Transcriptional regulation of the *Arabidopsis thaliana* flowering-time gene *GIGANTEA*

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Markus Christian Berns
aus Moers

Köln, November 2012

Die vorliegende Arbeit wurde am Max-Planck-Institut für Pflanzenzüchtungsforschung in Köln in der Abteilung für Entwicklungsbiologie der Pflanzen (Direktor Prof. Dr. George Coupland) angefertigt.

Berichterstatter:	Prof. Dr. George Coupland Prof. Dr. Ute Höcker
Prüfungsvorsitzender:	Prof. Dr. Martin Hülskamp

Tag der mündlichen Prüfung: 14.11.2011

Nothing in biology makes sense except in the light of evolution

Theodosius Dobzhansky, 1973

Abstract

Plants adjust their developmental programmes to the surrounding environment, which allows them to colonise almost every habitat on earth. One key player in regulating different developmental processes in response to the environment in *Arabidopsis thaliana* is GIGANTEA (GI), a circadian-clock regulated protein that is most abundant in the evening. The precise timing of *GI* transcription is proposed to be crucial for it to fulfil its different functions such as the regulation of flowering time, raising the question of how *GI* itself is transcriptionally regulated.

A combination of phylogenetic and genome-wide bioinformatic analysis as well as the study of transgenic promoter-reporter and complementation lines demonstrated that a highly conserved 700bp block within the *GI* promoter is important for many aspects of *GI* regulation and function. These include the response to light and temperature, control of hypocotyl growth and the regulation of flowering time. Moreover, conserved Evening Element (EE) motifs within this block were shown to be important for several specific features of *GI* transcription. Having shown the importance of EEs within the *GI* promoter, all EEs were mapped on a genome-wide level and co-occurrences with other circadian-clock related *cis*-regulatory elements were determined. This analysis revealed striking patterns between EEs and between other *cis*-elements that gave insights into the general transcriptional code in plants.

Taken together, this thesis demonstrates that the pleiotropic functions of *GI* in light signalling, the circadian clock, freezing tolerance and the regulation of flowering time are reflected within its promoter. This work not only contributed to understanding the complex transcriptional regulation of *GI* and its function in the plant, but also provided novel insights into the regulation of co-expressed genes and the general transcriptional code in plants.

Zusammenfassung

Pflanzen passen ihre gesamte Entwicklung an ihre natürliche Umgebung an, eine Fähigkeit, durch die sie fast jedes Habitat auf der Erde erobert haben. Ein Schlüsselregulatur in der Steuerung verschiedener umweltabhängiger Entwicklungsprozesse in *Arabidopsis thaliana* ist GIGANTEA (GI), ein durch die zirkadiane Uhr reguliertes Protein, welches am Abend in der größten Abundanz vorliegt. Es wird vermutet, dass die genaue Steuerung der *GI* Transkription entscheidend für die verschiedenen Funktionen von GI ist, wie z.B. die Regulation des Blühzeitpunktes. Dies wirft die Frage auf, wie *GI* selbst transkriptionell reguliert wird.

Eine Kombination aus phylogenetischer und genom-weiter bioinformatischer Analyse sowie das Studium von Promoter – Reporter und Komplementationslinien hat gezeigt, dass ein hoch konservierter 700bp langer Block innerhalb des *GI* Promoters wichtig für viele Aspekte der *GI*-spezifischen Regulation und Funktion ist. Dazu gehören die Antwort auf Licht- und Temperatursignale, die Kontrolle des Hypokotylwachstums sowie die Regulation des Blühzeitpunktes. Außerdem wurde gezeigt, dass bestimmte *cis*-regulatorische Elemente, die sogenannten Evening Elemente innerhalb dieses Blockes wichtig für verschiedene spezifische Merkmale der *GI* Transkription sind. Nachdem die Wichtigkeit dieser Evening Elemente für den *GI* Promoter gezeigt wurde, wurden die Positionen sämtlicher Evening Elemente und anderer *cis*-regulatorischer Elemente sowie deren Auftreten untereinander im gesamten Genom bestimmt. Diese Analyse zeigte bemerkenswerte Muster zwischen Evening Elementen und anderen *cis*-regulatorischen Elementen und vermittelte somit Einblicke in den generellen transkriptionellen Code in Pflanzen.

Zusammenfassend zeigt diese Doktorarbeit, dass sich die vielfältigen Funktionen von *GI* in Lichtsignaltransduktionswegen, der zirkadianen Uhr, der Kältetoleranz und der Regulation des Blühzeitpunktes auch in seinem Promoter widerspiegeln. Diese Arbeit trägt nicht nur dazu bei, die komplexe Regulation GIGANTEAs und dessen Funktion für die Pflanze zu verstehen, sondern gewährt darüber hinaus auch neue Einblicke in die Regulation co-exprimierter Gene und dem generellen transkriptionellen Code in Pflanzen.

Introduction

1.1 Preamble

Plants provide the basis for almost every other life form on earth. Plants as photoautotrophic organisms fix carbon dioxide from the atmosphere and produce carbon-rich compounds such as sugars or starch, thus providing the basis for the nutrition of all heterotrophic organisms. Doing so, plants have also colonised almost every environment on earth, from the dense growing tropical rainforests to the hostile plains of Antarctica. Unlike most other organisms, plants are sessile and cannot move from their original location, therefore they have to cope with all environmental conditions to which they are exposed at that site. To this end plants have evolved an enormous capacity to respond to a wide range of environmental factors. This plasticity allows plants to adjust their entire development to the surrounding environment, such as over-growing their neighbours to forage more light, synchronizing reproduction with the changing seasons or coping with environmental stresses such as freezing temperatures.

Genetic approaches in the model plant species *Arabidopsis thaliana* have proven a powerful means of identifying the molecular mechanisms underlying plant responses to their environment. Generally distinct, specific pathways confer responses to each environmental stimulus. However, some mutants show surprisingly pleiotropic effects, suggesting that the affected gene contributes to multiple pathways that were previously assumed to be independent.

One of those is *gigantea (gi)*, which was identified in the 1960s as a mutant that shows delayed flowering under long day conditions (Redei, 1962). This late flowering phenotype is accompanied by massive growth of these plants and hence the name of the mutant (Fig. 1.1). During the last 50 years, the underlying gene has been cloned (Fowler et al., 1999; Park et al., 1999) and even though the precise biochemical function of GIGANTEA protein still remains elusive, it clearly regulates many other responses including circadian rhythms, freezing tolerance and sugar accumulation in addition to flowering time.

This thesis studies the transcriptional regulation of *GI* and how this contributes to these responses.

1.2 The functions of GIGANTEA

1.2.1 GI and its role in flowering time regulation

The initiation of flowering is a key step in the life cycle of all higher plants, marking the transition from the vegetative to the reproductive state. Therefore, this process has been studied by plant scientists and breeders for at least 100 years (Bresinsky et al., 2008; Jung and Müller, 2009). In the last two decades research has focused on the application of genetics to the model species *A.thaliana* and a wealth of knowledge was gained. The decision to flower in this facultative long-day plant is regulated by a number of different, but partially overlapping flowering time pathways. These include the vernalisation pathway, the gibberelin pathway, the aging pathway, the autonomous pathway and the photoperiodic flowering pathway (Boss et al., 2004; Fornara et al., 2010). All these pathways respond to internal or external cues such as the age of the plant, regulation by hormones as gibberellins, the presence of a cold period or the length of the day. These pathways converge on transcriptional regulation of so called floral integrators such as *FLOWERING LOCUS T (FT)* or *SUPRESSOR OF OVEREXPRESSION OF CONSTANS (SOC1)* (Fornara et al., 2010). These genes integrate the signals from different floral pathways, which finally results in the reprogramming of a vegetative to a reproductive meristem. Ultimately this complex network of different flowering time pathways enables plants to flower at the most appropriate time of the year and therefore to secure optimal conditions for seed development and maturation.

The photoperiodic flowering time pathway has been extensively studied at the genetic and molecular levels. *A. thaliana* is a facultative long-day plant: long day conditions induce rapid flowering whereas exposure to short days results in a pronounced delay in flowering. A number of mutants have been isolated that are late flowering under long day conditions, but flower almost like wt under short day conditions, such as *co, ft, gi* and *fkf1* (Koornneef et al., 1991). An external coincidence model (Pittendrigh, 1966) was proposed and validated for the induction of flowering depending on the photoperiod. According to this model, *CONSTANS* (*CO*) transcription is regulated by the circadian clock so that its mRNA is abundant in the late evening from around 12h after dawn (Putterill et al., 1995; Suarez-Lopez et al., 2001). This timing of mRNA transcription means that CO mRNA accumulates in the light under long days but only in darkness under short days. As CO protein is rapidly degraded in the dark, no CO can accumulate under short days, thus preventing flowering under these conditions. Under long days, however, the peak of *CO* mRNA occurs in the light phase, thus CO protein can accumulate and activate FT transcription which in turn leads to the induction of flowering

(Suarez-Lopez et al., 2001; Valverde et al., 2004). Recent studies suggest that FT is the long sought florigen, the proposed 'flowering-hormone' that transmits the signal for flowering from the leaves to the shoot apical meristem (Corbesier et al., 2007).

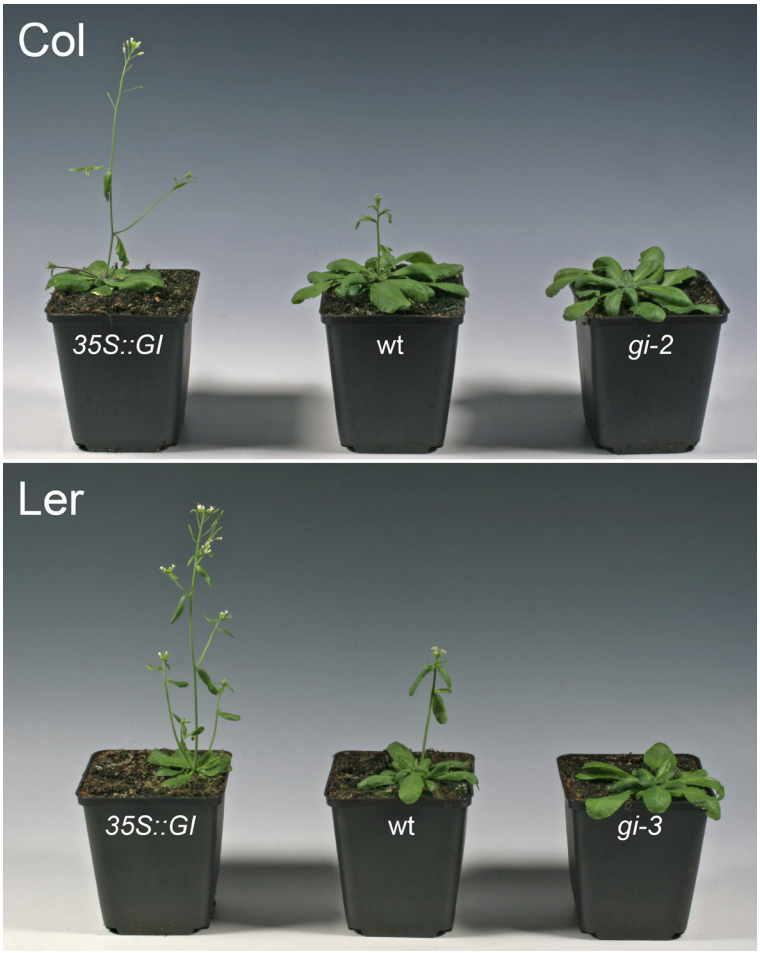


Fig. 1.1 Flowering phenotype of *gi* mutants compared to *35S::GI* and wildtype plants. Pictures show five week old plants grown under long days (16h light / 8h darkness), either in the Columbia (upper panel) or the Landsberg *erecta* (lower panel) background.

GI was proposed to have a crucial function in the promotion of the photoperiodic flowering pathway, as different screens for late-flowering mutants yielded a number of different *gi* alleles and the late flowering of these mutants was always dependent on the photoperiod (Redei, 1962; Koornneef et al., 1991; Fowler et al., 1999; Park et al., 1999) (Fig.1.1). The over-expression of *GI* leads to early flowering irrespective of the photoperiod (Mizoguchi et al., 2005). By genetic analysis *GI* could be placed upstream of *CO* and *FT* (Mizoguchi et al., 2005), but because the GI protein showed no homology to proteins of known function its biochemical role within the photoperiodic flowering time pathway remained elusive.

FKF1 encodes the F-box containing protein FLAVIN-BINDING; KELCH-DOMAIN; F-BOX PROTEIN 1 (FKF1)(Nelson et al., 2000; Imaizumi et al., 2003). Both *GI* and *FKF1* are clock regulated and display a peak in mRNA abundance in the evening. The GI and FKF1 proteins interact

in the nucleus when plants are exposed to blue light and FKF1 is stabilized by this interaction (Sawa et al., 2007), which suggested a mechanism by which GI might regulate *CO* transcription.

The GI-FKF1 complex triggers ubiquitination of CYCLING OF DOF FACTORs (CDFs), which are then degraded in the proteasome. CDFs are transcriptional repressors that bind directly at the *CO* promoter (Sawa et al., 2007). This degradation of CDFs occurs in the evening, when the GI-FKF1 complex is most abundant, which then leads to the accumulation of *CO* mRNA under long day conditions and finally to the expression of *FT* mRNA leading to induction of flowering (Sawa et al., 2007; Fornara et al., 2009). Thus the abundance of GI at the precisely right time of the day is crucial for its function in inducing *CO* mRNA at a time that distinguishes LD and SD. This timing of GI expression is primarily achieved via the exact timing of *GI* mRNA accumulation in the evening (Fig.1.2).

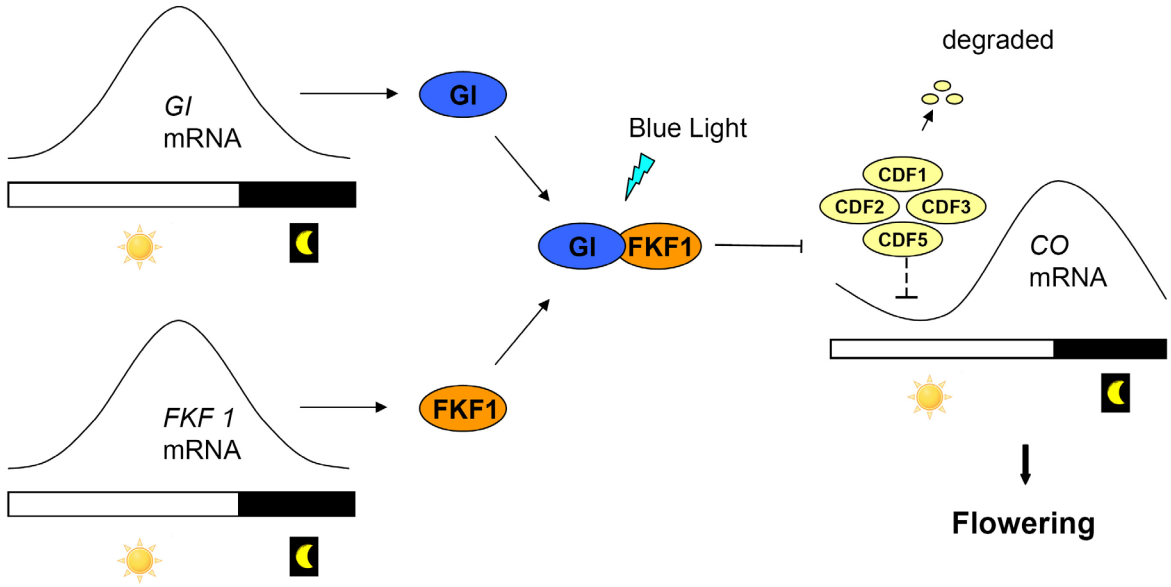


Fig. 1.2 The function of *GI* in the control of photoperiodic flowering in A. thaliana. *GI* and *FKF1* mRNA as well as the respective proteins are most abundant in the evening and form a complex under the influence of blue light. This complex targets CDFs for degradation, thus allowing *CO* transcription in the evening.

However, this evening-abundance of GI is not only achieved by the regulation at the transcriptional level, but also on the posttranscriptional level. Use of tagged versions of GI demonstrated that the protein accumulates in the evening even when being expressed from a strong constitutive promoter (David et al., 2006a). Therefore plants have apparently evolved two layers of regulation in order to ensure the evening-abundance of GI.

Recently it was shown that a quintuple mutant, lacking GI and four CDFs, still retains diurnal expression of *CO* at a similar level as in wildtype plants (Fornara et al., 2009). This surprising result

demonstrates that other factors than GI and the CDFs contribute to the activation of CO transcription within the photoperiodic flowering time pathway, thus adding another layer of complexity to the network of photoperiodic flowering.

GI has also been proposed to regulate the induction of flowering independently from CO. In *gi-2* plants, the level of *miR172* is severely decreased (Jung et al., 2007). This Micro-RNA, in turn, represses TARGET OF EAT1 (TOE), an AP2 like protein that represses *FT* transcription. The mechanism by which GI regulates miR172 levels has not been described. However this creates a CO-independent layer of FT regulation (Jung et al., 2007).

Taken together, the exact timing of *GI* mRNA expression in the evening seems to be crucial for plants to measure day-length and therefore for the decision to flower at the right time of the year.

1.2.2 GI and its role in the plant circadian clock

The earth rotates around its axis every 24h, thus creating day and night. For most organisms this daily change of light and dark periods has major consequences for their metabolism, physiology and behavior. This is especially true for phototrophic organisms such as plants, which directly rely on the exposure to light to ensure their nourishment.

First observations of diurnal leaf movements in tropical tree species were made as early as the 4th century BC by the Greek chronicler Androsthenes, who was exploring the Arabian peninsula in order to prepare its conquest by Alexander the Great (Androsthenes; Bretzl, 1903). Since that time, such leaf movements were discovered and investigated in a number of different plant species. However, it took as long as the 1930s, until Erwin Brünning first proposed that the cause for such plant movements is an internal oscillator with a cycle time of approximately 24h (Bünning, 1936). Together with work from Aschhoff and Pittendrigh this set the basis for a completely new discipline in biology, which was called chronobiology (Dunlap et al., 2003). Soon it became clear that almost every organism on earth contains an internal pacemaker and that this inner pacemaker is regulating an enormous amount of physiological processes, from the leaf movements in plants (Dunlap et al., 2003) to sleep rhythms of humans (Hogenesch and Ueda, 2011).

Circadian clocks enable organisms to match their internal rhythms to external rhythms and thus allow them to anticipate periodic events such as sunrise in the morning or sunset in the evening, an ability that was proposed to confer a selective advantage. In an elegant study by Dodd et al this was demonstrated for *A. thaliana*. Mutant plants with different period lengths were grown under different

day-night cycles of 20h, 24h or 28h. This study revealed that each plant performed best when the pace of its inner clock matched the external photoperiod (Dodd et al., 2005). Wildtype plants outcompeted mutant plants under 24h cycles, demonstrating the fitness advantage of a proper circadian clock in 24h cycles (Dodd et al., 2005). Due to the importance of circadian clocks, a large proportion of transcripts is under circadian control in most organisms. In *Arabidopsis*, between approximately 6% (Harmer et al., 2000a) and 36% (Michael and McClung, 2003) of all transcripts are directly controlled by the circadian clock and up to 90% of all transcripts (Michael et al., 2008b) cycle at least under one environmental condition.

All circadian clocks are characterised by a number of general features (McClung, 2006). First, circadian rhythms have a period of approximately 24h, which is the basis of their name (circa = about, dies = day in latin). Second, the internal rhythm persists under constant conditions, such as continuous light or dark and constant temperature. Furthermore, circadian clocks show temperature compensation, a feature that enables them to operate over a range of physiological temperatures without changing the speed of the oscillator. Finally, circadian clocks can be re-set by strong external stimuli, such as light or temperature pulses. This resetting ensures that the clock can be continuously matched to the environment, a process that is called entrainment. Often the sensitivity of the clock to environmental signals is dependent on the time of day of this signal.

In plants circadian clock research focuses mainly on the model plant *A.thaliana*. In numerous screens for circadian clock mutants many components of the plant circadian clock have been discovered. As for circadian clocks in fruitfly or mammals, a transcriptional feedback loop was proposed for the *A. thaliana* circadian clock. It was proposed that this central loop consists of three major components, TIMING OF CAB EXPRESSION 1 (TOC1) and the two MYB like transcription factors CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY) (Alabadi et al., 2001). Within this central oscillator, TOC1 activates the expression of *LHY/CCA1* in the morning. This happens via interaction of TOC1 with the TCP transcription factor CCA1 HIKING EXPEDITION (CHE) and the resulting complex directly binds the *CCA1* promoter (Pruneda-Paz, 2009). Later during the day LHY/CCA1 represses the expression of *TOC1*, a process that is mediated by direct binding of LHY/CCA1 to the *TOC1* promoter via Evening Elements (EEs) (Alabadi et al., 2001; Perales and Mas, 2007). At dusk, when LHY/CCA1 levels decrease, *TOC1* is activated again, thus completing the loop (Fig.3) (Alabadi et al., 2001).

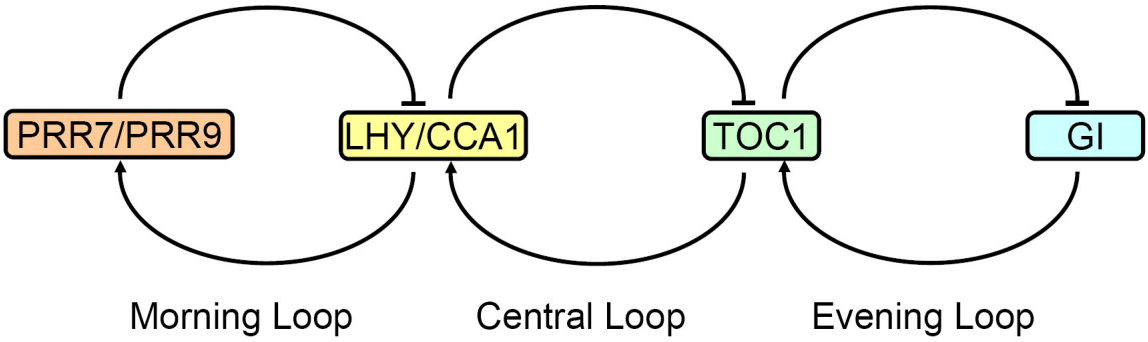


Fig. 1.3 Simplified 3-loop model of the *Arabidopsis* circadian clock.

However, the *lhy/cca1* double mutant still retains rhythmicity under constant conditions (Mizoguchi et al., 2002b), indicating that additional factors play a role in clock function. To address this problem, a model with three interlocked feedback loops was proposed, comprising of the described central loop as well as morning and evening loops (Locke et al., 2006a) (Fig. 1.3). The morning loop contains PSEUDO RESPONSE REGULATOR 7 and 9 (PRR7 and PRR9) (Farre et al., 2005) that are positively regulated by *LHY* and *CCA1* in the morning and in turn negatively regulate *LHY/CCA1* transcription during the day (Locke et al., 2006a).

The evening loop was predicted to comprise TOC1 and the hypothetical factor Y (Locke et al., 2006a). According to the model by Locke et al, Y is repressed by LHY/CCA1 throughout the day and is a positive regulator of *TOC1* in the evening, thus maintaining robust rhythmicity in the plant circadian clock. Furthermore the model proposed that Y is a transducer of light signaling, therefore making the clock sensitive to dusk entrainment and enabling it to measure day length. GIGANTEA was suggested to a likely candidate for factor Y (Locke et al., 2006a).

GI interacts with ZEITLUPE (ZTL), a blue light photoreceptor and F-Box ubiquitin ligase related to FKF1 (Kim et al., 2007). This GI-ZTL complex is blue light dependent and targets TOC1 for degradation in the evening, thus directly regulating a core clock component from the central oscillator. Interestingly, *ZTL* is constitutively transcribed during the day, but ZTL protein shows an evening abundance similar to GI. Therefore, the correct timing of *GI* expression determines the abundance of the GI-ZTL complex in the evening and thus the precise negative regulation of TOC1 (Kim et al., 2007).

This indeed shows the importance of GI in maintaining rhythmicity and sharpening the circadian clock output. However, it also shows that *GI* is not the proposed positive regulator of *TOC1*, as suggested in the model of Locke and colleagues (Locke et al, 2006). Collectively these data indicate that the

proposed evening loop of the *A.thaliana* circadian clock is more complex than initially thought and consequently recent clock models separate the function of GI and a still unknown component Y (Imaizumi, 2010).

Apart from degradation of TOC1, GI also plays a role in temperature compensation of the circadian clock (Gould et al., 2006). The *gi* mutant shows impaired rhythmicity under low temperatures and a shorter period under higher temperatures, indicating that temperature compensation is impaired in these plants. However, the evening-phased expression pattern of *GI* mRNA seems to be almost unaffected between 12°C and 28°C (Gould et al., 2006).

Interestingly, the effects of GI on the circadian clock and the regulation of flowering are separable. Whereas overexpression and mutation have opposite effects on flowering time (i.e. early flowering in the overexpressor and extreme late flowering in the mutant (Fig. 1.1) (Mizoguchi et al., 2005), both cause a short period and low amplitude phenotype (Mizoguchi, 2005). These transgenic data were further supported by the isolation of the *gi-200* allele (Martin-Tryon et al., 2007). This mutant allele has the characteristic short period phenotype of all *gi* mutants, whereas it flowers like wildtype in LD. In SD, gi-200 flowers even earlier than wt, which is caused by a shift of *CO* mRNA expression towards the morning and therefore high *FT* mRNA expression occurs at the end of the day, a pattern that can be explained by the short period phenotype in *gi-200* (Martin-Tryon et al., 2007). However, these data show that, although the timing of *CO* expression is changed in gi-200, its amplitude of expression is not reduced, as observed in other *gi* mutants. This proves the separable functions of GI within the circadian clock and *CO* activation. These different functions may be partly explained by the interaction of GI with two different F-Box proteins. Whereas the interaction between GI and ZTL controls clock function, interaction between GI and FKF1 controls flowering time.

