 - 6 controlled temperature-dependent *FLM* expression **(FIG. 21)**. Through phylogenetic footprinting, a conserved 373 bp region in intron 1 was identified as essential for basal expression of *FLM* **(FIG. 34)**. Through association analysis using genomic sequence information, regulatory regions of *FLM* (PRO2+ and INT6+) were identified, which modulated basal as well as temperature-controlled *FLM* expression in a haplotype-specific manner **(FIG. 33)**. The INT6+[CAA] variants were strongly compromised in temperature-sensitive *FLM* regulation but most of PRO2+ and INT6+ variants caused differential basal *FLM* abundance but temperature-sensitive gene expression was maintained **(FIG. 33)**. Thus, natural *FLM* PRO2+ and INT6+ alleles were identified that modulate basal as well as temperature-sensitive FLM expression

3.2.2. *FLM-ß* is the major physiologically relevant *FLM* isoform

FLM-ß almost perfectly correlated (R^2 = 0.94) with changes in flowering time among a broad range of vegetative growth (15 - 30 rosette leaves) when analyzed at 15°C and tested in a homozygous background **(FIG. 35)**. Several of the experiments strongly suggested that the *FLM-ß* variant has a stronger effect in the control of flowering time than *FLM-δ*. These findings opposed a previously suggested model that the *FLM-ß* repressive function may be antagonized in warmer temperatures by increased expression of *FLM-δ* (Posé et al., 2013). Data presented here shows that the abundance of *FLM-δ*, just like the abundance of *FLM-ß*, correlated positively with the delay in flowering, which is in conflict with the proposed antagonistic function of the two splice variants **(FIG. 35)**. Further, *FLM-δ* is estimated to be expressed at very low levels **(FIG. 27)**. This supports the interpretation of more recent studies suggesting that, at least under the conditions tested here, *FLM-δ* is physiologically not relevant (Sureshkumar et al., 2016).

3.2.3. Hypotheses on the regulatory function of *FLM* PRO2+

The first 500 bp of gene upstream sequences of *A. thaliana* are especially sensitive to treatment with DNaseI indicating that chromatin regions are accessible and proteins that act as transcription factors are able to directly physically associate. Further, these DNaseI hypersensitive sites (DHS) are enriched with conserved non-coding sequences with implied *cis*-regulatory function (Zhang et al., 2012; John et al., 2013; Van de Velde et al., 2016; Zhang et al., 2016). In *FLM*, the PRO2+ is covered by a DHS of 287 bp and it is likely that transcription factors act on *FLM* expression by binding to DNA recognition motifs in this region (Zhang et al., 2016). Interestingly, binding of

MADS-box proteins is not restricted to CArG-box (5'-CC[A/T]$_6$GG-3') DNA recognition elements only (de Folter and Angenent, 2006). Depending on the *FLM* PRO2+ haplotype, PRO2+ harbors a non-CArG MADS-box transcription factor binding site, which contains a longer PolyA stretch than the 5'-CC[A/T]$_6$GG-3' motif (Deng et al., 2011; Chow et al., 2016; Matys et al., 2006; Mathelier et al., 2014; Pajoro et al., 2014). Given the fact that MADS-box transcription factors auto-regulate themselves and genetically and physically interact with each other, it is likely that *FLM* expression is specifically regulated by MADS-box transcription factors by binding at the PRO2+ site with transcriptional activity being triggered by the PRO2+ haplotype (de Folter et al., 2005; Kaufmann et al., 2009; Gu et al., 2013; Smaczniak et al., 2012; Posé et al., 2013; Pajoro et al., 2014). New methods that facilitate the isolation of binding proteins through DNA probes could be used to identify putative direct regulators of *FLM* and how PRO2+ haplotypes are involved in determination of the identity of transcription factor binding sites (Fujita et al., 2013; Vierstra et al., 2015). Furthermore, it would be interesting to precisely map the footprints of proteins that directly bind to *FLM* chromatin in a targeted manner by high-resolution mapping of DHSs (Zhang et al., 2012; Sullivan et al., 2014).

3.2.4. Hypotheses on the regulatory function of *FLM* INT6+ and PolyA motifs

Natural haplotypes of the nucleotide triplet in intron 6 (INT6+) modified relative *FLM-ß* abundance around fourfold when tested in transgenic experiments. By considering the linear model presented in **FIG. 35**, this difference may strongly decelerate flowering by the formation of an additional 31.5 rosette leaves until the initiation of the generative phase, suggesting that INT6+ harbors extensive potential to trigger flowering time. It would be interesting to evaluate the upper saturating level of *FLM-ß* and thereby to validate the extended linearity of the model. The INT6+ site is directly flanked by a PolyA motif and, in humans, short PolyA motifs are recognition sites for hnRNP (heterogeneous nuclear ribonucleoprotein) splicing factors (Swanson and Dreyfuss, 1988; Han et al., 2010; Piva et al., 2012). Depending on the INT6+ haplotype, different binding patterns for hnRNP splicing factors were predicted and their binding, in interaction with other splicing regulators and the transcription machinery, may lead to the observed changes in splicing (Piva et al., 2009; Carrillo Oesterreich et al., 2011). Interestingly, deletions of short PolyA motifs at the *FLM* locus also led to substantial changes in *FLM* splicing, reduced *FLM-ß* abundance and early flowering time. Further, by mutating the INT6+ site from INT6+[AAT] (Col-0 reference allele) to INT6+[AAA], and thereby extending the PolyA motif adjacent to INT6+, strong increases in *FLM-ß* but not *FLM-δ* abundance were generated **(FIG. 34)**. Thus, the extension of the PolyA motif in intron 6 may lead to changes in splicing preference in favor of *FLM-ß*. Although information about regulators of splicing and the identity of splicing recognition sites is still limited, hnRNPs are known to be important splicing factors in plants. Several hnRNPs have a reported role in flowering time regulation in *A. thaliana* and wheat (Wang and Brendel, 2004; Fusaro et al., 2007; Reddy, 2007; Streitner et al., 2012; Kippes et al., 2015; Xiao et al., 2015). Concurrently, alternative splicing coupled to transcript degradation via NMD through modulations of the INT6+ region may be an important posttranscriptional mechanism controlling *FLM-ß* abundance, as also proposed in the first part of the present thesis (Kalyna et al., 2012; Sureshkumar et al., 2016). Despite their potential role as splice factor binding sites, PolyA motifs ([A]$_{>9}$) prevent nucleosomes from binding DNA and these nucleosome exclusion sites are

associated with active chromatin modifications and interfere with the splicing process (Suter et al., 2000; Reddy, 2007; Wijnker et al., 2013). Whether such changes of chromatin architecture through *FLM* sequence variation have a role for *FLM* transcriptional regulation and alternative splicing needs to be further elucidated.

3.3. Conclusion

The aim of this thesis was to elucidate the role of *FLM* for flowering time regulation and the contribution of naturally occurring allelic variation to flowering time variation of *A. thaliana* accessions that were collected from the whole natural habitat.

This thesis reports the first large structural expression-modifying allele for *FLM* and a molecular mechanism for the control of flowering time in ambient temperature through natural variation in *FLM* intron 1 is proposed **(Fɪɢ. 36)**. The genetic and demographic distribution of this allele was consistent with a recent adaptive selective sweep. Thus, the structural variations of *FLM* intron 1 as described here may represent an important mechanism for the adaptation of flowering time to specific climatic conditions during *A. thaliana* evolution.
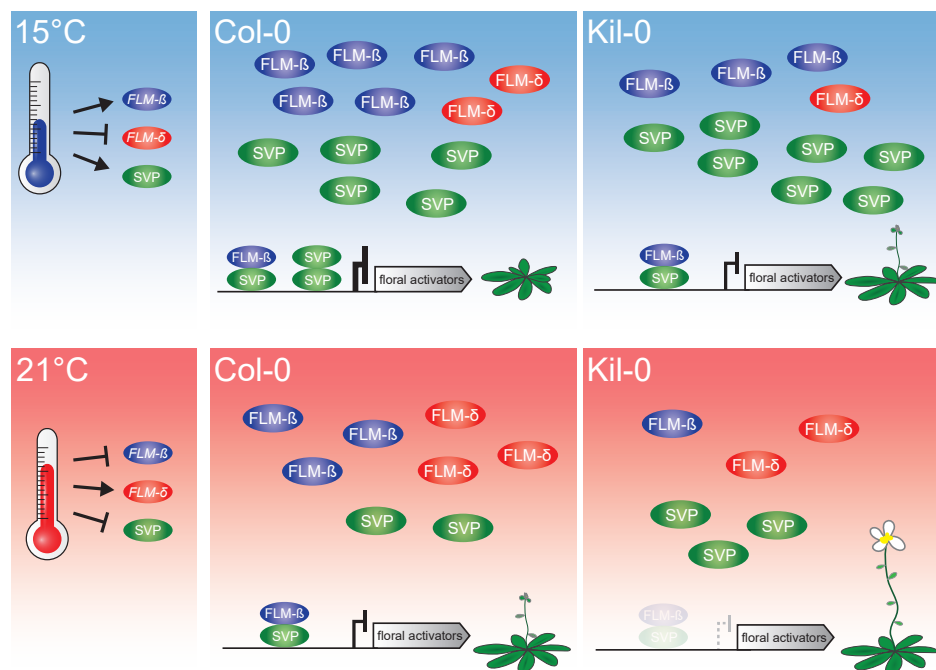
FIGURE 36: Model of the proposed effect of the *FLM^Kil-0* allele on *FLM* abundance. At 15°C, flowering is delayed due to an active repression of floral activators by FLM-ß-SVP heterodimers. An increased ambient temperature (21°C) results in decreased levels of the transcriptional repressive FLM-ß-SVP protein complexes and floral activator genes are expressed. Flowering in Kil-0 (right) might be accelerated in comparison to Col-0 due to a decreased abundance of *FLM-ß* at 15°C as well as at 21°C while *SVP* is unaltered in the different temperatures and between Kil-0 and Col-0. Note that the results of the present thesis suggest a minor role of FLM-δ in the control of flowering time in the lines and conditions examined. The present model assumes that protein abundance follows transcript abundance except for SVP, which is known to be degraded in response to increasing temperature (Lee et al., 2013).

Studies that describe the functions of causative regulatory non-coding regions are still comparatively rare, most likely due to the fact that identification and proving causality is much more challenging than for coding mutations. Mutations in *cis*-regulatory regions have important roles in adaptation and phenotypic evolution of *A. thaliana* as well as during domestication of wild species. They allow the precise modulation of transcript abundance while the probability to generate negative pleiotropic effects is low. In contrast, coding mutations often lead to deleterious alleles (Koornneef et al., 2004; Wray, 2007; Alonso-Blanco et al., 2009; Meyer and Purugganan, 2013; Swinnen et al., 2016).

The results presented in the present thesis allow proposing a model according to which genetic variations of the PRO2+ and INT6+ regions harbor the potential to readjust flowering time to buffer changes of ambient temperature conditions **(Fig. 37)**. Hence, these small non-coding regulatory regions may be interesting examples for plant breeding as *FLM-ß* precisely modulates flowering without - at least during our investigations - negative pleiotropic effects. In combination with the identification of putative *FLM* orthologues from other *Brassicaceae* species and new technologies of genome editing, the results presented in this thesis may help to fine-tune flowering time of crop cultivars in a dynamic manner and over a broad range of temperature conditions to buffer negative effects of climate change on agricultural systems and to develop future climate change-resilient crops (Thuiller et al., 2005; Wheeler and von Braun, 2013; Pachauri and Meyer, 2014; Moore and Lobell, 2015).
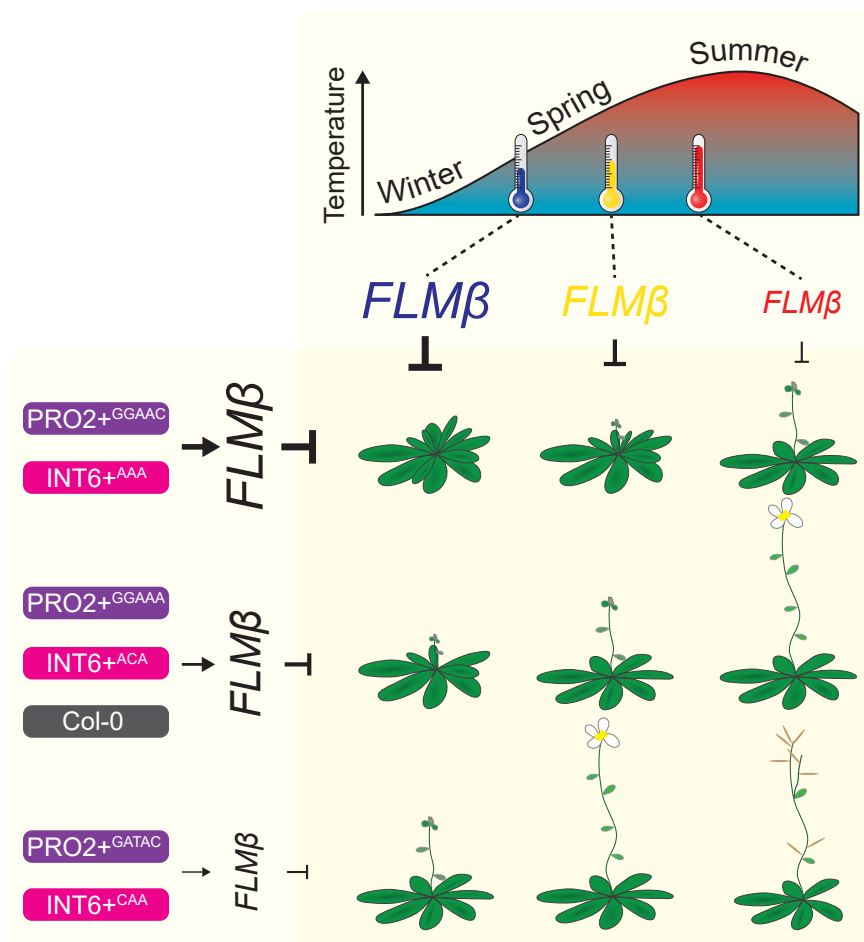
FIGURE 37: Model of the proposed effect of PRO2+ and INT6+ haplotypes and temperature on *FLM-ß* abundance and flowering. The abundance of the flowering repressor *FLM-ß* decreases in response to higher temperature and flowering is consequently accelerated (This thesis; Lee et al., 2013; Posé et al., 2013). Note, that results, which are described in the first part of the present thesis imply an especially prominent effect of *FLM* in a range from 9°C to 21°C (long day photoperiod). On the genetic level, *FLM-ß* abundance and therefore flowering time is triggered by the PRO2+ (purple) and/or the INT6+ (pink) haplotype. Among the PRO2+/INT6+ combinations tested, the Col-0 (grey) reference alleles showed intermediate *FLM-ß* levels. Results presented in the present thesis propose that changes in flowering time, which are caused by changing ambient temperature can be buffered by modifying the PRO2+ and INT6+ regions, as illustrated by similar plant symbols.

4.   MATERIALS AND METHODS

4.1.   Biological material

The following *A. thaliana* accessions and genotypes were used in this study and were provided by the Nottingham Arabidopsis Stock Centre (NASC; Nottingham, UK): The *A. thaliana* accessions Killean-0 (Kil-0), Columbia-0 (Col-0), Nossen (No-0), Niederzenz-1 (Nd-1) as well as Coimbra-1 (Co-1). All nine additional *FLM^{LINE}* accessions were also provided by the NASC and are listed in **TABLE 1**. The insertion mutants *flm-3* (SALK_141971; Col-0), GABI-KAT GK487H01 (GABI; Col-0), SALK_068360 (SALK; Col-0) were provided by the NASC and RIKEN-13-4593-1(RIKEN; No-0) was provided by the RIKEN Stock Center (RIKEN BioResource Center Tsukuba, Japan). In each case, the positions of the insertions were determined by DNA sequencing. PCR primers for genotyping were obtained from the SIGnAL T-DNA Primer Design tool (Salk Institute Genomic Analysis Laboratory, http://signal.salk.edu/tdnaprimers.2.html) and are provided as **TABLE 2**. *flc-3* and the line *FRI^{SF-2}FLC* were kindly provided by Franziska Turck and George Coupland (Max Planck Institute of Plant Breeding Research, Cologne, Germany) (Michaels and Amasino, 1999). The lines pFLM::gFLM (pDP34), pFLM::FLM(ß) (pD79), and pFLM::FLM(δ) (pDP80) were described in Posé et al. (2013) and kindly provided by David Posé and Markus Schmid (Max Planck Institute for Developmental Biology, Tübingen, Germany).

4.2.   Physiological experiments

For flowering time analyses, plants were randomly arranged in trays and grown under constant white light (70 - 90 µmol m-2 s-1) or in long day conditions with 16 hrs white light (110 - 130 µmol m-2 s-1)/8 hrs dark photoperiod in MobyLux GroBanks (CLF Plant Climatics, Wertingen, Germany) or MLR-351 SANYO growth chambers (Ewald, Bad Nenndorf, Germany). Trays were rearranged every two days and water was supplied by subirrigation. Analysis of large plant sets was performed in a walk-in chamber with constant white light conditions as described above (Gewächshauslaborzentrum Dürnast, Germany). Flowering time was quantified by determining the time until the macroscopic appearance of the first flower bud (days to bolting, DTB) or rosette and cauline leaf numbers (RLN, CLN). Student's t-tests, ANOVA, Tukey HSD, and Wilcoxon rank sum tests were calculated with Excel (Microsoft) and R (http://www.r-project.org/), respectively.

4.3.   Kil-0 genome sequencing and pool sequencing

For the resequencing of the Kil-0 genome or the late and early flowering pools of F3 segregants of the Kil-0 x Col-0 hybrid, libraries were prepared from 600 ng genomic DNA following the standard protocol of the TruSeq DNA Sample Preparation Kit v2 (Illumina, San Diego, CA). Paired-end sequencing with a read length of 100 bp was performed on a HiSeq 2500 (Illumina, San Diego, USA). Sequencing was performed at the Chair of Animal Breeding of the Technical University Munich (Freising, Germany). Post-sequencing quality trimming was performed with the CLC Genomics Workbench (v. 7.0) using the following parameters: low quality limit = 0.05; ambiguous nucleotide = maximum 1; length minimum = 15. Post-trimming, a read number of 15×10^6 and 20×10^6 reads was obtained for the early and late flowering samples, respectively. Read mapping was performed using

the Col-0 TAIR10 reference sequence (www.arabidopsis.org) with the stringent settings: mismatch cost = 2; insertion cost = 2; deletion cost = 2; length fraction = 0.9; similarity = 0.9. An average 57- or 79-fold coverage was obtained from the early and late flowering DNA pools. Variant calling was performed using the probabilistic variant calling tool of the CLC Genomics Workbench v. 7.0 (Qiagen, Hilden, Germany) and default settings. SNPs with a 30 − 120-fold coverage and frequency $f > 20\%$ as well as a presence call in both pooled samples were selected for allele frequency mapping. From those, SNPs that showed a frequency of < 80% in the resequencing analysis of the homozygous Kil-0 parental line were discarded. Smoothing using locally weighted scatterplot smoothing (LOESS) of SNP frequency values was performed with R (http://www.r-project.org/). 95% confidence intervals, $\Delta f > 25\%$, and Δf_{max} were calculated from the LOESS values. The *de novo* assembly of Kil-0 resequencing reads was performed using the CLC Genomics Workbench v. 7.0 with default settings. Contigs were identified by simple search and reassembled to the Col-0 genomic *FLM* sequence.

4.4.   LINE insert and *FLM* locus characterization

The characterization of the inserted sequence in FLM^{Kil-0} was performed in cooperation with Heidrun Gundlach (Helmholtz Zentrum, München, Germany). The LINE sequence within the Kil-0 insert was identified with RepeatMasker (http://www.repeatmasker.org) against PGSP-REdat, v_9.3_Eudicot (http://pgsb.helmholtz-muenchen.de/plant/recat/). The Pfam domains where annotated with hmmsearch of hmmer3 in all 6 reading frames against PfamA v27 (Finn et al., 2011). A detailed description of the inserted sequence is provided as **TABLE A1**. For phylogenetic analyses of the *FLM* locus, sequences of 88 randomly selected accessions were extracted from the 1001 *A. thaliana* Genomes Project GEBrowser tool (http://signal.salk.edu/atg1001/3.0/gebrowser.php) and aligned with sequences of the ten FLM^{LINE} accessions and Col-0. ClustalW2 alignments and neighbor-joining trees (HKY genetic distance model, 1000 bootstrap replicates) were constructed with MEGA5 (Tamura et al., 2011).

4.5.   Genetic mapping and backcrossing

To identify the causative locus for early flowering in Kil-0, the FT15 locus was mapped with polymorphic markers as selected from a previously described marker collection (Pacurar et al., 2012). Additional SSLP (single sequence length polymorphism) markers were generated by searching the publicly available Kil-0 genomic sequence or the genome sequence that was generated as part of this thesis for indel (insertion/deletion) polymorphisms (The 1001 Genomes Consortium, 2016). PCR primers spanning these sites were designed with Primer3 and tested on Col-0 and Kil-0 genomic DNA (Untergasser et al., 2007). The SSLP amplicons were generally between 200 and 700 bp long and were separated on 2 - 3.5% agarose gels or using the QIAxcel Advanced Capillary Electrophoresis High-Resolution Kit (Qiagen, Hilden, Germany). For the rough mapping of FT15, ten early and ten late flowering F2 plants were selected from the extreme phenotypic borders of a F2 population (n = 124). Genotypes were assessed and QTL analysis was performed using the scanone function of the R/qtl package (Broman et al., 2003). A 5% LOD threshold was calculated by 1,000 permutations. Further markers were used to fine-map a 968 kb interval using recombinant F3 plants from the F2 genotypic screen with the flanking markers M1 and M2 as described in the text. The genetic marker distances were calculated from the genotype data of all screened F2 plants with JoinMap

v.4.1 (Kyazma B.V., Wageningen, Netherlands). Markers and the respective primers are provided as **Table 3**. The backcross lines Col$^{FLM-Kil}$ and Kil^{flm-3} were generated by marker-assisted backcrosses. Heterozygous F2 plants were genotyped with the primers LP1, LP2, and RP1 to examine the lines for the presence of the Col-0 or Kil-0 *FLM* allele and with SALK_141971 forward and reverse primers to test for the *flm-3* knock-out insertion, respectively. Primer sequences are provided as **Table 2**.

4.6.   Quantitative real-time PCR

For qRT-PCR analyses, total RNA was isolated from three biological replicates using the NucleoSpin RNA Plant kit (Machery-Nagel, Düren, Germany). DNA was removed by an on-column treatment with rDNase (Machery-Nagel, Düren, Germany). Two to three 3 µg of total RNA were reverse transcribed with an oligo(dT$_{18}$) primer and M-MuLV Reverse Transcriptase (Fermentas, St. Leon-Rot, Germany) and the cDNA equivalent of 30 - 50 ng total RNA was used in a ten µl PCR reaction with SsoAdvanced™ Universal SYBR® Green Supermix (BioRad, München, Germany) with three biological replicates in a CFX96 Real-Time System Cycler (BioRad, München, Germany). The relative quantification was calculated with the ΔΔCt method with *ACTIN8* as a control (Pfaffl, 2001). In case of the pooled T2 transgenic lines, each 25% of the total number of available independent lines per construct (18 to 45) was split on one of four so-called replicate pools resulting in pools that comprise of four to eleven lines. Around 1,000 seeds were used for pooling per line. One RNA sample was extracted per replicate pool and processed as described above. The large-scale expression experiments were performed in a previously described 96-well format (Box et al., 2011). DNA was digested with DNaseI (Thermo Fisher Scientific) and reverse transcription was performed as described above. Quantitative PCR reactions were performed as described above using a CFX384 Real-Time System Cycler (BioRad, München, Germany). *ACTIN8* (AT1G49240) and *BETA-TUBULIN-4* (AT5G44340) were used as reference genes. Student's t-tests were calculated with Excel (Microsoft). Primer sequences are provided as **Table 2**.

4.7.   3'-RACE-PCR

Polyadenylated transcripts were amplified according to Scotto-Lavino et al. (2006). Transcript pools were analyzed on an agarose gel and the bands were purified from all samples, subcloned into pCR2.1-TOPO (Life Technologies, Carlsbad, USA) and transformed to E. coli strain DH5α. Bacteria were cultured in a 96 deep-well plate and plasmid DNA was extracted using the NucleoSpin 96 Plasmid Kit (Machery-Nagel, Düren, Germany). The vector insert was determined by sequencing. Primer sequences are provided as **Table 2**.

4.8.   Determination of *FLM* pre-mRNA abundance

Nuclei were isolated as previously described with two to three biological replicates per line (Kaufmann et al., 2010). RNA from the nuclei pellet was extracted according to Box et al. (2011) and DNA was digested using DNaseI (Life Technologies, Carlsbad, USA). cDNA synthesis and qRT-PCR were performed as described above. Primer sequences are provided as **Table 2**.