The circadian clock in *A.thaliana* is proposed to be cell autonomous so that every cell contains the same multiloop clock. However, surprisingly the clock in *A.thaliana* roots is different from the above described shoot clock (James et al., 2008). The root clock comprises only the morning loop with LHY/CCA1 and PRR7/9. Interestingly this was proposed to be due to the inability of LHY to bind EEs in roots, at least in the promoter of the *CATALASE3* (*CAT3*) gene (James et al., 2008). Therefore LHY/CCA1 were proposed not to be able to bind EEs in the promoters of the evening-phased clock genes *TOC1* and *GI*, which uncouples the morning loop from the central and the evening loop in roots (James et al., 2008).

In summary, GI is proposed to be an essential component of the *A. thaliana* circadian clock in the shoot, whose precise pattern of expression in the evening is proposed to be crucial for it to fulfill its proper function.

1.2.3 Further functions of GIGANTEA in plants

Plant development is characterised by a number of physiological processes, such as growth, the metabolism of sugars or the deposition of starch. Moreover plants have to cope with different kinds of stress during their life cycle, such as the response to extreme temperatures.

Interestingly, GI seems to be involved in all the aforementioned processes, either directly or indirectly. As the biochemical function of GI within all of these processes is still elusive, predictions for GI function largely rely on the analysis of *gi* mutants.

One well documented feature of the *gi* mutant is its long hypocotyl phenotype that occurs in red, but not in far-red light and is observed for different *gi* mutant alleles (Huq et al., 2000b). Whereas phytochrome A mediates the plant responses to far red light, phytochrome B is the mediator of hypocotyl elongation in red light (Quail et al., 1995). Therefore it was concluded that the *gi* mutation causes a reduction in PHYB signalling, but does not affect PHYA signalling (Huq et al., 2000b). However, no differences in phyA or phyB abundance were observed in *gi*, indicating that *gi* affects the PHYB signalling cascade downstream of the photoreceptor (Huq et al., 2000b).

Furthermore it was shown that different *gi* mutants display a long hypocotyl phenotype also in blue and white light, but not in constant dark (Martin-Tryon et al., 2007). Together with the finding that red and blue light affect the amplitude of the CCR2::Luc marker in *gi* compared to wildtype plants, it was proposed that GI acts both in blue and in red light signalling to the clock (Martin-Tryon et al., 2007). Moreover, even though GI does not affect PHYA signalling under high light conditions, a role in the PHYA-mediated very-low-fluence response was described (Oliverio et al., 2007).

For sessile organisms like plants a response to cold or freezing temperatures is crucial for their fitness and survival in environments where low temperatures occur. Therefore many plants from temperate regions mount a protective response to freezing temperatures when they are exposed to cold. This process called cold acclimation involves a massive reprogramming of the *A.thaliana* transcriptome (Fowler and Thomashow, 2002). The best characterised regulators of cold acclimation are the C-REPEAT BINDING FACTORs (CBFs), CBF1, CBF2 and CBF3. These are AP2-type transcription factors that are immediately upregulated if the plant is exposed to cold temperatures (Gilmour et al., 1998). CBFs in turn induce the expression of COLD REGULATED (COR) genes via direct binding to the C-Repeat promoter element ((Baker et al., 1994; Gilmour et al., 2000). The transcriptional response of the CBFs to cold is gated by the circadian clock (Fowler et al., 2005) and this gating is

most probably dependent on direct binding of the core clock components LHY/CCA1 to EEs and CBS in the promoters of *CBF1, 2* and *3* (Dong et al., 2011).

Surprisingly, also *GI* transcripts were found to be upregulated five- to ten-fold in response to low temperatures (Fowler and Thomashow, 2002). These microarray data were confirmed in another study by semiquantitative PCR by the observation that *gi-3* mutants were less sensitive to freezing temperatures (Cao et al., 2005). The increase in *GI* mRNA levels upon cold temperatures was also observed in *Medicago truncatula* (Paltiel et al., 2006) and *Brachypodium distachyon* (Hong et al., 2010). Therefore not only circadian-clock mediated evening expression, but also the induction of expression by exposure to cold temperatures is likely to be a conserved feature of *GI* in different plant species.

Crosstalk between regulation of flowering time and resistance to cold temperatures has been reported for other mutants. The *sfr6* mutant (Knight et al., 1999) is not only sensitive to freezing temperatures, but also displays a late flowering phenotype under LD conditions. The late flowering correlates with low levels of *GI* and *FKF1* mRNA, which lead to low expression of *CO* mRNA and to no expression of *FT* (Knight et al., 2008). In order to identify downstream targets of SFR6, microarray analysis was performed and a set of 209 genes were found to be down-regulated at least 2 fold in *sfr6*. Interestingly, EEs were over-represented in the promoters of these putative targets of SFR6 (Knight et al., 2008), thus making it a candidate for an upstream regulator of *GI*. However, cloning of the sfr6 locus revealed that *SFR6* encodes a 1268 amino acid protein without any known function or homology to any known protein (Knight et al., 2009).

A number of additional phenotypes have been described for the *gi* mutant: It is more resistant to paraquat, a herbicide that causes oxidative stress (Kurepa et al., 1998). Moreover, *gi* accumulates starch in its leaves, not only at the adult stage, but also in seedlings (Eimert et al., 1995). GI was also reported to affect fruit set in *A.thaliana* and therefore to have an important ecological role (Brock et al., 2007). A recent study suggests that GI is involved in the wall ingrowth deposition of phloem parenchyma cells of leaf minor veins in *A.thaliana*, a process that is induced by high light and cold temperatures (Edwards et al., 2010a). Finally GI physically interacts with SPINDLY (SPY), a negative regulator of gibberellin (GA) signalling. This study suggests that the GI-SPY complex in concert regulates hypocotyl growth and flowering time in *A. thaliana* (Tseng et al., 2004).

1.2.4 Summary

Taken together, the *gi* mutant exhibits a complex pleiotropic phenotype. Apart from its well documented functions within the circadian clock and in the regulation of flowering time, it has additional roles in plant architecture, physiology, metabolism and stress responses. In how far all these different functions reflect related or independent traits remains to be elucidated. However, at least for its flowering time and circadian clock functions, the precise timing of *GI* abundance proposed to be crucial for its function, as this leads to the stabilisation of the related ubiquitin ligases FKF1 and ZTL at the appropriate time of the day.

1.3 Transcriptional regulation

1.3.1 General mechanism

Nearly every cell of a multi cellular organism has the same genetic information encoded in the DNA of its nucleus. However, multi-cellular organisms consist of different cell types that are specialised for different tasks. A major source of this variety is the different transcriptional programming of distinct cell types. Transcriptional programming in multicellular organisms does not only change during the development of the cell, but also upon the response to internal and external stimuli, such as hormonal regulation or the response to environmental signals.

This variation in gene transcription is often regulated by its 5' upstream sequence, commonly defined as the promoter of a gene, but can also involve intronic or 3' end sequences. Promoters comprise a core promoter region close to the transcriptional start site and further distal promoter elements (Alberts et al., 2007).

Core-promoters are regions that are located close to the transcriptional start site und contain the region where the initial binding of the RNA polymerase II (Pol II) complex takes place. The best characterised core promoter element is the TATA-Box (Alberts et al., 2007). However, a number of other core-promoter motifs have been described, such as TC-rich repeats in plants (Bernard et al., 2010). The distal promoter typically consists of a number of so called *cis*-regulatory elements, short stretches of DNA that are bound by transcription factors. Such *cis*-regulatory elements often occur clustered within promoters and are called *cis*-regulatory modules (Nguyen and Xu, 1998; Zinzen et al., 2009).

The discovery of gene regulation emerged from the pioneering work of Jakob and Monod, who described the regulation of the Lac operon in *Escherichia coli* (Jacob and Monod, 1961). Subsequent research has revealed the basic concepts of transcription, so that a widely established model of the general transcriptional machinery has been described. Generally, transcription can be divided into 4 phases: pre-initiation, initiation, elongation and termination (Alberts et al., 2007). In the pre-initiation phase, the Pol II complex assembles at the core promoter, often in association with specific transcription factors bound to *cis*-regulatory elements. In the initiation phase Pol II complex then binds to the core promoter and transcription can start. In the elongation step the polymerase moves along the template DNA strand and produces a messenger RNA in 5'→3' direction. Finally, in the termination phase, the polymerase uncouples from the DNA strand and the produced mRNA is released (Alberts et al., 2007).

1.3.2 Recent concepts of transcription

The basic concepts of transcription have been described to some detail in the second half of the 20th century. However, a complete picture including the interplay of combinatorial, temporal and spatial transcriptional regulation is still missing even on the level of single genes. However, new techniques such as Next Generation Sequencing or ChIP-seq have yielded an enormous amount of genome-wide *in vivo* data and have increased our understanding of transcription tremendously.

One outcome of such genome-wide studies is that transcription is occurring in most eukaryotic genomes even outside of protein-coding genes. In mammals, for instance, 80% of the genome is being transcribed whereas only 1% of the genome consists of protein coding regions (Kim et al., 2005; Heard et al., 2010). Similar observations were reported for the genomes of rice (Li et al., 2006) and yeast (David et al., 2006b) The function of much of this pervasive transcription is still not clear. In most cases short stretches of non-coding RNA are being transcribed, especially in 5' upstream regions of the transcriptional start site (Dutrow et al., 2008). Transcription of such non-coding RNAs can change the chromatin state of the promoter region and thus prepare the protein-coding gene for transcription (Hirota et al., 2008). Therefore it has been suggested that the pervasive transcription of the genome has an important regulatory function in transcription of classical protein-coding genes.

Furthermore transcription occurs in well defined locations within the nucleus, which were named transcription factories (Osborne et al., 2004; Cook, 2010) and were found in different species, for instance in human erythroid cells (Fraser et al., 2010) or in fission yeast (Tanizawa et al., 2010). These transcription factories occur in different numbers in different cell types specialised transcription factories might transcribe co-regulated genes (Tanizawa et al., 2010). Indeed, such transcription

factories contain several polymerases and are capable of transcribing at least three different genes from three different chromosomes in parallel (Fraser et al., 2010).
In plants, transcription factories have not been described so far. However, as the entire transcriptional machinery is well conserved in all eukaryotes, transcription factories are also likely to occur in plants.

1.3.3 Circadian clock related *cis*-regulatory elements in *Arabidopsis*

In plants, a number of clock-related *cis*-regulatory elements have been discovered and described. One well-characterised one is the Evening Element (EE). This 9bp element (AAAATATCT) was shown to be overrepresented in the promoters of evening-phased genes and if multimerised within an artificial promoter construct confers evening expression in luciferase-marker based assays (Harmer et al., 2000b; Harmer and Kay, 2005). Related elements are the CCA1 binding site (CBS, AAAAAATCT) (Michael and McClung, 2002) and a shorter version of the EE, called the short EE (SEE, AATATCT) (Mikkelsen and Thomashow, 2009). EEs are bound by Myb-like transcription factors such as LHY and CCA1. The best characterised example is the binding of LHY and CCA1 to an EE in the *TOC1* promoter, an interaction that is proposed to be crucial for circadian clock function (Alabadi et al., 2001; Perales and Mas, 2007). Furthermore it was demonstrated that LHY and CCA1 can form homo- and hetero- dimers (Lu et al., 2009; Yakir et al., 2009) and that the binding of this LHY/CCA1 complex to the EE is temperature compensated (O'Neill et al., 2011).

However, LHY and CCA1 are not the only factors that bind EEs. A study using protein microarrays coupled to mass-spectrometric analysis revealed a number of candidate transcription factors that might be capable of binding EEs (Rawat et al., 2011). Most of these candidates belong to the REVEILLE family, Myb-like transcription factors that are related to LHY and CCA1 (Rawat et al., 2009). However, this screen also indentified a number of putative binding partners from other transcription factor families, such as the WRKY and WHIRLY families (Rawat et al., 2011).

A number of other *cis*-elements that play a role in clock-mediated transcriptional regulation or related light responses have been described. Most of these elements were defined based on genome-wide analysis of transcriptomic data and identification of co-regulated genes. In these studies, a number of time course microarray-experiments were analysed and genes were clustered according to the peak expression of their transcripts (Michael et al., 2008a; Michael et al., 2008b). Subsequently the promoters of these genes were investigated for overrepresented motifs (Michael et al., 2008a; Michael et al., 2008b).

One *cis*-regulatory element discovered in these screens is the Hormone up at down Box (HUD Box, CACATG), which is a variant of the core binding site for bHLH transcription factors. This HUD-Box was found to be overrepresented in genes with a peak in expression around dawn (Michael et al., 2008a). Multimerised versions of this element inserted adjacent to the luciferase open reading frame together with a minimal promoter conferred a dawn expression pattern, demonstrating that this HUD Box is indeed capable of driving diurnal expression. Moreover, the HUD Box is implicated to play a role in PHYB-mediated responses (Michael et al., 2008a).

Another element discovered in these screens is the Protein Box (PBX, ATGGGCC), a *cis*-regulatory element that is overrepresented in the promoters of genes with the GO annotation 'protein synthesis' and confers peak expression in the early night in artificial luciferase constructs (Michael et al., 2008b). Two other elements, the Starch Box (SBX, AAGCCC) and the Telobox (TBX, AAACCCT) are over-represented in promoters that confer night-specific expression as well, thus the PBX/TBX/SBX was defined as a night specific module. The SBX and the TBX, however, were only predicted based on computational analysis and no experimental proof has been reported regarding the biological function of these elements (Michael et al., 2008b).

Two of the largest transcription factor classes in plants comprise bZIP and bHLH transcription factors that bind to a number of partially overlapping *cis*-regulatory elements, the so called E-Box, the G-Box and the C-Box (Quail, 2000). The best characterised of these three is the G-Box (CACGTG) (Giuliano et al., 1988), a *cis*-regulatory element that binds bHLH transcription factors such as PHYTOCHROME INTERACTING FACTOR 3 (PIF3) (Quail, 2000).

Related to the G-Box and other bHLH-binding *cis*-regulatory elements are bZIP binding sites.
These include the ABA Response element-like element (ABREL, CACGT), a variant of the core binding site for bZIP transcription factors (Suzuki et al., 2005). It was shown that these elements in concert with EEs drive the cold-dependent expression of *COL1* and *COR27* (Mikkelsen and Thomashow, 2009), suggesting that EEs and ABRELs might act in concert.

Name (Abbreviation)	Consensus	Description	Reference
Evening Element (EE)	AAAATATCT	Over-represented in evening phased genes, multimerised version confers evening expression, can be bound by LHY /CCA1 and other Myb-like transcription factors	(Harmer et al., 2000b; Harmer and Kay, 2005)
CCA1 Binding Site (CBS)	AAAAAATCT	Very similar to the EE, presumably with the same function	(Michael and McClung, 2002)
Short Evening Element (SEE)	AATATCT	Two nucleotides shorter than the EE.	(Mikkelsen and Thomashow, 2009)
LUX Binding Site (LBX)	CGAATC	Consensus sequence that binds LUX ARRHYTHMO with highest efficiency	(Kay et al., 2011)
ABA Response Element-like (ABREL)	CACGT	Core binding site for bZIP transcription factors; regulates cold responses of several genes together with EEs	(Suzuki et al., 2005; Mikkelsen and Thomashow, 2009)
G-Box	CACGTG	Core binding-site for bHLH transcription factors	(Giuliano et al., 1988)
Morning Element (ME)	CCACAC	Overrepresented in the promoters of morning-phased genes	(Michael et al., 2008b)
Hormone Up at Dawn (HUD Box)	CACATG	Implicated in PHYB signalling	(Michael et al., 2008a)
Protein-Box (PBX)	GGCCCAT	Overrepresented in the promoters of night-phased genes	(Michael et al., 2008b)
Telobox (TBX)	AAACCCT	Overrepresented in the promoters of night-phased genes	(Michael et al., 2008b)
Starch-Box (SBX)	AAGCCC	Overrepresented in the promoters of night-phased genes	(Michael et al., 2008b)

Tab. 1 Overview of circadian-clock related *cis*-regulatory elements in *Arabidopsis thaliana*

1.3.4 Combinatorial approaches

Different *cis*-regulatory elements are often clustered within promoters and together form so-called *cis*-regulatory modules. Depending on their regulatory functions, such modules are also termed as enhancer or silencer. *Cis*-regulatory modules have been investigated extensively in different species. In *D. melanogaster*, for instance, it was demonstrated that more then 2000 sites in the genome are targeted by more than eight different transcription factors und therefore these modules were called high occupancy target (HOT) regions (Negre et al., 2011). Similar patterns were discovered for *C.elegans* (Gerstein et al., 2010), suggesting that HOT regions are a common feature of eukaryotic promoters. Moreover, in *Drosophila* that such *cis*-regulatory modules can act redundantly, for instance to confer phenotypic robustness under non-optimum conditions (Frankel et al., 2010).

Also in *Arabidopsis thaliana* such regulatory modules and the combinatorial action of different *cis*-elements has been demonstrated, but only for a limited number of promoters and lesser detail. For CRABS CLAW (CRC), it was shown that a combination of conserved *cis*-regulatory modules drives the spatial expression pattern of *CRC* (Lee et al., 2005), whereas a combination of different *cis*-regulatory elements is crucial for the correct expression pattern of *COL1* (Mikkelsen and Thomashow, 2009), *COR27* (Mikkelsen and Thomashow, 2009) and *ELF4* (Li et al., 2011).

One implication from such combinatorial approaches is that a gene that responds to many different stimuli would have a more complex promoter with a higher number of *cis*-regulatory elements. Indeed such a correlation has been shown for *D.melanogaster* (Nelson et al., 2004), *C.elegans* (Nelson et al., 2004) and *A. thaliana* (Walther et al., 2007). Interestingly, the study in *Arabidopsis* revealed that EEs and ABRELs were over-represented in the set of genes that respond to a large number of stimuli, with the ABREL having highest statistical significance from all *cis*-regulatory elements in this analysis (Walther et al., 2007). This indicates that these two elements are involved in a number of different developmental traits or environmental responses.

1.3.5 *GI* transcriptional regulation

The transcriptional regulation of *GI* has not been studied extensively. It shows characteristic evening expression, which was observed in every plant species investigated so far, such as *Arabidopsis thaliana (Fowler et al., 1999), Medicago truncatula (Paltiel et al., 2006)*, *Pisum sativum (Hecht et al., 2007)*, *Hordeum vulgare* (Dunford et al., 2005), *Brachypodium distachyon (Hong et al., 2010)* or *Oryza sativa (Hayama et al., 2002)*. Therefore it is likely that the evening expression of *GI* is an integral part of its function. In *A.thaliana* it was shown that this peak in expression depends on the

photoperiod and that the peak of *GI* expression tracks dusk ((Edwards et al., 2010b); Toth and Cremer, unpublished). Moreover, *GI* is strongly upregulated in light and upon the exposure to cold temperatures, a pattern that has been described in *Arabidopsis* (Fowler and Thomashow, 2002; Cao et al., 2005), *Medicago* (Paltiel et al., 2006) and *Brachypodium* (Hong et al., 2010). Collectively this suggests that *GI* is similarly regulated in a wide range of species and therefore similar *cis*- and *trans*-regulatory factors might mediate these transcriptional responses. However, even though the expression pattern of *GI* has been described in some detail, the underlying *cis*- and *trans*- acting factors causing this complex transcriptional expression pattern of *GI* have remained completely elusive so far.

1.3.5 Summary

The regulation of transcriptional responses is key for all living organisms on earth. Although the basic concepts of transcriptional regulation have been described in the last decades, we still do not understand transcription in all its different aspects, neither on the level of single promoters nor on the level of entire cells or organisms. Although several promoters have been described in some detail, a complete picture about all binding sites and all transcription factors that regulate a single gene is still lacking. In plants a very limited number of promoters have been studied in detail. Due to the high conservation of the clock- light- and temperature-dependent expression pattern across different species *GI* is an ideal candidate for intensively studying the *cis*- and *trans*- regulatory factors that determine its precise expression pattern.

1.4 Phylogenetic analysis

Comparative analyses of gene sequences between species are widely used to identify evolutionary conserved and mechanistically important segments in proteins. This concept is also used to study promoter sequences, which evolve faster than protein sequences. This method is called phylogenetic footprinting or phylogenetic shadowing. Both concepts are based on the idea that functionally important sequences are maintained during evolution und thus will be more conserved between related species than functionally non-important DNA.

The term "phylogenetic footprinting" was introduced by Tagle and coworkers, who compared the 5' upstream sequences of globulin encoding genes from a wide range of mammals in order to identify conserved *cis*-regulatory elements (Tagle et al., 1988). The term phylogenetic footprinting was chosen as the same conserved *cis*-regulatory elements were discovered with this approach as in previous DNase footprinting experiments (Tagle et al., 1988). Since then the term phylogenetic footprinting has been used for the comparison of promoter sequences from more distantly related species. The term "phylogenetic shadowing" was introduced in 2003 by Boffelli and co-workers who compared the human genome to orthologous sequences from different ape and monkey species (Boffelli et al., 2003). Due to the close relationship of the species used this comparison revealed larger conserved regions instead of single conserved *cis*-regulatory elements, thus this method was termed shadowing instead of footprinting.

Both methods have been successfully applied in many species, for single promoters as well as on the genome-wide level. Genome-wide phylogenetic analysis was conducted in yeast (Kellis et al., 2003), fruitfly (Negre et al., 2011) and mammals (Waterston et al., 2002), revealing a huge number of novel and known binding sites and giving insights into the combinatorial action of such *cis*-regulatory elements. In plants, such genome-wide studies have not been reported so far, probably due to the lack of different sequenced plant genomes from the same family. Plant-related genome-wide approaches focus more on evolutionary aspects, such as the evolution of genome size (Hu et al., 2011) or the evolution of genome duplications (Van de Peer et al., 2009). However, for a few single loci intensive phylogenetic analyses were used in order to elucidate evolutionarily conserved *cis*-regulatory elements. Examples of such studies are for the first intron of *AGAMOUS* (AG) (Hong et al., 2003) or the promoters of *CRC* (Lee et al., 2005), *LHY* (Spensley et al., 2009) or *FT* (Adrian et al., 2010).

1.5 Brassicaceae

Flowering plants are classified into families based on their evolutionary relatedness. A number of these families are of particular interest to research and agriculture, such as the Poaceae (including the world's three most important crops rice, maize and wheat) or the Solanaceae (including tomato, potato and tobacco). Another example is the mustard family (Brassicaceae / Crucifera), a taxon with more than 300 genera and around 3700 species (Bailey et al., 2006; Franzke et al., 2011). These include not only the well studied model plant *Arabidopsis thaliana*, but also a number of important crop plants such as rapeseed and the different cabbage varieties. Characteristic for the Brassicaceae family are flowers with four sepals, four cross-like arranged petals and six stamens of which two are shorter than the others. The pistil comprises two fused carpels, whereas the fruit is a capsule that is called silique or siliqule, depending on the ratio its length to its width (Bresinsky et al., 2008). Another characteristic is the production of glucosinolates, secondary metabolites that play a crucial role in defence against herbivores.

Most Brassicaceae are annual, biennial or perennial herbs, but small shrubs or lianas have evolved in a number of genera (Franzke et al., 2011). Geographically the Brassicaceae occur around the world, with the highest diversity across Europe and Asia, especially in Asia Minor (Koch et al., 2006).
The familiy Brassicaceae comprises the two major lineages Aethionemeae and the Core Brassicaceae, respectively. Recent reports estimate the divergence time between these two tribes to be 40 to 50 mya (Clauss, 2006). This coincides with a whole genome duplication that took place approximately 40 mya years ago (Franzke et al., 2011). Therefore it was suggested that this last genome duplication did not only divide the Aethionemeae from the Core Brassicaceae, but also provided the genetic basis for the adaptive radiation that occurred within the Core Brassicaceae after that duplication event (Franzke et al., 2011). The cradle of the Brassicaceae is still under debate with suggestions ranging from North America to Europe and Asia. However, some reports suggest an origin in Turkey, as the diversity of both Aethionemae and Core Brassicaceae is still highest here, coupled with extreme diverse ecosystems on a small scale, thus providing optimal conditions for such an adaptive radiation as took place for the Brassicaceae (Franzke et al., 2011).
The sister family of the Brassicaceae within the order of Brassicales is the Cleomaceae (Schranz and Mitchell-Olds, 2006), a family that comprises many tropical species and reaches its highest diversity in South America. The order Brassicales contains a number of other agronomically important families, such as the Capparaceae (includes the caper bush), the Caricaceae (including Papaya, which has been sequenced) and the Resedaceae (including Dyer's rocket, which was used for dying cloth in the past) (Schranz and Mitchell-Olds, 2006; Bresinsky et al., 2008).

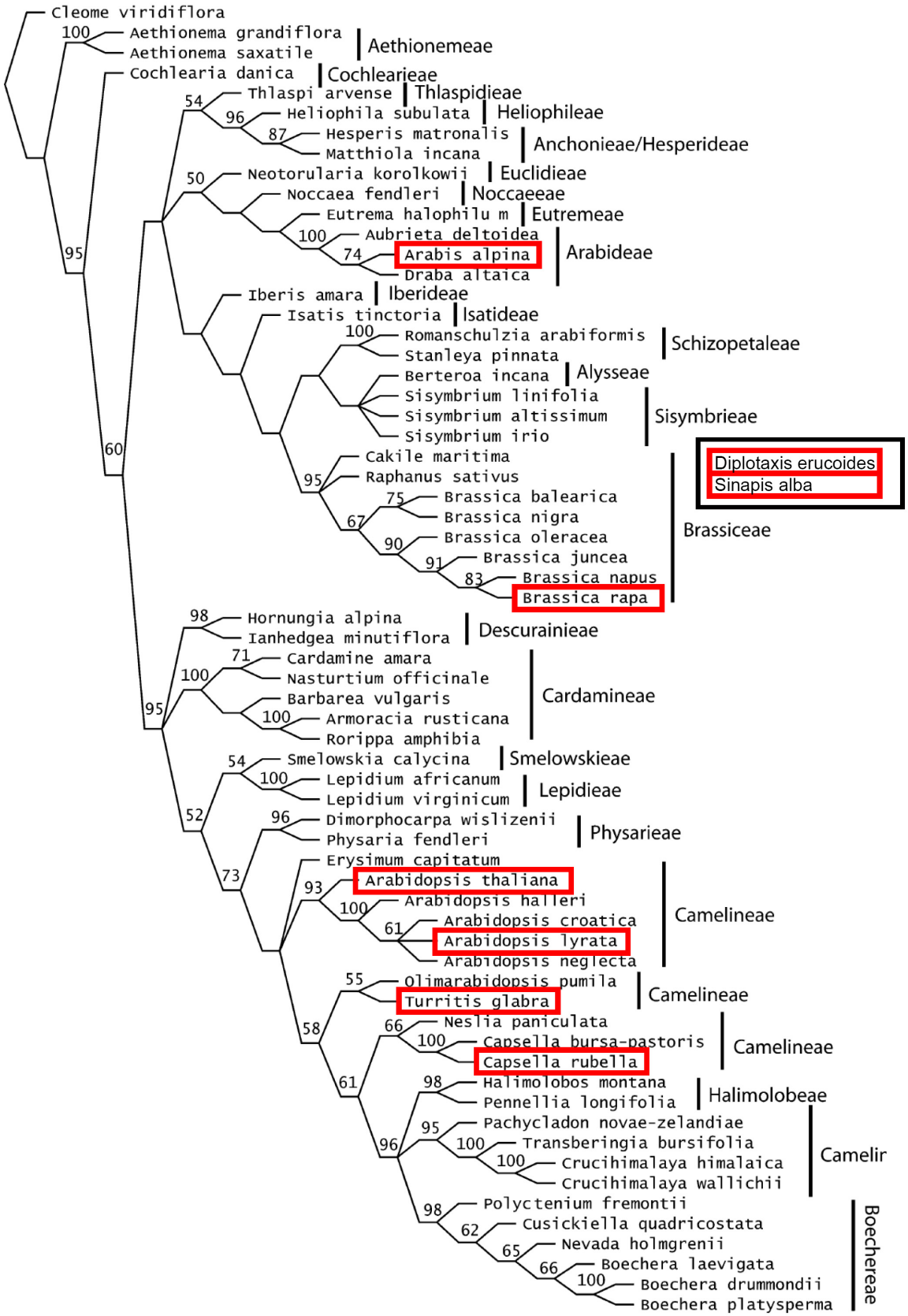


Fig.1.4 Phylogeny of the Brassicaceae, adapted from Bailey et al, 2007. Species that were used for the phylogenetic analysis of the *GI* promoter (chapter 1) are highlighted with red boxes. *Diplotaxis erucoides* and *Sinapis alba* were added to the clade Brassiceae in a recent phylogeny (Franzke et al., 2011).

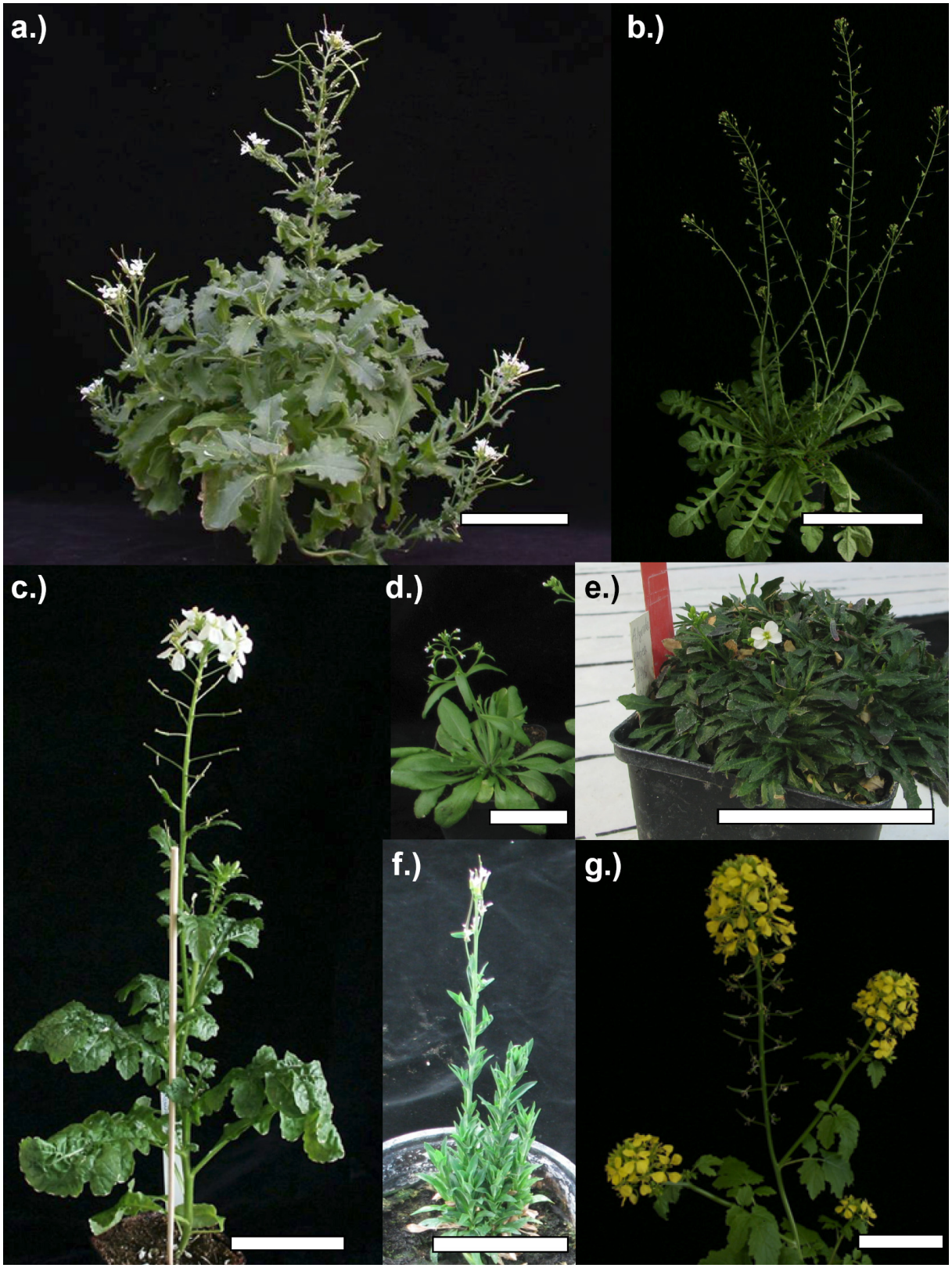


Fig. 1.5 Overview of Brassicaceae species that were used for the phylogenetic analysis of the *GI* promoter. a.) *Arabis alpina* Pajares; b.) *Capsella rubella* Monte Gargano; c.) *Diplotaxis erucoides*; d.) *Arabidopsis thaliana* Columbia; e.) *Arabidopsis lyrata*; f.) *Turritis glabra*; g.) *Sinapis alba*; not included: *Brassica rapa ssp. pekiniensis*; White bars = 10cm (a, b, c, f) or 5cm (d, e., f), respectively.

Arabidopsis. thaliana is the most intensively studied Brassicaceae species. It is an annual or biennial rosette plant that is native to the entire northern hemisphere (Kaul et al., 2000). Its rapid life cycle, small genome (125Mb (Kaul et al., 2000)) and an increasing number of available and easy-to-use genomic tools made it the model plant of choice and therefore a wealth of knowledge was accumulated in this model species.

Arabidopsis lyrata is one of the closest relatives of *A. thaliana* (divergence time approximately 5mya (Clauss and Koch, 2006)) and has colonised temperate to mild regions of the northern hemisphere (Clauss and Koch, 2006). Unlike *A. thaliana*, it occurs in diploid as well as in polyploid forms and is self-incompatible. This makes it less accessible to genetic experiments, but makes it a good model to study these traits in comparison to *A. thaliana*. The genome of diploid *A. lyrata* is fully sequenced and has a size of 207Mb (Hu et al., 2011).

The genus *Capsella* is one of the closest to the genus *Arabidopsis* within the Brassicaceae with a divergence time of 10-14mya (Clauss and Koch, 2006). *Capsella rubella* is an annual weed that is, as *A. thaliana*, diploid and self compatible. Together with its close relative, the self-incompatible tetraploid *Capsella bursa-pastoris*, it is an invasive weed that has colonised different habitats around the world (Hurka and Neuffer, 1997; Hintz et al., 2006). The genome size of *Capsella rubella* was not reported so far.

The different varieties of *Brassica rapa* are important vegetable and oil crops. Within the genus Brassica, it has the smallest genome size (529 Mb (Mun et al., 2010)) and is therefore a good model for studying the polyploid genome structure in Brassica. Its divergence time from *Arabidopsis* was calculated, depending on the study, to a time between 13 – 17 mya (Mun et al., 2010) and 16 – 21 mya (Clauss and Koch, 2006).

Arabis alpina is a perennial that is native to arctic-alpine habitats in Europe (Koch et al., 2006) and is diploid, self-fertile and susceptible to Agrobacterium-mediated transformation, thus making it an ideal model organism for perennial plant species (Wang et al., 2009). Calculated divergence times to *Arabidopsis* range from 19 – 25mya (Koch et al., 2001), to 16 - 50mya (Clauss and Koch, 2006), or based on fossile records approximately 43mya (Beilstein et al., 2010). Collectively this might suggest that the divergence time between *Arabidopsis* and *Arabis* could be approximately 30 million years. The genome size of *A. alpina* is approximately 370Mb (Nordstöm et al., unpublished).

Other emerging model organisms within the Brassicaceae are *Arabidopsis halleri* (to study heavy metal tolerance (Hanikenne et al., 2008)), *Capsella bursa-pastoris* (to study flowering time and loss of

petals (Hintz et al., 2006; Ziermann et al., 2009)) *or Cardamine hirsuta* (to study leaf shape, (Hay and Tsiantis, 2006)).

Taken together, there is a variety of different species from the mustard family that is used to study various different aspects of plant biology. This offers a rich source of comparative approaches within the mustard family, especially on the level of genome sequences. The availability of the full genomic sequence or large parts of it from *A. thaliana*, *A. lyrata*, *C. rubella*, *B. rapa* and *A. alpina* now makes it possible to use this information for different kinds of phylogenetic analyses.

Moreover, the rough divergence time of 5mya (*A. lyrata*), 10mya (*C. rubella*), 20mya (*B. rapa*) and 30mya (*A. alpina*) from *A. thaliana* provides an ideal 'phylogenetic clockwork' that allows comparisons at multiple levels.

1.6 Aim of this study

GIGANTEA has important functions in controlling flowering time and related environmental responses in Arabidopsis.

The precise temporal transcriptional pattern of *GI* is highly conserved and has been proposed to be crucial for its function within the circadian clock, the regulation of flowering time and probably for its response to environmental stresses. The main scope of this study is to elucidate the *cis*- and *trans*-acting factors that control the precise temporal expression pattern of *GI* and to test whether these are required for the biological function of *GI*.

I made use of a phylogenetic shadowing approach in order to indentify functional important *cis*-regulatory modules and elements within the *GI* promoter. Subsequently, candidate modules and *cis*-elements from this comparative analysis were tested for their biological function by subcloning or mutating respective fragments and using the luciferase reporter system to analyse the expression patterns conferred by these sequences under different conditions. In order to further investigate the contribution of these *cis*-regulatory modules and elements, *GI* was mis-expressed using these different promoter constructs in the background of a *gi* mutant. The resulting transgenic plants were analysed for a number of GI-specific traits, such as flowering time under different photoperiods or the response to freezing temperatures.

Finally, knowledge gained from the *GI* promoter analysis was applied to study *cis*-regulatory elements in related promoters and ultimately on a genome-wide level. Therefore, statistical analyses of the co-occurance of different *cis*-regulatory elements were conducted on a whole-genome level. This approach helped to better understand the mode of *GI* transcriptional regulation and provided insights into the general transcriptional code in plants.

Phylogenetic analysis of the *GI* promoter

2.1 Perspective

GIGANTEA is one of the key players in the regulation of flowering time and the circadian clock in *Arabidopsis thaliana (Fowler et al., 1999)*. In all plant species investigated so far, *GI* shows an evening-phased expression, suggesting that this precise temporal expression pattern is an integral part of GI function. To reveal the molecular basis of the evening phase of *GI* transcription, I made use of a phylogenetic shadowing approach with the goal of identifying conserved regions and *cis*-regulatory elements in the *GI* promoter that are crucial for the precise light- and clock- dependent temporal pattern of *GI* transcription.

2.2 Isolation of orthologous GI promoters

Sequences of orthologous *GI* promoters from Brassicaceae species were isolated from databases or by amplification using PCR and subsequent sequencing as described in the Materials and Methods section. The length of promoter was defined as sequence between the translational start site and the last exon of the upstream gene, encoding *POLYADENYLATE BINDING PROTEIN 3 (PAB3, At1g22760*).
Briefly, the sequences of *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Brassica rapa ssp. pekiniensis, Capsella rubella*, and *Arabis alpina* were identified from databases, whereas the promoters from *Diplotaxis erucoides*, *Sinapis alba* and *Turritis glabra* (Bürstel and Cremer, unpublished) were amplified from plant material and then sequenced.

2.3 Evolutionary conservation of the GI promoter in Brassicaceae

Multiple sequence alignments are a valuable tool for comparing different numbers of orthologous sequences. However, it has been also shown that multiple sequence alignments can be misleading, especially for noncoding sequences (Tompa and Chen, 2010). Indeed several attempts to generate multiple sequence alignments at the beginning of this study with all eight Brassicacean promoters using the standard multiple alignment tools ClustalW (Larkin et al., 2007) and DIALIGN (Morgenstern, 2004) did not lead to satisfactory results (data not shown).

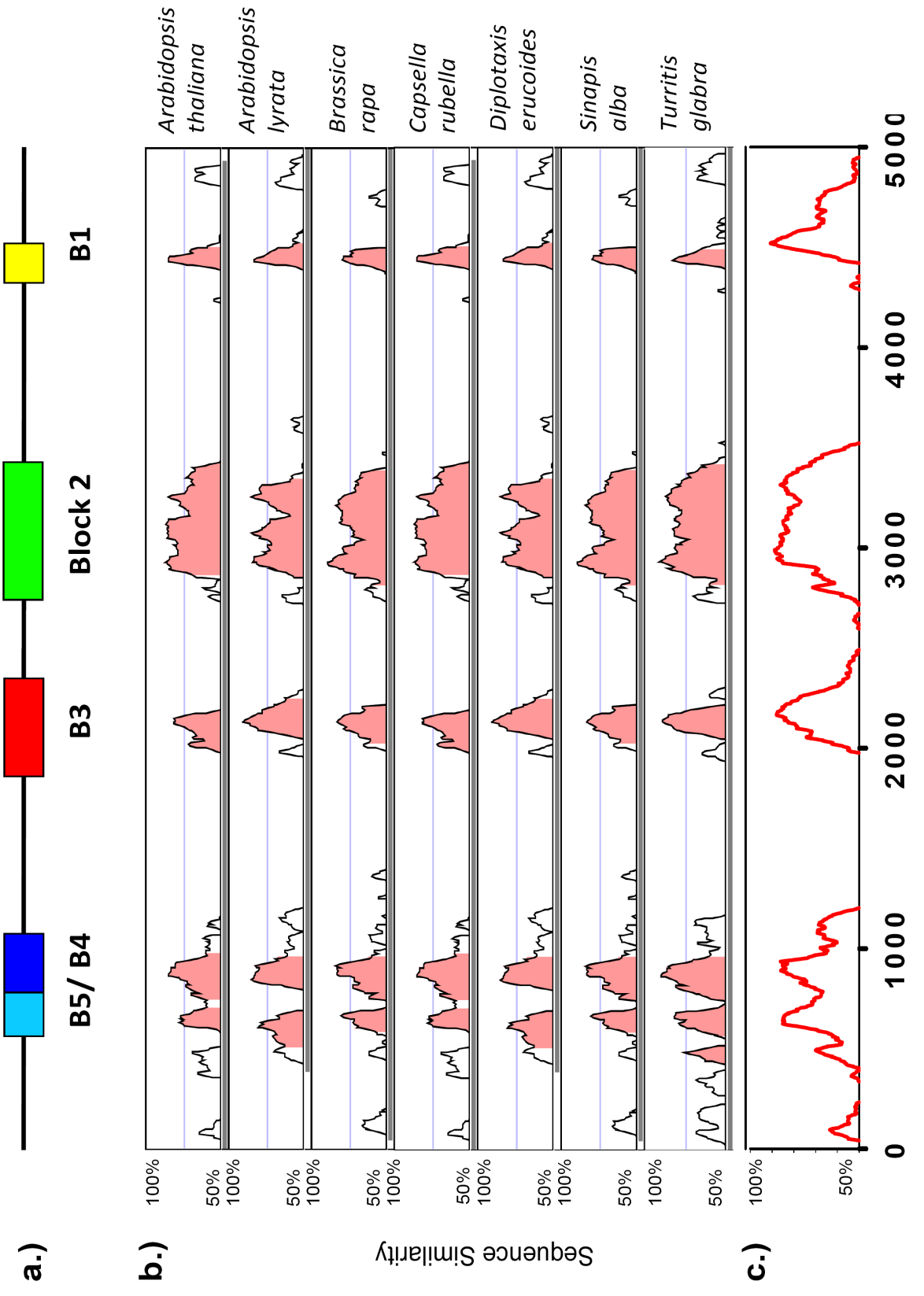

a.)
B5/ B4
B3
Block 2
B1
b.)
Arabidopsis thaliana
Arabidopsis lyrata
Brassica rapa
Capsella rubella
Diplotaxis erucoides
Sinapis alba
Turritis glabra
100%
50%
Sequence Similarity
c.)
0
1000
2000
3000
4000
5000

Fig. 2.1 Evolutionary conservation of the *GI* promoter in Brassicaceae

a.) Conserved blocks 1-5, defined on the basis of high conservation between the *GI* promoters of *A. thaliana* and *A.alpina*. Yellow = Block 1, Green = Block 2, Red = Block 3, Dark Blue = Block 4, Blue = Block 5
b.) Pairwise alignments of eight Brassicaceae *GI* promoters. Alignments performed with CHAOS / Shuffle-LAGAN and visualised with VISTA Browser. Sliding window=100bp; Conservation=70%; Red colour indicates regions where a sliding window of 100bp shows at least 70% of conservation. All sequences aligned to the 5kb *GI* promoter of *A.alpina.*
c.) Multiple sequence alignment with seven *GI* promoter sequences aligned to the *GI* promoter of *A.alpina;* same scalebar for all figures.

Pairwise alignments using LAGAN (Brudno et al., 2003c) can efficiently identify highly conserved regions within orthologous plant promoters, as demonstrated for the *CRABS CLAW* gene (Lee et al., 2005). However, whereas the accuracy of a pairwise alignment is generally better than a multiple sequence alignment (Brudno et al., 2003a), a multiple sequence alignment contains more sequence information. In order to combine the strengths of both methods while minimising the disadvantages, I applied a sequential pipeline of first using a pairwise alignment tool to identify and extract highly conserved regions. In a second step these conserved regions were used for multiple sequence alignments and in a final step putatively functional *cis*-regulatory elements were selected based on a candidate list or high conservation of a >10bp motif.

First, pairwise alignments of all *GI* promoters from eight Brassicaceae were made using the alignment tools CHAOS and Shuffle LAGAN. CHAOS first detects highly conserved regions between two sequences and uses them as anchor for the subsequent alignment. Shuffle-LAGAN then fills the gaps between the anchors and finishes the alignment (Brudno et al., 2003a). The resulting pairwise alignments were visualised using the VISTA Browser (Mayor et al., 2000; Brudno et al., 2007) with a sliding window of 100bp and a conservation threshold of 70%. Finally all pairwise alignments were piled up for each species, such as Fig 2.1b for *Arabis alpina*. VISTA Plots for the remaining seven species can be found in the Supplements.

This analysis revealed high conservation within the *GI* promoter that occurs in five distinct blocks that were named Block1, Block2, Block3 and Block4/5 (Fig. 2.1a). The *A.alpina* promoter was used as a reference sequence for comparisons for several reasons. First, the *A.alpina* sequence is the most diverged one in this set of eight sequences (see Fig. 2.6), therefore making it ideal for comparisons to all other species. Second, it was shown that the 5kb *GI* promoter of *A. alpina,* fused to the luciferase open reading frame and stably transformed into *Arabidopsis thaliana* plants, confers the same expression pattern as the 2,5kb *A. thaliana GI* promoter (see chapter 4).

Therefore I concluded that important sequence information for the light- and clock-dependent transcriptional regulation of *GI* must be present in the conserved sequence between these two species. Thus the exact size of the blocks was defined on the basis of conservation between *A.thaliana* and *A. alpina*. (For detailed sequence information and VISTA plots of all species see Supplements).

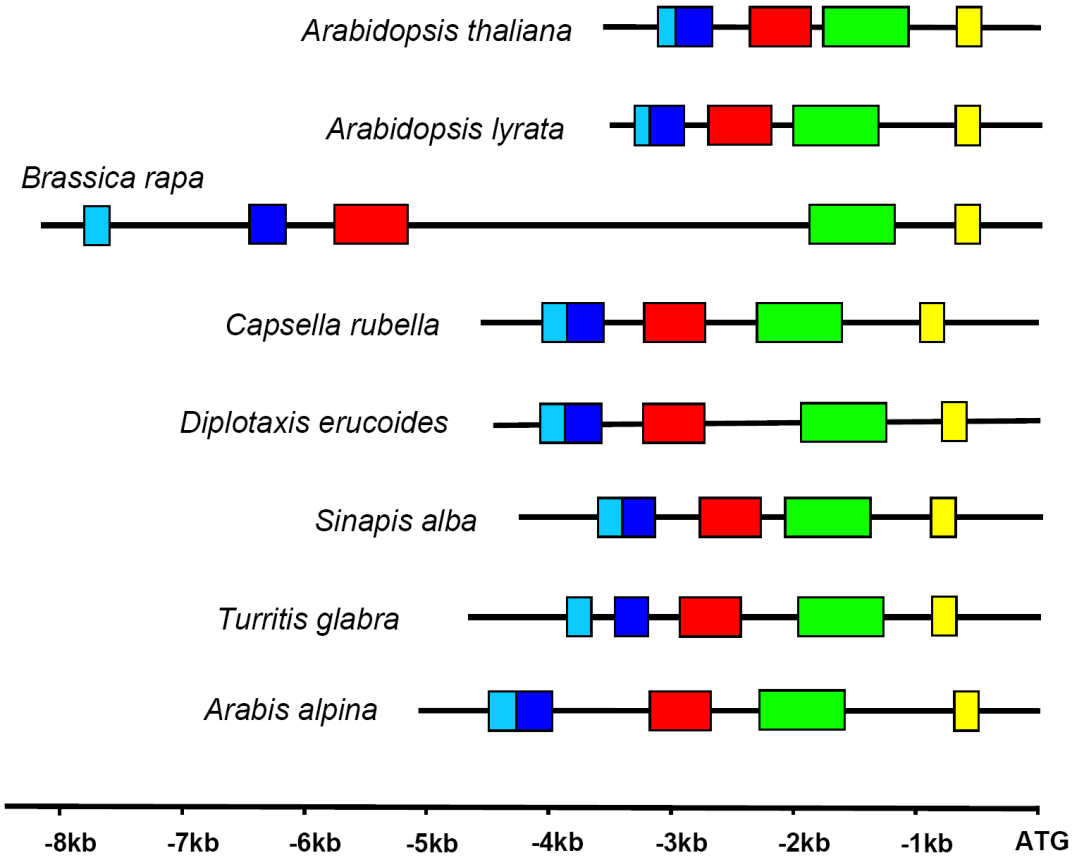


Fig. 2.2 Distribution of five conserved blocks within orthologous *GI* promoters of eight different Brassicaceae species. Yellow = Block 1, Green = Block 2, Red = Block 3, Dark Blue = Block 4, Blue = Block 5

Next, the spatial distribution of the previously defined conserved blocks was elucidated. Therefore all conserved blocks were marked within the respective sequences and a pile up performed for comparison (Figure 2.2). This analysis showed that the length of the different orthologous *GI* promoters was highly variable, reaching from 3,6kb in *A.thaliana* up to more than 8kb in *B.rapa*. Whereas size and synteny of the five conserved blocks were well conserved in all eight species, location and spacing were quite variable (Fig. 2.2). Interestingly, Block4/5 occurred as one contiguous block in most species apart from *T. glabra* and *B. rapa*, where it is separated.

As a 2,5kb fragment upstream of the translational start site in *A.thaliana* is sufficient to phenocopy the expression pattern of the endogenous *GI* promoter (Cremer et al., unpublished) and to complement the *gi*-mutant phenotype when fused to the *GI* cDNA (chapter 5), subsequent analysis focused on this 2,5kb fragment, which includes the conserved blocks 1, 2 and 3.