4.9. RNA-sequencing

To investigate significant gene expression differences between Col-0 and Kil-0, RNA was prepared from three biological replicate samples from ten day old seedlings (21°C, long days) using the NucleoSpin RNA Plant kit (Machery-Nagel, Düren, Germany). DNA was removed by an on-column treatment with rDNase (Machery-Nagel, Düren, Germany). Sequencing libraries were prepared following the standard protocol of the TruSeq RNA Sample prep v2 Kit (Illumina, San Diego, USA). Paired-end sequencing with a read length of 100 bp was performed on a HiSeq 1000 (Illumina, San Diego, USA) at the Department for Cell Biology and Plant Biochemistry at the University of Regensburg (Regensburg, Germany). RNAseq data analysis was performed in cooperation with Mathias Pfeiffer (Helmholtz Zentrum, München, Germany). RNAseq reads from each biological replicate were then aligned against the TAIR10 release of the *A. thaliana* reference genome (www.arabidopsis.org) using Bowtie (version 2.1.0) and Tophat2 (version 2.0.8) with default parameters (Trapnell and Salzberg, 2009; Kim et al., 2013). On the basis of the structural gene annotation for Arabidopsis TAIR10, gene expression was quantified using HTSeq and differential expression tests were performed by using DESeq2 (Love et al., 2014; Anders et al., 2015). Significantly differentially expressed genes were defined as genes with Benjamini-Hochberg-adjusted p-values < 0.01. Exon-based expression tests were conducted using DEXSeq (Anders et al., 2012). A list of genes with a role in flowering time regulation was generated by searching the TAIR database (www.arabidopsis.org) for the term "flowering time". The resulting list was curated manually and was published as supplemental data in Lutz et al. (2015) (see Appendix). To subsequently analyze gene expression of the *FLM* locus at a nucleotide resolution, the corresponding Col-0 gene sequences were extracted from the TAIR10 Col-0 reference genome (www.arabidopsis.org) and the Kil-0 genome sequence as determined as part of this thesis. Subsequently, RNA-seq reads were aligned against the Kil-0 and Col-0 genomic sequences using Bowtie and Tophat2 as described above. The number of mapped RNA-seq reads from Col-0 and Kil-0 per nucleotide were counted with the toolset Bedtools and subsequently normalized to range between 0 (no expression) and 1 (maximum expression). For visualization, the mean expression level and 5 % and 95 % confidence intervals were determined across the biological replicates of one sample.

4.10. Cloning procedures

To obtain a genomic *FLM* fragment from Kil-0 with a deletion of the 5.7 kb LINE insertion (gFLM(Kil-0)ΔLINE), a 6981 bp deletion fragment was amplified by overlap extension PCR from Kil-0 genomic DNA. Fragment 1 (bp - 2367 to bp + 631) was amplified with primers ULC-1 and ULC-2 and fragment 2 (bp + 632 to bp + 4156 and 251 bp 3'-UTR plus 207 bp downstream sequence) with primers ULC-3 and ULC-4. After purification using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, USA) of the two subfragments, the full fragment was generated by overlap-extension PCR reaction with primers ULC-1 and ULC-4, purified and subcloned to pCR2.1-TOPO (Life Technologies, Carlsbad, USA) to generate pUL1. The insert of the pUL1 full-length fragment was sequenced and subcloned as a BamHI and XhoI fragment into pGreen0229 to generate pUL2 (Hellens et al., 2000). To insert the LINE insertion into the previously described construct with the Col-0 genomic *FLM* fragment pFLM::gFLM (pDP34; Posé et al., 2013), pFLM::gFLM (pDP43) was mutagenized by PCR with the primer ULC-12 to replace the sequence ATTGTTCA (bp +632 to bp

+640) with a unique AscI restriction site GGCGCGCC to generate pUL3 (Sawano and Miyawaki, 2000). The LINE insertion was then amplified from Kil-0 genomic DNA with primers ULC-16 and ULC-17, purified and subcloned into pCR2.1-TOPO (Life Technologies, Carlsbad, CA) to generate pUL7 and inserted as an AscI fragment into the modified pUL3 to generate pUL10.

The constructs described in the second part of the present thesis were generated by recombining a previously described construct with the Col-0 genomic *FLM* fragment pFLM::gFLM template (pDP34; Posé et al., 2013) into a pDONR201 destination vector using the Gateway system (Life Technologies, Carlsbad, USA) to generate pUL11. This vector was used as a template to generate mutations using either a single phosphorylated primer or a combination of forward and reverse primers (Sawano and Miyawaki, 2000; Hansson et al., 2008). In case of variants with multiple modifications, individual mutations were introduced sequentially. The mutated inserts were recombined to the pFAST-R07 expression vector using the Gateway system (Life Technologies, Carlsbad, USA) (Shimada et al., 2010). The Col-0 reference construct was directly recombined to the pFAST-R07 expression vector without mutations (pUL12). All inserts of expression constructs were verified by digestion and sequencing.

All expression constructs were transformed into *Agrobacterium tumefaciens* strain GV3101+pSOUP (pUL2, 3, and 10) or GV3101 (pUL12 - 78) into Kil-0 (pUL2 and pDP34), *flm-3* (pUL10), or Nd-1 (pUL12 - 78) using the floral dip transformation (Clough and Bent, 1998). Empty pGreen0229 plasmids were transformed to GV3101+pSOUP to generate Col-0 BAR, Kil-0 BAR, and *flm-3* BAR empty vector controls (Clough and Bent, 1998; Hellens et al., 2000). Depending on the expression vector used, T1 transformants were selected either by spraying soil-grown plants with 0.1% BASTA (Bayer CropScience AG, Monheim, Germany) or based on the red fluorescence of seeds (Shimada et al., 2010). All primers used for cloning are provided as **Table 4**. All generated expression constructs are provided as **Table 5**.

4.11.   *FLM* locus analysis of *A. thaliana* accessions

In order to test for the presence of the Kil-0 *FLM* allele or a related structural polymorphism in a large collection of *A. thaliana* accessions, PCRs with the diagnostic primers LP1, LP2 and RP1 **(Fig. 9B)** were performed on *A. thaliana* genomic DNA using Phusion (New England Biolabs, Frankfurt, Germany) or TaKaRa LA Taq polymerases (Takara Bio, Saint-Germain-en-Laye, France). Selected PCR products were analyzed by DNA sequencing following gel extraction with Wizard SV Gel and PCR Clean-Up System (Promega, Madison, USA). Sequencing primers are provided as **Table 1**.

Additionally, in cooperation with Jörg Hagmann and Congmao Wang (Max Planck Institute for Developmental Biology, Tübingen, Germany), the Kil-0 genome sequence information was used to identify accessions with a Kil-0-specific LINE insertion in *FLM* by analyzing genome sequence data that was determined in the frame of the 1001 *A. thaliana* Genomes Project (The 1001 Genomes Consortium, 2016). To render read mapping specific for the Kil-0 allele, the insertions at the experimentally retrieved insertion sites flanked by 140 bp sequence were extracted for each of the two insertion breakpoints and defined as target sequences for mapping. Subsequently, all reads of the individual *A. thaliana* accessions were mapped against the target sequences using SHORE and genomemapper allowing for up to 5% single base pair differences including gaps with regard to the read length (Schneeberger et al., 2009). To ensure that each supporting read spans at least 25% of

its sequence across the insertion breakpoint, only reads that overlapped an insertion site with 50% of the read length were counted as supporting an insertion breakpoint as core-mapping reads. An insertion was defined as present if there were at least two core-mapping reads at each left and right insertion breakpoints.

4.12. Phylogenetic analysis and multiple alignments

For phylogenetic analyses, sequences were aligned using ClustalW alignments or MUSCLE and neighbor-joining trees (Maximum Composite Likelihood method, 1000 bootstrap replicates) were constructed with Geneious vR7.0.5 (Biomatters Limited, Auckland, New Zealand) and MEGA7.0.14 (Tamura et al., 2011).

4.13. Phylogenetic footprinting

The phylogenetic footprinting analysis was performed in cooperation with Manuel Spannagl (Helmholtz Zentrum, München, Germany). Putative *FLM* orthologs were identified first by OrthoMCL (V2.0) clustering (PMID: 12952885; standard parameters, inflation value = 1.5, BLAST e-value-cutoff = e-05), incorporating predicted protein sequences from *Arabidopsis thaliana* (TAIR10; AT1G77080.4, AT5G65060.1), *Arabidopsis lyrata* (V1.0; fgenesh2_kg.2__2000__ AT1G77080, fgenesh2_kg.8__2543__AT5G65050, fgenesh2_kg.233__4__AT5G65050), *Boechera stricta* (V1.2; Bostr.4104s0001.1), *Brassica rapa* (V1.3; Brara.B03928.1.p, Brara.F02378.1.p), *Brassica oleracea* (V2.1; Bo2g166500.1, Bo2g166560.1), *Capsella grandiflora* (V1.1; Cagra.0450s0030.1, Cagra.0917s0081.1), *Capsella rubella* (V1.0; Carubv10020979m, Carubv10027327m), *Gossypium raimondii* (V2.1), *Medicago truncatula* (V4.0), *Oryza sativa* (MSU7), *Populus trichocarpa* (V3.0) and *Solanum lycopersicum* (ITAG2.3). Data sets were downloaded from Phytozome, PGSB PlantsDB and Ensembl Plants (Goodstein et al., 2012; Kersey et al., 2016; Spannagl et al., 2016). Cluster(s) containing *FLM* were extracted and, in case of the presence of multiple group members from individual species, filtered further using the best bi-directional BLAST hit criterion. Genomic sequences and 2 kb upstream region were aligned with ClustalW2 (Larkin et al., 2007). Conserved regions in the alignment were manually annotated.

4.14. Analysis of linkage disequilibrium

Linkage (R^2) between the sites that were used as input for the simple single locus association test was calculated using the LD function in GGT v2.0 (van Berloo, 2008). Only the 109 of 119 diallelic SNPs were considered.

4.15. Data retrieval from the 1001 *A. thaliana* Genomes Project

The genomic sequences of a comprehensive set of 1,135 *A. thaliana* accessions have been released just recently (The 1001 Genomes Consortium, 2016). Earlier, sequence information from datasets of in total 776 and 840 accessions, respectively, have been available and were used in this study. Genomic *FLM* sequences (7 kb; Chr1: 28953637 – 28960296; including 2 kb upstream and 0.2 kb downstream sequence) were extracted from 776 accessions from the 1001 *A. thaliana* Genomes portal (http://1001genomes.org/datacenter/) using the Wang dataset (343 accessions), the GMI

dataset (180 accessions), the Salk dataset (171 accessions) and the MPI dataset (80 accessions) (Cao et al., 2011; Long et al., 2013; Schmitz et al., 2013; The 1001 Genomes Consortium, 2016). 45 SNPs with a MAF > 5% were extracted and used for haplotype analysis of the *FLM* locus. 31 SNPs were polymorphic between the 54 selected accessions represented in the *FLM* haplotype set.

To obtain an extended set of sequences that include not only SNPs but also information about insertions and deletions we extracted *FLM* genomic sequences (including 2 kb upstream and 0.38 kb downstream sequence) from the GEBrowser resource (v3.0, http://signal.salk.edu/atg1001/3.0/gebrowser.php). A set of 850 sequences was manually curated and aligned using MEGA7.0.14 (Tamura et al., 2011). Some sequences were excluded since they possessed a high number of ambigious bases. The core sequence set consisted of 840 sequences. This sequence set was used for association analysis and all further analysis on *FLM* PRO2+ and INT6+ variation. Accession identifiers, geographic data, and ADMIXTURE group membership (k=9) were obtained from the 1001 *A. thaliana* Genomes tool collection (http://tools.1001genomes.org/; The 1001 Genomes Consortium, 2016).

4.16.   Analysis of nucleotide diversity of *FLM*

Nucleotide diversity was calculated with DNaSP using the sequence data retrieved from the GEBrowser (described above) (Rozas and Rozas, 1995). In case of sliding window analysis, a window size of 20 and a step size of 10 was used. Alignment gaps were ignored and all polymorphic sites were considered. Haplotype analyses was performed with DNaSP using either the SNPs of the set of 776 sequences or the data retrieved from the 1001 GEBrowser as indicated in the text. Sites with alignment gaps were considered and invariable sites were removed. Networks were generated using FLUXUS network software in cooperation with Thomas Nussbaumer (Helmholtz Zentrum, München, Germany) (Bandelt et al., 1999).

4.17.   Analysis of nucleotide diversity in the type II MADS-box family

The relative frequency of polymorphic sites for the *FLM* gene model (AT1G77080.4) was compared to 54 type II MADS-box genes (Martinez-Castilla and Alvarez-Buylla, 2003). To count the number of polymorphic sites per element type, all splice variants and 2 kb upstream sequence of the *FLM* gene locus were exported from the previously described 776 sequences and for each element type (upstream, 5'-UTR, exon, intron, 3'-UTR).

4.18.   Kruskal-Wallis test of association

This analysis was performed in cooperation with Thomas Nussbaumer (Helmholtz Zentrum, München, Germany). To detect variants with significant effects on *FLM* transcript levels (simple single locus association test), 119 variants were extracted from a set of 840 sequences which were retrieved from the 1001 Genomes GEBrowser resource, as described above. To examine whether a variant showed a significant effect on *FLM* expression, the qRT-PCR expression values of the 54 accessions from the *FLM* haplotype set were used as input data to run Kruskal-Wallis tests using R (http://www.r-project.org/). P-values from all comparisons were corrected following the Benjamini-Hochberg multiple testing correction and resulting values were -log(10) transformed and plotted along the gene model.

4.19.   Primers and markers

4.19.1. List of primers for genotyping, qRT-PCR, sq-PCR, 3'-RACE-PCR, and sequencing

TABLE 2: qRT-PCR, quantitative real-time PCR; sq-PCR, semi-quantitative PCR; 3'-RACE, three prime rapid amplification of cDNA ends.

Name	Purpose	Fw. Sequence	Rev. Sequence	Source
FLM CDS	sq-PCR	ATGGGAAGAAGAAAAATCGAG	TAATTGAGCAGCGGGAGAGT	Balasubramanian, 2006
GABI_vector	Genotyping	ATATTGACCATCATACTCATTGC	-	*1
GABI-Kat 487H01	Genotyping	TGAAACTCTGATCATCTGGTGC	TGTCGTGGATATTTTTAACCCC	*1
LINE-ins. flanking (LP1, RP1)	Genotyping	TGGCCAAGATATGTTCTTGATG	TGTGTTAACCAATTTGAACCAA	This study
LINE-ins. specific (LP2)	Genotyping	ACAGCAACGCTCCGAGTAAAG	-	This study
RIKEN 13-4593-1	Genotyping	TTCATCAAGATTTTGTCTTTTGGA	TATCAACTTCGCAGGCCAAT	*1
RIKEN Vector, 3-1a	Genotyping	-	GGTTCCCGTCCGATTTCGACT	*1
SALK_068360	Genotyping	CACTTTGCATAGTGCCACTTTG	GCAGATGCGTAAGCTTTAACC	*1
SALK_141971 (flm-3)	Genotyping	GAGCACCAGATGATCAGAGTTTC	TCATCCTCATCCAAATTCGAC	*1
ACT8	qRT-PCR	GCAGCATGAAGATTAAGGTCGTGG	TGTGGACAATGCCTGGACCTGCT	This study
FLC	qRT-PCR	GCTACTTGAACTTGTGGATAGCAA	GGAGAGGGCAGTCTCAAGGT	This study
FLM-ß	qRT-PCR	CTATGACTCTTCCTCCGGTGAC	GATCTAAGGCTCTAAGTTCATCAGCATG	Lee, 2013; Posé, 2013
FLM-δ	qRT-PCR	CTATGACTCTTCCTCCGGTGAC	GCTTGAACAGCGCTTCTATC	Lee, 2013; Posé, 2013
FT	qRT-PCR	GGTGGAGAAGACCTCAGGAA	GGTTGCTAGGACTTGGAACATC	This study
MAF1.2	qRT-PCR	CCTCAATGTTTTGAACTCGATC	TCGACATTTGGTTCTTCAAGCTTGC	Severing, 2012
MAF1.4	qRT-PCR	GATCGTTATGAAATACAACATGC	GTATTCTTTCCCATCTGGCTAGC	Severing, 2012
SVP	qRT-PCR	ATGGCGAGAGAAAAGATTCAG	CTAACCACCATACGGTAAGCC	Posé, 2013
E1_Fw	qRT-PCR	GCGAATCGAGAACAAAAGCA	-	This study
E1I1p_Frag.2_Rev	qRT-PCR	-	CGAGTGAGACCCAGAAAAGG	This study
E1I1p_Frag.3_Rev	qRT-PCR	-	AAGCCCTAGATTTGAAGCTTCT	This study
E1I1p_Frag.4_Rev	qRT-PCR	-	GGTTAAAGCTTACGCATCTGC	This study
E1I1p_Frag.5_Rev	qRT-PCR	-	CACGAAACGAACATCAAGAACA	This study
E1I1p_Frag.6_Rev	qRT-PCR	-	AAATGAAGGTAAAAAAAGAGA	This study
*ACT8*_intron2	qRT-PCR	ATTAATGTCGAGAAAATGAACAAA	AAGAAAATCATTTCACTGTGATCAATA	This study
*ACT8*_intron3	qRT-PCR	TGGACTGTCATTTCTAAGTTTTTGC	CTGATTACACACACACACACACA	This study
*FLM*_intron2	qRT-PCR	GAGTTACAGTTATTGTGTCTACTGGAA	GCACATGAAGAAAAGGAATGG	This study
*FLM*_intron6	qRT-PCR	AAAAGTTGAAGGACTTTGATGGA	TCATCGGTGAGACATACCTTTC	This study
*FLM*_3'-downstream	sq-PCR	-	CGAGCAGTGAGTTTGGCATA	This study
GSP1	3'-RACE-PCR	CGAGATCAAGCGAATCGAGAA	-	This study
GSP2	3'-RACE-PCR	AATCCTCCGTCGCTGTTGTC	-	This study
ULS-17	Sequencing	-	TGTGTTAACCAATTTGAACCAA	This study
ULS-25	Sequencing	TTCAAGAAACTTTTCTACACGTGG	-	This study
ULS-26	Sequencing	CCAGTTTATCTTCCCCACGGAA	-	This study
ULS-27	Sequencing	GGAGACATTTCAGGAAGTGAGA	-	This study
ULS-29	Sequencing	AAACCATCTTTGGGAACGGTTA	-	This study
ULS-23	Sequencing	-	CTTTACTCGGAGCGTTGCTGT	This study
ULS-24	Sequencing	-	CATGTTTTGTGTAATGCCATCCC	This study
ULS-28	Sequencing	-	TCTTTGCAGATCTCTTTCGTCA	This study
ULS-30	Sequencing	-	TGAAGAAAAGAGAATCGTCCGC	This study
ULS-31	Sequencing	-	CTCATTCCAAAAACGTTGGTAG	This study
ULS-32	Sequencing	-	GGACTTGATTCTTGCTTCTGCA	This study
ULS-51	Sequencing	-	CACGTTTGATTGTGCCATCGAA	This study
ULS-20	Sequencing	AATTTGCCTCTTAGTGGCTTCA	-	This study
ULS-22	Sequencing	ACCGGTCTAACCGACTGGAT	-	This study
ULS-34	Sequencing	GCCAACCAATTTCCAACCTA	-	This study
ULS-33	Sequencing	TTAATCCGGTTAGGGTCGTG	-	This study
ULS-52	Sequencing	AGTGGTCAAGTTACTAATCACTTT	-	This study
ULS-3	Sequencing	TGAGAGATTATTTGGGGTTAGTGG	-	This study
ULS-5	Sequencing	TGCTTGATCATGAAAGCAAA	-	This study
ULS-7	Sequencing	TAATCGAACTCATGGGGATG	-	This study
ULS-9	Sequencing	TCCAAAGGTTAGCAGTACGACA	-	This study
ULS-11	Sequencing	GCCAGGTAACAATGACCACA	-	This study
ULS-35	Sequencing	-	TCGATGACGATGAGTAAATGTTTACA	This study
* SIGnAL T-DNA Primer Design tool, http://signal.salk.edu/tdnaprimers.2.html				

4.19.2. List of markers and respective primers

TABLE 3: Chr., chromosome; SSLP, simple sequence length polymorphism; CAPS, cleaved amplified polymorphic sequence.

Marker name	Purpose	Fw. sequence	Rev. sequence	Chr.	~ Chr. position	Type	Source
UPSC_1-1221	Genome-wide mapping	TTTAGGGATGGGTCATGGTC	TTGGTTCTTCTTTCGGATTTTC	I	1220831	SSLP	Pacurar et al. (2011)
UPSC_1-9236	Genome-wide mapping	CCCTTCCATTGAACAGAGAATGA	TAAATCTGTCCCGCCAAA	I	9235991	SSLP	Pacurar et al. (2011)
UL_1-17270	Genome-wide mapping	GCAAACATGTTCGGTTTACAA	TCCATCCAATCGTCCTTCTG	I	17269783	SSLP	This study
UL_1-24703	Genome-wide mapping	GGTCCACGACAAGTTTCGTTA	CTGCTGCGTTTCAGGTACAA	I	24702685	SSLP	This study
UL_1-27097	Genome-wide mapping	CCGTCTGCACGGATACTCAA	CATGATGGTCCCTTCCGTTA	I	27097065	SSLP	This study
UL_1-27547	Genome-wide mapping	GCAACGCCACACATTTTCAT	TGACCGAAGTTTCGTCCAAG	I	27546976	SSLP	This study
UL_1-28035	Genome-wide mapping	CATCCCCAAATGTCACGTA	CGAAATTGGCCCATTAAAGC	I	28035179	SSLP	This study
UL_1-28544	151 kb interval finemapping (M1)	AAGCTTCGCAGCCATCATCT	TCTCGATGGGTTGCAAGAGA	I	28543975	SSLP	This study
UL_1-28607	151 kb interval finemapping	TCATCGTGATGCAACCTTCC	GGCGTGGTGATGATGTGAAT	I	28607319	SSLP	This study
UL_1-28731	151 kb interval finemapping	CGACTGGATGTGTCCTCCTG	TCCAAGGCGTTTACAAATCC	I	28731359	SSLP	This study
UL_1-28816	Genome-wide mapping	AGGTCCTTAAGCGGGTGTCCT	TCGAGAACCAGTGAAACAACAC	I	28815679	SSLP	This study
UL_1-28838	151 kb interval finemapping	TCACGTTTTCAGGTAGGAAAAA	TTCCTTTTCTTCTCCTTCTTCTTT	I	28838479	CAPS (AgeI)	This study
UL_1-28919	151 kb interval finemapping	AGCCATCTCCGTGGACCTTA	CAGTGGCGGAGGTAGCATAA	I	28918595	SSLP	This study
UL_1-28996	151 kb interval finemapping	TCCCATGAGGTTTTGGAGAG	TGCCCTTACAACCTTTTCTCA	I	28996270	SSLP	This study
UL_1-29070	Genome-wide mapping	TGTTCTGCCAATTTGGTTCG	GTCGAATCAACCGACGATGA	I	29069831	SSLP	This study
UL_1-29208	151 kb interval finemapping	TGGAGCTGTGTTTGTGATGG	AAACACCGCCTTAGTTTTGGATG	I	29208152	SSLP	This study
UL_1-29512	151 kb interval finemapping (M2)	CAACCATGACCTGTCAACC	TGTGAGAGCACCCCAACAGT	I	29511772	SSLP	This study
UL_1-30037	Genome-wide mapping	TGCTGTAGCTGGCCATGTTT	TCCCTCTTCTCCTCGCACTC	I	30037343	SSLP	This study
UL_1-30411	Genome-wide mapping	CAAGATCGCGTGGAAATCAA	CCTCCGGTGCTTGTCTGTAGG	I	30411182	SSLP	This study
UPSC_2-1401	Genome-wide mapping	GTTTGGATCAGTCCCAGCTC	TGAAAAAGTGGTGGAACCAA	II	1401050	SSLP	Pacurar et al. (2011)
UPSC_2-10706	Genome-wide mapping	TCATGCAAATCAATGCAGAAA	GGATGGCTATTTCCATGCAG	II	10705939	SSLP	Pacurar et al. (2011)
UPSC_2-19331	Genome-wide mapping	ACGTATGCACCGCAACAAT	GGCGAGGGATACGAAAATGT	II	19330877	SSLP	Pacurar et al. (2011)
UPSC_3-3716	Genome-wide mapping	TAATGGTGGCCCAATCTCAT	AATTCCAAATGGAGCCACAA	III	3716008	SSLP	Pacurar et al. (2011)
UPSC_3-8894	Genome-wide mapping	CCTTCGTTGGGTCAAACAGT	CGGCAAAATGGTGAGAAGTT	III	8894047	SSLP	Pacurar et al. (2011)
UPSC_3-22022	Genome-wide mapping	TGGTCTCCAAATTATGTGTCA	TCATGAGGTAGGAGGGGGTA	III	22022476	SSLP	Pacurar et al. (2011)
UPSC_4-934	Genome-wide mapping	ATCATGTGACTGCGTTTTGC	GAAACTAAAGCCATGAATGGTGA	IV	934405	SSLP	Pacurar et al. (2011)
UL_4-9196	Genome-wide mapping	TTGGGTGTGACTCCAACTGA	CATACCGCACACAGCATCTC	IV	9195602	SSLP	This study
UPSC_4-12273	Genome-wide mapping	TCAACCAATCGCCTTAGTCA	TTTCCATTTGCATGCTCGTA	IV	12273372	SSLP	Pacurar et al. (2011)
UPSC_4-17110	Genome-wide mapping	ATCGCTAACCCTCTCACGAA	TGGCTGTGAGTGAGTGAAGA	IV	17110453	SSLP	Pacurar et al. (2011)
UPSC_5-4171	Genome-wide mapping	TCACATGAACTTCCACACCTG	TTGCAAGGTCACAAATGGAA	V	4171285	SSLP	Pacurar et al. (2011)
UPSC_5-17224	Genome-wide mapping	CGATCGCTTCTATTCACCTTC	CCCATCTCAAACACACCAAA	V	17223573	SSLP	Pacurar et al. (2011)
UPSC_5-17304	Genome-wide mapping	TCGTAGAGAAGAAAATCTCGGAAA	TTGCGCTCAACTAATGTTTCT	V	17304218	SSLP	Pacurar et al. (2011)
UL_5-26876	Genome-wide mapping	GCCCATGTTGTTCCACTCAAT	ATTCGTTTTCGCAACACCAG	V	26875626	SSLP	This study

4.19.3. List of primers for cloning

TABLE 4: 5'-phos, 5'-phosphorylation.

Fw. No.	Rev. No.	Fw. sequence	Rev. sequence	5'-phos.	Template	Construct Name
ULC-1	ULC-2	GATGAAAACGTGTTTGGATTCT	AAATCTCAAAATTTAACAAATCTAAACATATGTTTCTTATATGAAAAA	no	KiI-0 gDNA	gFLM(KiI-0)ΔLINE
ULC-3	ULC-4	AGAAACAATATGTTTAGATTGTTAAAATTTGAGATTGGTAATT	AATATCTTATATAGTTTGGAGCTCT	no	KiI-0 gDNA	gFLM(KiI-0)ΔLINE
ULC-12		CATATAAGAAACAATATGTTTAGGGCCGCGCCAATTTGAGATTTGGTAATTATA		yes	pDP34	pDP34-AscI
ULC-16	ULC-17	GGCGCGCCATCGGATTACCAAAAACAAG	GGCGCGCCATCTAAACATATGTTTCTTTTTT	no	KiI-0 gDNA	KiI-0-LINE
ULC-20		ATTTTTAATTTGTGAAATTTGCACACAAAAAAAAAATGCTTTCTACAAGA		yes	pUL11	INT6+CAC
ULC-21		ATTTTTAATTTGTGAAATTTGCACAAAAAAAAAAAATGCTTTCTACACAAGA		yes	pUL11	INT6+CAA
ULC-22		ATTTTTAATTTGTGAAATTTGCAACAAAAAAAAATGCTTTCTACAAGA		yes	pUL11	INT6+ACA
ULC-23		ATTTTTAATTTGTGAAATTTGCAAAAAAAAAAAAATGCTTTCTACAAGATAGATTAATT		yes	pUL11	INT6+AAA
ULC-26		CGACTGGATTAAAAAGTCTTAACAACGACAAGCTTAAAA		yes	pUL11	PRO1ΔT
ULC-27		CGAGTTGGAGGGAAAAAAAAACCCAAGAAAAAGAATAAAAAGCA		yes	pUL11	PRO2+AAACC
ULC-28	ULC-29	TTTTAATTTGTGAAATTTTTCTACAAGATAGATTAATTTCTTG	TAATCTATCTGTAGAAAAATTTCACAAATTAAAAATACCG	no	pUL11	INT6+Δ17bp
ULC-30	ULC-31	AAAAAAAAACGAGTTGGAGAAAAAGAATAAAAAGCAAAA	TGCTTTTTATTCTTTTTCTCCAACTCGTTTTTTTTTG	no	pUL11	PRO2+Δ16bp
ULC-40	ULC-41	TTATTTTCTCTATCTCGTTTAGTTGGAGAAAAAAGAATA	ATTCTTTTTTCTCCAACTAAACGAGATAGAGAAAATAATAGTA	no	pUL31	PolyA_2xΔa
ULC-28	ULC-29	TTTTTAATTTGTGAAATTTTTCTACAAGATAGATTAATTTCTTG	TAATCTATCTGTAGAAAAAATTTCACAAATTAAAAATACCG	no	pUL34	PolyA_3xΔ
ULC-28	ULC-29	TTTTTAATTTGTGAAATTTTTCTACAAGATAGATTAATTTCTTG	TAATCTATCTGTAGAAAAAATTTCACAAATTAAAAATACCG	no	pUL29	PolyA_2xΔb
ULC-36	ULC-37	GCAATGTTTAGTAGCTGATGTAAGAAAGTTTCTTACAGAT	TGTAAGAAACTTTCTTCACATCAGCTACTAAACATTGCAAGAA	no	pUL35	PolyA_4xΔ
ULC-32	ULC-33	GGTGAAACCCTAATGAGAGCAGTTTAAAACTCTTGATCA	CAAGAGTTTAAAACTGCTCTCATTAGGGTTTCACCTG	no	pUL37	PolyA_5xΔ
ULC-38	ULC-39	TTTCTCCCCTCTTCTTCTTTTTTTTGTCTACTTGTG	CAAGTAGACAAAAGAAAAAAAAGAAGAAGAGAGGGGAGAAAAAT	no	pUL38	PolyA_6xΔ
ULC-42	ULC-43	CCGGTGACGAGTAAGAAGATTGCTAGTTCCATCTGTTG	TGAACAACAGAGATGGAACTAGCAATCTTCTTACTCGTCACC	no	pUL11	INT1Δ373bp
ULC-44	ULC-45	GATGGTTTATTGTAGTGCTGACCATACAAGATACATTAT	AGACATAATGTATCTTGTGATGGTCCAGCACTACAATAAAC	no	pUL11	INT1Δ101bp
ULC-46	ULC-47	AACCCTCCAAATAAAAGAGCAAAAAGCATTGAGCGTC	GACGCTCAATGCTTTTTGCTCTTTTTATTTGGAGGGTTTAAAA	no	pUL11	PRO_Δ225bp
ULC-48		CGAGTTGGAGGGGAAAAAAAAACAAGAAAAAGAATAAAAAGCA		yes	pUL11	PRO2+GGAAA
ULC-49		CGAGTTGGAGGGGAAAAAAAAACCAAGAAAAAGAATAAAAAGCA		yes	pUL11	PRO2+GGAC
ULC-50		CGAGTTGGAGGGAAATAAAAAACCAAGAAAAAGAATAAAAAGCA		yes	pUL11	PRO2+GATAC
ULC-57	ULC-58	TTCGCCCTTGCTTGGGATTTAAAGAAAAACCTTAAAAACCAATTCTG	AACAGAAATTGGTTTTTAAGGTTTTCTTTAAAAATCCCAAGCAAGGGC	no	pUL11	PRO_254bp

4.20. List of constructs

TABLE 5: Kan, kanamycin; Spec, spectinomycin; Kil-0, Killean-0; Nd-1, Niedersenz-1.

Plasmid No.	Name	Vector	E. coli resistance	Agrobact. strain	Plant resistance	Transgenic background
1	gFLM(Kil-0)ΔLINE	pCR2.1-TOPO	Kan	-	-	-
2	gFLM(Kil-0)ΔLINE	pGREEN0229	Spec	GV3101+psoup	BASTA	Kil-0
3	gFLM(Col)-AscI	pDP34/pGREENIIs	Spec	GV3101+psoup	BASTA	Kil-0 / *flm-3*
7	Kil-0-LINE	pCR2.1-TOPO	Kan	-	-	-
10	gFLM(Col)+LINE	pDP34/pGREENIIs	Spec	GV3101+psoup	BASTA	*flm-3*
11	Col^{FLM}	pDONR201	Kan	-	-	-
20	INT6+^{CAC}	pDONR201	Kan	-	-	-
21	INT6+^{CAA}	pDONR201	Kan	-	-	-
22	INT6+^{ACA}	pDONR201	Kan	-	-	-
23	INT6+^{AAA}	pDONR201	Kan	-	-	-
26	PRO1^{ΔT}	pDONR201	Kan	-	-	-
27	PRO2+^{AAACC}	pDONR201	Kan	-	-	-
28	INT6+^{Δ17bp}	pDONR201	Kan	-	-	-
29	PRO2+^{Δ16bp}	pDONR201	Kan	-	-	-
34	PolyA_2xΔa	pDONR201	Kan	-	-	-
35	PolyA_3xΔ	pDONR201	Kan	-	-	-
36	PolyA_2xΔb	pDONR201	Kan	-	-	-
37	PolyA_4xΔ	pDONR201	Kan	-	-	-
38	PolyA_5xΔ	pDONR201	Kan	-	-	-
39	PolyA_6xΔ	pDONR201	Kan	-	-	-
40	INT1^{Δ373bp}	pDONR201	Kan	-	-	-
41	INT1^{Δ101bp}	pDONR201	Kan	-	-	-
42	PRO^{Δ225bp}	pDONR201	Kan	-	-	-
67	PRO2+^{GGAAA}	pDONR201	Kan	-	-	-
68	PRO2+^{GGAAC}	pDONR201	Kan	-	-	-
69	PRO2+^{GATAC}	pDONR201	Kan	-	-	-
70	PRO^{254bp}	pDONR201	Kan	-	-	-
12	Col^{FLM}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
45	INT6+^{CAC}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
46	INT6+^{CAA}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
47	INT6+^{ACA}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
73	INT6+^{AAA}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
50	PRO1^{ΔT}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
51	PRO2+^{AAACC}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
52	INT6+^{Δ17bp}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
53	PRO2+^{Δ16bp}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
58	PolyA_2xΔa	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
59	PolyA_3xΔ	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
60	PolyA_2xΔb	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
61	PolyA_4xΔ	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
62	PolyA_5xΔ	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
63	PolyA_6xΔ	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
64	INT1^{Δ373bp}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
65	INT1^{Δ101bp}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
66	PRO^{Δ225bp}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
75	PRO2+^{GGAAA}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
76	PRO2+^{GGAAC}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
77	PRO2+^{GATAC}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1
78	PRO^{254bp}	pFAST-R07	-	GV3101	FAST (Shimada et al., 2009)	Nd-1

5.   References

Airoldi, C.A., McKay, M., and Davies, B. (2015). MAF2 Is regulated by temperature-dependent splicing and represses flowering at low temperatures in parallel with FLM. PLoS One **10,** e0126516.

Alberto, F.J., Aitken, S.N., Alia, R., Gonzalez-Martinez, S.C., Hanninen, H., Kremer, A., Lefevre, F., Lenormand, T., Yeaman, S., Whetten, R., and Savolainen, O. (2013). Potential for evolutionary responses to climate change - evidence from tree populations. **Glob Chang Biol 19,** 1645-1661.

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

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., and Ecker, J.R. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science **301,** 653-657.

Amasino, R. (2010). Seasonal and developmental timing of flowering. Plant J **61,** 1001-1013.

Amasino, R.M., and Michaels, S.D. (2010). The timing of flowering. Plant Physiol **154,** 516-520.

Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome Res **22,** 2008-2017.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics **31,** 166-169.

Anderson, J.T., Lee, C.R., and Mitchell-Olds, T. (2011). Life-history QTLs and natural selection on flowering time in Boechera stricta, a perennial relative of Arabidopsis. Evolution **65,** 771-787.

Angel, A., Song, J., Dean, C., and Howard, M. (2011). A Polycomb-based switch underlying quantitative epigenetic memory. Nature **476,** 105-108.

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

Balasubramanian, S., and Weigel, D. (2006). Temperature induced flowering in Arabidopsis thaliana. Plant Signal Behav **1,** 227-228.

Balasubramanian, S., Sureshkumar, S., Lempe, J., and Weigel, D. (2006). Potent induction of Arabidopsis thaliana flowering by elevated growth temperature. PLoS Genet **2,** e106.

Bandelt, H.J., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol **16,** 37-48.

Blazquez, M.A., Ahn, J.H., and Weigel, D. (2003). A thermosensory pathway controlling flowering time in Arabidopsis thaliana. Nat Genet **33,** 168-171.

Blumel, M., Dally, N., and Jung, C. (2015). Flowering time regulation in crops-what did we learn from Arabidopsis? Curr Opin Biotechnol **32,** 121-129.

Bouche, F., Lobet, G., Tocquin, P., and Perilleux, C. (2016). FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. Nucleic Acids Res **44,** D1167-1171.

Box, M.S., Coustham, V., Dean, C., and Mylne, J.S. (2011). Protocol: A simple phenol-based method for 96-well extraction of high quality RNA from Arabidopsis. Plant Methods **7,** 7.

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics **19,** 889-890.

Caicedo, A.L., Stinchcombe, J.R., Olsen, K.M., Schmitt, J., and Purugganan, M.D. (2004). Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. Proc Natl Acad Sci U S A **101,** 15670-15675.

Cao, J., Schneeberger, K., Ossowski, S., Gunther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., Wang, X., Ott, F., Muller, J., Alonso-Blanco, C., Borgwardt, K., Schmid, K.J., and Weigel, D. (2011). Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet **43,** 956-963.

Capovilla, G., Schmid, M., and Posé, D. (2015). Control of flowering by ambient temperature. J Exp Bot **66,** 59-69.

Carrillo Oesterreich, F., Bieberstein, N., and Neugebauer, K.M. (2011). Pause locally, splice globally. Trends Cell Biol **21,** 328-335.

Chen, X., and Li, B.L. (2003). Effect of global climate change and human disturbances on tree diversity of the forest regenerating from clear-cuts of mixed broadleaved Korean pine forest in Northeast China. Chemosphere **51,** 215-226.

Choi, J., Hyun, Y., Kang, M.J., In Yun, H., Yun, J.Y., Lister, C., Dean, C., Amasino, R.M., Noh, B., Noh, Y.S., and Choi, Y. (2009). Resetting and regulation of FLOWERING LOCUS C expression during Arabidopsis reproductive development. Plant J **57,** 918-931.

Choi, K., Kim, J., Hwang, H.-J., Kim, S., Park, C., Kim, S.Y., and Lee, I. (2011). The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors. The Plant Cell **23,** 289-303.

Chow, C.N., Zheng, H.Q., Wu, N.Y., Chien, C.H., Huang, H.D., Lee, T.Y., Chiang-Hsieh, Y.F., Hou, P.F., Yang, T.Y., and Chang, W.C. (2016). PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. Nucleic Acids Res **44,** D1154-1160.

Clark, R.M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T.T., Fu, G., Hinds, D.A., Chen, H., Frazer, K.A., Huson, D.H., Scholkopf, B., Nordborg, M., Ratsch, G., Ecker, J.R., and Weigel, D. (2007). Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. Science **317,** 338-342.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16,** 735-743.

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

Coustham, V., Li, P., Strange, A., Lister, C., Song, J., and Dean, C. (2012). Quantitative modulation of polycomb silencing underlies natural variation in vernalization. Science **337,** 584-587.

Crevillen, P., and Dean, C. (2011). Regulation of the floral repressor gene FLC: the complexity of transcription in a chromatin context. Curr Opin Plant Biol **14,** 38-44.

De Bodt, S., Raes, J., Florquin, K., Rombauts, S., Rouze, P., Theissen, G., and Van de Peer, Y. (2003). Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J Mol Evol **56,** 573-586.

de Folter, S., and Angenent, G.C. (2006). trans meets cis in MADS science. Trends in plant science **11,** 224-231.

de Folter, S., Immink, R.G., Kieffer, M., Parenicova, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M., Davies, B., and Angenent, G.C. (2005). Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. Plant Cell **17,** 1424-1433.

De Lucia, F., Crevillen, P., Jones, A.M., Greb, T., and Dean, C. (2008). A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. Proc Natl Acad Sci U S A **105,** 16831-16836.

Deng, W., Ying, H., Helliwell, C.A., Taylor, J.M., Peacock, W.J., and Dennis, E.S. (2011). FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of Arabidopsis. Proc Natl Acad Sci U S A **108,** 6680-6685.

Distelfeld, A., Li, C., and Dubcovsky, J. (2009). Regulation of flowering in temperate cereals. Curr Opin Plant Biol **12,** 178-184.

Duc, C., Sherstnev, A., Cole, C., Barton, G.J., and Simpson, G.G. (2013). Transcription termination and chimeric RNA formation controlled by Arabidopsis thaliana FPA. PLoS Genet **9,** e1003867.

Dunlap, J.C. (1999). Molecular bases for circadian clocks. Cell **96,** 271-290.

el-Lithy, M.E., Bentsink, L., Hanhart, C.J., Ruys, G.J., Rovito, D., Broekhof, J.L., van der Poel, H.J., van Eijk, M.J., Vreugdenhil, D., and Koornneef, M. (2006). New Arabidopsis recombinant inbred line populations genotyped using SNPWave and their use for mapping flowering-time quantitative trait loci. Genetics **172,** 1867-1876.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol **14,** 2611-2620.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res **39,** W29-37.

Finnegan, E.J., and Dennis, E.S. (2007). Vernalization-induced trimethylation of histone H3 lysine 27 at FLC is not maintained in mitotically quiescent cells. Curr Biol **17,** 1978-1983.

Fitter, A.H., and Fitter, R.S. (2002). Rapid changes in flowering time in British plants. Science **296,** 1689-1691.

Fowler, S., Lee, K., Onouchi, H., Samach, A., Richardson, K., Coupland, G., and Putterill, J. (1999). GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. Embo J **18,** 4679-4688.

Franks, S.J., and Weis, A.E. (2008). A change in climate causes rapid evolution of multiple life-history traits and their interactions in an annual plant. J Evol Biol **21,** 1321-1334.

Fujita, T., Asano, Y., Ohtsuka, J., Takada, Y., Saito, K., Ohki, R., and Fujii, H. (2013). Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). Sci Rep **3,** 3171.

Fusaro, A.F., Bocca, S.N., Ramos, R.L., Barroco, R.M., Magioli, C., Jorge, V.C., Coutinho, T.C., Rangel-Lima, C.M., De Rycke, R., Inze, D., Engler, G., and Sachetto-Martins, G. (2007). AtGRP2, a cold-induced nucleo-cytoplasmic RNA-binding protein, has a role in flower and seed development. Planta **225,** 1339-1351.

Gazzani, S., Gendall, A.R., Lister, C., and Dean, C. (2003). Analysis of the molecular basis of flowering time variation in Arabidopsis accessions. Plant Physiol **132,** 1107-1114.

The 1001 Genomes Consortium (2016). 1,135 genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. Cell.

Geraldo, N., Baurle, I., Kidou, S., Hu, X., and Dean, C. (2009). FRIGIDA delays flowering in Arabidopsis via a cotranscriptional mechanism involving direct interaction with the nuclear cap-binding complex. Plant Physiol **150,** 1611-1618.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res **40,** D1178-1186.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet **17,** 333-351.

Gramzow, L., Ritz, M.S., and Theissen, G. (2010). On the origin of MADS-domain transcription factors. Trends Genet **26,** 149-153.

Gu, X., Le, C., Wang, Y., Li, Z., Jiang, D., and He, Y. (2013). Arabidopsis FLC clade members form flowering-repressor complexes coordinating responses to endogenous and environmental cues. Nat Commun **4,** 1947.

Han, S.P., Tang, Y.H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. Biochem J **430,** 379-392.

Hansson, M.D., Rzeznicka, K., Rosenback, M., Hansson, M., and Sirijovski, N. (2008). PCR-mediated deletion of plasmid DNA. Anal Biochem **375,** 373-375.

Harmer, S.L., Panda, S., and Kay, S.A. (2001). Molecular bases of circadian rhythms. Annu Rev Cell Dev Biol **17,** 215-253.

Hartmann, U., Hohmann, S., Nettesheim, K., Wisman, E., Saedler, H., and Huijser, P. (2000). Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. Plant J **21,** 351-360.

Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S., and Mullineaux, P.M. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. Plant Mol Biol **42,** 819-832.

Helliwell, C.A., Wood, C.C., Robertson, M., James Peacock, W., and Dennis, E.S. (2006). The Arabidopsis FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. Plant J **46,** 183-192.

Hemming, M.N., and Trevaskis, B. (2011). Make hay when the sun shines: the role of MADS-box genes in temperature-dependant seasonal flowering responses. Plant Sci **180**, 447-453.

Henschel, K., Kofuji, R., Hasebe, M., Saedler, H., Munster, T., and Theissen, G. (2002). Two ancient classes of MIKC-type MADS-box genes are present in the moss Physcomitrella patens. Mol Biol Evol **19,** 801-814.

Hepworth, J., and Dean, C. (2015). FLOWERING LOCUS C's lessons: conserved chromatin switches underpinning developmental timing and adaptation. Plant Physiol **168,** 1237-1245.

Hoffmann, M.H. (2002). Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). Journal of Biogeography **29,** 125 - 134.

Hong, R.L., Hamaguchi, L., Busch, M.A., and Weigel, D. (2003). Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. Plant Cell **15,** 1296-1309.

Huang, C.R., Burns, K.H., and Boeke, J.D. (2012). Active transposition in genomes. Annu Rev Genet **46,** 651-675.

Imaizumi, T. (2010). Arabidopsis circadian clock and photoperiodism: time to think about location. Curr Opin Plant Biol **13,** 83-89.

Imaizumi, T., and Kay, S.A. (2006). Photoperiodic control of flowering: not only by coincidence. Trends in plant science **11,** 550-558.

Imaizumi, T., Tran, H.G., Swartz, T.E., Briggs, W.R., and Kay, S.A. (2003). FKF1 is essential for photoperiodic-specific light signalling in Arabidopsis. Nature **426,** 302-306.

Immink, R.G., Posé, D., Ferrario, S., Ott, F., Kaufmann, K., Valentim, F.L., de Folter, S., van der Wal, F., van Dijk, A.D., Schmid, M., and Angenent, G.C. (2012). Characterization of SOC1's central role in flowering by the identification of its upstream and downstream regulators. Plant Physiol **160,** 433-449.

Isken, O., and Maquat, L.E. (2007). Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes Dev **21,** 1833-1856.

Jeon, J., and Kim, J. (2011). FVE, an Arabidopsis homologue of the retinoblastoma-associated protein that regulates flowering time and cold response, binds to chromatin as a large multiprotein complex. Mol Cells **32,** 227-234.

Johanson, U. (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science **290,** 344-347.

John, S., Sabo, P.J., Canfield, T.K., Lee, K., Vong, S., Weaver, M., Wang, H., Vierstra, J., Reynolds, A.P., Thurman, R.E., and Stamatoyannopoulos, J.A. (2013). Genome-scale mapping of DNase I hypersensitivity. Curr Protoc Mol Biol **Chapter 27,** Unit 21 27.

Jung, J.H., Seo, P.J., Ahn, J.H., and Park, C.M. (2012). Arabidopsis RNA-binding protein FCA regulates microRNA172 processing in thermosensory flowering. J Biol Chem **287,** 16007-16016.

Jung, J.H., Lee, S., Yun, J., Lee, M., and Park, C.M. (2014). The miR172 target TOE3 represses AGAMOUS expression during Arabidopsis floral patterning. Plant Sci **215-216,** 29-38.

Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol Cell **12,** 5-14.

Kalyna, M., Simpson, C.G., Syed, N.H., Lewandowska, D., Marquez, Y., Kusenda, B., Marshall, J., Fuller, J., Cardle, L., McNicol, J., Dinh, H.Q., Barta, A., and Brown, J.W. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic Acids Res **40,** 2454-2469.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., and Weigel, D. (1999). Activation tagging of the floral inducer FT. Science **286,** 1962-1965.

Kaufmann, K., Muino, J.M., Osteras, M., Farinelli, L., Krajewski, P., and Angenent, G.C. (2010). Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nat Protoc **5,** 457-472.

Kaufmann, K., Muino, J.M., Jauregui, R., Airoldi, C.A., Smaczniak, C., Krajewski, P., and Angenent, G.C. (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biol **7,** e1000090.

Kersey, P.J., Allen, J.E., Armean, I., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C., Humphrey, J., Kerhornou, A., Khobova, J., Aranganathan, N.K., Langridge, N., Lowy, E., McDowall, M.D., Maheswari, U., Nuhn, M., Ong, C.K., Overduin, B., Paulini, M., Pedro, H., Perry, E., Spudich, G., Tapanari, E., Walts, B., Williams, G., Tello-Ruiz, M., Stein, J., Wei, S., Ware, D., Bolser, D.M., Howe, K.L., Kulesha, E., Lawson, D., Maslen, G., and Staines, D.M. (2016). Ensembl Genomes 2016: more genomes, more complexity. Nucleic Acids Res **44,** D574-580.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nat Rev Mol Cell Biol **13,** 700-712.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14,** R36.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science **303,** 521-523.

Kippes, N., Debernardi, J.M., Vasquez-Gross, H.A., Akpinar, B.A., Budak, H., Kato, K., Chao, S., Akhunov, E., and Dubcovsky, J. (2015). Identification of the VERNALIZATION 4 gene reveals the origin of spring growth habit in ancient wheats from South Asia. Proc Natl Acad Sci U S A **112,** E5401-5410.

Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P., and Weisshaar, B. (2012). GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. Nucleic Acids Res **40,** D1211-1215.

Kobayashi, Y., and Weigel, D. (2007). Move on up, it's time for change - mobile signals controlling photoperiod-dependent flowering. Genes Dev **21,** 2371-2384.

Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., and Araki, T. (1999). A pair of related genes with antagonistic roles in mediating flowering signals. Science **286,** 1960-1962.

Koncz, C., Dejong, F., Villacorta, N., Szakonyi, D., and Koncz, Z. (2012). The spliceosome-activating complex: molecular mechanisms underlying the function of a pleiotropic regulator. Front Plant Sci **3,** 9.

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

Koornneef, M., Alonso-Blanco, C., Peeters, A.J.M., and Soppe, W. (1998). Genetic control of flowering time in arabidopsis. Annu Rev Plant Phys **49,** 345-370.

Kumar, S.V., and Wigge, P.A. (2010). H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. Cell **140,** 136-147.

Kumar, S.V., Lucyshyn, D., Jaeger, K.E., Alos, E., Alvey, E., Harberd, N.P., and Wigge, P.A. (2012). Transcription factor PIF4 controls the thermosensory activation of flowering. Nature **484,** 242-245.