2.4 Block 2 contains a number of known cis-regulatory elements

In order to discover important *cis*-regulatory elements I first compiled a candidate list of known elements that have been shown to play a role in light- and circadian clock-mediated transcriptional regulation (for overview see introduction, Table 1.1).

Visual inspection of the full length *A.thaliana GI* promoter revealed that it contains many of the elements present in the candidate list. Even more strikingly, all these elements were almost exclusively located within conserved Block 2, suggesting that this 700bp region is of particular importance for the light- and clock mediated transcriptional regulation of the *GI* promoter.

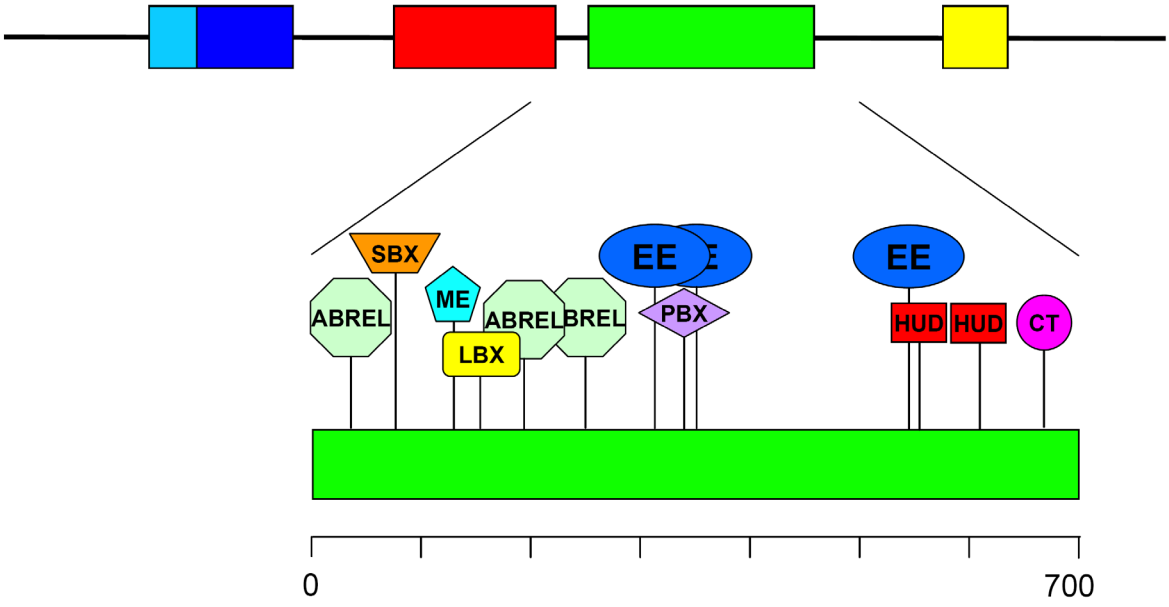


Fig. 2.3 Conserved *cis*-regulatory elements within Block2 of the *GI* promoter.
ABREL= ABA Response Element-like; SBX=Starch Box; LBX=LUX ARRYTHMO Binding Site; ME=Morning Element; EE=Evening Element; PBX=Protein Box; HUD= Hormone Up at Dawn Box; CT = CT Element

Conservation shown as WEBLOGO (all eight species included)	Name / Consensus	Position / Strand	Reference
	ABREL CACGT	33-37 -	Mikkelsen et al, Plant Journal (2009)
	Starch-Box AAGCCC	75-80 +	Michael et al PLOS Genetics (2008)
	Morning Element CCACAC	135-140 +	Michael et al PLOS Genetics (2008)
	LUX Binding Site CGAATC	146-151 -	Helfer et al., Current Biology (2011)
	ABREL CACGT	194-198 -	Mikkelsen et al, Plant Journal (2009)
	ABREL CACGT	255-259 -	Mikkelsen et al, Plant Journal (2009)
	Evening Element AAAATATCT	317-326 +	Harmer et al, Nature (2000)
	Protein-Box GGCCCAT	341-346 -	Michael et al PLOS Genetics (2008)
	Evening Element AAAATATCT	350-358 -	Harmer et al, Nature (2000)
	Evening Element AAAATATCT	536-544 +	Harmer et al, Nature (2000)
	HUD Box CACATG	550-555 +	Michael et al, PLOS Biology (2008)
	HUD Box CACATG	604-609 +	Michael et al, PLOS Biology (2008)
	CT element	702-707 +	Bernard et al. BMC Genomics (2010)

Tab. 1 Overview about conserved *cis*-regulatory elements
Overview of conserved *cis*-regulatory elements that have been described to play a role in light- or clock-mediated transcriptional regulation and that are found within Block 2; Degree of conservation is shown as WEBLOGO.

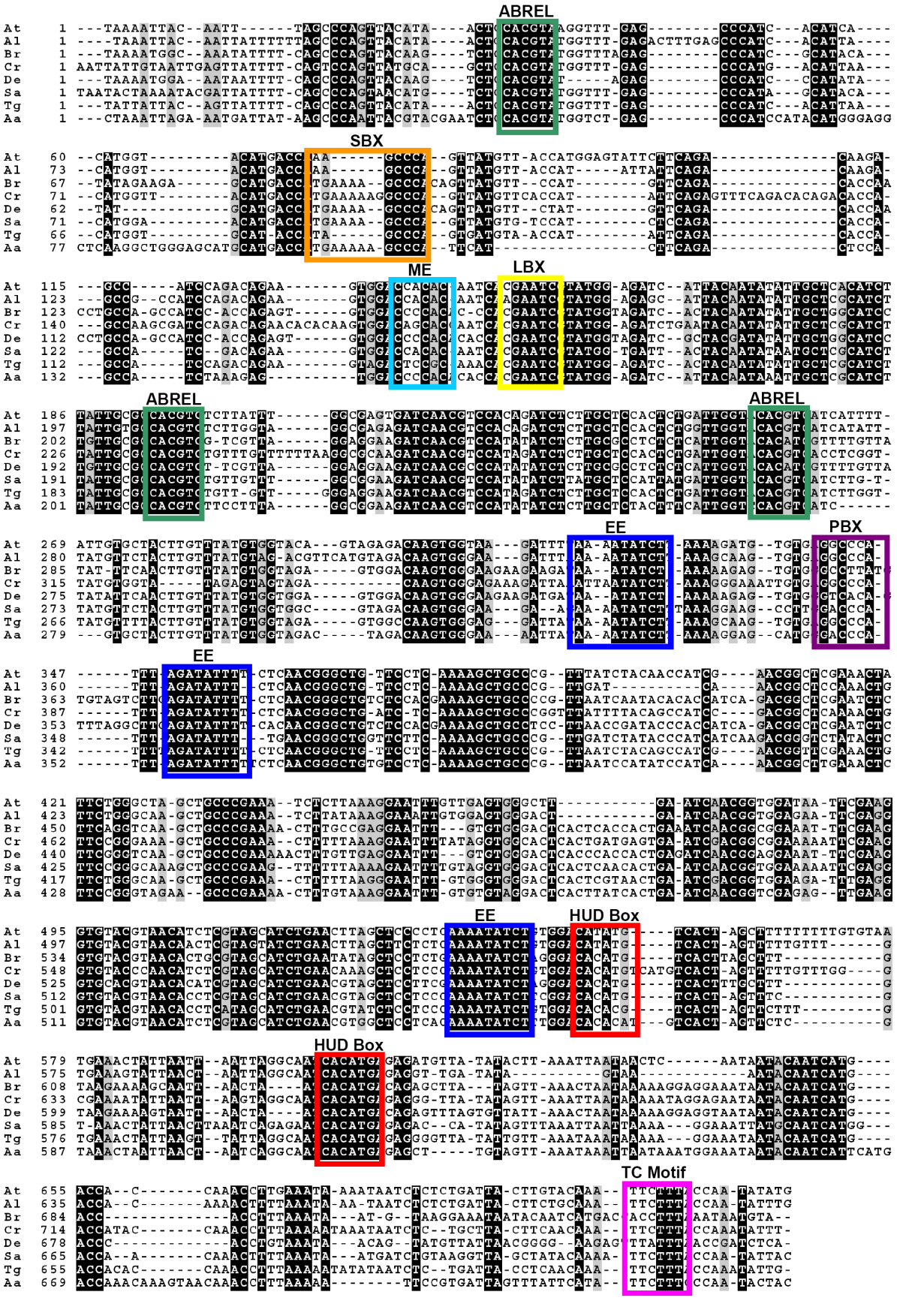


Fig. 2.4 Multiple sequence alignment of conserved Block 2 from eigth orthologous *GI* promoters
At=*Arabidopsis thaliana*; Al=*Arabidopsis lyrata*; Br=*Brassica rapa*; Cr=*Capsella rubella*; De=*Diplotaxis erucoides*; Sa=*Sinapis alba*; Tg=*Turritis glabra*; Aa=*Arabis alpina*

Conserved *cis*-regulatory elements are highlighted with coloured boxes.
ABREL= ABA Response Element-like; SBX=Starch Box; LBX=LUX ARRYTHMO Binding Site; ME=Morning Element; EE=Evening Element; PBX=Protein Box; HUD= Hormone Up at Dawn Box; CT = CT Element

In order to further elucidate the conservation of specific *cis*-regulatory elements and to finish the phylogenetic shadowing pipeline, a multiple sequence alignment was made with conserved Block 2 (Fig 2.4). For better representation, conservation of *cis*-regulatory elements was visualised using WEBLOGO (Crooks et al., 2004) and putative important conserved *cis*-regulatory elements are shown in Table 2.1 according to their appearance in Block 2.

Among those conserved *cis*-regulatory elements were 2 Hormone Up at Dawn-Boxes (HUD-Box), elements that are proposed to play a role in phytochrome B mediated signalling and are a variant of the core recognition site for bHLH transcription factors (Michael et al., 2008a). Other elements present in Block 2 are a Morning Element (ME), a Protein Box (PBX) and a Starch Box (SBX), elements that have been identified as important for clock control in genome-wide screens (Chory et al., 2008). Furthermore, a LUX ARRYTHMO Binding Site (LBX) was identified (Kay et al., 2011), suggesting that the clock-related transcription factor LUX might directly bind the *GI* promoter. Moreover, three ABA Response Element-like elements (ABRELs) were found in Block2, cis-elements that are bound by bZIP transcription factors and have been shown to be important for the regulation of the cold response in two different promoters (Mikkelsen and Thomashow, 2009)(Fig. 2.3, Fig. 2.4; Table 2.1)

Most notably, three Evening Elements (EEs) were found in Block 2 (Fig. 2.4). The EE was initially identified by analysing promoters of evening-expressed genes, where the EE was found to be over-represented. Artificial promoter constructs with multimerised versions of the EE adjacent to a minimal promoter were able to confer evening-expression in luciferase-based assays, demonstrating that the EE is sufficient to drive evening expression (Harmer et al., 2000b; Harmer and Kay, 2005). Interestingly, a recent study reported that EEs and ABRELs in concert regulate the response to cold temperatures of the two evening-phased genes *COL1* and *COR27* (Mikkelsen and Thomashow, 2009). As both ABRELs and EEs occur three times within Block 2 of the *GI* promoter and as all of them show extreme conservation (Fig 2.3; Table 2.1), these particular elements seemed to be promising candidates for further biological analysis.

In addition, highly conserved sequences within Block 2 were identified that do not have a previously described function, but might also be functionally important. Therefore stretches of at least six bases that are absolutely conserved in all species were aligned with surrounding sequence and are shown as WEBLOGOS in Table 2.2. Interestingly, these stretches contain also variants of already described *cis*-regulatory elements as a DOF Binding Site (Yanagisawa and Schmidt, 1999), MYC binding sites that

are capable of binding INDUCER OF CBF EXPRESSION1 (ICE1) (Shirsat et al., 1989; Dunn et al., 1998; Chinnusamy et al., 2003) and several others (Solano et al., 1995; Simpson et al., 2003).

Conservation	Position	Comments / Putative Function/ Reference
	214-226	Contains ACGT, core motif for bZIP binding (Simpson et al., 2003)
	300-311	Contains CANNTG, core motif for MYC binding, can be bound by ICE (Chinnusamy et al., 2003)
	359-370	Contains CNGTTR , bound by different Myb transcription factors (Solano et al., 1995)
	377-387	Contains AAAG, DOF binding site (Yanagisawa et al., 1999)
	431-441	Contains CCGAAA, low temperature element in barley (Dunn et al., 1998)
	509-523	Contains CANNTG, core motif for MYC binding, can be bound by ICE (Chinnusamy et al., 2003)
	643-654	Contains CAAT Box (Shirsat et al., 1989)

Tab. 2 Overview of highly conserved sequence fragments within Block 2 with no described function. Degree of conservation is shown as WEBLOGO.

2.5 Core promoter elements at the end of Block 2

Curiously, a conserved TC element was discovered at the end of Block 2. TC elements are over-represented upstream of the transcriptional start sites of plant genes and appear to replace TATA-Boxes in TATA-less promoters (Bernard et al., 2010). However, the TC element in Block 2 of the *GI* promoter is approximately 500bp from the described transcriptional start site of *GI* mRNA. In order to further elucidate the end of conserved Block 2, the last 80bp of this region were compared in a multiple alignment and the entire region is shown as WEBLOGO (Fig.2.5). Due to lower conservation of this region, only the sequences of the three closest relatives to *A.thaliana* were used for this alignment (*A.lyrata*, *C.rubella* and *T.glabra*). However, this analysis shows that this region contains the TC element as well as a TATA Box and a CT repeat that was proposed to play a role in plant gene

expression (Pauli et al., 2004) and two different versions of the CAAT – Box. Therefore the end of Block 2, at least in *A.thaliana* and its close relatives contains all the features of a core promoter and thus transcription might initiate within this block.

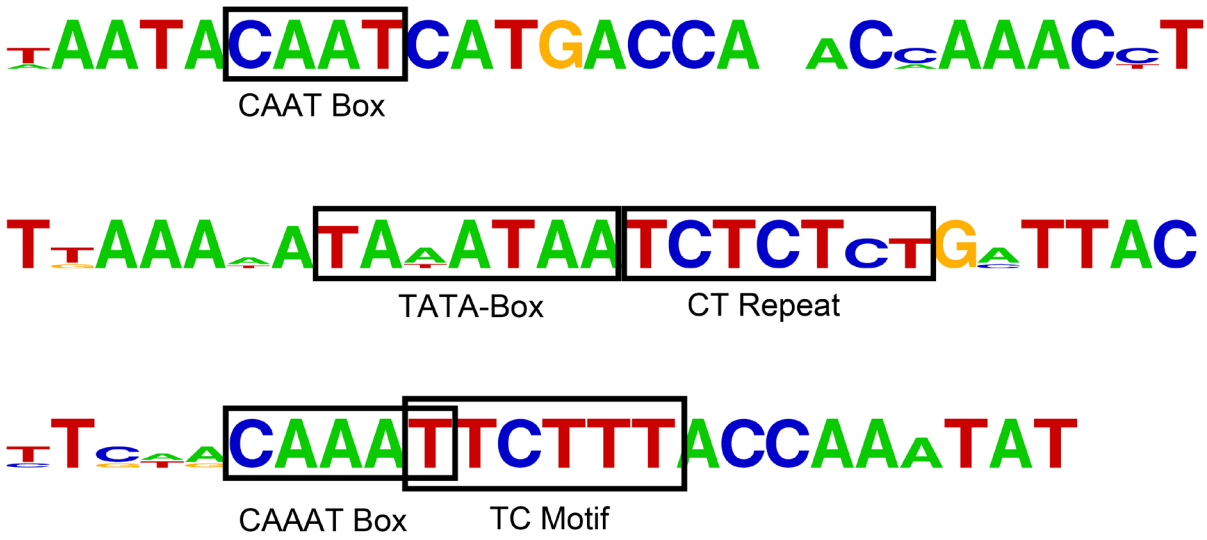


Fig. 2.5 Conservation of the last 80bp of Block 2. Conservation shown as WEBLOGO, sequences compared are from *A.thaliana, A.lyrata, C.rubella* and *T.glabra*. Conserved core promoter elements are highlighted with black boxes.

2.6 Evolutionary conservation of the GI promoter in grass species

Multiple sequence alignments of the *GI* promoters of Brassicaceae with more distant dicotyledonous species such as *Medicago truncatula* or *Populus trichocarpa* did not yield results because the sequences could not be aligned, although *GI* mRNA in *M.truncatula* is expressed in a similar temporal pattern to *A. thaliana* (data not shown). Also using the conserved Block 2 to detect conservation in *GI* promoters of more distant species was not successful, demonstrating that phylogenetic shadowing, at least for the *GI* promoter, is restricted to evolutionary closely related plant species.

Therefore I analysed the *GI* promoters of the five grass species *Oryza sativa, Hordeum vulgare, Setaria italica, Sorghum bicolor* and *Zea mays*. All but the *H.vulgare* sequence were isolated from publically available databases. The *H.vulgare* promoter was isolated and sequenced from a BAC containing Hv*GI* (Bürstel and Cremer, unpublished). The *Brachypodium distachyon* sequence was not used due to gaps in the sequence upstream of Bd*GI*. Interestingly, two copies of *GI* were discovered in *Z.mays* and therefore both promoters were analysed.

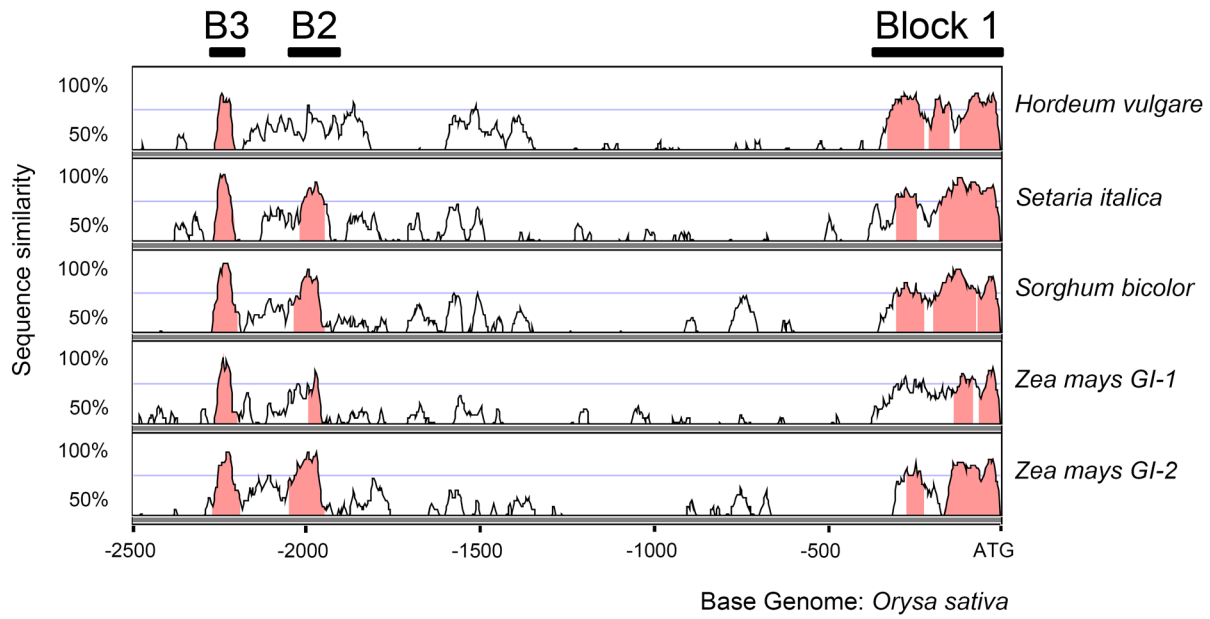


Fig. 2.6 Evolutionary conservation of the grass *GI* promoter
Pairwise alignments of six grass *GI* promoters. Alignments performed with CHAOS / Shuffle-LAGAN and visualised with VISTA Browser. Sliding window=50bp; Conservation=80%; Red colour indicates regions where a sliding window of 80bp shows at least 80% of conservation. All sequences aligned to the 2,5kb upstream region of the *Oryza sativa GI*.

Conservation of known cis-regulatory elements in six grass species	Location	Comments / Function
	Block1	ABREL, binds bZIP transcription factors; two further ABRELs in Block1 with low conservation
	Block2	CCA1 Binding Site; similar to Evening Element; can be bound by the Myb-like transcription factor CCA1
	Block3	Consensus 1equence of the Evening Element, but last base is changed (A instead of T)

Tab. 3 Overview of conserved *cis*-regulatory elements with a light- or clock- function and are found in the *GI* promoters of six grass species and that were previously described in Arabidopsis. Conservation is shown as WEBLOGO.

Similarly to the analysis of the *GI* promoters in the Brassicaceae, high conservation was discovered in distinct regions of the grass *GI* promoters. However, in contrast to the Brassicaceae the highest conservation occurred in a 400bp region directly upstream of the translational start site (Block1), whereas conservation in more distal parts of the promoters was restricted to much smaller patches (Block2 and Block3) (Fig. 2.7). As for the Brassicaceae, conserved sequences were isolated and

analysed for known *cis*-regulatory elements. Strikingly, this analysis revealed one CBS, one ABREL and one EE-related motif that were conserved within five grass species (Table 2.3), demonstrating that similar *cis*-regulatory elements are conserved between such distant clades as the Brassicaceae and the Poaceae, although their position, spacing and order are unrelated.

2.7 Evolutionary conservation of the PRR9 promoter

In order to use the described phylogenetic shadowing pipeline for a limited number of already sequenced Brassicaceae, I analysed the promoter of the clock gene *PRR9* from *A.lyrata*, *C.rubella* and *A.alpina*. This analysis revealed high conservation of a region starting 200bp upstream to approximately 80bp downstream of the transcriptional start site of *PRR9* (Fig 2.8, Fig 2.9). Further inspection of this conserved region also revealed an absolutely conserved EE as well as a highly conserved LBS (Fig 2.9). These two elements appear in the same configuration as in the *GI* promoter, with the LBS being located approximately 150 bp upstream of the EE. Furthermore two conserved G-Boxes were found in this region, *cis*-regulatory elements can be bound by bHLH transcription factors and that have been shown to play a role in light-mediated transcriptional responses. Notably, also the *PRR9* promoter contains conserved CT motifs, both upstream and downstream of the transcriptional start site (Fig 2.9).

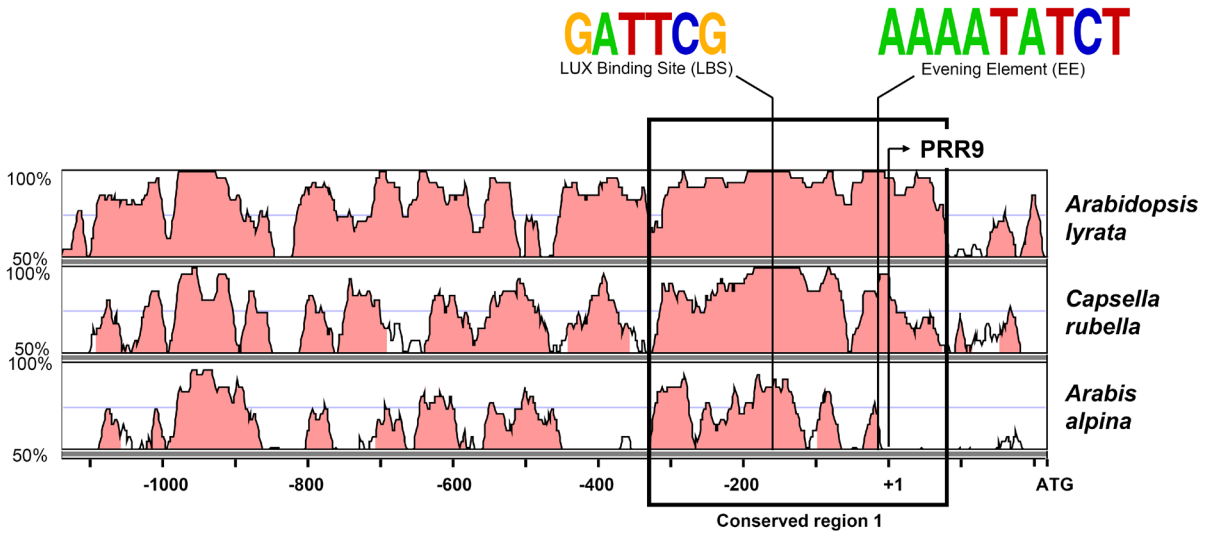


Fig. 2.7 Evolutionary conservation of the *PRR9* promoter

Pairwise alignments of the *A.thaliana PRR9* promoter to orthologous *Arabidopsis lyrata, Capsella rubella* and *Arabis alpina* promoters shown as VISTA plots. Red color indicates regions where a sliding window of 30 base pairs has at least 70% conservation. Conserved Region 1 is highlighted with a black frame. Vertical bars indicate the position of highly conserved LBS / EE and the transcriptional start site. Conservation of LBS and EE is shown as WEBLOGO.