Kuromori, T., Hirayama, T., Kiyosue, Y., Takabe, H., Mizukado, S., Sakurai, T., Akiyama, K., Kamiya, A., Ito, T., and Shinozaki, K. (2004). A collection of 11 800 single-copy Ds transposon insertion lines in Arabidopsis. Plant J **37,** 897-905.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics **23,** 2947-2948.

Lee, H., Yoo, S.J., Lee, J.H., Kim, W., Yoo, S.K., Fitzgerald, H., Carrington, J.C., and Ahn, J.H. (2010). Genetic framework for flowering-time regulation by ambient temperature-responsive miRNAs in Arabidopsis. Nucleic Acids Res **38,** 3081-3093.

Lee, J.H., Yoo, S.J., Park, S.H., Hwang, I., Lee, J.S., and Ahn, J.H. (2007). Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes Dev **21,** 397-402.

Lee, J.H., Ryu, H.S., Chung, K.S., Posé, D., Kim, S., Schmid, M., and Ahn, J.H. (2013). Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. Science **342,** 628-632.

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

Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C.A., Ito, T., Meyerowitz, E., and Yu, H. (2008). A repressor complex governs the integration of flowering signals in Arabidopsis. Dev Cell **15,** 110-120.

Li, P., Filiault, D., Box, M.S., Kerdaffrec, E., van Oosterhout, C., Wilczek, A.M., Schmitt, J., McMullan, M., Bergelson, J., Nordborg, M., and Dean, C. (2014). Multiple FLC haplotypes defined by independent cis-regulatory variation underpin life history diversity in Arabidopsis thaliana. Genes Dev **28,** 1635-1640.

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

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., Carrington, J.C., Doerge, R.W., Colot, V., and Martienssen, R. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature **430,** 471-476.

Lisch, D. (2009). Epigenetic regulation of transposable elements in plants. Annu Rev Plant Biol **60,** 43-66.

Liu, C., Zhou, J., Bracha-Drori, K., Yalovsky, S., Ito, T., and Yu, H. (2007). Specification of Arabidopsis floral meristem identity by repression of flowering time genes. Development **134,** 1901-1910.

Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010). Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. Science **327,** 94-97.

Liu, L., Liu, C., Hou, X., Xi, W., Shen, L., Tao, Z., Wang, Y., and Yu, H. (2012). FTIP1 is an essential regulator required for florigen transport. PLoS Biol **10,** e1001313.

Long, Q., Rabanal, F.A., Meng, D., Huber, C.D., Farlow, A., Platzer, A., Zhang, Q., Vilhjalmsson, B.J., Korte, A., Nizhynska, V., Voronin, V., Korte, P., Sedman, L., Mandakova, T., Lysak, M.A., Seren, U., Hellmann, I., and Nordborg, M. (2013). Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nat Genet **45,** 884-890.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15,** 550.

Lukowitz, W., Gillmor, C.S., and Scheible, W.R. (2000). Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you. Plant Physiol **123,** 795-805.

Marquardt, S., Raitskin, O., Wu, Z., Liu, F., Sun, Q., and Dean, C. (2014). Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription. Mol Cell **54,** 156-165.

Marquez, Y., Brown, J.W., Simpson, C., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Res **22,** 1184-1195.

Martinez-Castilla, L.P., and Alvarez-Buylla, E.R. (2003). Adaptive evolution in the Arabidopsis MADS-box gene family inferred from its complete resolved phylogeny. Proc Natl Acad Sci U S A **100,** 13407-13412.

Mascarenhas, D., Mettler, I.J., Pierce, D.A., and Lowe, H.W. (1990). Intron-mediated enhancement of heterologous gene expression in maize. Plant Mol Biol **15,** 913-920.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H., Lim, J., Shyr, C., Tan, G., Zhou, M., Lenhard, B., Sandelin, A., and Wasserman, W.W. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res **42,** D142-147.

Mathieu, J., Warthmann, N., Kuttner, F., and Schmid, M. (2007). Export of FT protein from phloem companion cells is sufficient for floral induction in Arabidopsis. Curr Biol **17,** 1055-1060.

Mathieu, J., Yant, L.J., Murdter, F., Kuttner, F., and Schmid, M. (2009). Repression of Flowering by the miR172 Target SMZ. PLoS Biol **7,** -.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A.E., and Wingender, E. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res **34,** D108-110.

McCue, A.D., and Slotkin, R.K. (2012). Transposable element small RNAs as regulators of gene expression. Trends Genet **28,** 616-623.

Mendez-Vigo, B., Martinez-Zapater, J.M., and Alonso-Blanco, C. (2013). The flowering repressor SVP underlies a novel Arabidopsis thaliana QTL interacting with the genetic background. PLoS Genet **9,** e1003289.

Mendez-Vigo, B., de Andres, M.T., Ramiro, M., Martinez-Zapater, J.M., and Alonso-Blanco, C. (2010). Temporal analysis of natural variation for the rate of leaf production and its relationship with flowering initiation in Arabidopsis thaliana. J Exp Bot **61,** 1611-1623.

Meyer, R.S., and Purugganan, M.D. (2013). Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet **14,** 840-852.

Michaels, S.D., and Amasino, R.M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell **11,** 949-956.

Michaels, S.D., and Amasino, R.M. (2001). Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization. Plant Cell **13,** 935-941.

Michaels, S.D., He, Y., Scortecci, K.C., and Amasino, R.M. (2003). Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. Proc Natl Acad Sci U S A **100,** 10102-10107.

Mockler, T., Yang, H.Y., Yu, X.H., Parikh, D., Cheng, Y.C., Dolan, S., and Lin, C.T. (2003). Regulation of photoperiodic flowering by Arabidopsis photoreceptors. Proc Natl Acad Sci U S A **100,** 2140-2145.

Moore, F.C., and Lobell, D.B. (2015). The fingerprint of climate trends on European crop yields. Proc Natl Acad Sci U S A **112,** 2670-2675.

Mora, C., Caldwell, I.R., Caldwell, J.M., Fisher, M.R., Genco, B.M., and Running, S.W. (2015). Suitable Days for Plant Growth Disappear under Projected Climate Change: Potential Human and Biotic Vulnerability. PLoS Biol **13,** e1002167.

Nakano, Y., Higuchi, Y., Sumitomo, K., and Hisamatsu, T. (2013). Flowering retardation by high temperature in chrysanthemums: involvement of FLOWERING LOCUS T-like 3 gene repression. J Exp Bot **64,** 909-920.

Ng, M., and Yanofsky, M.F. (2001). Function and evolution of the plant MADS-box gene family. Nat Rev Genet **2**, 186-195.

Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N.A., Shah, C., Wall, J.D., Wang, J., Zhao, K., Kalbfleisch, T., Schulz, V., Kreitman, M., and Bergelson, J. (2005). The pattern of polymorphism in Arabidopsis thaliana. PLoS Biol **3,** e196.

O'Malley, R.C., Huang, S.S., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. Cell **165,** 1280-1292.

O'Neill, C.M., Morgan, C., Kirby, J., Tschoep, H., Deng, P.X., Brennan, M., Rosas, U., Fraser, F., Hall, C., Gill, S., and Bancroft, I. (2008). Six new recombinant inbred populations for the study of quantitative traits in Arabidopsis thaliana. Theor Appl Genet **116,** 623-634.

Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N., and Weigel, D. (2008). Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res **18,** 2024-2033.

Ossowski, S., Schneeberger, K., Lucas-Lledo, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D., and Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. Science **327,** 92-94.

Pachauri, R.K., and Meyer, L.A. (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In IPCC, R.K. Pachauri and L.A. Meyer, eds (Geneva, Switzerland), pp. 151.

Pacurar, D.I., Pacurar, M.L., Street, N., Bussell, J.D., Pop, T.I., Gutierrez, L., and Bellini, C. (2012). A collection of INDEL markers for map-based cloning in seven Arabidopsis accessions. J Exp Bot **63,** 2491-2501.

Pajoro, A., Madrigal, P., Muino, J.M., Matus, J.T., Jin, J., Mecchia, M.A., Debernardi, J.M., Palatnik, J.F., Balazadeh, S., Arif, M., O'Maoileidigh, D.S., Wellmer, F., Krajewski, P., Riechmann, J.L., Angenent, G.C., and Kaufmann, K. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome Biol **15,** R41.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res **29,** e45.

Piva, F., Giulietti, M., Nocchi, L., and Principato, G. (2009). SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. Bioinformatics **25,** 1211-1213.

Piva, F., Giulietti, M., Burini, A.B., and Principato, G. (2012). SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. Hum Mutat **33,** 81-85.

Posé, D., Yant, L., and Schmid, M. (2012). The end of innocence: flowering networks explode in complexity. Curr Opin Plant Biol **15,** 45-50.

Posé, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G.C., Immink, R.G., and Schmid, M. (2013). Temperature-dependent regulation of flowering by antagonistic FLM variants. Nature **503,** 414-417.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics **155,** 945-959.

Putterill, J., Laurie, R., and 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., and Coupland, G. (1995). The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to Zinc-finger transcription factors. Cell **80,** 847-857.

Quadrana, L., Bortolini Silveira, A., Mayhew, G.F., LeBlanc, C., Martienssen, R.A., Jeddeloh, J.A., and Colot, V. (2016). The mobilome and its impact at the species level. Elife **5**.

Questa, J.I., Song, J., Geraldo, N., An, H., and Dean, C. (2016). Arabidopsis transcriptional repressor VAL1 triggers Polycomb silencing at FLC during vernalization. Science **353,** 485-488.

Ratcliffe, O.J., Nadzan, G.C., Reuber, T.L., and Riechmann, J.L. (2001). Regulation of flowering in Arabidopsis by an FLC homologue. Plant Physiol **126,** 122-132.

Ratcliffe, O.J., Kumimoto, R.W., Wong, B.J., and Riechmann, J.L. (2003). Analysis of the Arabidopsis MADS AFFECTING FLOWERING gene family: MAF2 prevents vernalization by short periods of cold. Plant Cell **15,** 1159-1169.

Reddy, A.S. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol **58,** 267-294.

Reich, P.B., and Oleksyn, J. (2008). Climate warming will reduce growth and survival of Scots pine except in the far north. Ecol Lett **11,** 588-597.

Rose, A.B. (2008). Intron-mediated regulation of gene expression. Curr Top Microbiol Immunol **326,** 277-290.

Rosloski, S.M., Singh, A., Jali, S.S., Balasubramanian, S., Weigel, D., and Grbic, V. (2013). Functional analysis of splice variant expression of MADS AFFECTING FLOWERING 2 of Arabidopsis thaliana. Plant Mol Biol **81,** 57-69.

Roux, F., Touzet, P., Cuguen, J., and Le Corre, V. (2006). How to be early flowering: an evolutionary perspective. Trends in plant science **11,** 375-381.

Rozas, J., and Rozas, R. (1995). DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. Comput Appl Biosci **11,** 621-625.

Sakharkar, M.K., Chow, V.T., and Kangueane, P. (2004). Distributions of exons and introns in the human genome. In Silico Biol **4,** 387-393.

Salome, P.A., Bomblies, K., Laitinen, R.A., Yant, L., Mott, R., and Weigel, D. (2011). Genetic architecture of flowering-time variation in Arabidopsis thaliana. Genetics **188,** 421-433.

Sawano, A., and Miyawaki, A. (2000). Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. Nucleic Acids Res **28,** E78.

Schauer, S.E., Schluter, P.M., Baskar, R., Gheyselinck, J., Bolanos, A., Curtis, M.D., and Grossniklaus, U. (2009). Intronic regulatory elements determine the divergent expression patterns of AGAMOUS-LIKE6 subfamily members in Arabidopsis. Plant J **59,** 987-1000.

Schmid, K.J., Torjek, O., Meyer, R., Schmuths, H., Hoffmann, M.H., and Altmann, T. (2006). Evidence for a large-scale population structure of Arabidopsis thaliana from genome-wide single nucleotide polymorphism markers. Theor Appl Genet **112,** 1104-1114.

Schmitz, R.J., Schultz, M.D., Urich, M.A., Nery, J.R., Pelizzola, M., Libiger, O., Alix, A., McCosh, R.B., Chen, H., Schork, N.J., and Ecker, J.R. (2013). Patterns of population epigenomic diversity. Nature **495,** 193-198.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jorgensen, J.E., Weigel, D., and Andersen, S.U. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Methods **6,** 550-551.

Schneeberger, K., Ossowski, S., Ott, F., Klein, J.D., Wang, X., Lanz, C., Smith, L.M., Cao, J., Fitz, J., Warthmann, N., Henz, S.R., Huson, D.H., and Weigel, D. (2011). Reference-guided assembly of four diverse Arabidopsis thaliana genomes. Proc Natl Acad Sci U S A **108,** 10249-10254.

Schultz, S.T., Lynch, M., and Willis, J.H. (1999). Spontaneous deleterious mutation in Arabidopsis thaliana. Proc Natl Acad Sci U S A **96,** 11393-11398.

Scortecci, K., Michaels, S.D., and Amasino, R.M. (2003). Genetic interactions between FLM and other flowering-time genes in Arabidopsis thaliana. Plant Mol Biol **52,** 915-922.

Scortecci, K.C., Michaels, S.D., and Amasino, R.M. (2001). Identification of a MADS-box gene, FLOWERING LOCUS M, that represses flowering. Plant J **26,** 229-236.

Scotto-Lavino, E., Du, G., and Frohman, M.A. (2006). 3' end cDNA amplification using classic RACE. Nat Protoc **1,** 2742-2745.

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

Sheldon, C.C., Hills, M.J., Lister, C., Dean, C., Dennis, E.S., and Peacock, W.J. (2008). Resetting of FLOWERING LOCUS C expression after epigenetic repression by vernalization. Proc Natl Acad Sci U S A **105,** 2214-2219.

Shimada, T.L., Shimada, T., and Hara-Nishimura, I. (2010). A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana. Plant J **61,** 519-528.

Shore, P., and Sharrocks, A.D. (1995). The MADS-box family of transcription factors. Eur J Biochem **229,** 1-13.

Simon, M., Simon, A., Martins, F., Botran, L., Tisne, S., Granier, F., Loudet, O., and Camilleri, C. (2012). DNA fingerprinting and new tools for fine-scale discrimination of Arabidopsis thaliana accessions. Plant J **69,** 1094-1101.

Smaczniak, C., Immink, R.G., Muino, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q.D., Liu, S., Westphal, A.H., Boeren, S., Parcy, F., Xu, L., Carles, C.C., Angenent, G.C., and Kaufmann, K. (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proc Natl Acad Sci U S A **109,** 1560-1565.

Song, J., Irwin, J., and Dean, C. (2013). Remembering the prolonged cold of winter. Curr Biol **23,** R807-811.

Song, J., Angel, A., Howard, M., and Dean, C. (2012). Vernalization - a cold-induced epigenetic switch. J Cell Sci **125,** 3723-3731.

Spannagl, M., Nussbaumer, T., Bader, K.C., Martis, M.M., Seidel, M., Kugler, K.G., Gundlach, H., and Mayer, K.F.X. (2016). PGSB PlantsDB: updates to the database framework for comparative plant genome research. Nucleic Acids Res **44,** D1141-D1147.

Srikanth, A., and Schmid, M. (2011). Regulation of flowering time: all roads lead to Rome. Cell Mol Life Sci **68,** 2013-2037.

Staiger, D., and Brown, J.W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. Plant Cell **25,** 3640-3656.

Streitner, C., Koster, T., Simpson, C.G., Shaw, P., Danisman, S., Brown, J.W., and Staiger, D. (2012). An hnRNP-like RNA-binding protein affects alternative splicing by in vivo interaction with transcripts in Arabidopsis thaliana. Nucleic Acids Res **40,** 11240-11255.

Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., and Coupland, G. (2001). CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature **410,** 1116-1120.

Sullivan, A.M., Arsovski, A.A., Lempe, J., Bubb, K.L., Weirauch, M.T., Sabo, P.J., Sandstrom, R., Thurman, R.E., Neph, S., Reynolds, A.P., Stergachis, A.B., Vernot, B., Johnson, A.K., Haugen, E., Sullivan, S.T., Thompson, A., Neri, F.V., 3rd, Weaver, M., Diegel, M., Mnaimneh, S., Yang, A., Hughes, T.R., Nemhauser, J.L., Queitsch, C., and Stamatoyannopoulos, J.A. (2014). Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. Cell Rep **8,** 2015-2030.

Sureshkumar, S., Dent, C., Seleznev, A., Tasset, C., and Balasubramanian, S. (2016). Nonsense-mediated mRNA decay modulates FLM-dependent thermosensory flowering response in Arabidopsis. Nat Plants **2,** 16055.

Suter, B., Schnappauf, G., and Thoma, F. (2000). Poly(dA.dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. Nucleic Acids Res **28,** 4083-4089.

Swanson, M.S., and Dreyfuss, G. (1988). Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. Mol Cell Biol **8,** 2237-2241.

Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature **462,** 799-802.

Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from Domestication: Targeting Cis-Regulatory Elements for Crop Improvement. Trends in plant science **21,** 506-515.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28,** 2731-2739.

Thuiller, W., Lavorel, S., Araujo, M.B., Sykes, M.T., and Prentice, I.C. (2005). Climate change threats to plant diversity in Europe. Proc Natl Acad Sci U S A **102,** 8245-8250.

Toomajian, C., Hu, T.T., Aranzana, M.J., Lister, C., Tang, C., Zheng, H., Zhao, K., Calabrese, P., Dean, C., and Nordborg, M. (2006). A nonparametric test reveals selection for rapid flowering in the Arabidopsis genome. PLoS Biol **4,** e137.

Trapnell, C., and Salzberg, S.L. (2009). How to map billions of short reads onto genomes. Nat Biotechnol **27,** 455-457.

Trevaskis, B., Hemming, M.N., Dennis, E.S., and Peacock, W.J. (2007). The molecular basis of vernalization-induced flowering in cereals. Trends in plant science **12,** 352-357.

Tsuchiya, T., and Eulgem, T. (2013). An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. Proc Natl Acad Sci U S A **110,** E3535-3543.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res **35,** W71-74.

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

van Berloo, R. (2008). GGT 2.0: versatile software for visualization and analysis of genetic data. J Hered **99,** 232-236.

Van de Velde, J., Heyndrickx, K.S., and Vandepoele, K. (2014). Inference of transcriptional networks in Arabidopsis through conserved noncoding sequence analysis. Plant Cell **26,** 2729-2745.

Van de Velde, J., Van Bel, M., Van Eechoutte, D., and Vandepoele, K. (2016). A Collection of Conserved Non-Coding Sequences to Study Gene Regulation in Flowering Plants. Plant Physiol.

Verhage, L., Angenent, G.C., and Immink, R.G. (2014). Research on floral timing by ambient temperature comes into blossom. Trends in plant science **19,** 583-591.

Vierstra, J., Reik, A., Chang, K.H., Stehling-Sun, S., Zhou, Y., Hinkley, S.J., Paschon, D.E., Zhang, L., Psatha, N., Bendana, Y.R., O'Neil, C.M., Song, A.H., Mich, A.K., Liu, P.Q., Lee, G., Bauer, D.E., Holmes, M.C., Orkin, S.H., Papayannopoulou, T., Stamatoyannopoulos, G., Rebar, E.J., Gregory, P.D., Urnov, F.D., and Stamatoyannopoulos, J.A. (2015). Functional footprinting of regulatory DNA. Nat Methods **12,** 927-930.

Wang, B.B., and Brendel, V. (2004). The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. Genome Biol **5,** R102.

Wang, B.B., and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. Proc Natl Acad Sci U S A **103,** 7175-7180.

Weigel, D. (2012). Natural variation in Arabidopsis: from molecular genetics to ecological genomics. Plant Physiol **158,** 2-22.

Werner, J.D., Borevitz, J.O., Uhlenhaut, N.H., Ecker, J.R., Chory, J., and Weigel, D. (2005a). FRIGIDA-independent variation in flowering time of natural Arabidopsis thaliana accessions. Genetics **170,** 1197-1207.

Werner, J.D., Borevitz, J.O., Warthmann, N., Trainer, G.T., Ecker, J.R., Chory, J., and Weigel, D. (2005b). Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. Proc Natl Acad Sci U S A **102,** 2460-2465.

Wheeler, T., and von Braun, J. (2013). Climate change impacts on global food security. Science **341,** 508-513.

Wigge, P.A. (2013). Ambient temperature signalling in plants. Curr Opin Plant Biol **16,** 661-666.

Wijnker, E., Velikkakam James, G., Ding, J., Becker, F., Klasen, J.R., Rawat, V., Rowan, B.A., de Jong, D.F., de Snoo, C.B., Zapata, L., Huettel, B., de Jong, H., Ossowski, S., Weigel, D., Koornneef, M., Keurentjes, J.J., and Schneeberger, K. (2013). The genomic landscape of meiotic crossovers and gene conversions in Arabidopsis thaliana. Elife **2,** e01426.

Wilczek, A.M., Cooper, M.D., Korves, T.M., and Schmitt, J. (2014). Lagging adaptation to warming climate in Arabidopsis thaliana. Proc Natl Acad Sci U S A **111,** 7906-7913.

Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. Cold Spring Harb Perspect Biol **3**.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat Rev Genet **8,** 206-216.

Wu, X., Liu, M., Downie, B., Liang, C., Ji, G., Li, Q.Q., and Hunt, A.G. (2011). Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. Proc Natl Acad Sci U S A **108,** 12533-12538.

Xiao, J., Li, C.H., Xu, S.J., Xing, L.J., Xu, Y.Y., and Chong, K. (2015). JACALIN-LECTIN LIKE1 regulates the nuclear accumulation of GLYCINE-RICH RNA-BINDING PROTEIN7, influencing the RNA processing of FLOWERING LOCUS C antisense transcripts and flowering time in Arabidopsis. Plant Physiol **169,** 2102-2117.

Yanovsky, M.J., and Kay, S.A. (2002). Molecular basis of seasonal time measurement in Arabidopsis. Nature **419,** 308-312.

Yoo, S.K., Wu, X., Lee, J.S., and Ahn, J.H. (2011). AGAMOUS-LIKE 6 is a floral promoter that negatively regulates the FLC/MAF clade genes and positively regulates FT in Arabidopsis. Plant J **65,** 62-76.

Zapata, L., Ding, J., Willing, E.M., Hartwig, B., Bezdan, D., Jiao, W.B., Patel, V., Velikkakam James, G., Koornneef, M., Ossowski, S., and Schneeberger, K. (2016). Chromosome-level assembly of Arabidopsis thaliana Ler reveals the extent of translocation and inversion polymorphisms. Proc Natl Acad Sci U S A **113,** E4052-4060.

Zeller, G., Clark, R.M., Schneeberger, K., Bohlen, A., Weigel, D., and Ratsch, G. (2008). Detecting polymorphic regions in Arabidopsis thaliana with resequencing microarrays. Genome Res **18,** 918-929.

Zhang, T., Marand, A.P., and Jiang, J. (2016). PlantDHS: a database for DNase I hypersensitive sites in plants. Nucleic Acids Res **44,** D1148-1153.

Zhang, W., Zhang, T., Wu, Y., and Jiang, J. (2012). Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. Plant Cell **24,** 2719-2731.

Zheng, B., Chenu, K., and Chapman, S.C. (2016). Velocity of temperature and flowering time in wheat - assisting breeders to keep pace with climate change. Glob Chang Biol **22**, 921-933.

Zobell, O., Faigl, W., Saedler, H., and Munster, T. (2010). MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. Mol Biol Evol **27,** 1201-1211.

6. Cᴏɴᴛʀɪʙᴜᴛɪᴏɴs

Ulrich Lutz conceived, designed, performed the experiments, and analyzed the data, with the following exceptions:

Dr. Heidrun Gundlach (Laboratory of Prof. Dr. Klaus F. X. Mayer, Helmholtz Zentrum, München, Germany) characterized the LINE element in *FLM^{Kil-0}* **(Fɪɢ. 9C ᴀɴᴅ Tᴀʙʟᴇ A1)**.

Dr. Thomas Nussbaumer (Laboratory of Prof. Dr. Klaus F. X. Mayer, Helmholtz Zentrum, München, Germany) analyzed nucleotide diversity of genes of the type II MADS-box family **(Fɪɢ. 22C)**, extracted information of 45 SNPs from 776 accessions (available from the 1001 *A. thaliana* Genomes Project; The 1001 Genomes Consortium, 2016), performed *FLM* haplotype analysis **(Fɪɢ 23A, Fɪɢ. 24A ᴀɴᴅ B)**, and performed Kruskal-Wallis tests of association **(Fɪɢ. 28)**.

Dr. Mathias Pfeiffer (Laboratory of Prof. Dr. Klaus F. X. Mayer, Helmholtz Zentrum, München, Germany) performed RNA-seq data analysis **(Fɪɢ. 9A, Fɪɢ. 13A, B, ᴀɴᴅ D)**.

Dr. Manuel Spannagl (Laboratory of Prof. Dr. Klaus F. X. Mayer, Helmholtz Zentrum, München, Germany) identified putative *FLM* orthologs from six *Brassicacea* species **(Fɪɢ. 34A ᴀɴᴅ B)**.

Dr. Jörg Hagmann and Dr. Congmao Wang (Laboratory of Prof. Dr. Detlef Weigel, Max Planck Institute for Developmental Biology, Tübingen, Germany) identified *FLM^{LINE}* accessions by reanalyzing sequences of 1,128 *A. thaliana* accessions **(Fɪɢ. 16E)**.