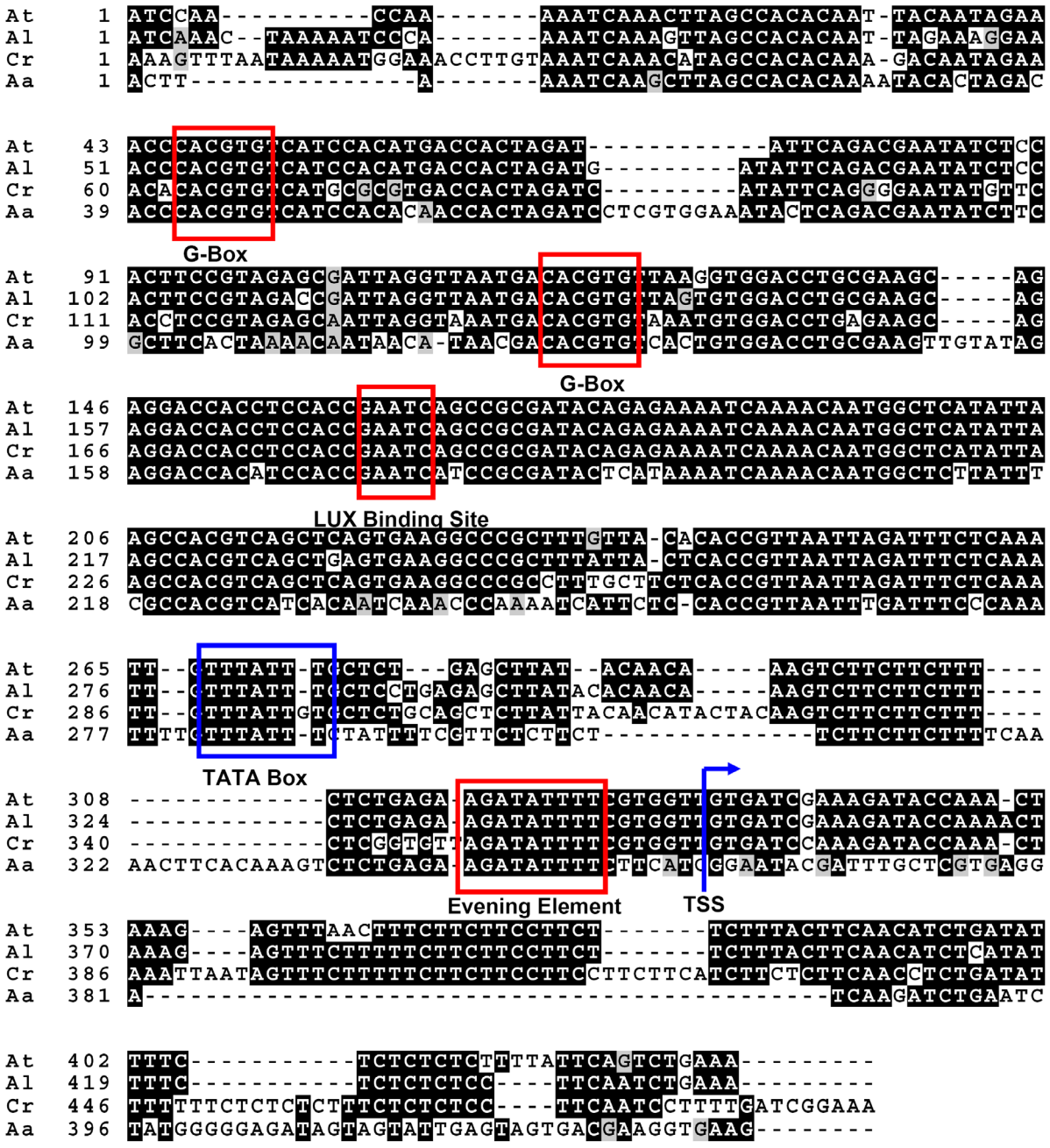


Fig. 2.8 Multiple sequence alignment of Conserved Region 1 of the *PRR9* promoter. Highly conserved *cis*-regulatory elements that are known to play a role in light- or clock-mediated transcriptional regulation are highlighted with red boxes. The predicted TATA Box and the transcriptional start site (based on TAIR annotation in Arabidopsis thaliana) are highlighted in blue. At=Arabidopsis thaliana, Al=*Arabidopsis lyrata*; Cr =*Capsella rubella*; Aa=*Arabis alpina* TSS=transcriptional start site

2.8 Discussion

Using a phylogenetic shadowing approach, I identified five highly conserved blocks within the orthologous *GI* promoters of eight Brassicaceae, including Block 2, a 700bp region with very high conservation.
Further inspection of Block 2 showed that it contains a number of highly conserved *cis*-regulatory elements proposed to play roles in light- and clock-mediated transcriptional regulation. Such blocks containing combinations of different *cis*-regulatory elements regulate complex transcriptional responses in a wide range of different organisms including yeast (Barbaric et al., 1999), Drosophila (Negre et al., 2011) and *Arabidopsis* (Mikkelsen and Thomashow, 2009), suggesting that such a combinatorial mechanism also mediates the precise light- and clock dependent regulation of *GI*.

Among the conserved elements were 3 EEs, *cis*-regulatory motifs that can be bound by Myb-like transcription factors such as LHY/CCA1 (Alabadi et al., 2001; Perales and Mas, 2007) or REVEILLE (REV) (Gong et al., 2008). Moreover, the peak in *GI* mRNA is shifted earlier to the morning in *lhy/cca1* double mutants (Mizoguchi et al., 2002a; Mizoguchi et al., 2005). Such results support mathematical models proposing that LHY and CCA1 repress *GI* transcription in the morning by binding to their EEs (Locke et al., 2005; Locke et al., 2006b).
Furthermore, EEs are overrepresented in evening-phased genes and confer evening expression in an artificial promoter system (Harmer et al., 2000b), thus making them ideal candidates for further analysis in biological assays of intact promoters. However, further interesting candidates for such an analysis are also the bZIP binding sites ABRELs as well as the bHLH binding HUD Boxes. These HUD Boxes are implied to play a role in phytochromeB-mediated light signalling, a process where it was shown that GI plays an important role (Huq et al., 2000b). Studying such elements *in situ* in full length promoters allows study of how they combine to confer complex diurnal patterns of regulation.

Phylogenetic analysis of the *PRR9* promoter revealed high conservation directly upstream of the transcriptional start site. Interestingly, as in the *GI* promoter, both a highly conserved EE and a LBS were found in the *PRR9* promoter. Recently, it was proposed that a complex comprising the clock proteins ELF3/ELF4/LUX directly binds to the promoters of *PRR9* and *GI* (Herrero, Kolmos, unpublished). The occurrence of the same highly conserved *cis*-regulatory elements within these two promoters provides further evidence for this hypothesis, although direct binding of ELF3 to the *GI* promoter could not be shown (Dixon et al., 2011) .

These examples also demonstrate the power of the presented shadowing method in detecting evolutionary conserved blocks and *cis*-elements. The sequential pipeline of first performing pairwise

alignments with CHAOS and Shuffle-LAGAN, then using the identified conserved blocks for multiple alignments using ClustalW or DIALIGN followed by detection of conserved *cis*-regulatory motifs by visual inspection based on a candidate list provided a set of functionally important motif candidates. Some of these were subsequently validated both for the *GI* promoter (chapter 4 and 5) and the *PRR9* promoter (Herrero, Kolmos, unpublished).

The power of related shadowing approaches was demonstrated previously (Lee et al., 2005; Adrian et al., 2010). These studies, as well as this study of the *GI* promoter, relied largely on PCR – based amplification and subsequent sequencing of orthologous promoters from related species, a process that is time-consuming and not always straight forward. Recent advances in sequencing whole genomes of various plant species will make it possible to directly carry out such a phylogenetic shadowing without any wet-lab step, as demonstrated here for the *PRR9* promoter in the Brassicaceae or the *GI* promoter of five grass species, respectively.

In particular this study highlights the usefulness of the *A. alpina* genome for comparative genomics with *A.thaliana*, especially on the level of transcriptional regulation. Given the evolutionary distance of approximately 30 million years between these two species (see chapter 1.5)(Wang et al., 2009), this provides an excellent evolutionary resolution for phylogenetic analyses. This was not just demonstrated for the promoters *GI* and *PRR9* as shown here, but also for the promoters of the flowering-time genes *CONSTANS* (Simon et al., unpublished), *FLOWERING LOCUS T* (Adrian et al., 2010) and *LEAFY* (Wagner et al., unpublished).

The shadowing pipeline developed here is well-suited to computer-based automation and is not restricted to particular plant species. Further development and automation could lead to powerful tools for detecting *cis*-regulatory elements in the future, not only for any plant promoter of interest, but also for analysing the transcriptional code of entire biological networks.

In summary, the bioinformatic analysis of *GI* promoters provided excellent candidates for testing their biological significance in experimental assays, both on the level of larger regions (conserved Block 2) as well as on the level of *cis*-regulatory elements (EE, ABRELs, HUD-Box, LBS).

Genome-wide bioinformatic analysis

3.1 Perspective

Phylogenetic shadowing of the *GI* promoter has revealed a number of evolutionary conserved *cis*-elements within one highly conserved block (chapter two), suggesting that the combinatorial interplay of those elements is crucial for the precise temporal regulation of *GI*. For a wide range of eukaryotic organisms including yeast (Lam et al., 2008) or Drosophila (Negre et al., 2011) it has been proposed that transcriptional regulation is mediated by enhancers containing multiple transcription factor binding sites and that these enhancers are bound by different combinations of transcription factors (Negre et al., 2011). Genome-wide analysis of promoter elements in co-regulated genes of *A.thaliana* is one approach to identify clusters of conserved elements and propose how such combinatorial regulation might work in plants.

3.2 Identification and analysis of promoters enriched for EEs

As the phylogenetic shadowing analysis (Chapter 2) suggested an important role for EEs within the *GI* promoter and this importance was confirmed by biological experiments (see chapter 4 and 5) I was interested in studying the significance of EEs and related *cis*-regulatory elements on a genome-wide level. First I asked the question how EEs are distributed within the genome of *A. thaliana*. All EEs and CBS in the *A. thaliana* genome were mapped and assigned a position relative to transcribed regions (calculations done by Karl Nordström). Not surprisingly, both the EE and the CBS showed high overrepresentation in intergenic regions, reflecting their role as *cis*-regulatory elements (Fig. 3.1). In coding regions, in contrast, the EE, and the CBS, appears at a frequency similar to that expected by chance. Interestingly, also 5' UTRs and introns are enriched in EEs and in the CBS (Fig.3.1).

Next, all EEs in intergenic regions in the *A.thaliana* genome were mapped and assigned to promoters of genes (lists generated by Karl Nordstöm). Promoters were defined as the 3kb upstream of the translational start site. This analysis showed that 15,7% of all promoters contain EEs, most of them a single one (Fig 3.2).

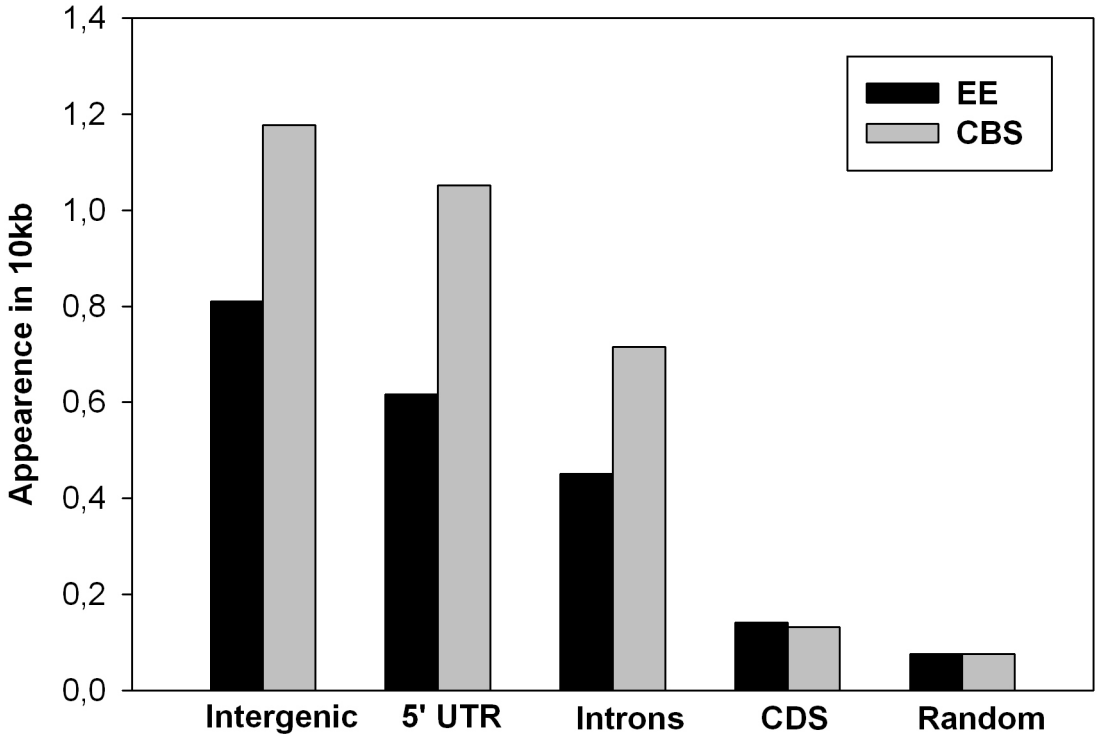


Fig. 3.1 Genomic Distribution of Evening Elements (EE) and CCA1-Binding-Sites (CBS) in *A.thaliana*. Random value describes the expected occurrence of a 9bp element if all bases would be distributed by chance.

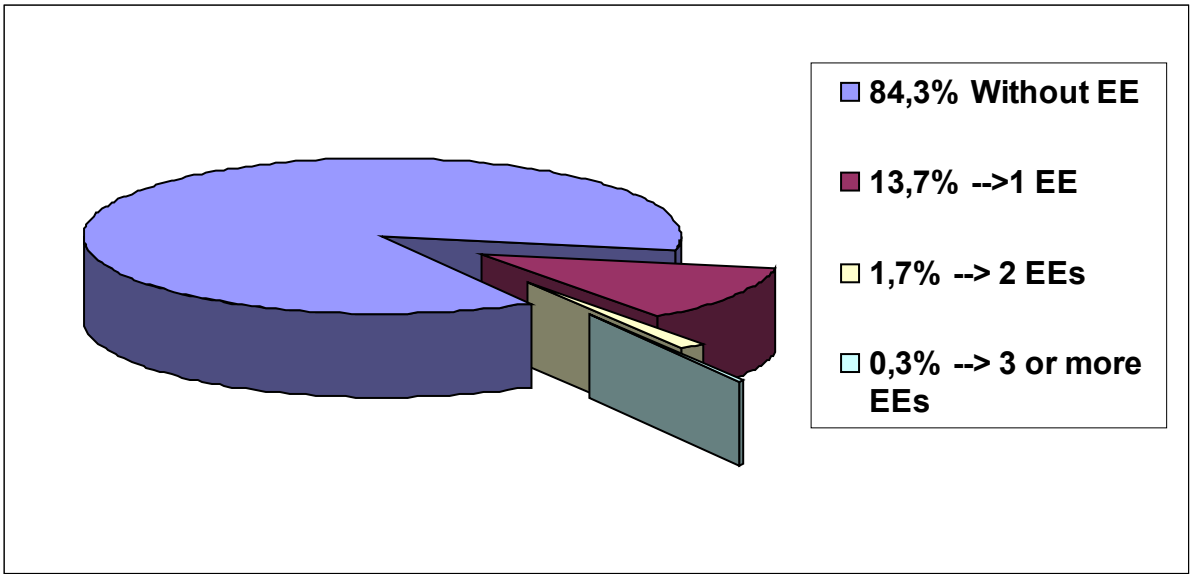


Fig. 3.2 Genome-wide distribution of the EE (AAAATATCT) within 3kb upstream region of the translational start site

ID	No of EEs	Name
AT1G10760	3	SEX1
AT1G22770	3	**GI (GIGANTEA)**
AT1G65360	3	AGL23
AT1G68050	3	**FKF1**
AT2G21660	4	**CCR2**
AT3G07650	3	COL9
AT4G25470	4	CBF2
AT4G25480	3	CBF3
AT4G25490	3	CBF1
AT5G24470	5	**PRR5**

Tab. 4 Well characterised genes with 3 or more EEs in their promoters. Circadian-clock-related genes are marked in bold

EE-enriched promoters containing at least 3 EEs in their promoters were focused on because these have a similar configuration as the *GI* promoter. Ninety nine promoters were originally identified to contain at least 3 EEs. Double counts and small annotated ORFs within promoters were removed, which resulted in a final list of 71 genes with at least 3 EEs in their promoters (see in Suppl. Material). Manual inspection of this list revealed a number of well-characterised clock-related genes, such as *GI, FKF1, PRR5* or *CCR2* (Tab.4). Next gene ontology term (GO term) analysis of the 71 EE-enriched promoters was performed using FatiGO from the Babylomix server (Al-Shahrour et al., 2004; Al-Shahrour et al., 2007). As expected, an over-representation of the GO term 'circadian rhythm' was discovered in the list of EE-enriched promoters, reflecting the previous findings of clock-related genes in that list. However, more surprisingly, an over-representation of cold stress related GO terms was also found, with higher statistical significance than for the circadian rhythm related GO-term. Further GO term analysis with a number of different tools revealed the same GO terms being over-represented (data not shown), supporting the results obtained with FatiGO. Genes whose promoters contain only 1 or 2 EEs did not show overrepresentation for cold- and clock-related GO terms (data not shown).

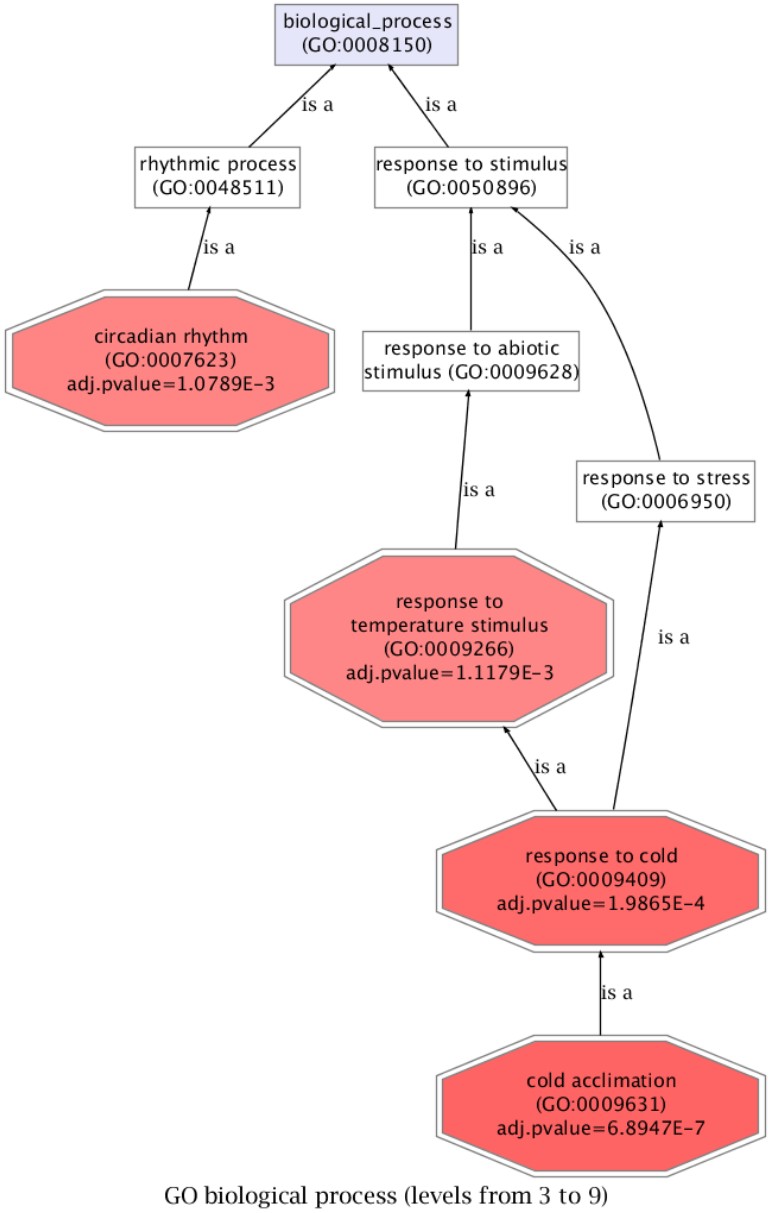


Fig. 3.3 GO-term analysis of genes with 71 EE-enriched promoters using FatiGO. p-value for statistical significance of over-representation (highlighted with red colour): 0,005.

The list with EE enriched promoters contains a number of cold-related genes, including the well characterised transcription factors CBF1, 2 and 3 (Tab 3.1). This suggests that EEs do not only have a prominent role in the regulation of clock related processes, but also in the regulation of cold-stress.

Some members of the list of 71 genes are well characterised regarding their role in clock- and / or cold-stress-related networks. However, for most of the 71 genes no function has been proposed for one of these processes. To further investigate the roles of these genes, the complete list of EE-enriched genes was compared to publicly available microarray data. This analysis revealed that almost all genes with at least 3 EEs in their promoters have a circadian expression pattern and that the peak-time almost exclusively occurs in the late afternoon or evening (Fig. 3.4).

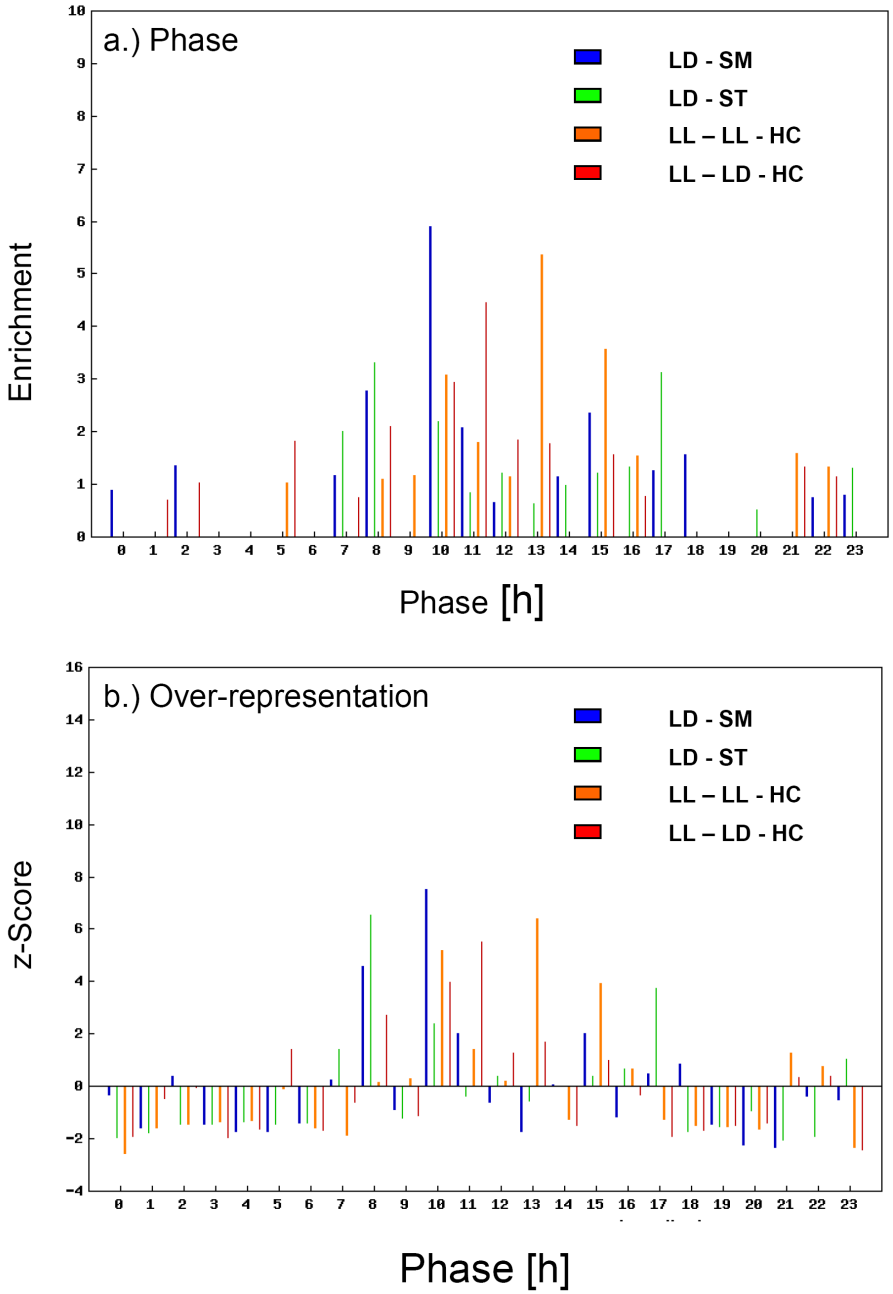


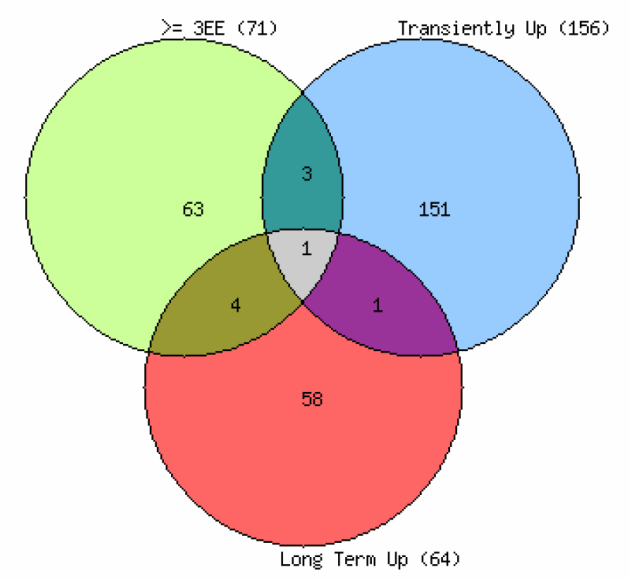
Fig. 3.4 Phase analysis of EE-enriched promoters. 71 genes whose promoters contain at least 3 EEs were subjected to phase analysis using Phaser (Mockler et al., 2007)
a.) Phase of the 71 genes calculated from 4 different microarray datasets
b.) Statistical analysis of the same data. Overrepresentation at distinct time points is indicated by positive values, under-representation by negative values.