Dr. David Posé performed expression analysis and flowering time experiments with a subset of transgenic plants generated as part of an independent but related analysis performed in the laboratory of Markus Schmid (Max Planck Institute for Developmental Biology, Tübingen, Germany) **(Fɪɢ. 18)**.

7. Acknowledgments/Danksagung

8. Appendix

8.1. Annotation of the LINE insertion identified in the first intron of *FLM*[Kil-0]

TABLE A1. ATLINE1_8 covers 84% of the Kil-0 insert; the first 257 bp and the last 678 bp after are not covered.

Source	pfam_acc	Name	Description	e_val	Score	seq_start	Element length	Strand
RepeatMasker vs mipsREdat v9.3 eudicot	-	ATLINE1_8	-	-	4046	258	4793	F
hmm3 vs PfamA v27	PF14111	DUF4283	Domain of unknown function (DUF4283)	2.10E-40	137.2	375	441	F
hmm3 vs PfamA v27	PF14392	zf-CCHC_4	Zinc knuckle	5.00E-15	54.9	819	141	F
hmm3 vs PfamA v27	PF03372	Exo_endo_phos	Endonuclease/exonuclease/phosphatase family	2.60E-19	70.1	1571	657	F
hmm3 vs PfamA v27	PF03304	Mlp	Mlp lipoprotein family	0.042	13.8	2237	162	F
hmm3 vs PfamA v27	PF00078	RVT_1	Reverse transcriptase (RNA-dependent DNA polymerase)	2.50E-48	164.3	3044	759	F
hmm3 vs PfamA v27	PF13966	zf-RVT	Zinc-binding in reverse transcriptase	1.70E-18	66.7	4688	198	F
hmm3 vs PfamA v27	PF13456	RVT_3	Reverse transcriptase-like	5.40E-17	61.3	5342	249	F

Modulation of Ambient Temperature-Dependent Flowering in *Arabidopsis thaliana* by Natural Variation of *FLOWERING LOCUS M*

Ulrich Lutz[1], David Posé[2¤a], Matthias Pfeifer[3¤b], Heidrun Gundlach[3], Jörg Hagmann[2], Congmao Wang[2¤c], Detlef Weigel[2], Klaus F. X. Mayer[3], Markus Schmid[2¤d], Claus Schwechheimer[1]*

1 Plant Systems Biology, Technische Universität München, Freising, Germany, 2 Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany, 3 Plant Genome and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

¤a Current address: Institute for Mediterranean and Subtropical Horticulture, University of Málaga, Málaga, Spain
¤b Current address: Roche Diagnostics, Penzberg, Germany
¤c Current address: Institute of Digital Agriculture, Zhejiang Academy of Agriculture Sciences, Hangzhou, Zhejiang, China
¤d Current address: Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Umeå, Sweden
* claus.schwechheimer@wzw.tum.de

Abstract

Plants integrate seasonal cues such as temperature and day length to optimally adjust their flowering time to the environment. Compared to the control of flowering before and after winter by the vernalization and day length pathways, mechanisms that delay or promote flowering during a transient cool or warm period, especially during spring, are less well understood. Due to global warming, understanding this ambient temperature pathway has gained increasing importance. In *Arabidopsis thaliana*, *FLOWERING LOCUS M* (*FLM*) is a critical flowering regulator of the ambient temperature pathway. *FLM* is alternatively spliced in a temperature-dependent manner and the two predominant splice variants, *FLM-ß* and *FLM-δ*, can repress and activate flowering in the genetic background of the *A. thaliana* reference accession Columbia-0. The relevance of this regulatory mechanism for the environmental adaptation across the entire range of the species is, however, unknown. Here, we identify insertion polymorphisms in the first intron of *FLM* as causative for accelerated flowering in many natural *A. thaliana* accessions, especially in cool (15°C) temperatures. We present evidence for a potential adaptive role of this structural variation and link it specifically to changes in the abundance of *FLM-ß*. Our results may allow predicting flowering in response to ambient temperatures in the *Brassicaceae*.

Author Summary

Plants control their flowering time in response to the temperatures of their environment, e.g. in response to the experience of winter or in response to cold and warm ambient temperatures experienced during spring. The knowledge about the evolutionary adaptation of plants to changing ambient temperatures is at present very limited. Understanding the latter is, however, becoming increasingly important due to the temperature changes associated with global warming and the anticipated changes in flowering time in ecosystems and agricultural systems. Here, we uncover an evolutionarily conserved molecular mechanism employed by *Arabidopsis thaliana* ecotypes for the adaptation of flowering time to cool temperatures. This structural change in the architecture of the gene *FLOWERING LOCUS M* can be found in multiple *A. thaliana* natural accessions and the knowledge gained in our study may be used to predict or modify flowering time in plants related to *A. thaliana* in the future.

Introduction

In plants, fertilization and reproduction are directly linked to the seasonal onset of flowering. Plants enter the reproductive phase when environmental conditions are favorable for seed set and thus reproduction. Since day length and temperature as well as temperature changes throughout the year provide the crucial information about the passage of the seasons and the environment, plants sense these cues for the adjustment of their flowering time [1]. Proper flowering time and reproductive success of a given species or ecotype, on the one side, and the differences in flowering time between species or ecotypes, on the other, are the result of the differential integration of temperature and day length information.

The vernalization and the ambient temperature pathways control temperature-dependent flowering in plants. Whereas vernalization requires long periods (weeks) of cold, usually below 10°C, as experienced during the winter [2], the ambient temperature pathway modulates flowering in response to short-term (days) temperature changes in the range between 12°C and 27°C [3–5]. In *A. thaliana*, the central mechanism of accelerating flowering in response to prolonged cold is achieved by repression of the negative regulator *FLOWERING LOCUS C* (*FLC*), a MADS-box transcription factor [6–9]. Different mechanisms than in *A. thaliana* control vernalization in cereal crops such as wheat and barley, and the activity or inactivity of the vernalization pathway determines the flowering behavior of their winter and spring varieties [10, 11]. To date, the understanding of the vernalization pathway in *A. thaliana* is already well advanced and it is possible to make predictions on the vernalization requirement based on the plants' genotypes [12, 13].

In contrast, the complexities of ambient temperature sensing are just beginning to be understood [5, 14, 15]. The finding that loss-of-function mutations of the gene *FLM* (*FLOWERING LOCUS M*) reduce the temperature-sensitivity of flowering in *A. thaliana* accessions suggested that this MADS-box transcription factor acts as a repressor in the ambient temperature pathway [16–18]. The molecular understanding of *FLM* is complicated by the fact that the *FLM* gene is alternatively spliced into at least four splice forms [18]. *FLM-ß* and *FLM-δ*, which result from the alternative use of the two exons 2 (*FLM-ß*) and 3 (*FLM-δ*), represent the two predominant splice variants in the Columbia-0 (Col-0) reference accession [19, 20]. The observation that the abundance of *FLM-ß* declines from 16°C to 27°C while the abundance of *FLM-δ* increases over the same temperature range has motivated experiments to examine the effects of the *FLM-ß* and *FLM-δ* isoforms in isolation in a *flm-3* loss-of-function background. These

experiments indicated that the expression of the low temperature-abundant *FLM-ß* and the warm temperature-abundant *FLM-δ* can repress and promote flowering, respectively, and consequently a model was established according to which changes in the relative abundance of *FLM-ß* and *FLM-δ* control flowering time in response to changes in ambient temperature [19].

FLM directly interacts with several other MADS-box transcription factors to control flowering through the expression of flowering time genes such as *FT* (*FLOWERING LOCUS T*) and *SOC1* (*SUPPRESSOR OF OVEREXPRESSION OF CO1*) [19–21]. SVP (SHORT VEGETATIVE PHASE) is an important FLM interaction partner and, in this context, the flowering-repressive activity of FLM-ß and the flowering-promoting activity of *FLM-δ* have been explained by the differential effects of the FLM-SVP interactions [19]: It was proposed that a DNA-binding heterodimer of FLM-ß with SVP represses flowering by repressing *FT* and *SOC1* expression. Conversely, FLM-δ could sequester SVP into an inactive complex that thereby indirectly promotes *FT* and *SOC1* expression and consequently flowering. Although this experimentally validated model is very intriguing, it is at present not known whether the alternative splicing of *FLM* plays a role in flowering time adaptation in natural accessions of *A. thaliana*.

There is increasing evidence for global warming due to climate change [8]. Temperature changes by only a few centigrade (°C) can already lead to ecological and physiological constraints that have negative impacts on agricultural production systems [22–24]. Thus, there is a need to better understand the ambient temperature pathway and to integrate this understanding in plant breeding programs [25]. Here, we identify a structural polymorphism in the first intron of *FLM* as being causative for the early flowering time of the *A. thaliana* accession Killean-0 (Kil-0). This structural polymorphism is present in several additional accessions and directly affects *FLM* transcript abundance, splicing, and flowering. We further correlate the abundance of the *FLM-ß* and *FLM-δ* splice variants with flowering behavior in several *A. thaliana* accessions and reveal an important role of intron 1 for *FLM* gene expression and a predominant role of *FLM* in flowering time control.

Results

Killean-0 is an early flowering accession

To understand the variation in flowering time in response to temperature, we compared the flowering behavior of a collection of *A. thaliana* accessions at 15°C and at 21°C. In this analysis, our attention was drawn to the Scottish accession Killean-0 (Kil-0), which flowered two weeks earlier than the Columbia-0 (Col-0) reference when grown at 15°C but only one week earlier at 21°C (Figs 1A, 1B, S1A and S1B). The vernalization pathway could potentially contribute to the early flowering behavior of Kil-0 at 15°C but we detected only minor flowering time effects after a six-week vernalization treatment (S1C Fig). These flowering time effects were similar to those observed in the reference Col-0 and confirmed also the results from previous surveys that had classified Kil-0 as a summer annual [12, 26]. Since the expression of the major vernalization-responsive gene *FLC* was also as strongly reduced in Kil-0 as in the summer annual accession Col-0 (S1D Fig), we concluded that the temperature-dependent early flowering phenotype of Kil-0 at 15°C was vernalization-independent.

FLM is the causative locus for early flowering in Kil-0

The prominent early flowering of Kil-0 at 15°C (hitherto FT15) reliably allowed distinguishing between Kil-0 and Col-0. Analyses of F_1 and F_2 Kil-0 x Col-0 plants indicated that the Kil-0 flowering phenotype was determined by a major-effect recessive locus (Fig 1C, 1D and 1E). We subsequently mapped the FT15 locus to a 968 kb genomic region (S2A and S2B Fig). After selfing F_2 plants with a recombination event in this interval, we identified 49 F_3 plants with an
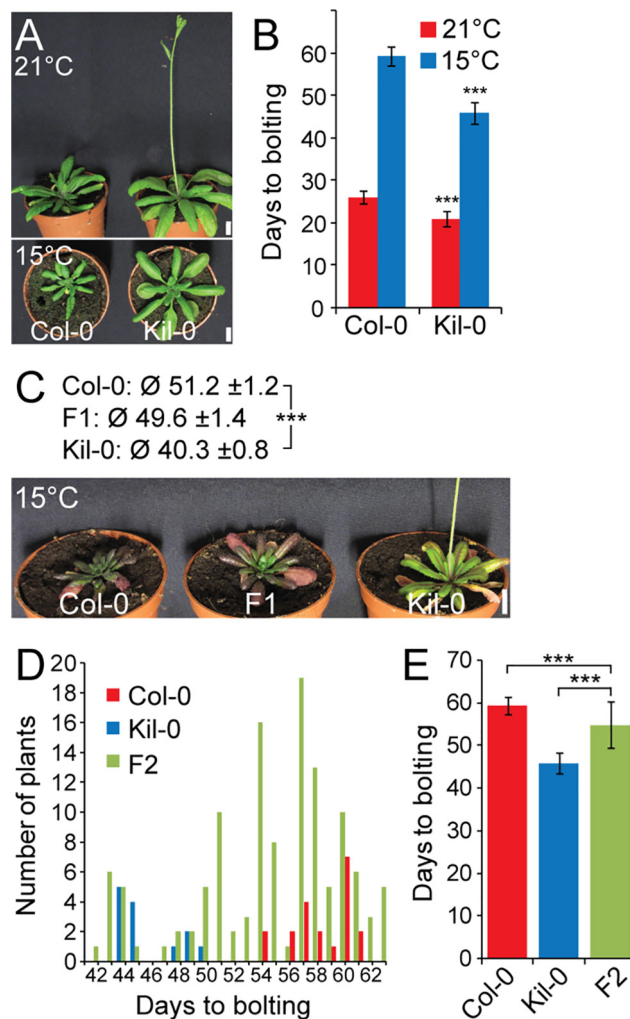
Fig 1. Early flowering is controlled by one major recessive locus in Kil-0. (A) Representative photographs and **(B)** quantitative flowering time analysis (days to bolting ± SD) of Col-0 and Kil-0 grown under continuous light at 21°C and 15°C. **(C)** Representative photographs of 50 day-old plants of Col-0, Kil-0, and their F_1 hybrid grown under continuous light at 15°C. The average days to bolting ± SD are indicated above the photograph. **(D)** Distribution of flowering time and **(E)** average days to bolting ± SD in a population of 124 F_2 segregants from a Col-0 x Kil-0 hybrid as well as among Col-0 (n = 20) and Kil-0 (n = 15) plants. Student's t-tests were performed in comparison to the wild type unless indicated otherwise in the figure: *** = $p \leq 0.001$. Scale bar = 1 cm.

early (Kil-0) and 41 F_3 plants with a late (Col-0) flowering behavior (S2C and S2D Fig). We further narrowed down the interval of interest to a 151 kb region between 28.9 and 29.1 Mb on chromosome 1 (S2E Fig) and sequenced pools of 15 early and 9 late flowering F_3 recombinants (S1 Table). Additionally, we sequenced the Kil-0 genome and identified 309 high confidence SNPs (single nucleotide polymorphisms) in the 151 kb mapping interval (S2E Fig). We smoothed the allele frequencies of the 309 SNPs of the two pools of early and late flowering F_3 plants using LOESS (locally weighted scatterplot smoothing) and calculated the difference (Δf) between them. A fraction of $\Delta f > 25\%$ defined a final mapping interval of 31.3 kb that comprised eleven annotated genes (S2E Fig). Since the flowering phenotype of Kil-0 segregated in a recessive manner (Fig 1C), we assumed that a potential candidate gene might show reduced transcript abundance in comparison to Col-0. When we investigated RNA-seq data from 10
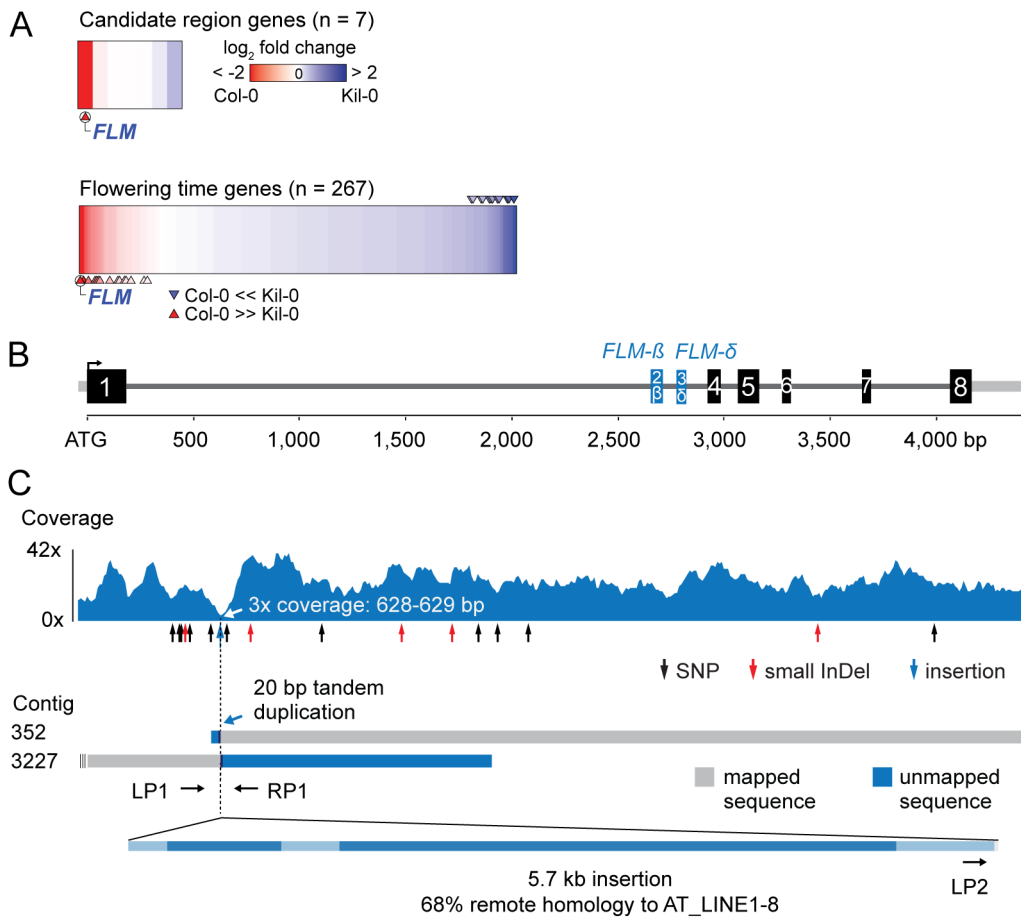
Fig 2. Identification of the Kil-0 early flowering locus. (A) Heatmap of differentially expressed genes in Kil-0 versus Col-0 in 21°C as analyzed by RNA-seq. The upper panel shows the fold change expression between Kil-0 and Col-0 of the seven expressed genes in the 31.3 kb mapping interval depicted in S2E Fig. The lower panel shows expression values of 267 flowering time genes. Fold changes are $\log_2$ transformed, blue represents upregulation and red downregulation in Kil-0 compared to Col-0. Significantly differentially ($p < 0.01$) expressed genes are marked with arrows. (B) Gene model of the FLM^{Kil-0} locus of the Col-0 reference gene with the alternatively used exons 2 and 3 that give rise to FLM-ß and FLM-δ are shown in blue. Black boxes indicate exons, grey lines are 5'- or 3'-untranslated regions or introns. (C) Read coverage of the Kil-0 genomic data on the Col-0 reference gene model. Small sequence polymorphisms are indicated below the graph. Distribution of two independently assembled contigs from Kil-0 genomic sequence reads reveal the presence of a 5.7 kb insertion in FLM intron 1 with homology to the A. thaliana transposon At_LINE1-8. Sequence elements with homology to the LINE element are shown in dark blue in the lowermost panel. The respective positions of the primers used for the screening of further A. thaliana accessions are depicted.

day-old Kil-0 and Col-0 plants grown at 21°C, we identified *FLOWERING LOCUS M* (*FLM*) as the only gene within the 31.3 kb region that was expressed at a significantly lower level in Kil-0 than in Col-0 (Fig 2A and S2 Table). *FLM* was also the most strongly downregulated gene in Kil-0 when we specifically analyzed 267 genes with a role in flowering time regulation (Fig 2A and S3 Table). These data suggested that *FLM* (*FLM^{Kil-0}*) may be causative for early flowering in Kil-0.

Kil-0 *FLM* harbors a LINE retrotransposon

Since the gene expression analyses had indicated that *FLM* may be the causative locus for early flowering im Kil-0, we compared the *FLM* genomic loci from Kil-0 and Col-0 at the molecular level. We found, however, no SNPs in the *FLM* coding sequence and only a few SNPs in the *FLM* promoter or introns. Interestingly, read coverage was greatly reduced at the beginning of the first *FLM* intron in Kil-0 (Fig 2B and 2C). Since a structural polymorphism could cause
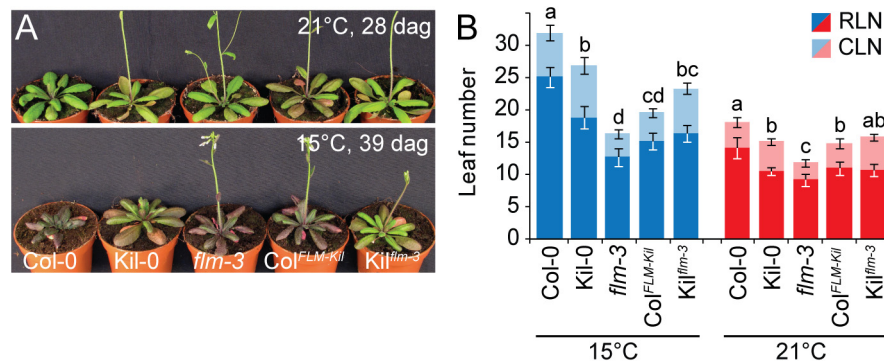
Fig 3. *FLM^{Kil-0}* accelerates flowering. (A) Photographs of representative 28 and 39 day-old plants of the Col$^{FLM-Kil}$ and Kil^{flm-3} backcross population in comparison to Col-0, Kil-0, and *flm-3* grown at 15°C and 21°C in long day photoperiod. (B) Averages ± SD of the quantitative flowering time analysis (RLN, CLN; rosette and cauline leaf number) of the genotypes shown in (A). Similar letters indicate no significant difference of total leaf number (Tukey HSD, $p < 0.05$).

doi:10.1371/journal.pgen.1005588.g003

such a read coverage reduction, we *de novo* assembled the Kil-0 genomic sequencing reads and identified an insertion in *FLM^{Kil-0}* that was flanked by two contigs corresponding to two halves of the *FLM^{Col-0}* locus at the predicted insertion site (Fig 2C). We PCR-amplified this region and confirmed the presence of a 5.7 kb insertion in the first intron of *FLM^{Kil-0}* (Fig 2C). The inserted sequence showed similarity (68% identity, 86% coverage) to *A. thaliana* ATLINE1_8 (hereafter called LINE insert), a non-LTR (NON-LONG TERMINAL REPEAT LONG INTERSPERSED NUCLEAR ELEMENTS) retrotransposon with six typical retrotransposon domains as well as a perfect copy of the second exon of a *RIBONUCLEASE H-LIKE* (AT1G04625) gene from chromosome 1 of the Col-0 genome (S4 Table and Figs 2C and S3). We considered this *FLM* insertion polymorphism as the most likely cause for early flowering in Kil-0 and subsequently conducted genetic experiments to confirm this hypothesis.

FLM^{Kil-0} is a functional temperature-sensitive *FLM* allele

Loss of *FLM* results in early flowering in the *flm-3* allele in 23°C-grown Col-0 [17, 18]. To be able to compare the effect of the *FLM^{Kil-0}* locus with that of *FLM^{Col-0}* and the *flm-3* loss-of-function allele, we backcrossed *FLM^{Kil-0}* six times into Col-0 to establish Col$^{FLM-Kil}$. Col$^{FLM-Kil}$ flowered much earlier than Col-0 but not as early as the *flm-3* knock-out mutant (Fig 3). We also introgressed *flm-3* with four backcrosses into Kil-0 to obtain Kil^{flm-3}. This introgression line flowered earlier than Kil-0. Although we performed marker assisted backcrossing, we cannot exclude the possibility that a background effect has an additional influence on flowering time regulation in the two described backcross lines. However, since the results between the two backcross lines are consistent with their effects in the original backgrounds, we considered it highly likely that the genetic data reflect *FLM* activity and concluded that the activity of *FLM^{Kil-0}* was intermediate between that of a functional *FLM^{Col-0}* allele and the *flm-3* loss-of-function allele (Fig 3).

The phenotypic differences between Col-0 and Col$^{FLM-Kil}$ as well as those between Col-0 and *flm-3* were more pronounced at 15°C than at 21°C (Fig 3). This was in agreement with our analysis of *flm-3* in a range of growth temperatures, which had revealed that *FLM* makes a particularly prominent contribution to flowering time control at 15°C (S4 Fig). Since 15°C is closer to the average temperature in the native range of the species than the commonly used 21°C, the strong effect of *FLM* on flowering at 15°C should be considered a physiologically and ecologically relevant phenotype [11, 27].

Gene expression and alternative splicing of *FLM* are differentially regulated in Kil-0

FLM produces two major splice isoforms, *FLM-ß* and *FLM-δ*, in the Col-0 accession [19, 20]. *FLM-ß* is the predominant splice form in cooler and *FLM-δ* is the predominant splice form in warmer temperatures [19]. Since the *FLM*[Kil-0] allele had weaker effects on flowering time than the *flm-3* loss-of-function allele and since *FLM*[Kil-0] did not have polymorphisms in the *FLM* coding region, we hypothesized that the temperature-sensitive flowering of Kil-0 may be caused by changes in *FLM* expression or *FLM* alternative splicing. To examine this, we transferred seven day-old 21°C-grown Col-0 and Kil-0 plants for three days to 9°C, 15°C, 21°C or 27°C and examined the effects on *FLM-ß* and *FLM-δ* transcript abundance. In agreement with published data, we observed a respective decrease of *FLM-ß* and an increase of *FLM-δ* in response to warmer temperatures (Fig 4A) [19, 20]. Importantly, temperature-dependent changes in the abundance of the *FLM* isoforms were maintained in Kil-0 as well as in Col[FLM-Kil] but the overall *FLM* transcript abundance was strongly reduced compared to Col-0. For example, when comparing the values at 21°C, Kil-0 and Col[FLM-Kil] had six-times less *FLM-ß* and 27-times less *FLM-δ* than Col-0 (Fig 4A).

It was previously shown that *FLM-ß* represses and that *FLM-δ* promotes flowering at 16°C when introduced as transgenes in the *flm-3* background [19–21]. It was further proposed that FLM-ß forms heterodimers with SVP (SHORT VEGETATIVE PHASE) to prevent flowering by direct DNA-binding to repress the transcription of *FT* and *SOC1*. Conversely, *FLM-δ* would form inactive heterodimers with SVP and would thereby indirectly induce flowering by relieving the repression from *FT* and *SOC1*. To understand the effects of this differential regulation, we measured *SVP*, *FT*, and *SOC1* expression levels. Whereas *SVP* expression was similar between the different genotypes, *FT* and *SOC1* were expressed more strongly in the early flowering Kil-0 or Col[FLM-Kil] than in Col-0 (Figs 4A and S5). We thus concluded that the reduced expression of *FLM* was likely the cause for the early flowering of Kil-0 and Col[FLM-Kil] and that the temperature-dependent differential accumulation of *FLM-ß and FLM-δ* could be the basis of the temperature-sensitive flowering time in Kil-0. Since Kil-0 flowered earlier than Col-0, we assumed that the effect of the downregulation of the repressive *FLM-ß* isoform was dominant over the downregulation of the flowering activating isoform *FLM-δ*.

The LINE insertion affects *FLM* splice isoform abundance

To test whether the large insertion in *FLM*[Kil-0] was the causative polymorphism for low *FLM* expression and the specific reduction in the *FLM-ß* isoform, we transformed Kil-0 with a genomic fragment of *FLM*[Col-0] (including 2 kb promoter plus 5'-UTR [untranslated region] and 0.5 kb 3'-UTR sequence) as well as an *FLM*[Kil-0] genomic variant with an engineered deletion of the LINE-insertion (*FLM*[Kil-0ΔLINE]). We found that *FLM*[Col-0] as well as *FLM*[Kil-0ΔLINE] delayed flowering in Kil-0 (Fig 4B). We also tested whether the 5.7 kb insertion had a comparable effect on *FLM* transcript abundance when engineered into the *FLM*[Col-0] reference and introduced a *FLM*[Col-0] transgene with an engineered 5.7 kb LINE insertion into the *flm-3* loss-of-function mutant. Indeed, the LINE insertion reduced *FLM* expression and changed *FLM* splicing, similar to the *FLM*[Kil-0] allele (Fig 4C). We thus considered it very likely that the LINE insertion in the first intron of *FLM* was the causative polymorphism for reduced *FLM* transcript abundance, differential *FLM* splice isoform accumulation, and early flowering in Kil-0.

To gain an understanding of the molecular effect of the LINE insertion on *FLM* transcription and splicing, we compared the exon usage of Col-0 and Kil-0 using RNA-seq data. In Col-0, we detected a strong differential use of alternative exons 2 (*FLM-ß)* and 3 (*FLM-δ*) that define the two dominant *FLM* isoforms (Fig 4D), with the *FLM-ß*-specific exon 2 being more
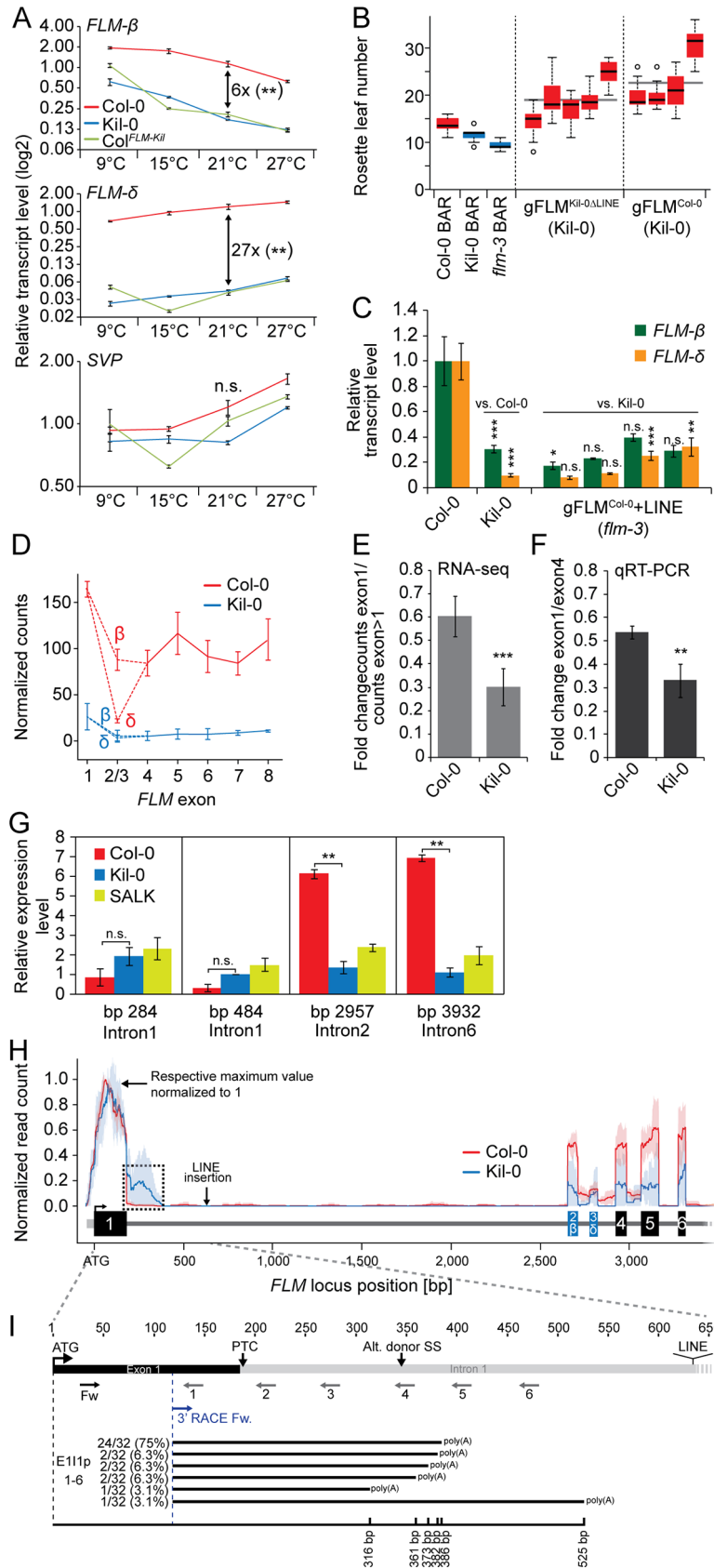
Fig 4. The large insertion in the first intron affects splicing efficiency and gene expression of *FLM*. (A) qRT-PCR analysis of *FLM-ß*, *FLM-δ*, and *SVP* expression in ten day-old Col-0, Kil-0, and Col*$^{FLM-Kil}$* plants. Seven day-old seedlings grown at 21°C were transferred to the indicated temperature and grown for further three days. The fold change differences between Col-0 and Kil-0 at 21°C are indicated in the graph. Error bars represent SE of three biological replicates. **(B)** Quantitative flowering time analysis of four—five independent T_2 lines in the Kil-0 background that are either hemi- or homozygous for the respective transgene construct. 14–20 transgenic plants grown at 21°C under long day photoperiod were analyzed for each genotype. Transgenic T_2 lines expressing the empty pGREEN0229 vector control construct (BAR+) in the Col-0, Kil-0, and *flm-3* lines were analyzed for comparison. Outliers were determined based on 1.5 x IQR (interquartile range). **(C)** qRT-PCR analysis of *FLM-ß and FLM-δ* expression of four independent homozygous T_3 lines of the indicated construct in the *flm-3* background. Plants were grown for ten days under 21°C long day photoperiod. Error bars represent SE of three biological replicates. **(D)** Graph displaying normalized read counts for the seven *FLM* exons in Col-0 and Kil-0, including the differentially spliced exon 2 that gives rise to the *FLM-ß* and *FLM-δ* isoforms. **(E)** Quantification of the normalized read counts shown in (D). The average ratios and SD from three biological replicates of the read counts of exon 1 compared to the read counts of exon 2 to exon 7 are shown. **(F)** qRT-PCR-based verification of the RNA-seq result shown in (E) using fragments corresponding to *FLM* exon 1 and exon 4. The ratios ± SE of the respective expression values are shown from three biological replicates. **(G)** qRT-PCR quantification of unspliced pre-mRNA from isolated nuclei. Primers are located in *FLM* introns 1, 2, and 6 and *ACT8* introns 2 and 3, respectively. Primer sequences are provided in S7 Table. The position of the respective reverse primer relative to the ATG start codon is indicated. Fold changes are averages ± SE of two biological replicates are indicated. Student's t-tests: ** = p ≤ 0.01; n.s., not significant. **(H)** Normalized read counts for *FLM* as detected by RNA-seq in Col-0 and Kil-0 mapped to the genomic *FLM* locus from exon 1 through exon 5. For both accessions, the number of mapped RNA-seq reads was normalized to range from 0 (no expression) to 1 (maximum expression). Lines indicate the mean expression level and light blue and light red areas indicate the 5% and 95% confidence intervals that were determined across the biological replicates of one genotype. Exons are represented as boxes, untranslated regions and introns as lines. Note the relative increase in intron 1 reads in Kil-0 as indicated with a dotted line. The position of the LINE insertion in *FLM^{Kil-0}* is indicated by an arrow. Student's t-tests were performed as indicated: * = p ≤ 0.05; ** ≤ 0.01; *** = p ≤ 0.001; n.s., not significant. **(I)** Schematic representation of the primers used for the qRT-PCR quantification of *FLM* intron 1 sequence-containing transcripts and the 3'-RACE PCR as indicated by arrows below the gene model; PTC, premature termination codon. A schematic representation of the intron 1 sequence-containing polyadenylated transcripts is indicated in the lower part of the panel. Numbers on the left indicate the frequency of each transcript among the 32 individual sequenced clones. The black line indicates the length of each transcript respective to the ATG start codon.

doi:10.1371/journal.pgen.1005588.g004

abundant than the *FLM-δ*-specific exon 3. In Kil-0, we detected generally fewer reads for all exons including exon 1, indicating that the LINE insertion between exon 1 and exon 2 controls the overall *FLM* transcript abundance (Fig 4D). Furthermore, we noted that the read coverage of the exons following intron 1, including the alternatively used exons 2 (ß) and 3 (δ) was more strongly reduced in Kil-0 than in Col-0, suggesting that the inserted LINE element may also negatively control *FLM* splicing efficiency (Fig 4D, 4E and 4F). To approximate the rate of *FLM* transcription, we determined the levels of unspliced *FLM* pre-mRNA using primers located in intron 1 (upstream of the LINE insertion), and intron 2 and 6. No significant differences in pre-mRNA levels were detected when we amplified intron 1 indicating a comparable transcription rate between the two accessions. However, the relative abundance of intron 2 and intron 6, which are located downstream from the insertion was strongly reduced in Kil-0, indicative for a partial premature termination of *FLM* transcription (Figs 4G, S6A and S6B). We next mapped the RNA-seq reads from Col-0 and Kil-0 to the *FLM^{Col-0}* genomic locus paying particular attention to the exon-intron junctions (Fig 4H). Whereas the read coverage dropped sharply at the exon 1-intron 1 junction in Col-0, reads covering the beginning of intron 1 could be readily retrieved in Kil-0 (Fig 4H). This finding suggested that the LINE insertion resulted in aberrant splicing within Kil-0 intron 1. This change in the splicing pattern had an impact on the abundance of the *FLM-ß* and *FLM-δ* full-length transcripts but, importantly, the respective full-length mRNAs were still generated as confirmed by semi-quantitative RT-PCR (S6C Fig). We then used qRT-PCR with primers spanning only exon 1 or exon 1 and parts of intron 1 to validate the occurrence of transcripts containing intron 1 sequences

(Fig 4I). In Col-0 as well as in Kil-0, we found transcripts including intron 1 sequences until 350 bp from the start codon (Figs 4I and S6D). The abundance of intron 1 sequence-containing reads was much higher in Kil-0 than in Col-0 indicating that the insertion may indeed promote premature transcription termination possibly in combination with aberrant splicing. Since transposon insertions were reported to induce alternative polyadenylation [28–30], we examined whether the corresponding intron 1 sequence-containing transcripts were polyadenylated and performed 3'-RACE (rapid amplification of cDNA ends) experiments. The RACE PCR yielded two abundant fragments, one corresponding in size to the full length transcript and one smaller fragment that was much more abundant in Kil-0 than in Col-0 (S6E Fig). We cloned and sequenced products and determined six different polyadenylated transcripts containing intron 1 sequences. One of these fragments represented the most abundant species (75%) among the 32 independent sequences (Fig 4I). Premature translation termination codons (PTC) located distantly from splice sites frequently trigger the degradation of aberrant transcripts through the NMD- (non-sense mediated decay-) pathway [31, 32]. We identified a PTC in intron 1, just two bases downstream from the exon1-intron1 border and hypothesized that the aberrant FLM^{Kil-0} transcripts may be NMD targets (Fig 4I). When we used the translation inhibitor CHX (cycloheximide) to mimic the molecular phenotype of NMD-defective mutants [32, 33], we detected indeed a significant increase in the abundance of two aberrant transcripts containing intron 1 sequences (S6F and S6G Fig). In summary, we concluded that the early flowering of Kil-0 correlated with the presence of a LINE insertion in *FLM* intron 1. Further, the LINE insertion did not affect *de novo* FLM transcription initiation but partially impaired the formation of full length transcripts, possibly as a result of premature transcription termination and the formation of aberrantly spliced polyadenylated transcripts that are targeted for NMD. Alternative molecular mechanisms that were not examined here may of course also be suitable to explain the overall reduction in *FLM* transcript abundance in Kil-0.

The geographic distribution of the FLM^{Kil-0} allele is consistent with a recent adaptive selective sweep

To gain information about the distribution of the FLM^{Kil-0} structural polymorphism across the native range of the species, we screened a genetically highly diverse set of accessions from a previously published HapMap population, which we supplemented with selected laboratory accessions to obtain a final population with 419 accessions (S6 Table and Figs 2E and S7) [34]. Through PCR-based screening, we identified nine additional accessions with a LINE insertion in the same position as in Kil-0 from Scotland (Kil-0) and Sweden (Ull2-3) to Germany (8 accessions) (S5 Table and Fig 5A and 5B). We subsequently also analyzed genome sequences from 1128 *A. thaliana* accessions (www.1001genomes.org) and confirmed by this approach seven FLM^{LINE} accessions and identified El-0 as an additional FLM^{LINE} accession from Germany (S7 Fig).

Sequence analyses of the LINE insertions revealed a high sequence similarity between the ten FLM^{LINE}-accessions (S9 Fig). Taking into account a spontaneous mutation rate of 6 to 7×10^{-9} per site per generation, approximately one to three seed generations per year, and the absence of any selective pressure on the insertion, we calculated that the common ancestor probably originated only 8.000 to 30.000 years ago [35, 36]. We found that the FLM^{LINE} accessions belonged to genetically differentiated clades and were thus truly independent [37]. Additionally we performed a phylogenetic analysis using the genomic sequence of the *FLM* locus of these ten FLM^{LINE} accessions together with Col-0 and 88 randomly selected accessions, which revealed that the FLM^{LINE} lines clustered into one clade when the *FLM* locus was analyzed in isolation (Figs 5B and S8).
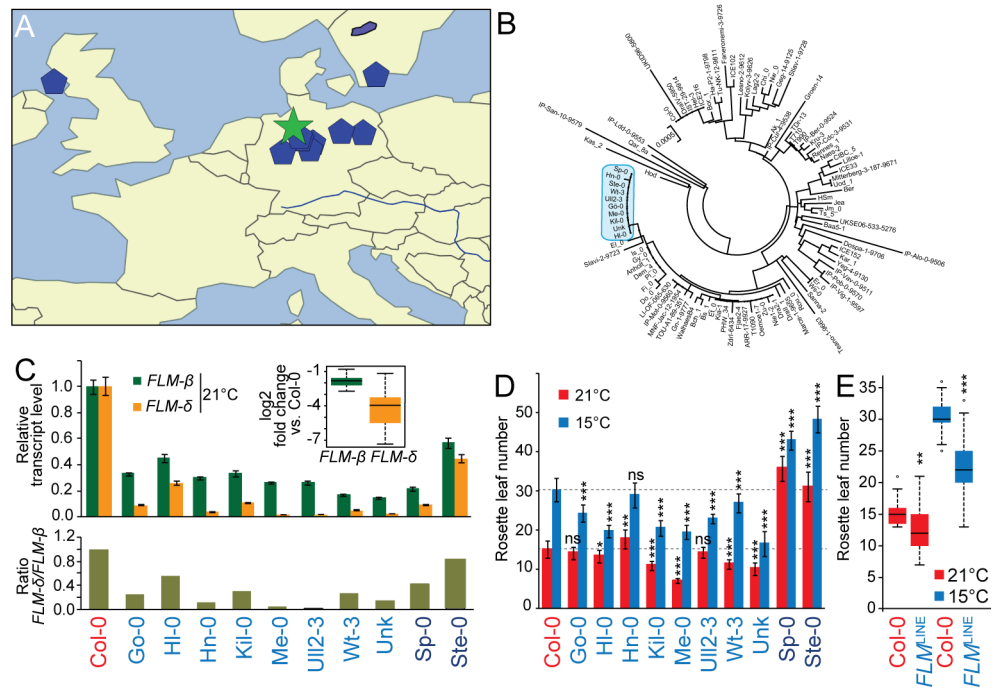
Fig 5. Geographical distribution and molecular analysis of FLM^{LINE} accessions. (A) Geographical distribution of the FLM^{LINE} accessions (blue pentagons) with the centroid (green star). Please note that one the accession with an ambiguous name was excluded due to the unavailability of geographic information. **(B)** Neighbour-Joining tree showing the genetic relationship among FLM sequences. The FLM sequences of the 10 FLM^{LINE} accessions were analyzed together with Col-0 and further 88 randomly selected sequences. The clade harboring all FLM^{LINE} alleles is marked in blue; see S8 Fig for a detailed representation of the entire tree. **(C)** qRT-PCR analyses of FLM-β and FLM-δ transcript abundance of ten day-old seedlings of all FLM^{LINE} accessions when grown in 21°C long day photoperiod (upper panel). Averages ± SE of three measurements after normalization to Col-0 as a control are shown. The inserted graph shows the summarized log$_2$-transformed fold changes of FLM^{LINE} accessions only (Col-0 excluded). The lower panel displays the ratios of FLM-δ over FLM-β when normalized to Col-0. **(D)** Quantitative flowering time analysis of FLM^{LINE} accessions grown at 21°C and 15°C under long day photoperiod. Dashed lines mark the respective values for Col-0. Please note that Sp-0 and Ste-0 are vernalization-sensitive accessions. **(E)** The right graph shows the summarized rosette leaf number of the eight vernalization-insensitive FLM^{LINE} accessions. Student's t-tests were performed in comparison to Col-0 unless indicated otherwise: * = p ≤ 0.05; ** ≤ 0.01; *** = p ≤ 0.001; n.s., not significant. El-0 was identified relatively late in this study and not included in this detailed analysis.

doi:10.1371/journal.pgen.1005588.g005

Similarly to Kil-0, all FLM^{LINE}-accessions had low expression of the FLM-β and FLM-δ iso-forms (Fig 5C). Apart from two late-flowering vernalization-dependent accessions (Sp-0 and Ste-0), seven of the remaining eight FLM^{LINE} accessions flowered earlier than Col-0 at 15°C (Figs 5D, S9A and S9B) [12, 26]. Consistent with the fact that these accessions come from genetically highly diverse groups, these lines showed a substantial variation in flowering time between them. However, when flowering time data of the vernalization-independent FLM^{LINE} accessions was averaged, we measured a significant reduction in flowering time at 21°C and 15°C in comparison to Col-0 (Fig 5E). Thus, the LINE insertion correlated with early flowering in summer annual accessions. Our data thus suggests that the FLM^{Kil-0} allele arose comparatively recently and subsequently spread geographically to contribute to flowering time regulation in a background- and temperature-dependent manner.

When we analyzed the ten FLM^{LINE}-accessions for FLM expression by qRT-PCR, we noted a prominent variation in the abundance of the FLM-δ isoform but relatively stable expression levels of FLM-β (Fig 5C). We therefore asked whether FLM polymorphisms other than the

LINE insertion could explain these differences. However, all ten FLM^{LINE} accessions were highly similar in this region and we identified only four different additional polymorphic sites (S9C Fig). Since a Mann-Whitney test (p > 0.05) indicated that none of these polymorphisms was significantly associated with *FLM-ß* and *FLM-δ* abundance or the ratio between these two isoforms, we concluded that the regulation of the *FLM-ß* and *FLM-δ* transcript abundance may be regulated in *trans*.

The first intron carries important regions for isoform-specific *FLM* regulation

Specifically within the family of MADS-box transcription factor genes, various cases are known where structural polymorphisms within the first intron enhance or repress gene expression [10, 38–41]. To test if *FLM* intronic sequences contributed to *FLM* expression, we transformed *flm-3* mutants with constructs for the expression of the *FLM-ß* or *FLM-δ* coding sequences under control of a 2.1 kb *FLM* promoter fragment (S10 Fig). Importantly, none of the resulting *FLM-ß* or *FLM-δ* T_1 transformants expressed significant levels of the respective transgene or showed a suppression of the *flm-3* early flowering phenotype (S10 Fig). Since *FLM* was expressed from corresponding genomic constructs containing all introns and since it was previously shown that *FLM* is expressed from the above-described construct when only intron 1 is included [19], we concluded that intron 1 was essential for *FLM* expression.

To address whether sequence identity, sequence length or the specific insertion site within intron 1 conferred the effect of the LINE element on *FLM* expression and splicing, we examined the effects of T-DNA or DS (Dissociator) transposon element insertions in *FLM* intron 1 [42–44]. These intron 1 insertions were of similar size (4.5 kb to 5.3 kb) to the 5.7 kb LINE insertion and present in the Col-0 and No-0 (Nossen-0) accessions: Salk_068360 (Salk, Col-0), RATM13-4593-1 (RIKEN, No-0) and GK_487H01 (GABI, Col-0) (Fig 6A). Additionally, we included Co-1 that carries a ~1.5 kb insertion in *FLM* intron 1 (Fig 6A). When we determined the effects of these structural variants on *FLM* transcript accumulation and alternative splicing in plants grown at 15°C and 21°C, we found that the insertion correlated in each case with reduced *FLM* transcript abundance when compared to the controls (Fig 6B). Furthermore, *FLM-δ* was in each case more strongly reduced than *FLM-ß*, inviting the conclusion that increases in intron length, regardless of the molecular identities of the insertions, resulted in decreased *FLM-ß* and *FLM-δ* expression and changes in the ratio between the isoforms. This was further confirmed by the molecular analysis of pre-mRNA transcript abundance, the formation of aberrant polyadenylated transcripts and transcript targeting to the NMD pathway, which we performed in parallel for the Salk insertion line, Col-0, and Kil-0 (Figs 4G and S6). In this analysis, we identified in each case the same molecular defects in the Salk insertion line as in Kil-0. At the same time, we noted that the temperature-sensitive regulation of the *FLM* isoforms was maintained in all lines. Interestingly, insertions in the second half of the intron as present in the GABI and RIKEN lines caused a particularly strong reduction in the expression of *FLM-δ* expression (Fig 6B). We thus concluded that insertions in the second half of the intron may have additional effects on *FLM-δ* splicing.