LD-SM (12h Long Day, Smith dataset (Smith et al, 2004)); LD-ST (12h Long Day, Stitt dataset (Blasing et al., 2005) LL-LL-HC (continuous light, Chory dataset1) LL-LD-HC (continuous light, Chory dataset 2)

Comparison of EE enriched promoters with microarray data of cold-induced genes (Fowler, Thomasshaw) revealed an overlap of 8 genes (Fig. 3.5). Strikingly, these genes also include *GI*, suggesting that *GI* might be upregulated in an EE-dependent manner both by clock and temperature. Moreover, all 3 *CBFs* as well as *RESPONSIVE TO HIGH LIGHT (RHL41), CHLOROPLAST BETA AMYLASE (CT-BMY)* and two unknown genes overlap between EE-enriched promoters and the cold-induced gene lists. To further investigate the cold-induced expression pattern of EE-enriched promoters, the list of EE-containing genes was also compared to a second series of microarray experiments (Kilian et al., 2007). In this study different durations of cold treatment were applied, ranging from 30min to 24h. The 10 selected genes from Table 4 were investigated (CBF1 as representative for the 3 CBFs) and the cold-dependent expression pattern was visualised using the Bar EFP Browser (Winter et al., 2007). Strikingly, all ten genes were up-regulated in response to low temperatures, most of them after a continuous cold treatment of at least 6h (Fig. 3.6).

Taken together, this analysis suggests a link between clock and cold-stress regulation for genes containing multiple EE in their promoters. This link would exist not only for the well described genes on the list, but also for some for which a clock- or cold-dependent function has not been described as well as for a number of unknown genes.

a.)



b.)

ID	Description
Long-time up-regulated	
AT1G22770	GI (GIGANTEA)
AT4G25470	CBF2 (C-REPEAT/DRE BINDING FACTOR 2)
AT4G33980	unknown protein
AT5G59820	RHL41 (RESPONSIVE TO HIGH LIGHT 41)
Transiently up-regulated	
AT3G15450	unknown protein
AT4G25480	DREB1A (DEHYDRATION RESPONSE ELEMENT B1A); CBF3
AT4G25490	CBF1 (C-REPEAT/DRE BINDING FACTOR 1)
Both long-time and transiently up-regulated	
AT4G17090	CT-BMY (CHLOROPLAST BETA-AMYLASE); beta-amylase

Fig. 3.5 Cold induction of genes with EE-enriched promoters1
a.) List with 71 genes having EE-enriched promoters was compared to public available microarray data of transiently and long-time upregulated genes upon cold exposure; (Fowler et al., 1999) Venn Diagramm illustrated overlaps between these 3 lists
b.) Genes that overlap between lists as described in a.)

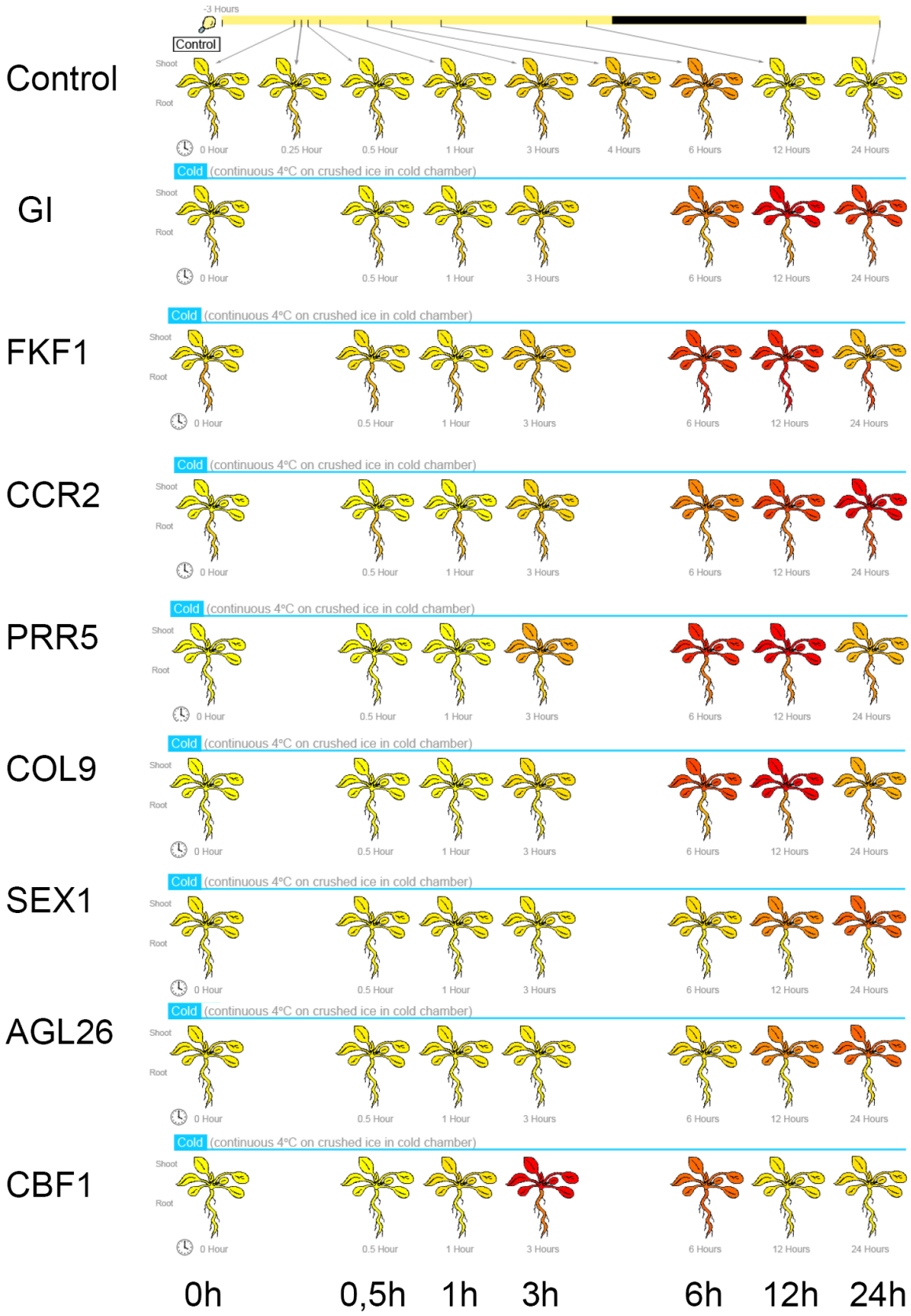


Fig. 3.6 Expression pattern after a cold treatment of the 10 genes with EE-enriched promoters from Tab. 3.1 . Pictures show transcript accumulation after cold exposure for 0,5h, 1h, 3h, 6h, 12h or 24h. Control is untreated *GI* sample. CBF1 was taken as representative for all three CBFs, who have a similar expression pattern.

3.3 Conservation of EEs in different promoters

As shown in Chapter 2, phylogenetic shadowing is a powerful tool to reveal functionally important promoter regions and *cis*-regulatory elements. Moreover, *A.alpina* was shown to be on an appropriate evolutionary distance from *A.thaliana* to allow efficient comparisons. Therefore the EE-enriched promoters of *FKF1*and *CCR2* were also subjected to a phylogenetic shadowing analysis and the conservation, position and orientation of all EEs were mapped and compared to the situation in the *GI* promoter. This analysis revealed that all EEs in the three investigated promoters were located within highly conserved regions (Fig 3.7). Moreover, all EEs found in the respective *A. thaliana* promoters were absolutely conserved in the *A. alpina* promoters. This observation suggests that the EEs in the promoters of *FKF1* and *CCR2* are likely to be functional. Furthermore, the distance and orientation of all EEs were also highly conserved between *A.thaliana* and *A.alpina* (Tab 3.2). Together with the genome-wide analysis of promoter elements (see Fig. 3.8 and 3.9) this highlights the importance of distance and orientation of *cis*-regulatory elements that are clustered within modules. Interestingly, 2 EEs in the promoters of *GI* and *FKF1* have exactly the same spacing (i.e. 177bp) in *A.thaliana*, suggesting that the same transcription factor complex might bind these two promoters to regulate their evening expression pattern.

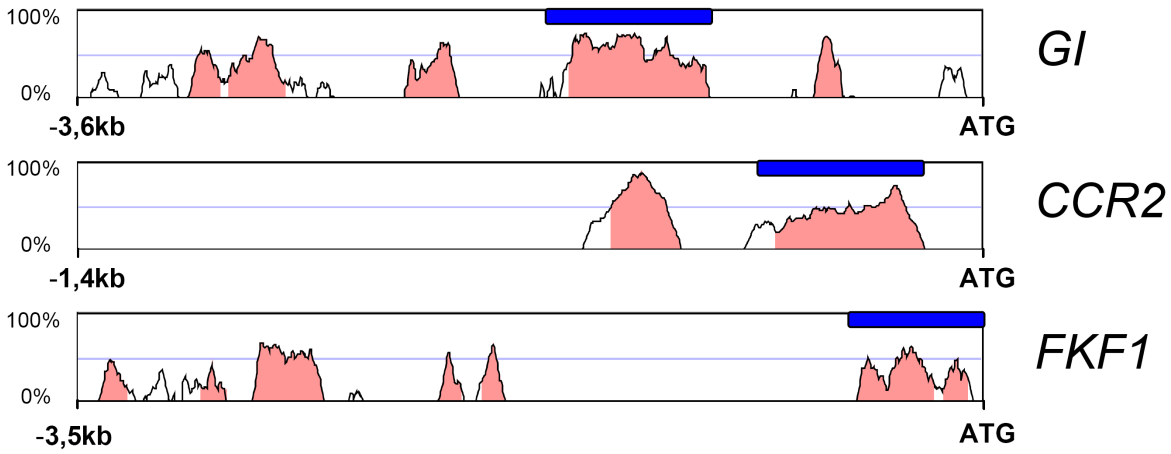


Fig. 3.7 Conservation of EE-enriched promoters
***A.thaliana* promoters of *GI, FKF1* and *CCR2* were compared to their respective *A.alpina* orthologues and visualised using VISTA Browser (Brudno). Sliding window: 100bp, Conservation: 70%; Base genome: *A.thaliana*; highly conserved regions containing EEs are highlighted with blue bars.**

		EE1	Distance [bp]	EE2	Distance [bp]	EE3	Distance [bp]	EE4	EEs in conserved Block
GI	**At**	For	23	Rev	177	For			yes
	Aa	For	23	Rev	188	For			yes
FKF1	**At**	For	178	Rev	177	Rev			yes
	Aa	For	151	Rev	182	Rev			yes
CCR2	**At**	Rev	12	Rev	15	For	136	For	yes
	Aa	Rev	13	Rev	23	For	130	For	yes

Tab. 6 Location and orientation of conserved EE in the promoters of *GI, FKF1* and *CCR2;* For: Orientation in 5'→3' direction Rev: orientation in 3' →5' direction; distance between the respective EE is given in bp; At = *A. thaliana*; *Aa = A. alpina*.

3.4 Genome-wide interactions between cis-regulatory elements

Both the phylogenetic shadowing of the *GI*-promoter as well as the analysis of genes with EE-enriched promoters suggests that clustering is a key feature of the EE. This leads to the question of how these multiple EEs might interact with each other.

To answer these questions all EEs in intergenic regions were further investigated regarding their location, orientation and the position with respect to neighbouring elements (calculations performed by Karl Nordström, unpublished). This analysis was performed separately for the EE (AAAATATCT), the CBS (AAAAAATCT), and the SEE (AATATCT) as well as for all three elements together. As the results for the single elements were comparable, data in Fig 3.8 is shown for all 3 elements considered together and for the rest of this chapter these three elements together will all be considered as EEs (data with separate analysis in Supplements).

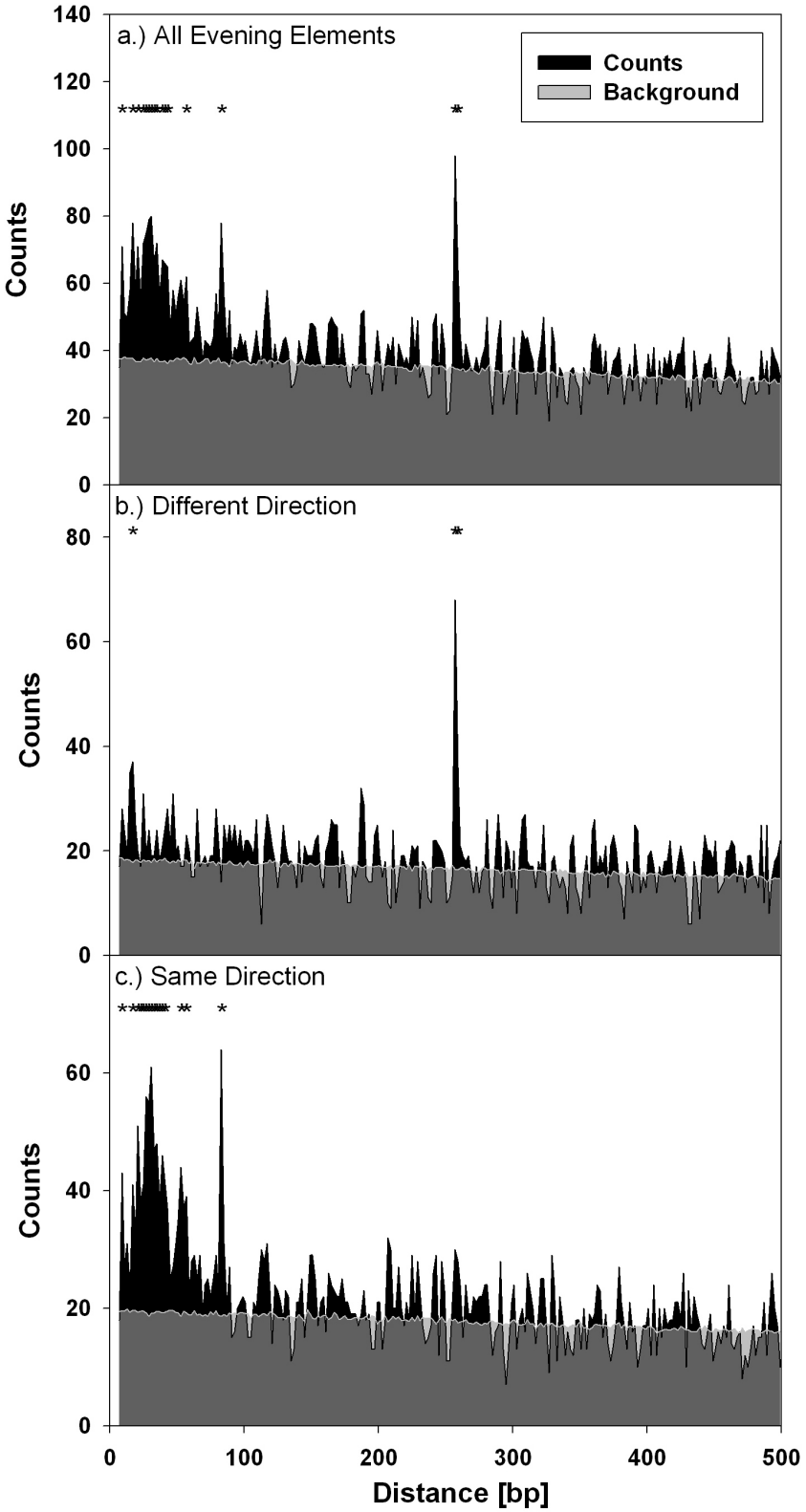


Fig. 3.8 Clustering of EEs within promoters

All EEs (AAAATATCT), SEE (AATATCT) and CBS (AAAAAATCT) were mapped in relation to neighbouring EEs. a.) Orientation of EE not considered; b.) Different orientation of neighbouring EEs; c.) Same orientation of neighbouring EEs; Asterisks indicate statistical significance p<0,05

This data clearly shows that EEs are clustered within the genome and that their preferential distance lies between 20 and 40bp from each other (Fig 3.8a). A second sharp peak appears at a distance of 260bp, indicating that different distances between EEs might play a role in EE-mediated regulation. Next the significance of EE orientation was investigated. This analysis showed that the 2 groups of over-representation were separable depending on the orientation of the EEs. Over-representation at a distance of about 260bp mainly occurs if the two elements are located in the opposite direction (Fig. 3.8b). In stark contrast, EEs that are clustered within a distance of 20 to 100 bp mainly show an over-representation if located on the same strand (Fig. 3.8c). Moreover, these closely spaced EEs can be divided into two classes: the first class appeared almost normally distributed around a distance of 30bp, whereas the second class is represented by a sharp peak around 90bp. These data suggest that orientation and spacing of EEs combine to confer important functional information on promoters.

The *GI* promoter does not only comprise 3 highly conserved EEs, but also a number of other conserved *cis*-regulatory elements such as ABRELs, HUD Boxes and a LBX (see chapter 2). This suggests that combinations of different transcription factors regulate the precise temporal expression pattern of *GI*. However, all these above mentioned elements are variants of elements that can be bound by classes of well-characterised transcription factors. The ABREL can be bound by bZIP transcription factors, the HUD Box is a variant of the core binding site for bHLH transcription factors and the LBX can be bound by GARP-type transcription factors such as LUX. Therefore, *cis*-elements were divided into 4 major classes according to their binding capacity for different transcription factors, i.e. Myb-binding (containing EE, CBS and SEE), bHLH-binding, bZIP binding and GARP-binding as indicated in table 3.3.

Class	Sequence	Name
Myb	AAAATATCT	Evening Element (EE)
	AAAAAATCT	CCA1 Binding Site (CBS)
	AATATCT	Short Evenin Element (SEE)
bZIP	GACGTC	C-Box (CBX)
	ACGTG	ABA Response Element-like (ABREL)
bHLH	CACATG	HUD-Box (HUD)
	CACGTG	G-Box (GBX)
GARP	GATTCG	LUX Binding Site (LBS)

Tab. 7 Different classes of transcription factor binding sites as defined for genome-wide analysis undertaken in Fig 3.9.

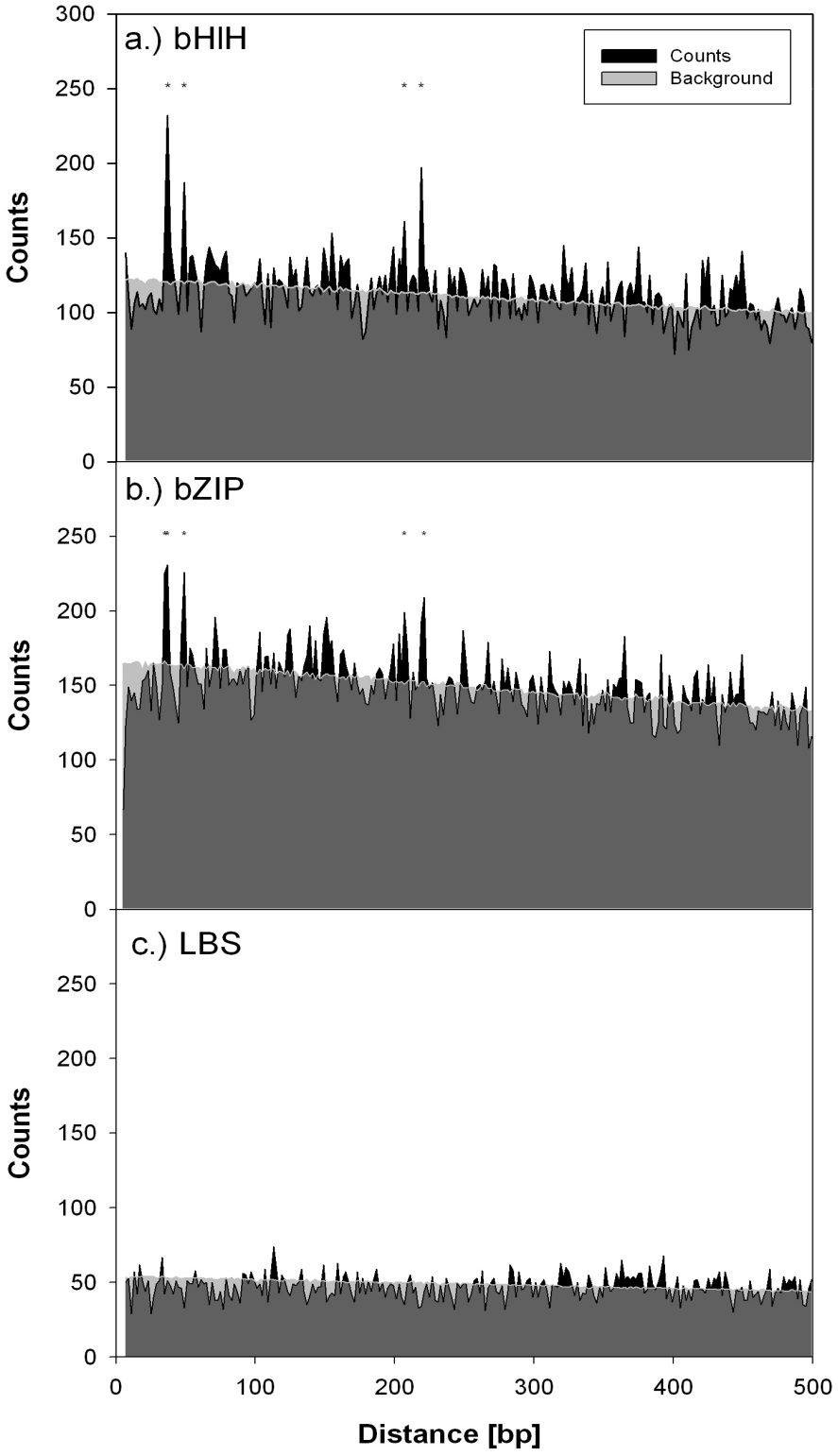


Fig. 3.9 Relationship of EEs to other *cis*-regulatory elements. All EEs, SEEs and CBS were clustered in relation to neighbouring *cis*-regulatory elements
a.) Relationship of EEs to the bHLH binding cis elements G-Box (CACGTG) and HUD-Box (CACATG)
b.) Relationship of EEs to the bZIP binding cis elements C-Box GACGTC) and ABREL (ACGTG)
c.) Relationship of EEs to the LUX Binding Site (LBX, GATTCG)Asterisks indicate statistical significance p<0,01

In order to analyse the relationship of EEs to the above mentioned binding sites, distances between all EEs and surrounding elements capable of binding bZIP or bHLH transcription factors or the GARP-transcription factor LUX were calculated. As the sequences of bZIP and bHLH binding sites are fully or partially palindromic, no differences of orientation of these elements could be calculated.

Comparing the positions of LUX binding sites to the location of EEs revealed no statistical significant over-representations (Fig. 3.9c). However, the analysis of the other combinations revealed some striking correlations. Both bZIP and bHLH binding *cis*-elements were over-represented at sharply defined distances from the EEs, (Fig. 3.9a and Fig. 3.9b). For both classes of binding sites, equal distances of over-representation were discovered at a distance of 37, 49, 207 and 219bp from an EE (Tab. 3.4), possibly due to the high similarity between bHLH and bZIP binding sites (Tab.3.3). Interestingly, the distance between the two more proximal peaks was the same as for the two more distal peaks: 12bp (Tab. 3.4). The significance of this pattern remains to be uncovered; however it suggests a combinatorial function of bHLH or bZIP binding sites and EEs.

	Peak1 [bp]	**Peak2 [bp]**	**Peak3 [bp]**	**Peak4 [bp]**	**Distance 1/2 [bp]**	**Distance 3/4 [bp]**
bHLH	37	49	207	219	12	12
bZIP	37/39	49	207	219	10/12	12

Tab. 8 Overrepresented distances of bHlH and bZIP binding sites in relation to EEs. Only statistical significant hits (p<0,01) from Fig. 3.9 are shown.

3.5 Discussion

Genome-wide investigation of EEs revealed that they occur in a large number of plant promoters. However, a relatively small fraction of genes, including *GI*, has 3 or more EEs in their promoters. Further analysis of these 71 promoters showed that the corresponding genes are regulated by the circadian clock and that the large majority show a peak in expression in the afternoon or in the evening. This is in agreement with how the EE was originally discovered by studying promoters of genes with a peak in expression in the evening to identify enriched motifs (Harmer et al., 2000b). However, unexpectedly, GO term analysis also revealed a high over-representation of cold-stress related terms. It has been shown recently, that EEs in concert with the bZIP binding ABRELs can regulate the cold-dependent up-regulation of *COL1* and *COR27* (Mikkelsen and Thomashow, 2009). This suggests that EEs are not only functional in clock-related processes as shown before, but also in the regulation of cold stress.

Comparison of genes with multiple EEs and genes that are up-regulated in the cold revealed that *GI* transcript was found to be increased in abundance after exposure to cold in *Arabidopsis* (Fowler and Thomashow, 2002; Cao et al., 2005) and other species (Paltiel et al., 2006; Hong et al., 2010). Moreover, *gi* mutants are less sensitive to freezing temperatures compared to wildtype plants (Cao et al., 2005), suggesting that *GI* indeed might play a role in cold-stress related processes.
The *GI* promoter does not only comprise 3 highly conserved EEs within its promoter, but also 3 highly conserved ABRELs (chapter 2), suggesting that these elements might act combinatorially to upregulate *GI* transcription during cold exposure.

So far this analysis highlighted the importance of multiple EEs within promoters. However, must EEs be present in multiple copies to be functional? Probably not, as a single EE in the promoter of *TOC1* plays a crucial role within the central loop of the circadian clock (Alabadi et al., 2001; Perales and Mas, 2007). One possibility is that multiple EEs simply provide redundancy within promoters. This is highlighted by recent findings that show loss of EE-mediated transcription only if multiple EEs are mutagenised (Mikkelsen and Thomashow, 2009). However, the massive over-representation of specific distances between EEs in my genome-wide analysis suggests that this pattern could have an important biological function. This is underlined by the finding that orientation and spacing of the EEs is highly conserved – at least in the promoters of *GI, FKF1* and *CCR2*. Therefore it seems more likely that multiple EEs provide promoters with additional regulatory functions, perhaps including cold regulation.

The genome-wide analysis of location and spacing of different *cis*-regulatory elements revealed relationships between EEs and bZIP and bHLH binding sites. Most strikingly, the overrepresented distances between EEs and bZIP or bHLH binding elements were the same. One explanation for this might simply be that bZIP and bHLH transcription factors bind to similar sequences and thus the same sequences are responsible for these peaks. This could be tested by biological experiments such as ChIP-seq analysis with different bZIP and bHLH transcription factors.
Another outcome of this analysis is that the distance between peaks of bZIP or bHLH binding sites was 12bp, both proximal and distal to the EEs. A possibility is that bZIP and bHLH binding sites at a distance of 37 and 49bp from the EEs form one regulatory unit and that the the two more distant peaks constitute the same pattern one ncleosome distant from the first peaks. Considering the average length of nucleosomes (178bp,(Alberts et al., 2007)) this would fit well to the distance of 170bp observed in this study (207 – 37 or 219 - 49 =170).

This analysis provided important insights into possible combinatorial action of different *cis*-regulatory elements and is one step in revealing the transcriptional code in clock and cold-regulated genes of

A.thaliana. However, many interesting questions remain to be resolved. Which genes have specific patterns of over-representation or interactions between *cis*-regulatory elements within their promoters? Apart from bZIP and bHLH binding sites do other *cis*-regulatory elements cluster or interact in a similar way as EEs? Could such an analysis discover previously undescribed *cis*-regulatory elements? Future analysis is required to answer these and related questions.

In summary, these data suggest that EEs are functional in multiple promoters and that they regulate both circadian clock and cold-stress related processes. This study provides evidence that such action requires multiple EEs as well as the combinatorial interaction of EEs with different *cis*-regulatory elements.

Analysis of luciferase reporter lines

4.1 Perspective

Phylogenetic shadowing is a powerful tool to predict functionally important *cis*-regulatory modules and elements. However, to verify the significance of such conserved *cis*-regulatory modules, biological experiments are necessary. One common way to do so is the use of promoter fragments fused to a reporter gene. Here I make use of the luciferase reporter system in order to study the significance of conserved modules and putative important *cis*-regulatory elements within the *GI* promoter of *A. thaliana* that were predicted by the phylogenetic analysis in chapter two. This will allow the *in planta* functional analysis of these elements and thereby reveal their contribution to different transcriptional features of *GI* such as the response to different photoperiods or the gated effect of a light pulse.

4.2 The GI promoters of A. thaliana and A. alpina confer similar expression patterns

One important assumption in doing phylogenetic shadowing approaches is that the orthologous sequences that are compared have the same function in the different organisms. To test this for the present study, a 5kb *GI* promoter fragment from the *A. alpina GI* promoter (that covers the entire intergenomic region between *GI* and its upstream gene *PAB3*) was cloned, fused to the luciferase open reading frame and stably introduced into *A. thaliana* by *Agrobacterium*-mediated transformation (done by Ingmar Bürstel and Frédéric Cremer, unpublished). The pattern of luciferase activity in these plants was compared to plants carrying a 2,5kb promoter fragment of the *A. thaliana GI* promoter fused to the luciferase open reading frame (done by Hailong An).

Three independent homozygous lines of the *AaGI::Luc* plants were selected and compared to the reference line of *AtGI::Luc*. This analysis revealed strong expression conferred by the *AaGI::Luc* transgene, at a comparable level to the *AtGI::Luc* reference line, demonstrating that both constructs confer a similar expression level in *A. thaliana* (Fig. 4.1a). Subsequently, the diurnal expression pattern under LD 16 conditions was determined. This analysis revealed the same diurnal expression of the three *AaGI::Luc* constructs compared to *AtGI::Luc* (Fig. 4.1b). Both constructs confer a strong peak in expression in the late afternoon and a small peak at the onset of light in the morning. This demonstrates that the *GI* promoters of *A. thaliana* and *A. alpina* confer the same expression pattern

and therefore conserved regions between these two sequences are likely to be important for generating *GI* evening expression in these two species.

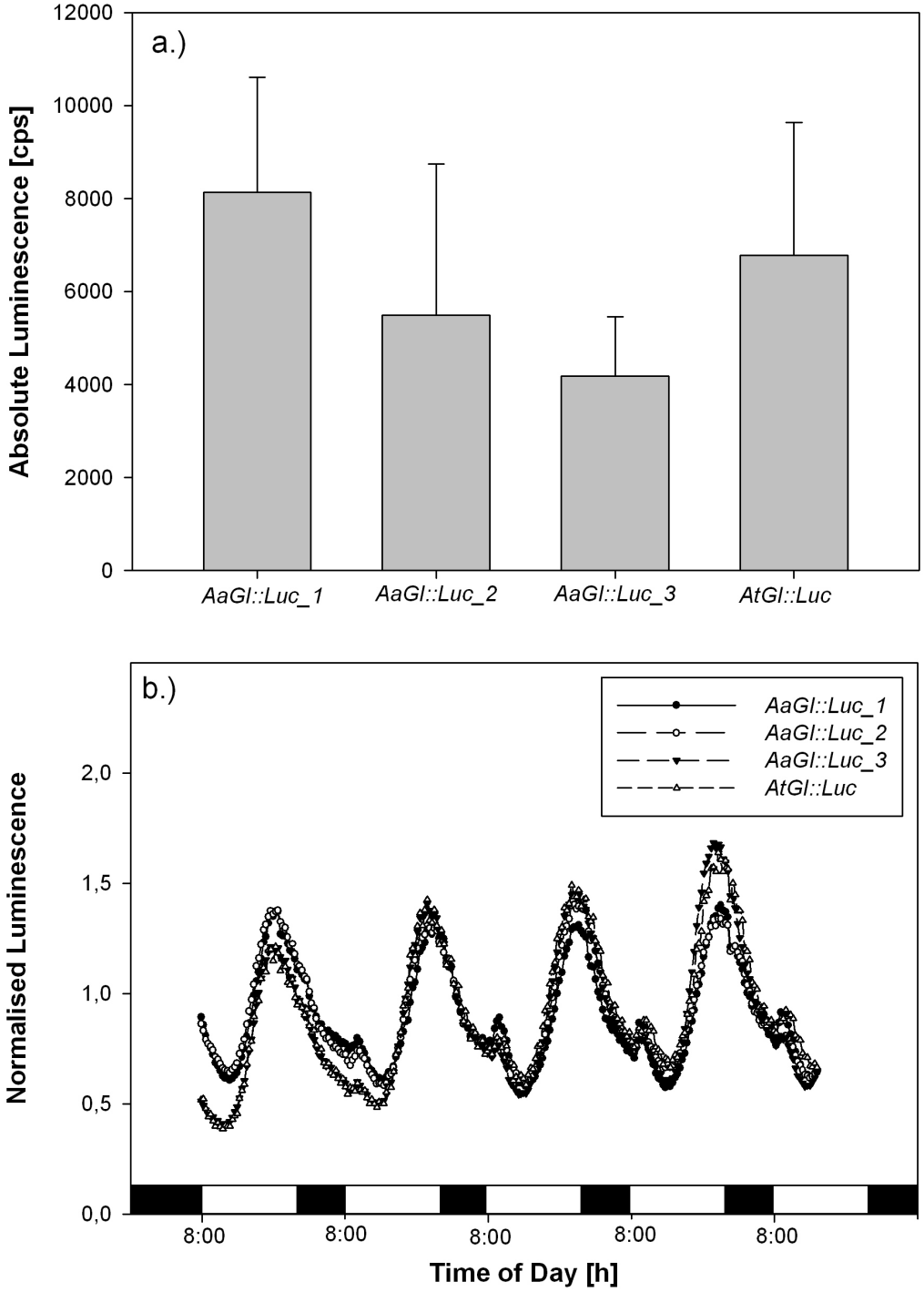


Fig. 4.1 The *GI* promoters of *A. thaliana* and *A. alpina* confer a comparable expression pattern
a.) Average luciferase expression per seedling of three independent lines with the *AaGI::Luc* construct compared to the reference line of *AtGI::Luc* over a four day period; Errorbars = Standard Deviation
b.) Diurnal expression pattern of three independent lines with the *AaGI::Luc* construct compared to the reference line of *AtGI::Luc*

4.3 Conserved Block 2 can drive a GI-like expression pattern

The phylogenetic shadowing analysis revealed 5 conserved blocks within the *GI* promoters of 8 Brassicaceae, 3 of them located within the 2,5kb *GI* promoter fragment used in this study (chapter 2).
In order to test the contribution of the three conserved blocks within the 2,5 kb promoter fragment of *GI*, conserved Blocks 1, 2 and 3 as well as the full 2,5kb *GI* promoter fragment were fused to a 105bp fragment carrying the *NOPALINE SYNTHASE (NOS)* minimal promoter and inserted upstream of the luciferase open reading frame (Puente et al., 1996). The Agrobacterium *NOS* promoter is one of the best described promoters that confers expression in plants and this 105bp minimal promoter has been used in previous circadian-clock associated studies in plants (Puente et al., 1996; Harmer and Kay, 2005). These four constructs were stably introduced into *A. thaliana*, homozygous lines were established and luciferase expression of three independent lines derived from each construct was analysed.

The 2,5kb *GI* promoter fragment in combination with the *pnos* minimal promoter drives expression in a similar pattern to *GI::Luc* without the minimal promoter, but with slightly reduced amplitude (not shown). Subsequently the expression pattern of this construct was compared to each of three independent homozygous lines of the *B1-pnos::Luc*, *B2-pnos::Luc* and *B3-pnos::Luc* construct. B1 and B3 conferred extremely low expression and none of the seedlings in those lines displayed rhythmic expression, demonstrating that these constructs are not capable of driving rhythmic expression on their own.
In contrast, all three B2 constructs conferred robust expression almost at the level of the *GI-pnos::Luc* control. In order to further examine the expression pattern of *B2-pnos::Luc*, plants were grown under diurnal conditions (LD 16) and then released to constant light. Under LD conditions, all three independent lines gave a peak in expression in the evening that was comparable to that of the *GI-pnos::Luc* control (Fig. 4.2b). Moreover, as for the 2,5kb promoter construct all lines displayed a rapid response to light in the morning, which was stronger in Line 1 compared to the other two *B2-pnos::Luc* lines and the *GI-pnos::Luc* control. Under LL conditions, the rhythmic expression in the evening continued in all three lines, indicating that this rhythmic expression is controlled by the plant circadian clock. In contrast, the morning peak disappeared in LL, confirming that it is a direct response to the dark/light transition. Overall the expression pattern conferred by the B2-*pnos::Luc* construct is similar to the expression pattern of the full length promoter, suggesting that Block 2 is important for *GI* regulation.
Collectively this data shows that conserved Block 2 – in combination with a *pnos* minimal promoter – can confer a robust *GI*-like expression pattern.

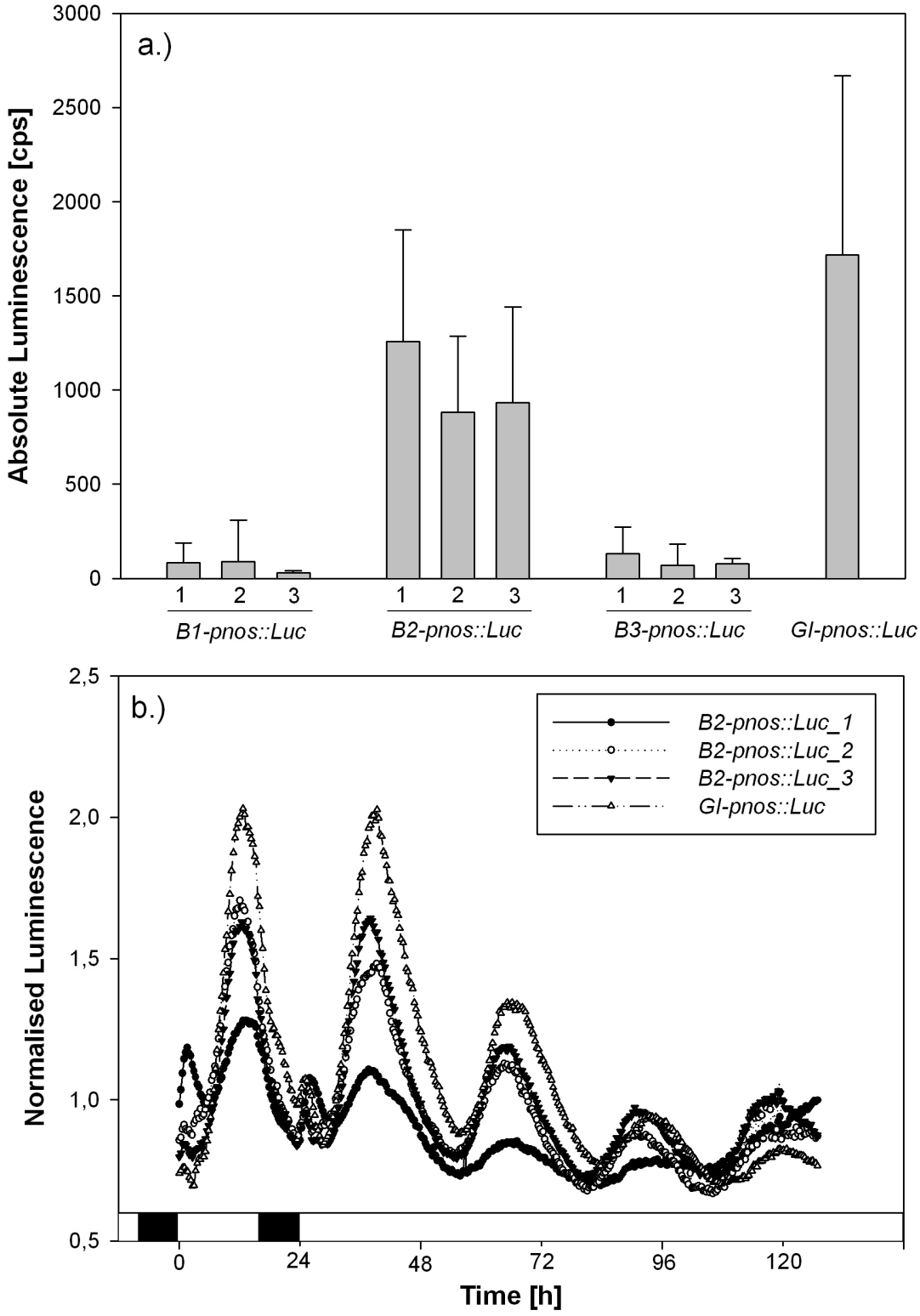


Fig. 4.2 Conserved Block 2 can drive a *GI*-like expression pattern

a.) Average luciferase expression per seedling of each three independent lines with the *B1-pnos::Luc, B2-pnos::Luc* and *B3-pnos::Luc* construct compared to one line of *AtGI-pnos::Luc* over a five days; Errorbars = Standard Deviation

b.) Expression pattern of three independent lines of the *B2-pnos::Luc* construct compared to *GI-pnos::Luc*. Plants were grown under long days (LD 16) and then released to continuous light conditions

4.4 Strategy for mutagenesis of conserved elements in the GI promoter

The fusions to the pnos minimal promoter demonstrated the importance of conserved Block 2. Moreover, the previous phylogenetic analysis revealed a number of conserved *cis*-regulatory elements within Block 2 that are implicated in light and clock-mediated transcriptional responses. Among these were three EEs, *cis*-regulatory elements that are over-represented in the promoters of evening expressed genes and can drive evening expression in artificial promoter systems (Harmer et al., 2000b).

To test the significance of these conserved EEs, all three were mutagenised by site-directed mutagenesis. In order to investigate the role of the EEs specifically within Block 2, the three EEs were also mutagenised only in the context of Block 2 and fused to the previously described 105bp fragment of the *pnos* minimal promoter. Both mutagenised fragments were cloned upstream of the luciferase open reading frame and stably introduced into *A. thaliana*. At least two homozygous lines were established for each line and characterised for aspects of expression characteristic of the *GI* promoter.

ID	Background	Description
GI::Luc	Ler	2,5kb *GI* promoter fused to luciferase
GI-pnos::Luc	Ler	2,5kb *GI* promoter +105bp nos minimal promoter, fused to luciferase
TEE::Luc	Ler	2,5kb *GI* promoter, fused to luciferase, 3 EEs mutated
B2-pnos::Luc	Ler	conserved Block 2 (700bp) +105bp nos minimal promoter, fused to luciferase
B2TEE-pnos::Luc	Ler	conserved Block 2 (700bp) +105bp nos minimal promoter, fused to luciferase, 3 EEs mutated

Tab. 9 Nomenclature and overview of some transgenic constructs frequently used in chapter 4

4.5 Mutating EEs results in altered GI expression

Different studies suggest that the precise timing of *GI* expression in the evening is crucial for its function within the photoperiodic flowering time pathway. In order to analyse the possible contribution of EEs and Block 2 to this evening specific expression pattern, 2 homozygous lines for each of the *TEE::Luc*, *B2-pnos::Luc* and *B2TEE-pnos::Luc* constructs were grown under LD 16 conditions with cool white light of approximately 70µE and luciferase expression was compared to the reference *GI::Luc.*

The *GI* promoter displays a small peak in the morning with a subsequent suppression of expression during the day and then reaches its peak in expression in the early evening around ZT 12 under LD 16 conditions. The *TEE::Luc* constructs, in contrast, behave differently. Upon the onset of light in the morning the expression level constantly rises until it reaches a plateau-like peak in the early afternoon (Fig. 4.3a). Repression in the morning as observed for the unmutated *GI* promoter does not take place. Apart from these differences in waveform, the peak time itself is approximately 2h earlier under long day conditions in the TEE::Luc lines.

As described earlier in this chapter, the *B2-pnos::Luc* construct confers similar expression characteristics as the full length *GI* promoter. However, the amplitude was reduced and the repression in the morning was weaker in these lines compared to the *GI::Luc* control (Fig 4.3b). The strongest phenotype was displayed by the *B2TEE-pnos::Luc* construct witch showed a severely disrupted rhythmicity. In the morning these lines display an immediate very low amplitude upregulation in Luc expression at the onset of light (Fig 4.3c). However, after approximately 2h, this increase in *GI::Luc* expression reaches a plateau during the rest of the day.

In summary, the three different types of *GI* promoter constructs all give a specific expression pattern under long day conditions. Mutation of three EEs in the full-length promoter leads to an earlier phase and a broader peak, Block 2 confers *GI*-like expression with lower amplitude and mutating the three EEs within Block 2 largely abolishes rhythmicity.

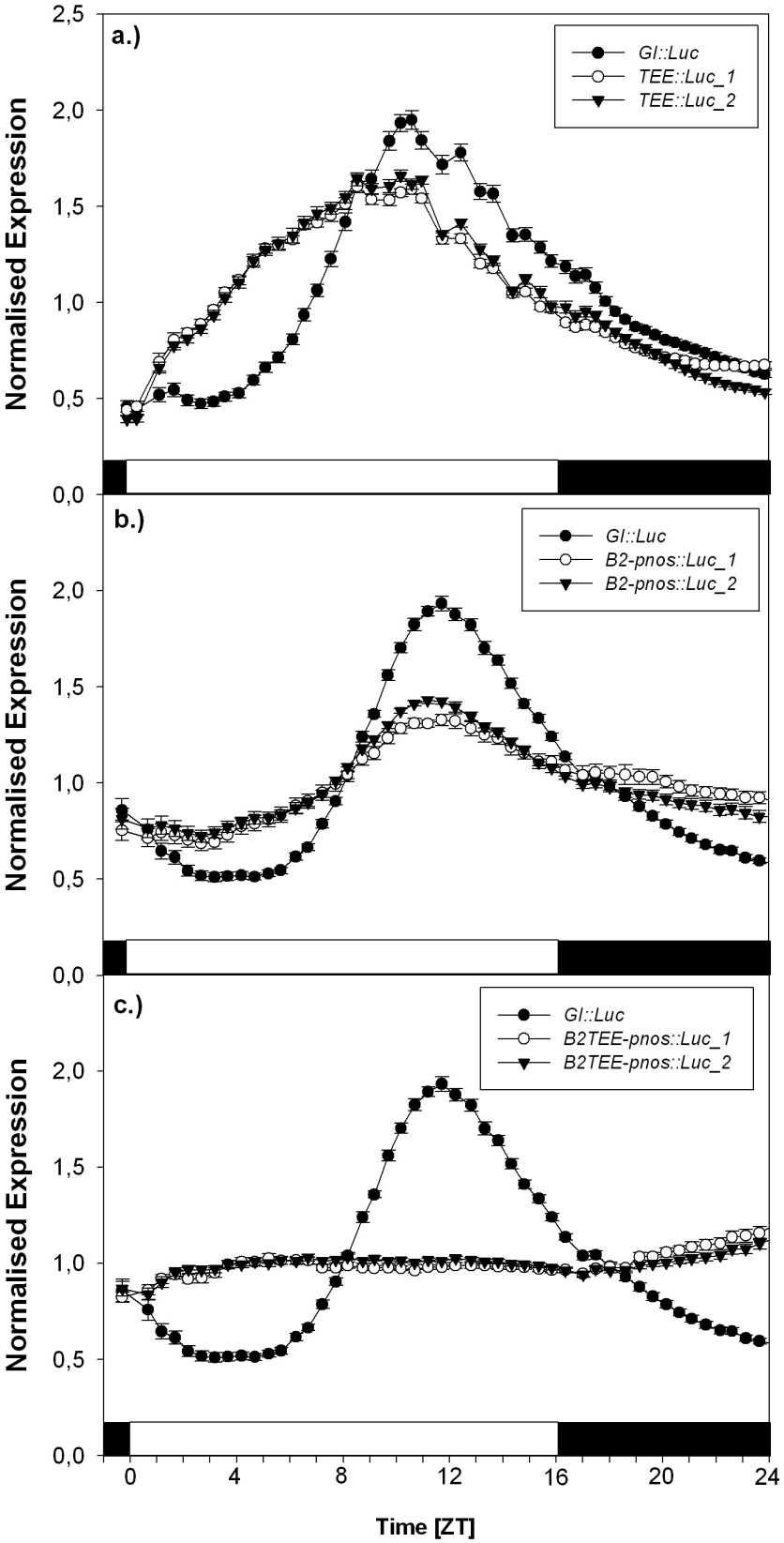


Fig. 4.3 Different *GI* promoter constructs confer different patterns of *Luc* expression
Two independent lines each of the *TEE::Luc*, *B2-pnos::Luc* and the *B2TEE-pnos::Luc* constructs were compared to the reference line of GI::Luc. All data was normalised to the average expression of the time course. Error Bars=Standard Error a.) *TEE::Luc* b.) *B2-pnos::Luc* c.) *B2TEE-pnos::Luc*

4.6 Tracking of dusk is impaired in transgenic GI promoter constructs

Under LD 16 conditions the four different *GI* promoter reporter fusions showed clearly different expression patterns, which is reflected by the different peak times in expression. The peak time of *GI* expression differs under SD 8 and LD 16 conditions (Fowler et al., 1999). Moreover, the *GI* peak in expression tracks dusk during different photoperiods (Edwards et al., 2010b). The current model of photoperiodic flowering suggests that this dusk-tracking of *GI* is crucial for the degradation of CDFs at the right time of the day and therefore for *CO* mRNA upregulation in long days.

In order to test the contribution of EEs and Block 2 on *GI::Luc* tracking of dusk, plants with the previously described luciferase constructs were grown under different photoperiods from 4h light / 20h darkness to 16h light / 8h darkness. Subsequently all peak times were calculated and compared.
As reported previously, *GI::Luc* expression tracked dusk at photoperiods between 4h and 12h, whereas between photoperiods of 12h and 16h peak time changed only mildly (Fig.4.4a). In contrast, two independent TEE lines showed a different pattern. For photoperiods between 10h and 16h of light a clear difference of approximately 1,5h in peak time was detected compared to *GI::Luc* plants, however between photoperiods of 4h and 8h of light the peak time was the same as for *GI::Luc* (Fig. 4.4a). This shows that the tracking of dusk is impaired in these plants, with a critical point between 8h light / 16h darkness and 10h light / 14h of darkness.

The *B2-pnos* construct, in contrast, displayed a delayed peaktime compared to the *GI::Luc* construct (Fig. 4.4b). This was most pronounced under short day conditions of SD4 and SD6. Under all other photoperiods that were studied this effect was more subtle, with a phase delay of approximately 0.5h.
As previously described, the *B2TEE-pnos* construct is largely arrhythmic. Nevertheless peak times were calculated for all photoperiods and revealed a tracking of dusk under SD4 and SD 6, whereas the remaining measured points display a huge variety (Fig 4.4b). This is due to the plateau-like level of *Luc* expression conferred by this construct, which distributes the highest level of expression in each seedling throughout the day.

Taken together, all three of these *GI* promoter constructs impair the correct tracking of dusk that is displayed by the *GI* promoter.