To examine the phenotypic consequences of the observed transcriptional changes of the *FLM* insertion lines, we evaluated their flowering at 15°C and 21°C. We thereby focused on the Salk and $Col^{FLM-Kil}$ as well as the GABI and RIKEN lines, which had contrasting phenotypes with regard to the abundance of the *FLM-δ* isoform. Regardless of the differences in expression of the *FLM-δ* isoform, all lines flowered earlier than the respective wild type, and this effect was particularly prominent at 15°C (Fig 6C and 6D). Importantly, we did not notice any pleiotropic effects on plant growth or plant height for the tested alleles suggesting that
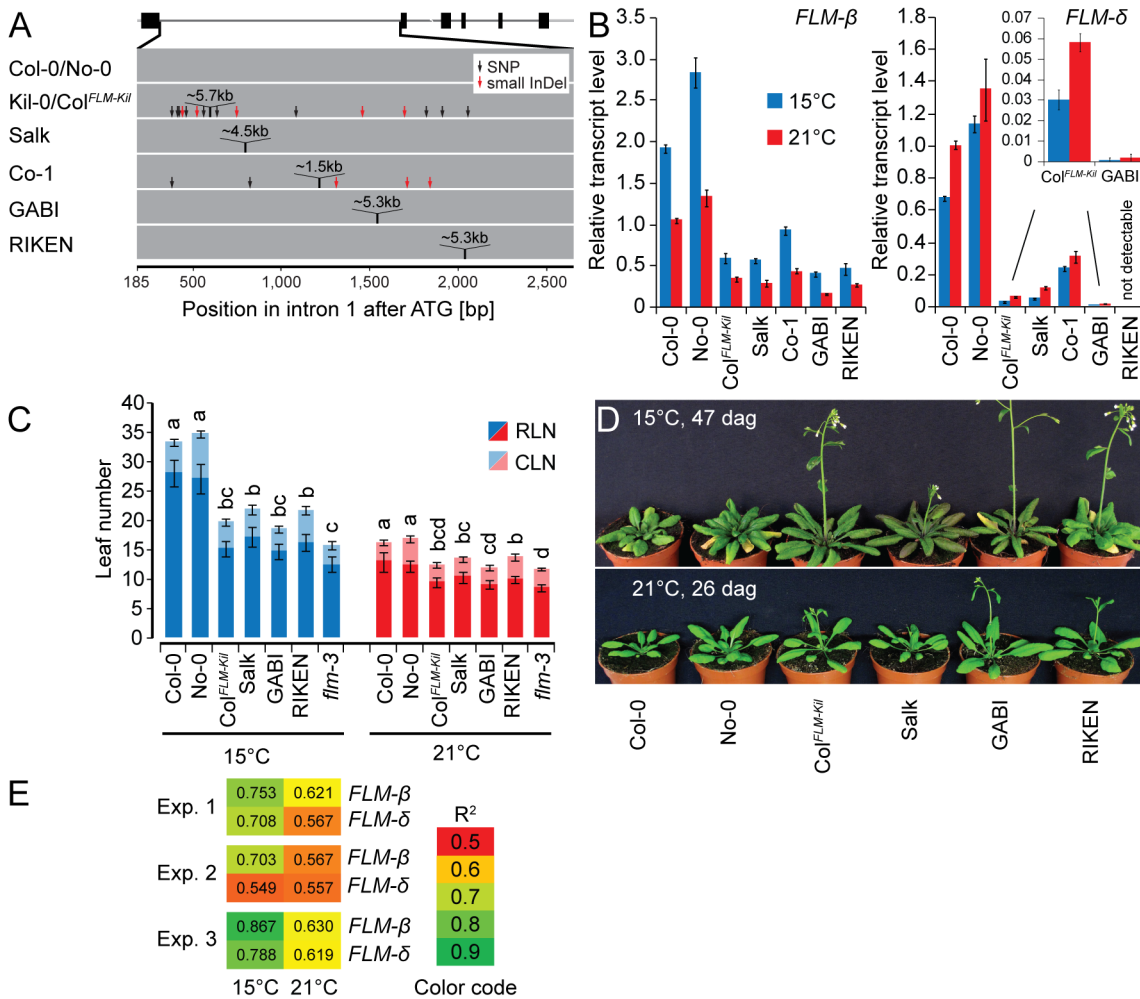
Fig 6. The first intron carries important regions for isoform-specific *FLM* abundance. (A) Schematic representation of the insertion lines used in this study. The *FLM-ß* gene model is depicted as shown in Fig 3C, intron 1 of the respective lines is depicted as a grey bar. **(B)** *FLM-ß* and *FLM-δ* expression of the lines shown in (A). Seven day-old seedlings were grown at 21°C and then transferred to 21°C or 15°C for further three days. Shown are averages and SE of three biological replicates. **(C)** Quantitative flowering time analysis of the lines shown in (A) with Col-0 and No-0 as wild type controls. Similar letters indicate no significant difference of total leaf number (Tukey HSD, *p < 0.05*). **(D)** Representative photographs of the lines shown in (A) at 26 and 47 days after germination (dag) when grown at 21°C and 15°C temperature and under long day photoperiod. **(E)** Correlation analysis of the flowering time experiment shown in Fig 6C and further two independent experiments (15°C and 21°C) with the expression values shown in Fig 6B. Note that flowering time values of the loss-of-function allele *flm-3* were also included in this analysis but that its *FLM* expression was set to 0 (no expression). R^2 values from linear regression are shown.

doi:10.1371/journal.pgen.1005588.g006

FLM acts specifically on flowering time regulation (Figs 6D and S11). We next evaluated to what extent changes in *FLM-ß* or *FLM-δ* abundance could explain the observed difference in flowering time. To this end, we correlated datasets on flowering time and *FLM* expression from three independent flowering time experiments (Fig 6E). In each experiment, expression of *FLM-ß* correlated much better with flowering time than *FLM-δ* and this effect was particularly pronounced at 15°C (Fig 6E). Thus, *FLM-ß* expression levels alone rather than the ratio between *FLM-ß* and *FLM-δ* have a prominent effect on flowering time regulation in *A. thaliana*.

Discussion

We have identified a recently evolved *FLM* allele from the accession Kil-0. The insertion of a LINE element in intron 1 of *FLM*[Kil-0] resulted in reduced *FLM* transcript abundance and correlated with an overall acceleration of flowering time that was particularly prominent at 15°C (Fig 7). We identified additional nine *FLM*[LINE] accessions that mainly represented lines collected from Germany. Although these *FLM*[LINE] accessions were highly homologous over the *FLM* locus, they represented accessions from genetically different clades indicating that *FLM*[LINE] was involved in recent adaptation to early flowering and that its rather narrow geographical distribution is likely due to the young demographic history of this allele.

The LINE element insertion of *FLM*[LINE] shares 68% homology with LINE class I retrotransposons from *A. thaliana*. Transposable elements are typically suppressed by epigenetic mechanisms and this suppression can also negatively interfere with the expression of neighboring genes [45–48]. Epigenetic regulation is also known to control the expression of MADS-box transcription factor genes. For example, the chromatin of *FLC* is modified during vernalization by lysine 27 methylation of histone 3 (H3K27me), a repressive mark, and several natural variants interfere with regulatory regions in the *FLC* intron 1 [41, 49–53]. Thus, the LINE insertion could interfere with the direct transcriptional regulation but, alternatively, also with the epigenetic control of the *FLM*[Kil-0] locus. However, previous genome-wide studies failed to identify epigenetic marks such as H3K27me on intron 1 of FLM [54] and we detected no
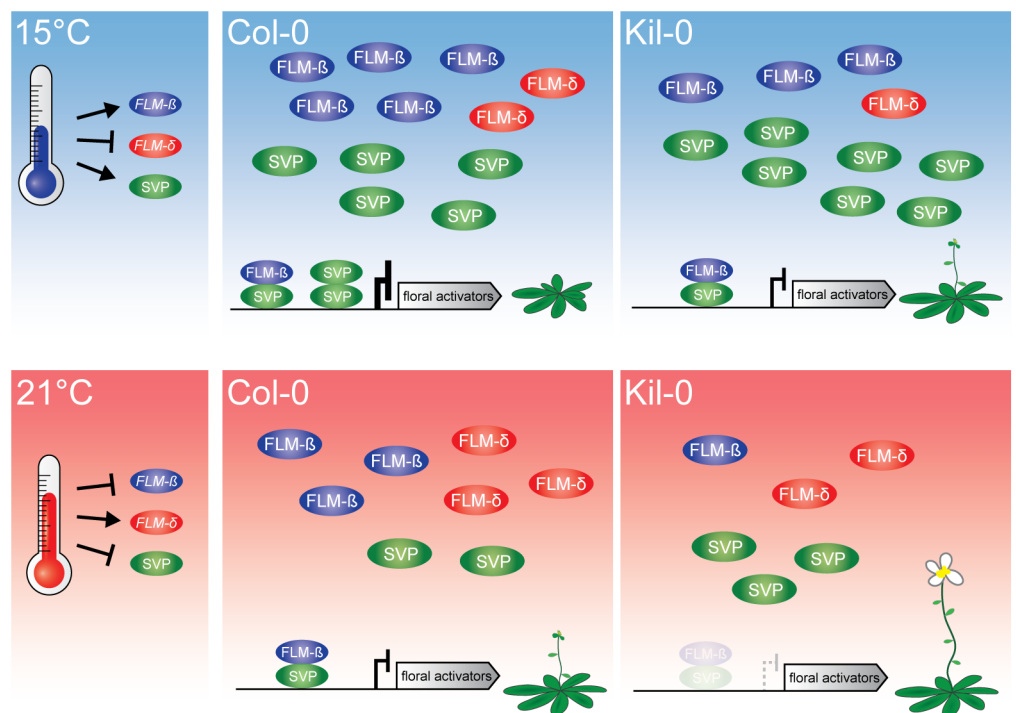

Fig 7. Model of the proposed mode of FLM action. At 15°C, flowering is delayed due to an active repression of floral activators by FLM-ß-SVP heterodimers. An increased ambient temperature (21°C) results in decreased levels of the transcriptional repressive FLM-ß-SVP protein complexes and floral activator genes are expressed. Flowering in Kil-0 (right) might be accelerated in comparison to Col-0 due to a decreased abundance of FLM-ß at 15°C as well as at 21°C while *SVP* is unaltered in the different temperatures and between Kil-0 and Col-0. Note that our study suggests a minor role of FLM-δ in the control of flowering time in the lines and conditions examined. The present model assumes that protein abundance follows transcript abundance except for SVP, which is degraded in response to increasing temperature [20].

doi:10.1371/journal.pgen.1005588.g007

differences in transcription rate between the insertion lines and the Col-0 reference. Furthermore, transposon and T-DNA insertions were reported to mediate alternative polyadenylation through the utilization of alternative polyadenylation sites [28–30]. Although we detected a higher abundance of short aberrant polyadenylated transcripts in the insertion lines, we consider it unlikely that the insertion itself provides *cis* elements that result in their synthesis. Aberrrant transcripts did not extend into the insertion and no differences in quantity and composition of these transcripts were detected between Kil-0 and the Salk line, which have molecularly distinct insertions. We concluded that the reduction of *FLM* full-length transcript abundance in Kil-0 is caused by a combination of partial premature termination of transcription and aberrant splicing due to the enlargement of the first intron.

Conversely, through experiments with *FLM* transgenes where intronic sequences were deleted, we could conclude that intron 1 was strictly required for *FLM* expression and activity. A contribution of intronic sequences in gene expression regulation, generally referred to as IME (intron-mediated enhancement), was previously reported for many genes, and in plants specifically for members of the MADS-box transcription factor family [55, 56]. Several studies have already identified corresponding intronic *cis*-regulatory elements, e.g. in intronic regions of the floral homeotic genes *AG* (*AGAMOUS*) and members of the *AGL6* (*AGAMOUS-LIKE6*)-subfamily [38–40, 57]. Several independent structural intron polymorphisms were also reported for the MADS-box factor and flowering time regulator *VRN1* (*VERNALIZATION1*) from wheat and barley. There, these structural differences in intron 1 composition can promote high *VRN1* expression and these differences are the main molecular cause for vernalization-independent flowering in many spring barley and wheat cultivars [10]. Thus structural intron polymorphisms, e.g. through transposon insertions, are a recurrent theme in the expression control of MADS-box transcription factors and the adaptation to the environment through these factors.

Interestingly, our expression analysis showed that the expression of the two *FLM* isoforms, *FLM-ß* and *FLM-δ* were differentially affected by the intron 1 insertions. Since this behavior was found in several accessions and was recapitulated by inserting a LINE-bearing *FLM* transgene in the Col-0 *FLM* allele, we judge that this regulation again is not related to the nature of the insertion in *FLM* but rather to the position of the insertion or the corresponding increase in intron length. It remains to be investigated, however, what the underlying molecular basis of the differential effect of the insertions on *FLM-ß* and *FLM-δ* abundance is.

Through investigations of insertion lines from different sources and ecotypes, we found that the position of the inserted sequence affected the relative abundance of *FLM-ß* and *FLM-δ*. In combination, the availability of these lines allowed the examination of flowering time and its correlation with the abundance of the two *FLM* splice variants.

Importantly, our study did not support a role for *FLM-δ* in flowering time control but rather suggests that *FLM-ß* abundance alone is the predominant determinant of flowering time in natural variants under the ecologically relevant temperature of 15°C. Thus, previous experiments exploring the contribution of *FLM-ß* and *FLM-δ* to flowering time control using transgenic approaches may have overestimated the contribution of *FLM-δ*, at least in the diverse genetic material and at the temperatures used in our study. At the same time, it cannot be ruled out that *FLM-δ* levels reached a subcritical level in the insertion lines analyzed in our study so that the contribution of *FLM-δ* could not be accurately determined. Along these lines, a recent study showed no contribution of the *FLM* locus to flowering time regulation in a large set of wild *A. thaliana* ectoypes in warm (27°C) temperature where *FLM-δ* is upregulated [58] indicating that even under temperature conditions that promote *FLM-δ* abundance no important regulatory role could be ascribed to it. Once a detailed understanding of factors controlling

FLM gene expression and splicing will be obtained, it will be important to reexamine this aspect in more detail.

The role of *FLM* in flowering time variation was previously established in the temperature-insensitive accessions Nd-1 and Ei-6 where *FLM* is deleted [16, 17]. In our study, we report the first gene expression variation-allele for *FLM* and describe a molecular mechanism for the control of flowering time in ambient temperature through structural changes in *FLM* intron 1. It is interesting to note that loss-of-function alleles of flowering time genes are generally very rare and are typically distributed within a small geographic region. This may indicate that adaptation through gene loss may be disadvantageous outside of the specific ecological niche simply because the loss of the gene will prevent its future reactivation [59, 60]. In the case of *FLM*, this is exemplified by the Nd-1 and Ei-6 accessions, but also among the many known *FLC* alleles, only a few null alleles have been reported whereas gene expression-modulatory *FLC* alleles are more common [6, 12, 49, 50, 61]. In this regard, we perceive at least a trend towards a similar distribution among the *A. thaliana FLM* alleles. A deeper analysis about the *FLM* coding sequence polymorphisms will be required to make conclusive statements about the importance of strong and weak *FLM* alleles during adaptation of flowering time.

We conclude that structural variations of *FLM* intron 1 as described here represent an adaptive mechanism for the control of flowering time in *A. thaliana* and possibly also in other closely related *Brassicaceae*. *FLM* might have a very specific role in flowering time regulation because modulations of its expression seemingly affect only flowering and not other plant growth traits. We, therefore, think that *FLM* is an excellent candidate gene to precisely and steadily modulate flowering time in a dynamic manner over a broad range of temperature conditions to overcome the impacts of climate change on flowering in plants.

Material and Methods

Biological material

The following *A. thaliana* accessions and genotypes were used in this study and, unless stated otherwise, provided by the Nottingham Arabidopsis Stock Centre (NASC; Nottingham, UK): The Arabidopsis accessions Killean-0 (Kil-0), Columbia-0 (Col-0), and Nossen (No-0); the insertion mutants *flm-3* (Salk_141971; Col-0), GABI-KAT GK487H01 (GABI; Col-0), Salk_068360 (Salk; Col-0) as well as RIKEN-13-4593-1 (RIKEN; No-0) from the RIKEN Stock Center. In each case, the positions of the insertions were verified by DNA sequencing. The transgenic line gFLM^{Col-0} (*flm-3*) was previously described [19], *flc-3* and the line *FRI^{SF-2} FLC* [6] were a gift from Franziska Turck and George Coupland (Max-Planck Institute of Plant Breeding Research, Cologne, Germany). A list of the *A. thaliana* accessions screened by PCR with the primers LP1, LP2, and RP1 for the *FLM^{LINE}* structural polymorphism is provided as S6 Table. Primer sequences are provided as part of S7 Table.

Physiological experiments

For flowering time analyses, plants were randomly arranged in trays and grown under constant white light (70–90 µmol m^{-2} s^{-1}) or in long day-conditions with 16 hrs white light (110–130 µmol m^{-2} s^{-1})/8 hrs dark in MobyLux GroBanks (CLF Plant Climatics, Wertingen, Germany) or MLR-351 SANYO growth chambers (Ewald, Bad Nenndorf, Germany). Trays were rearranged every two days and water was supplied by subirrigation. Analysis of large plant sets was performed in a walk-in chamber with constant white light as described above. Flowering time was quantified by determining the time until the macroscopic appearance of the first flower bud (days to bolting, DTB) or by counting rosette and cauline leaf numbers (RLN,

CLN). Student's t-tests, ANOVA, and Tukey HSD tests were calculated with Excel (Microsoft) and R (http://www.r-project.org/), respectively.

Kil-0 genome sequencing

For the resequencing of the Kil-0 genome or the late and early flowering F_3 recombinant pools, libraries were prepared from 600 ng genomic DNA following the standard protocol of the Tru-Seq DNA Sample Preparation Kit v2 (Illumina, San Diego, CA). Paired-end sequencing with a read length of 100 bp was performed on a HiSeq 2500 (Illumina, San Diego, CA). Post-sequencing quality trimming was performed with the CLC Genomics Workbench (v. 7.0) and the following parameters: low quality limit = 0.05; ambiguous nucleotide = maximum 1; length minimum = 15. Post-trimming, 15×10^6 and 20×10^6 reads were obtained for the early and late flowering samples, respectively. Read mapping was performed using the TAIR10 release of the *A. thaliana* reference (The Arabidopsis Genome Initiative, 2000) genome reference sequence with the stringent settings: mismatch cost = 2; insertion cost = 2; deletion cost = 2; length fraction = 0.9; similarity = 0.9. An average 57- or 79-fold coverage was obtained from the early and late flowering DNA pools. Variant calling was performed using the probabilistic variant calling tool of the CLC Genomics Workbench (v. 7.0) and default settings. SNPs with a 30–120-fold coverage and frequency $f > 20\%$ as well as a presence call in both pooled samples were selected for allele frequency mapping. From those, SNPs that showed a frequency of $< 80\%$ in the resequencing analysis of the homozygous Kil-0 parental line were discarded. Smoothing using locally weighted scatterplot smoothing (LOESS) of SNP frequency values was achieved with R (http://www.r-project.org/). 95% confidence intervals, $\Delta f > 25\%$, and Δf_{max} were calculated from the LOESS values. The *de novo* assembly of Kil-0 resequencing reads was performed using the CLC Genomics Workbench v. 7.0 with default settings. Contigs were identified by a simple search and were reassembled to the Col-0 genomic *FLM* sequence. The Kil-0 genomic sequence is available as LN866842 at www.ebi.ac.uk/ena.

Mapping and backcrossing

To identify the causative locus for early flowering in Kil-0, the FT15 locus was mapped with polymorphic markers selected from a previously described marker collection [62]. Additional SSLP (single sequence length polymorphisms) markers were generated by searching the publicly available Kil-0 genomic sequence (www.1001genomes.org) or the genome sequence that was determined as part of this project for InDel (insertion/deletion) polymorphisms. PCR primers spanning these sites were designed with Primer3 [63] and tested on Col-0 and Kil-0 genomic DNA. The PCR fragments were generally between 200 and 700 bp long and separated on 2–3.5% agarose gels or using the QIAxcel Advanced Capillary Electrophoresis high resolution kit (Qiagen, Hilden, Germany). For rough mapping of FT15, ten early and ten late flowering F_2 plants were selected from the extreme phenotypic borders of an F_2 population (n = 124). The genetic marker distances were calculated from the genotype data of all screened F_2 plants with JoinMap v.4.1 (Kyazma B.V.). A list of markers and the respective primers is provided as S8 Table.

Col$^{FLM-Kil}$ and Kil^{flm-3} were generated by marker-assisted backcrosses. In brief, heterozygous F_1 plants were genotyped with the primers LP1, LP2, and RP1 to examine the lines for the presence of the Col-0 or Kil-0 *FLM* allele and with Salk_141971 forward and reverse primers to test for the *flm-3* T-DNA insertion, respectively. Primer sequences are provided as part of S7 Table.

Quantitative real-time PCR

For qRT-PCR analyses, total RNA was isolated from three biological replicates using the NucleoSpin RNA Plant kit (Machery-Nagel, Düren, Germany). DNA was removed by an

on-column treatment with rDNase (Machery-Nagel, Düren, Germany). 2–3 µg total RNA were reverse transcribed with an oligo(dT) primer and M-MuLV Reverse Transcriptase (Fermentas, St. Leon-Rot, Germany) and the cDNA equivalent of 30–50 ng total RNA was used in a 10 µl PCR reaction with SsoAdvanced™ Universal SYBR Green Supermix (BioRad, München, Germany) in a CFX96 Real-Time System Cycler (BioRad, München, Germany). The relative quantification was calculated with the ΔΔCt method with *ACT8* as a control [64]. See S7 Table for a list of qRT-PCR primers.

RNA-sequencing

To investigate significant gene expression differences between Col-0 and Kil-0, RNA was prepared from three biological replicate samples from ten day-old seedlings (21°C, long days) as described above. Sequencing libraries were prepared following the standard protocol of the TruSeq RNA Sample prep v2 Kit (Illumina, San Diego, CA). Paired-end sequencing with a read length of 100 bp was performed on a HiSeq 1000 (Illumina, San Diego, CA). RNA-seq reads from each biological replicate were then aligned against the Col-0 TAIR10 release genome using Bowtie (version 2.1.0) and Tophat2 (version 2.0.8) with default parameters [65, 66]. On the basis of the structural gene annotation for *A. thaliana* TAIR10, exon- and gene-level transcription was quantified by using HTSeq, differential expression tests were performed by using DESeq2 [67, 68]. Significantly differentially expressed genes were defined as genes with Benjamini-Hochberg-adjusted p-values < 0.01. Tests for differential exon usage were conducted on basis of RNA-seq exon counts with DEXseq [69]. The RNA-seq data are publically available as PRJEB9470 at www.ebi.ac.uk/ena.

A list of genes with a role in flowering time regulation was generated by searching the TAIR database (www.arabidopsis.org) for the term "flowering time". The resulting list was reviewed manually and is presented in S3 Table. To subsequently analyze gene expression of the *FLM* locus at a nucleotide resolution, the corresponding Col-0 gene sequences were extracted from the Col-0 TAIR10 reference genome and the Kil-0 genome sequence as determined as part of this study. Subsequently, RNA-seq reads were aligned against the Kil-0 and Col-0 genomic sequences using Bowtie and Tophat2 as described above. The number of mapped RNA-seq reads from Col-0 and Kil-0 per nucleotide were counted with the toolset Bedtools and subsequently normalized to range between 0 (no expression) and 1 (maximum expression). For visualization, the mean expression level and 5% and 95% confidence intervals were determined across the biological replicates of one sample.

Cloning procedures

To obtain a genomic *FLM* fragment from Kil-0 with a deletion of the 5.7 kb LINE insertion, a 6,981 bp g*FLM*(Kil-0)ΔLINE deletion fragment was amplified by overlap extension PCR from Kil-0 genomic DNA. Fragment 1 (bp—2367 to bp + 631) was amplified with ULC-1 and ULC-2 and fragment 2 (bp + 632 to bp + 4156 and 251 bp 3'-UTR plus 207 bp downstream sequence) with ULC-3 and ULC-4. After purification of the two subfragments, the full fragment was generated by overlap-extension PCR reaction with ULC-1 and ULC-4. The insert of the full-length fragment in pCR2.1-TOPO (Life Technologies, Carlsbad, CA) was sequenced and subcloned as a *Bam*HI and *Xho*I fragment into pGreen0229 [70].

To insert the LINE insertion into pDP34, a previously described construct with the Col-0 genomic *FLM* fragment p*FLM*::g*FLM* [19], pDP34 was mutagenized by PCR with ULC-12 to replace the sequence ATTGTTCA (bp +632 to bp +640) with a unique *Asc*I restriction site (GGCGCGCC) [71]. The LINE insertion was then amplified from Kil-0 genomic DNA with ULC-16 and ULC-17 and inserted as an *Asc*I fragment into the modified pDP34. All constructs

were transformed into *Agrobacterium tumefaciens* strain GV3101 containing pSOUP and subsequently using floral dip transformation into Col-0, Kil-0, and *flm-3* [70, 72]. pDP79 and pDP80 were previously described [19]. T_1 transformants were selected by spraying soil-grown plants with 0.1% BASTA. The list of primers is provided in S7 Table.

FLM locus analysis of *A. thaliana* accessions

To test for the presence of the FLM^{Kil-0} allele in a large collection of accessions, we performed PCR with the primers LP1, LP2, and RP1 on genomic DNA using Phusion (New England Biolabs, Frankfurt, Germany) or TaKaRa LA Taq polymerases (Takara Bio, Saint-Germain-en-Laye, France). Selected PCR products were analyzed by DNA sequencing following gel extraction. Sequencing primers are listed in S7 Table. A complete list of all accessions examined is provided as S6 Table.

Kil-0 genome sequence information was used to identify accessions with a LINE insertion by analyzing genome sequences as determined in the frame of the 1001 Arabidopsis genome sequencing project (www.1001genomes.org). To render read mapping specific for the Kil-0 allele, the insertion at the experimentally retrieved insertion sites flanked by 140 bp sequence was extracted for each of the two insertion breakpoints and defined as target sequences for mapping. Subsequently, all reads of the individual *A. thaliana* accessions were mapped against the target sequences using SHORE and genomemapper allowing for up to 5% single base pair differences including gaps with regard to the read length [73, 74]. To ensure that each supporting read spans at least 25% of its sequence across the insertion breakpoint, only reads that overlapped an insertion site with the inner 50% of its read sequence were counted as supporting an insertion breakpoint (core-mapping read). Furthermore, only unique mapped reads were included. An insertion was defined as present if there were at least two unique core-mapping reads at the left and right insertion breakpoint.

LINE insert and *FLM* locus characterization

Several database searches with the Kil-0 LINE insert yielded no highly similar sequences, except for a 559 bp match to the second exon of Col-0 AT1G04625, which is most likely a transposon-related gene based on its ribonuclease H-like description. The LINE sequence within the Kil-0 insert was identified with RepeatMasker (http://www.repeatmasker.org) against PGSP-REdat, v_9.3_Eudicot (http://pgsb.helmholtz-muenchen.de/plant/recat/). The Pfam domains where annotated with hmmsearch of hmmer3 in all 6 reading frames against PfamA v27 [75, 76].

For phylogenetic analyses of the *FLM* locus, sequences of 88 randomly selected accessions were extracted from the *A. thaliana* 1001 genome project GEBowser (http://signal.salk.edu/atg1001/3.0/gebrowser.php) and aligned with sequences of the ten FLM^{LINE} accessions and Col-0. Alignments were calculated using ClustalW and Neighbor-Joining trees (Maximum Composite Likelihood method, 1000 bootstrap replicates) were constructed with MEGA5 [77].

3' RACE PCR

Polyadenylated transcripts were amplified according to [78]. Transcript pools were analyzed on an agarose gel and the upper and lower bands as depicted in S6E Fig were purified from all samples, subcloned into pCR2.1-TOPO (Life Technologies, Carlsbad, CA) and sequenced. Primers sequences are listed in S7 Table.

Determination of *FLM* pre-mRNA abundance

Nuclei were isolated as previously described [79] with two biological replicates per line. RNA from nuclei pellet was extracted as previously described [80] and DNA was digested using DNaseI (Life Technologies, Carlsbad, CA). cDNA synthesis and qRT-PCR were performed as described above. Primer sequences are provided in S7 Table.

Supporting Information

S1 Fig. Flowering time analysis in Kil-0 and Col-0. (A) and **(B)** Quantitative flowering time analysis of Col-0 and Kil-0 in different growth conditions. LD, long days; SD, short days; CC, continuous light. **(C)** Vernalization response of Col-0 and Kil-0. Total leaf number of vernalized plants (8 weeks at 4°C) compared to non-vernalized plants. **(D)** qRT-PCR analysis of *FLC* of 12 day-old non-vernalized plants grown in 21°C and long day photoperiod. Kas-1 was included as a vernalization-sensitive accession.
(TIF)

S2 Fig. Characterization of the Kil-0 x Col-0 mapping population. (A) Representative photographs of Kil-0 x Col-0 F_2 plants grown at 15°C and selected from the extreme phenotypic borders. **(B)** Rough mapping analysis using ten early and ten late flowering F_2 plants as shown in (A). The average Kil-0 allele frequency of the early or late flowering plants is shown for twenty SSLP markers. Markers were selected from [62] or generated as described. Significance of the two associations is indicated with $^* = p \leq 0.05$; ** $p \leq 0.01$. **(C)** Schematic representation of the FT15 mapping procedure. 43 recombinant plants with the genotypes M1$^{\text{Kil-0/Col-0}}$/M2$^{\text{Kil-0/Kil-0}}$ or M1$^{\text{Kil-0/Kil-0}}$/M2$^{\text{Kil-0/Col-0}}$ in the 968 kb interval were selected from 1049 F_2 plants. After selfing, 43 F_3 recombinant lines were obtained and four plants per F_3 family were individually genotyped and phenotyped at 15°C with additional markers identifying a 151 kb peak region. Bulks of early or late flowering F_3 plants with a recombination in this region were sequenced. **(D)** Quantitative flowering time measured in days to bolting of 127 F_3 recombinant plants from 43 families grown in 15°C and continuous light. Each bar represents one plant. Averages ± SD of 15 Col-0 and Kil-0 plants are indicated as blue and red bars, respectively. **(E)** Results of the mapping procedure described in (A). The 968 kb interval with flanking marker M1 and M2 and seven internal markers is shown as a vertical line with the genetic [cM] and physical [kb] marker distances as calculated based on the 1049 F_2 individuals and the chromosomal position. The middle panel shows the respective allele frequency of the selected 309 SNPs within either the early or the late flowering plant pool. Red and blue lines show the respective LOESS-smoothed allele frequency values with the 95%-confidence interval shown as a grey dotted line. Vertical black lines represent the final mapping interval of 31.3 kb with $\Delta f > 25\%$. The SNP with the highest difference between early and late pool (Δf_{max}) is marked with a black dotted line. The lower panel shows a detailed illustration of the eleven annotated gene models in the 31.3 kb interval. **(F)** Physical distribution of the 309 SNPs along the 151 kb mapping interval.
(TIF)

S3 Fig. Detailed description of the inserted sequence in intron 1 of $FLM^{\text{Kil-0}}$. The light grey bar indicates the complete 5.7 kb insertion. The dark blue bar indicates the partition of the insertion that shows 68% identity to the AT_LINE1-8 retrotransposon sequence.
(TIF)

S4 Fig. Flowering time analysis of *flm-3* in different ambient temperatures. Total leaf number [TLN] of *flm-3* plants (n = 20–30) was compared to Col-0 plants (n = 20–30). (TIF)

S5 Fig. Analysis of flowering time gene expression. (A) and **(B)** qRT-PCR analyses of *SVP* and *FT* transcript abundance at 21°C during a 24 h long day photoperiod. Fold changes are averages ± SE of three measurements. **(C)**, **(D)**, and **(E)** qRT-PCR analyses of *FT*, *SOC1*, and *SVP* transcript abundance at 21°C in long day photoperiod. Samples were taken at ZT16 from 6, 9, 13, 15 day-old plants. Fold changes are averages ± SE of three measurements. (TIF)

S6 Fig. Analysis of *FLM* processing (A) Images of agarose gels illustrating the absence of any spliced transcripts from the nuclear RNA preparations since the intron-spanning ACT8 primers only amplify a 254 bp fragment corresponding to the non-spliced fragment but not the 111 bp fragment corresponding to the spliced form. **(B)** The absence of genomic DNA contamination was determined by a primer located downstream from the FLM 3' UTR. Genomic DNA samples with different DNA template concentrations are shown as positive controls. **(C)** Semi-quantitative PCR with primers amplifying the full-length FLM gene using two cDNA samples each from ten days-old Col-0 and Kil-0 plants grown at 21°C. **(D)** qRT-PCR analyses of intron 1 sequence-containing transcripts, which were amplified using the reverse primers 1–6 as indicated in (Fig 4I). Fold changes are averages + SE of three biological replicates. Note that the expression value comparisons between the fragments are approximations since the primer efficiencies are not exactly identical. **(E)** Image of an agarose gel with the analysis of 3' RACE PCR products. The respective upper and lower fragments were isolated for sequencing. **(F)**, and **(G)** qRT-PCR analyses of ten day-old seedlings grown under 21°C long-day conditions without treatment or following a 5 hr mock or 20 μM CHX treatment, respectively. Fragment 3 and 4 correspond to fragment 3 and 4 depicted in Fig 4I. Fold changes are averages ± SE of three biological replicates. Student's t-tests were performed as indicated: $^* = p \leq 0.05$; $^{**} \leq 0.01$. All primer sequences are provided in S7 Table. (TIF)

S7 Fig. Geographic distribution of the accessions used (A) for the PCR-based screen and (B) for the analysis of unmapped reads. (TIF)

S8 Fig. Neighbour-Joining tree of *FLM*^LINE^ accessions and randomly selected accessions. (TIF)

S9 Fig. Phenotypic and molecular analysis of *FLM*^LINE^ accessions. (A) and **(B)** Representative photographs of the *FLM*^LINE^ accessions grown at 21°C and 15°C in long day photoperiod. **(C)** Alignment of the *FLM* locus and 1.8 kb upstream sequence of the *FLM*^LINE^ accessions. Polymorphic sites between the *FLM*^LINE^ accessions are indicated with an asterisk. The 5.7 kb insertion of *FLM*^Kil-0^ at position +631 is not represented here. **(D)** Multiple alignment of the 5.7 kb inserted sequence of the *FLM*^LINE^ accessions. Black vertical lines indicate SNPs, light grey horizontal lines indicate insertions and deletions. (TIF)

S10 Fig. *FLM* introns are essential for *FLM* expression. (A) Schematic representation of the constructs used for the analysis. **(B)** Averages ± SD from two biological replicates of qRT-PCR analyses of *FLM-ß and FLM-δ* from four independent T$_2$ lines transformed with the respective constructs. **(C)** Quantitative flowering time analysis of independent T$_1$ transformants with the

constructs described in (A).
(TIF)

S11 Fig. *FLM* alleles do not control plant height. Final plant height of plants grown at 21°C and 15°C under long day conditions.
(TIF)

S1 Table. List of plants with an early and late flowering phenotype that were pooled for next generation sequencing.
(XLS)

S2 Table. List of genes that are annotated in the 31.3 kb mapping interval.
(XLS)

S3 Table. List of genes with an implied role in flowering time regulation and analysis of their differential regulation by RNA-seq.
(XLS)

S4 Table. Annotation of the LINE insertion identified in the first intron of *FLM* from Kil-0.
(XLS)

S5 Table. List of accessions carrying the *FLM^{Kil-0}* allele.
(XLS)

S6 Table. List of accessions screened for the Kil-0 insertion.
(XLS)

S7 Table. List of primers used for cloning, genotyping, qRT, and sequencing.
(XLS)

S8 Table. List of primers for marker detection.
(XLS)

Acknowledgments

Author Contributions

Conceived and designed the experiments: UL DP MP HG JH KFXM DW MS CS. Performed the experiments: UL DP MP HG JH. Analyzed the data: UL DP MP HG JH. Contributed reagents/materials/analysis tools: UL DP MP HG JH CW DW KFXM MS CS. Wrote the paper: UL DP MP HG JH CW DW KFXM MS CS.

References

1. Andres F, Coupland G. The genetic basis of flowering responses to seasonal cues. Nat Rev Genet. 2012; 13(9):627–39. Epub 2012/08/18. doi: 10.1038/nrg3291 PMID: 22898651
2. Song J, Irwin J, Dean C. Remembering the prolonged cold of winter. Curr Biol. 2013; 23(17):R807–11. Epub 2013/09/14. doi: 10.1016/j.cub.2013.07.027 PMID: 24028964

3. Kumar SV, Wigge PA. H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. Cell. 2010; 140(1):136–47. Epub 2010/01/19. doi: 10.1016/j.cell.2009.11.006 PMID: 20079334

4. Balasubramanian S, Weigel D. Temperature Induced Flowering in Arabidopsis thaliana. Plant Signal Behav. 2006; 1(5):227–8. Epub 2006/09/01. PMID: 19704664

5. Wigge PA. Ambient temperature signalling in plants. Curr Opin Plant Biol. 2013; 16(5):661–6. Epub 2013/09/12. doi: 10.1016/j.pbi.2013.08.004 PMID: 24021869

6. Michaels SD, Amasino RM. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell. 1999; 11(5):949–56. Epub 1999/05/20. PMID: 10330478

7. Michaels SD, Amasino RM. Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization. Plant Cell. 2001; 13(4):935–41. Epub 2001/04/03. PMID: 11283346

8. Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C. Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science. 2000; 290(5490):344–7. Epub 2000/10/13. PMID: 11030654

9. Song J, Angel A, Howard M, Dean C. Vernalization—a cold-induced epigenetic switch. J Cell Sci. 2012; 125(Pt 16):3723–31. Epub 2012/09/01.

10. Distelfeld A, Li C, Dubcovsky J. Regulation of flowering in temperate cereals. Curr Opin Plant Biol. 2009; 12(2):178–84. Epub 2009/02/07. doi: 10.1016/j.pbi.2008.12.010 PMID: 19195924

11. Trevaskis B, Hemming MN, Dennis ES, Peacock WJ. The molecular basis of vernalization-induced flowering in cereals. Trends in plant science. 2007; 12(8):352–7. Epub 2007/07/17. PMID: 17629542

12. Lempe J, Balasubramanian S, Sureshkumar S, Singh A, Schmid M, Weigel D. Diversity of flowering responses in wild Arabidopsis thaliana strains. PLoS Genet. 2005; 1(1):109–18. Epub 2005/08/17. PMID: 16103920

13. Strange A, Li P, Lister C, Anderson J, Warthmann N, Shindo C, et al. Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in Arabidopsis accessions. PLoS One. 2011; 6(5):e19949. Epub 2011/06/01. doi: 10.1371/journal.pone.0019949 PMID: 21625501

14. Verhage L, Angenent GC, Immink RG. Research on floral timing by ambient temperature comes into blossom. Trends in plant science. 2014; 19(9):583–91. Epub 2014/05/02. doi: 10.1016/j.tplants.2014.03.009 PMID: 24780095

15. Capovilla G, Schmid M, Pose D. Control of flowering by ambient temperature. J Exp Bot. 2015; 66 (1):59–69. Epub 2014/10/19. doi: 10.1093/jxb/eru416 PMID: 25326628

16. Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, Chory J, et al. Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. Proc Natl Acad Sci U S A. 2005; 102(7):2460–5. Epub 2005/02/08. PMID: 15695584

17. Balasubramanian S, Sureshkumar S, Lempe J, Weigel D. Potent induction of Arabidopsis thaliana flowering by elevated growth temperature. PLoS Genet. 2006; 2(7):e106. Epub 2006/07/15. PMID: 16839183

18. Scortecci KC, Michaels SD, Amasino RM. Identification of a MADS-box gene, FLOWERING LOCUS M, that represses flowering. Plant J. 2001; 26(2):229–36. Epub 2001/06/08. PMID: 11389763

19. Pose D, Verhage L, Ott F, Yant L, Mathieu J, Angenent GC, et al. Temperature-dependent regulation of flowering by antagonistic FLM variants. Nature. 2013; 503(7476):414–7. Epub 2013/09/27. doi: 10.1038/nature12633 PMID: 24067612

20. Lee JH, Ryu HS, Chung KS, Pose D, Kim S, Schmid M, et al. Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. Science. 2013; 342(6158):628–32. Epub 2013/09/14. doi: 10.1126/science.1241097 PMID: 24030492

21. Gu X, Le C, Wang Y, Li Z, Jiang D, He Y. Arabidopsis FLC clade members form flowering-repressor complexes coordinating responses to endogenous and environmental cues. Nat Commun. 2013; 4:1947. Epub 2013/06/19. doi: 10.1038/ncomms2947 PMID: 23770815

22. Wheeler T, von Braun J. Climate change impacts on global food security. Science. 2013; 341 (6145):508–13. Epub 2013/08/03. doi: 10.1126/science.1239402 PMID: 23908229

23. Moore FC, Lobell DB. The fingerprint of climate trends on European crop yields. Proc Natl Acad Sci U S A. 2015; 112(9):2670–5. Epub 2015/02/19. doi: 10.1073/pnas.1409606112 PMID: 25691735

24. Mora C, Caldwell IR, Caldwell JM, Fisher MR, Genco BM, Running SW. Suitable Days for Plant Growth Disappear under Projected Climate Change: Potential Human and Biotic Vulnerability. PLoS Biol. 2015; 13(6):e1002167. Epub 2015/06/11. doi: 10.1371/journal.pbio.1002167 PMID: 26061091

25. Thuiller W, Lavorel S, Araujo MB, Sykes MT, Prentice IC. Climate change threats to plant diversity in Europe. Proc Natl Acad Sci U S A. 2005; 102(23):8245–50. Epub 2005/05/28. PMID: 15919825

26. Werner JD, Borevitz JO, Uhlenhaut NH, Ecker JR, Chory J, Weigel D. FRIGIDA-independent variation in flowering time of natural Arabidopsis thaliana accessions. Genetics. 2005; 170(3):1197–207. Epub 2005/05/25. PMID: 15911588

27. Weigel D. Natural variation in Arabidopsis: from molecular genetics to ecological genomics. Plant Physiol. 2012; 158(1):2–22. Epub 2011/12/08. doi: 10.1104/pp.111.189845 PMID: 22147517

28. Tsuchiya T, Eulgem T. An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. Proc Natl Acad Sci U S A. 2013; 110(37):E3535–43. Epub 2013/08/14. doi: 10.1073/pnas.1312545110 PMID: 23940361

29. Wu X, Liu M, Downie B, Liang C, Ji G, Li QQ, et al. Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. Proc Natl Acad Sci U S A. 2011; 108(30):12533–8. Epub 2011/07/13. doi: 10.1073/pnas.1019732108 PMID: 21746925

30. Duc C, Sherstnev A, Cole C, Barton GJ, Simpson GG. Transcription termination and chimeric RNA formation controlled by Arabidopsis thaliana FPA. PLoS Genet. 2013; 9(10):e1003867. Epub 2013/11/10. doi: 10.1371/journal.pgen.1003867 PMID: 24204292

31. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A. 2003; 100(1):189–92. Epub 2002/12/28. PMID: 12502788

32. Kalyna M, Simpson CG, Syed NH, Lewandowska D, Marquez Y, Kusenda B, et al. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic Acids Res. 2012; 40(6):2454–69. Epub 2011/12/01. doi: 10.1093/nar/gkr932 PMID: 22127866

33. Hori K, Watanabe Y. UPF3 suppresses aberrant spliced mRNA in Arabidopsis. Plant J. 2005; 43 (4):530–40. Epub 2005/08/16. PMID: 16098107

34. Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO. Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. Proc Natl Acad Sci U S A. 2010; 107(49):21199–204. Epub 2010/11/17. doi: 10.1073/pnas.1007431107 PMID: 21078970

35. Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, Shaw RG, et al. The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. Science. 2010; 327(5961):92–4. Epub 2010/01/02. doi: 10.1126/science.1180677 PMID: 20044577

36. Schultz ST, Lynch M, Willis JH. Spontaneous deleterious mutation in Arabidopsis thaliana. Proc Natl Acad Sci U S A. 1999; 96(20):11393–8. Epub 1999/09/29. PMID: 10500187

37. Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, et al. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature. 2010; 465(7298):627–31. Epub 2010/03/26. doi: 10.1038/nature08800 PMID: 20336072

38. Hong RL, Hamaguchi L, Busch MA, Weigel D. Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. Plant Cell. 2003; 15(6):1296–309. Epub 2003/06/05. PMID: 12782724

39. Schauer SE, Schluter PM, Baskar R, Gheyselinck J, Bolanos A, Curtis MD, et al. Intronic regulatory elements determine the divergent expression patterns of AGAMOUS-LIKE6 subfamily members in Arabidopsis. Plant J. 2009; 59(6):987–1000. Epub 2009/05/29. doi: 10.1111/j.1365-313X.2009.03928.x PMID: 19473325

40. Yoo SK, Wu X, Lee JS, Ahn JH. AGAMOUS-LIKE 6 is a floral promoter that negatively regulates the FLC/MAF clade genes and positively regulates FT in Arabidopsis. Plant J. 2011; 65(1):62–76. Epub 2010/12/24. doi: 10.1111/j.1365-313X.2010.04402.x PMID: 21175890

41. Coustham V, Li P, Strange A, Lister C, Song J, Dean C. Quantitative modulation of polycomb silencing underlies natural variation in vernalization. Science. 2012; 337(6094):584–7. Epub 2012/07/17. doi: 10.1126/science.1221881 PMID: 22798408

42. Kleinboelting N, Huep G, Kloetgen A, Viehoever P, Weisshaar B. GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. Nucleic Acids Res. 2012; 40(Database issue):D1211–5. Epub 2011/11/15. doi: 10.1093/nar/gkr1047 PMID: 22080561

43. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, et al. Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science. 2003; 301(5633):653–7. Epub 2003/08/02. doi: 10.1126/science.1086391 PMID: 12893945.

44. Kuromori T, Hirayama T, Kiyosue Y, Takabe H, Mizukado S, Sakurai T, et al. A collection of 11 800 single-copy Ds transposon insertion lines in Arabidopsis. Plant J. 2004; 37(6):897–905. Epub 2004/03/05. PMID: 14996221

45. Lisch D. Epigenetic regulation of transposable elements in plants. Annu Rev Plant Biol. 2009; 60:43–66. Epub 2008/11/15. doi: 10.1146/annurev.arplant.59.032607.092744 PMID: 19007329

46. McCue AD, Slotkin RK. Transposable element small RNAs as regulators of gene expression. Trends Genet. 2012; 28(12):616–23. Epub 2012/10/09. doi: 10.1016/j.tig.2012.09.001 PMID: 23040327

47. Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, et al. One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science. 2004; 303(5657):521–3. Epub 2003/11/25. PMID: 14631047

48. Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, et al. Role of transposable elements in heterochromatin and epigenetic control. Nature. 2004; 430(6998):471–6. Epub 2004/07/23. PMID: 15269773

49. Michaels SD, He Y, Scortecci KC, Amasino RM. Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. Proc Natl Acad Sci U S A. 2003; 100(17):10102–7. Epub 2003/08/09. PMID: 12904584

50. Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD. Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. Proc Natl Acad Sci U S A. 2004; 101(44):15670–5. Epub 2004/10/27. PMID: 15505218

51. He Y, Doyle MR, Amasino RM. PAF1-complex-mediated histone methylation of FLOWERING LOCUS C chromatin is required for the vernalization-responsive, winter-annual habit in Arabidopsis. Genes Dev. 2004; 18(22):2774–84. Epub 2004/11/03. PMID: 15520273

52. Angel A, Song J, Dean C, Howard M. A Polycomb-based switch underlying quantitative epigenetic memory. Nature. 2011; 476(7358):105–8. Epub 2011/07/26. doi: 10.1038/nature10241 PMID: 21785438

53. Castaings L, Bergonzi S, Albani MC, Kemi U, Savolainen O, Coupland G. Evolutionary conservation of cold-induced antisense RNAs of FLOWERING LOCUS C in Arabidopsis thaliana perennial relatives. Nat Commun. 2014; 5:4457. Epub 2014/07/18. doi: 10.1038/ncomms5457 PMID: 25030056

54. Zhang X, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, Goodrich J, et al. Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol. 2007; 5(5):e129. Epub 2007/04/19. PMID: 17439305

55. Mascarenhas D, Mettler IJ, Pierce DA, Lowe HW. Intron-mediated enhancement of heterologous gene expression in maize. Plant Mol Biol. 1990; 15(6):913–20. Epub 1990/12/01. PMID: 2103480

56. Rose AB. Intron-mediated regulation of gene expression. Curr Top Microbiol Immunol. 2008; 326:277–90. Epub 2008/07/18. PMID: 18630758

57. Sieburth LE, Meyerowitz EM. Molecular dissection of the AGAMOUS control region shows that cis elements for spatial regulation are located intragenically. Plant Cell. 1997; 9(3):355–65. Epub 1997/03/01. PMID: 9090880

58. Sanchez-Bermajo E, Zhu W, Tasset C, Eimer H, Sureshkumar S, Singh R, et al. Genetic architecture of natural variation in thermal responses of Arabidopsis thaliana. Plant Physiol. 2015. Epub 2015/07/22.

59. Koornneef M, Alonso-Blanco C, Vreugdenhil D. Naturally occurring genetic variation in Arabidopsis thaliana. Annu Rev Plant Biol. 2004; 55:141–72. Epub 2004/09/21.: doi: 10.1146/annurev.arplant.55.031903.141605 PMID: 15377217.

60. Alonso-Blanco C, Aarts MG, Bentsink L, Keurentjes JJ, Reymond M, Vreugdenhil D, et al. What has natural variation taught us about plant development, physiology, and adaptation? Plant Cell. 2009; 21(7):1877–96. Epub 2009/07/04. doi: 10.1105/tpc.109.068114 PMID: 19574434

61. Gazzani S, Gendall AR, Lister C, Dean C. Analysis of the molecular basis of flowering time variation in Arabidopsis accessions. Plant Physiol. 2003; 132(2):1107–14. Epub 2003/06/14. PMID: 12805638

62. Pacurar DI, Pacurar ML, Street N, Bussell JD, Pop TI, Gutierrez L, et al. A collection of INDEL markers for map-based cloning in seven Arabidopsis accessions. J Exp Bot. 2012; 63(7):2491–501. Epub 2012/01/28. doi: 10.1093/jxb/err422 PMID: 22282537

63. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 2007; 35(Web Server issue):W71–4. Epub 2007/05/09. PMID: 17485472

64. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001; 29(9):e45. Epub 2001/05/09. PMID: 11328886

65. Trapnell C, Salzberg SL. How to map billions of short reads onto genomes. Nat Biotechnol. 2009; 27(5):455–7. doi: 10.1038/nbt0509-455 PMID: 19430453

66. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14(4):R36. doi: 10.1186/gb-2013-14-4-r36 PMID: 23618408

67. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31(2):166–9. Epub 2014/09/28. doi: 10.1093/bioinformatics/btu638 PMID: 25260700

68. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15(12):550. Epub 2014/12/18. PMID: 25516281

69. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. Genome Res. 2012; 22(10):2008–17. Epub 2012/06/23. doi: 10.1101/gr.133744.111 PMID: 22722343

70. Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM. pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. Plant Mol Biol. 2000; 42(6):819–32. Epub 2000/07/13. PMID: 10890530

71. Sawano A, Miyawaki A. Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. Nucleic Acids Res. 2000; 28(16):E78. Epub 2000/08/10. PMID: 10931937

72. Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 1998; 16(6):735–43. Epub 1999/03/09. PMID: 10069079

73. Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, et al. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Methods. 2009; 6(8):550–1. Epub 2009/08/01. doi: 10.1038/nmeth0809-550 PMID: 19644454

74. Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D. Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res. 2008; 18(12):2024–33. Epub 2008/09/27. doi: 10.1101/gr.080200.108 PMID: 18818371

75. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011; 39(Web Server issue):W29–37. Epub 2011/05/20. doi: 10.1093/nar/gkr367 PMID: 21593126

76. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014; 42(Database issue):D222–30. Epub 2013/11/30. doi: 10.1093/nar/gkt1223 PMID: 24288371

77. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28(10):2731–9. Epub 2011/05/07. doi: 10.1093/molbev/msr121 PMID: 21546353

78. Scotto-Lavino E, Du G, Frohman MA. 3' end cDNA amplification using classic RACE. Nat Protoc. 2006; 1(6):2742–5. Epub 2007/04/05. PMID: 17406530

79. Kaufmann K, Muino JM, Osteras M, Farinelli L, Krajewski P, Angenent GC. Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nat Protoc. 2010; 5(3):457–72. Epub 2010/03/06. doi: 10.1038/nprot.2009.244 PMID: 20203663

80. Box MS, Coustham V, Dean C, Mylne JS. Protocol: A simple phenol-based method for 96-well extraction of high quality RNA from Arabidopsis. Plant Methods. 2011; 7:7. Epub 2011/03/15. doi: 10.1186/1746-4811-7-7 PMID: 21396125