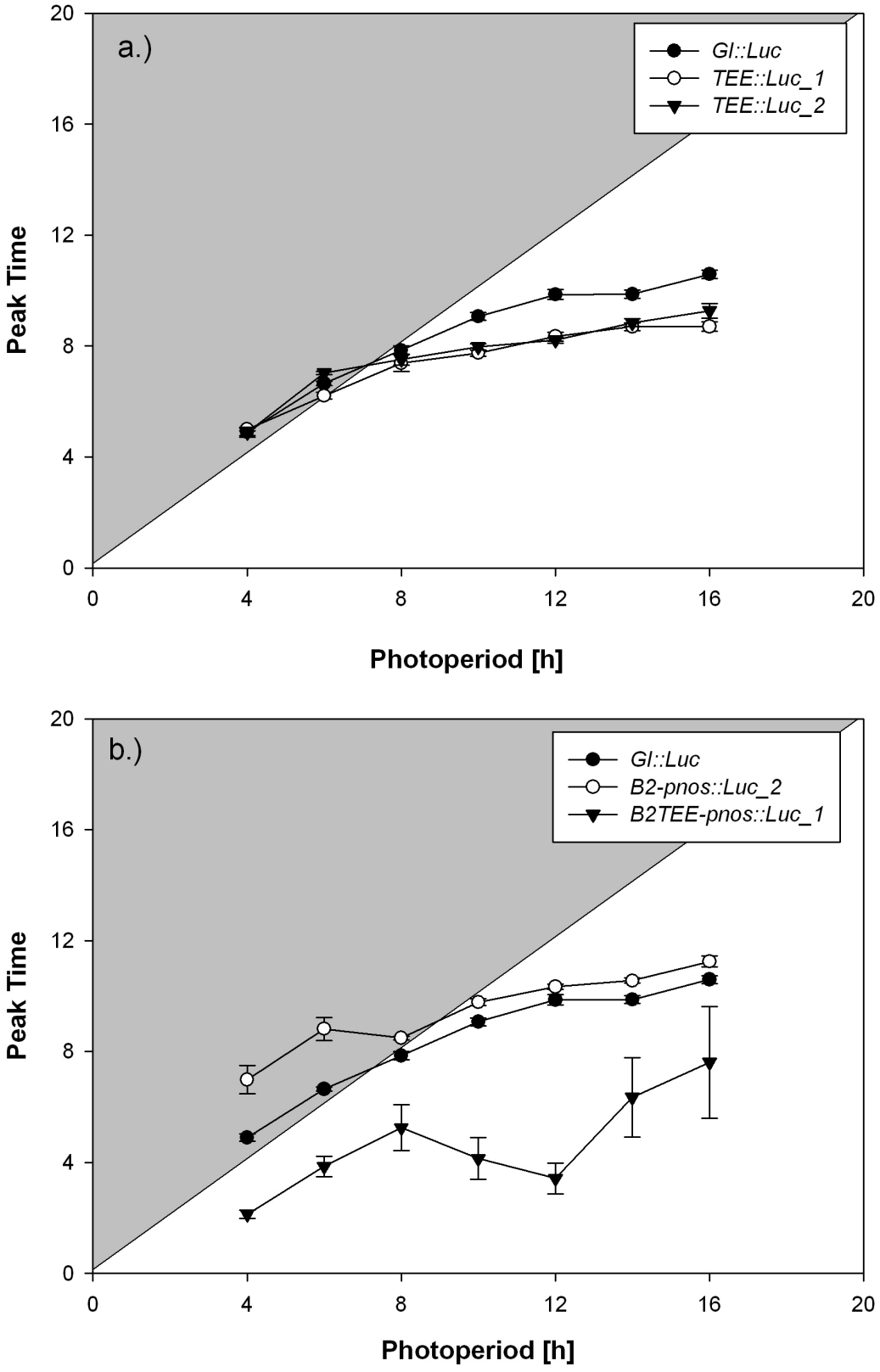


Fig.4. 4 Tracking of dusk is impaired in transgenic *GI* promoter constructs
Seedlings were grown under different photoperiods from 4h light / 20h dark to 16h light / 8h dark. Peak time of *GI* expression was calculated and plotted against the respective photoperiod. Error Bars = Standard Error; white and gray areas indicate light and dark, diagonal line indicates dusk.
a.) Two independent lines of the *TEE::Luc* construct compared to *GI::Luc*
b.) One line of the *B2-pnos::Luc* and the *B2TEE-pnos::Luc* construct compared to *GI::Luc*

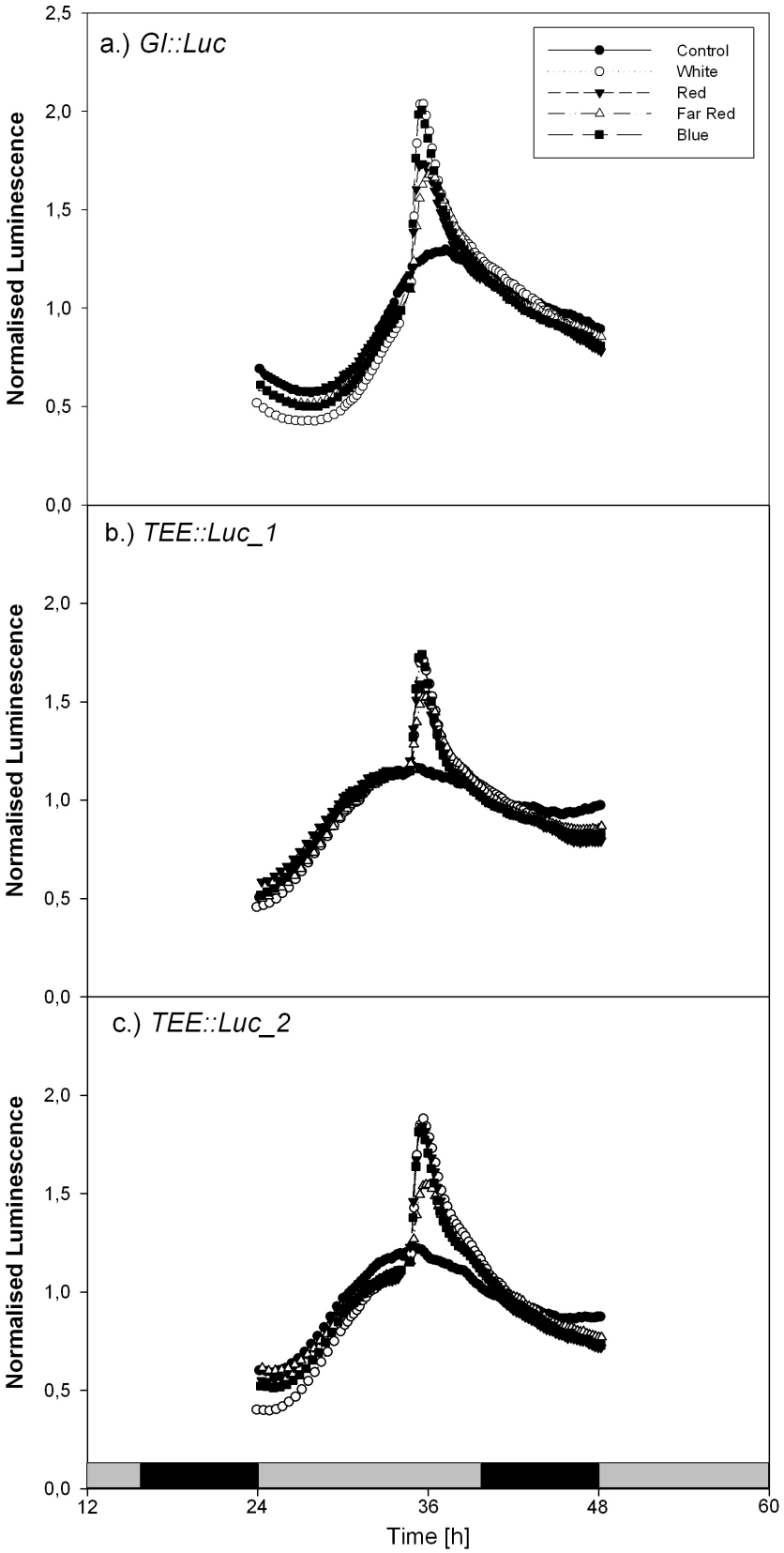


Fig. 4.5 TEE constructs show an acute response to different light qualities
Seedlings carrying the different constructs were grown under LD 16 conditions and then transferred to continuous dark. At ZT 36, plants were exposed to a light pulse of blue, red, far-red or white light. Control plants were kept in dark. Luminescence is presented as normalised values of at least 16 seedlings per line.a.) *GI::Luc* b.) TEE::Luc_1 c.) *TEE::Luc_2*

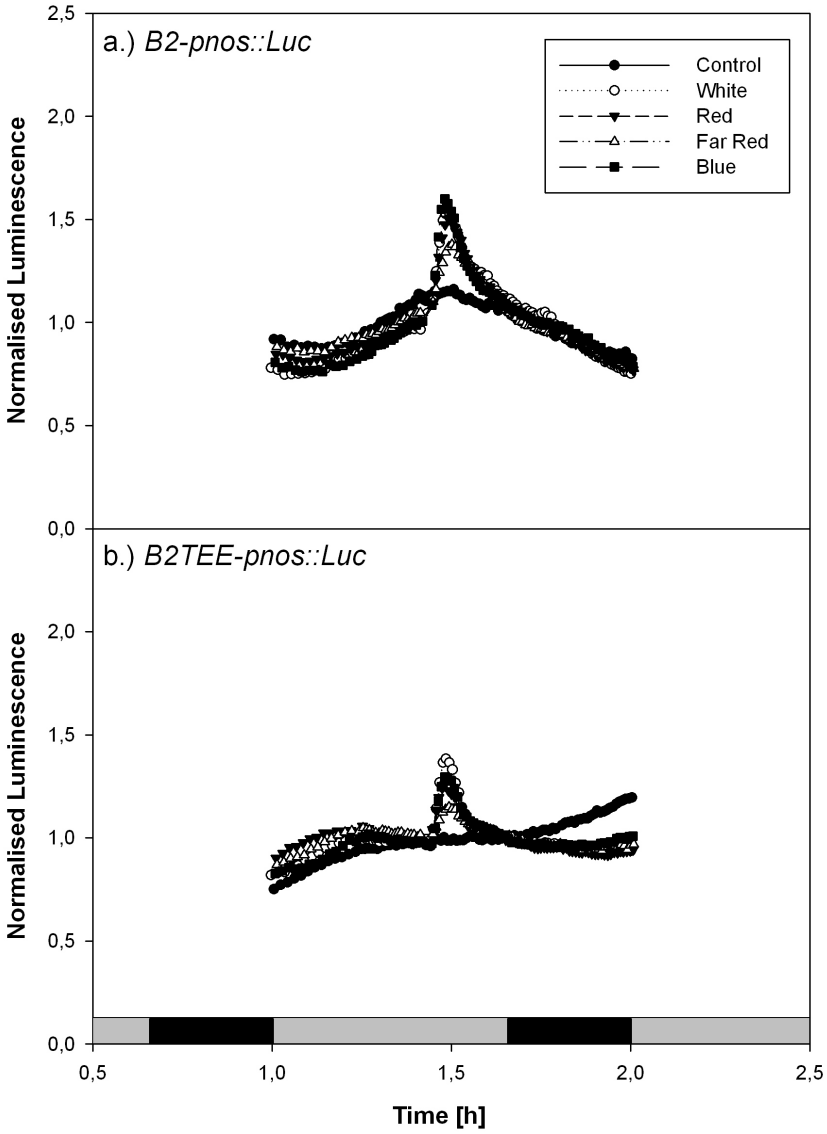


Fig. 4.6 *B2-pnos* and *B2TEE-pnos* constructs show an acute response to different light qualities
Seedlings of the different constructs were grown under LD 16 conditions for 7d and then transferred to continuous dark. At ZT 36, plants were exposed to a light pulse of blue, red, far-red or white light. Control plants were kept in dark. Luminescence is presented as normalised values of at least 16 seedlings per line.a.) *B2-pnos::Luc* b.) B2TEE-pnos::Luc

4.7 Mutation of EEs within the GI promoter impairs the acute response to light

One of the key features of the *GI* promoter is its induction by light at dawn and in the evening, which may contribute to clock entrainment (Toth et al., unpublished). This induction by light is gated during the day and can be caused by red, far-red or blue light (Toth et al., unpublished).

In order to test if this is still true for the EE depleted lines, all lines were grown in DD and a light pulse of 30min was applied in order to induce *GI*. As for the *GI::Luc* fusion, all constructs showed clear induction to all light conditions used. For all constructs, blue and white light gave the strongest response, whereas the far-red light pulse had the weakest effect (Fig. 4.5/ Fig. 4.6). Overall, however, this acute response to light was strongest in *GI::Luc*, whereas the *B2TEE-pnos::Luc* construct displayed a reduced response(Fig. 4.5a/ Fig. 4.6a,b). The *TEE::Luc* lines displayed an intermediate response (Fig. 4.5b,c). Collectively this data shows that Block 2 can mediate light induction and that the mutation of EEs might weaken this effect, but does not prevent it.

The acute response of the *GI* promoter is gated by the circadian clock (Toth et al., unpublished). As for other gated responses (Fowler et al., 2005), the effect of the light pulse is strongest at the time of maximum expression. As the *TEE::Luc* constructs confer an earlier peak time compared to the *GI::Luc* construct, this might also alter the timing of the gated acute response making it difficult to compare different constructs at the same time.
In order to overcome these difficulties, a full timecourse experiment was conducted. Seedlings were grown under LDs, transferred to the dark and subsequently replicate seedlings were subjected to a white light pulse of 30min every 2,5h. Plants were transferred back to darkness and luciferase expression was measured before and after the treatment. The acute response was then calculated by subtracting the pre-treatment values from those after the treatment with a light pulse.
The data clearly demonstrated the gated response to the light pulse for the *GI::Luc* construct (Fig. 4.7a). The maximum peak in induction correlated with the peak in expression of *GI*. Moreover, the maximum induction in the subjective evening was approximately sevenfold higher compared to the minimum induction in the subjective morning (Fig. 4.7a). In contrast, this gated response to a light pulse was much weaker in a representative *TEE::Luc* line (Fig. 4.7b) and almost depleted in a representative *B2TEE-pnos* line (Fig. 4.7c). Moreover, the differences between minimum and maximum induction during the day are much smaller compared to those observed with the full *GI* promoter.

Taken together, mutating EEs within the *GI* promoter reduces the gated response to a light pulse, most pronounced in the background of the Block 2 construct.

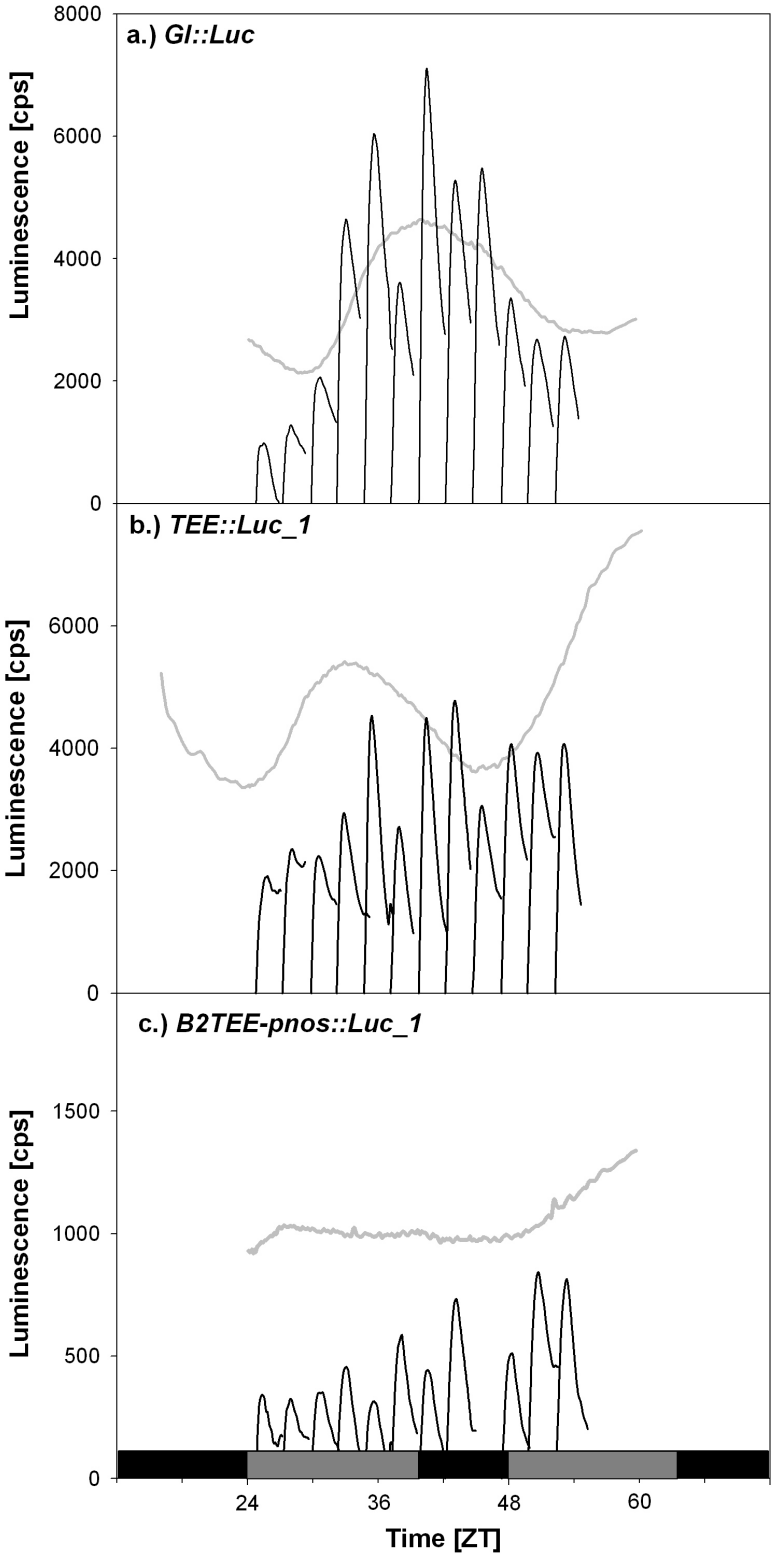


Fig. 4.7 Circadian clock-gated light induction is impaired in transgenic *GI* promoter lines
Seedlings of the *GI::Luc*, *TEE::Luc* and *B2TEE-pnos::Luc* constructs were grown under LD 16 conditions for 7d and then transferred to continuous dark. Replicate samples received a white light pulse of 30min at different times between ZT 25 and ZT 54 and luciferase expression was measured continuously before and after this treatment. Acute response was calculated by subtracting the luminescence values before the treatment from those after the treatment. Gray line represents the untreated dark-sample.
a.) *GI::Luc* b.) *TEE::Luc_1* c.) *B2TEE-pnos::Luc_1*

4.8 Other factors than LHY/CCA1 contribute to the EE-mediated transcriptional regulation of GI

It has been shown that the *lhy/cca1* double mutation causes a peak in *GI* expression in the morning (Mizoguchi et al., 2005). This led to the proposal that LHY/CCA1 binds the EEs in the *GI* promoter in the morning to repress *GI* transcription. During the day, LHY and CCA1 are degraded and therefore *GI* transcription can take place with a peak in the evening. Thus the phase-shift of *GI* expression towards the morning in the *lhy/cca1* double mutant could be caused by the loss of the direct repressors LHY and CCA1 that bind the EEs within the *GI* promoter in the morning.

In order to test these assumptions, the *TEE::Luc*, the *B2-pnos::Luc* and the *B2TEE-pnos::Luc* constructs were stably introduced into *lhy/cca1* double mutant plants and the luciferase expression of homozygous lines was measured and compared to an already established *GI::Luc* line in the *lhy/cca1* background (Cremer, unpublished). As previously reported for the RNA abundance of *GI* in a *lhy/cca1* double mutant (Mizoguchi et al., 2005), luciferase expression of *GI::Luc* in *lhy/cca1* peaks in the morning around ZT5 (Fig. 4.8a). The *TEE::Luc* constructs displayed a largely comparable expression pattern compared to *GI::Luc*. However, the peak in *Luc* expression in these transformants was slightly shifted earlier and as for the *TEE::Luc* constructs in the wildtype background, these constructs displayed a broader, flatter peak (Fig. 4.8a). *The B2-pnos::Luc* constructs displayed a similar expression pattern as the full length *GI::Luc*, with a slightly earlier expression peak (Fig. 4.8b). The *B2TEE-pnos::Luc* constructs, however, showed the strongest effect of all constructs in the *lhy/cca1* background. In these plants rhythmicity of luciferase expression was severely disrupted and only a weak peak in the morning was detected (Fig 4.8c).
If LHY/CCA1 were unique repressors of GI via direct binding to the EEs within the *GI* promoter, the deletion of these EEs would have no additional effect in the *lhy/cca1* background. As this is clearly not true for the *B2TEE-pnos::Luc* constructs compared to the *B2-pnos::Luc*, this shows that other factors regulate *GI* via EEs.

Taken together, all constructs in the *lhy/cca1* background showed a similar expression pattern as in the wt background, with the difference that the peak in expression was generally shifted towards the morning in the *lhy/cca1* background. Collectively this shows that other factors than LHY/CCA1 are important for the EE-mediated transcriptional response.

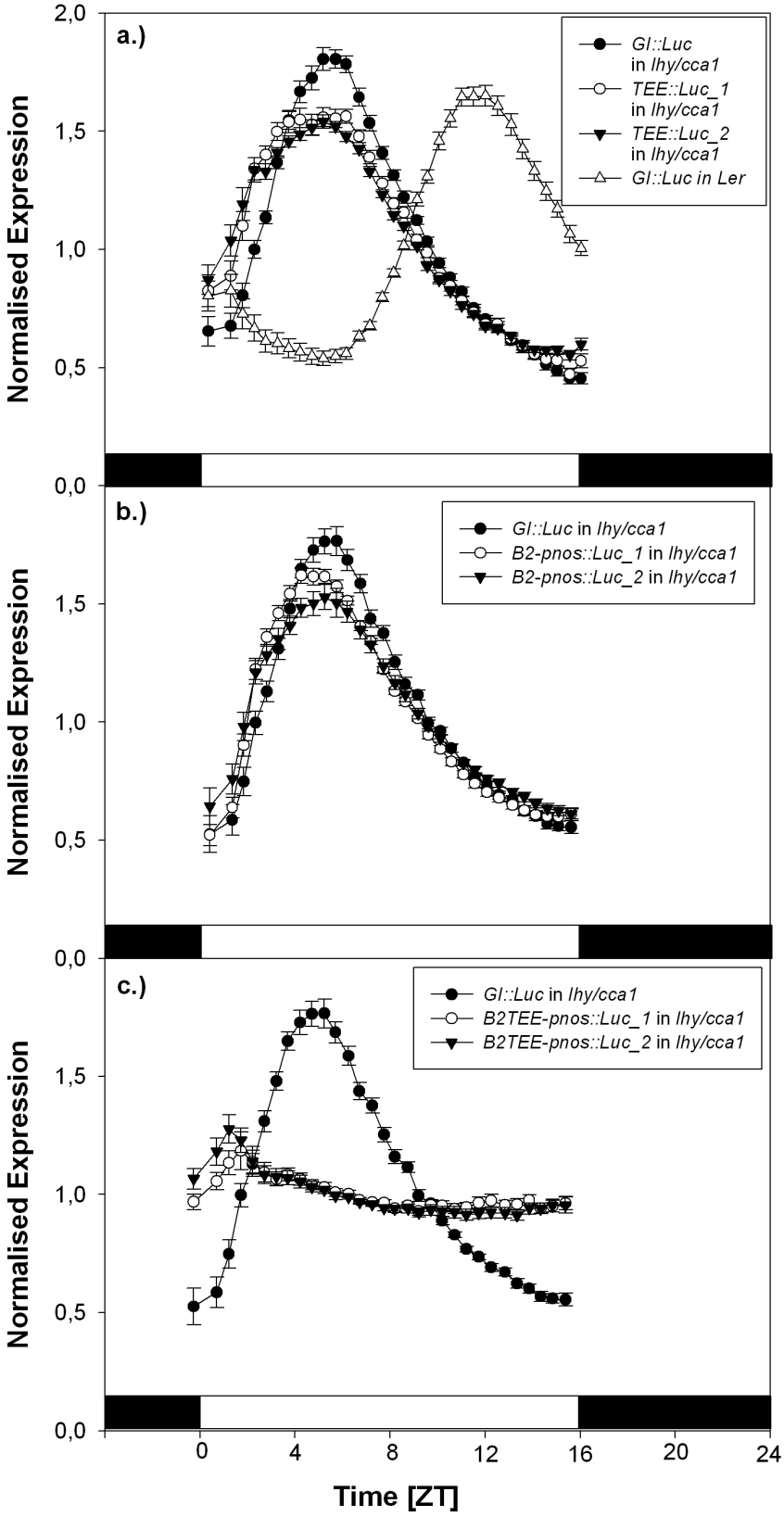


Fig. 4.8 Other factors than LHY/CCA1 contribute to the EE-mediated transcriptional regulation of *GI*
Each 2 independent lines of the *TEE::Luc*, *B2-pnos::Luc* and the *B2TEE-pnos::Luc* constructs in the background of the *lhy/cca1* double mutation were compared to the reference line of *GI::Luc* in *lhy/cca1*. All data was normalised to the average expression of the measured time frame. Error Bars=Standard Error a.) *TEE::Luc* b.) *B2-pnos::Luc* c.) *B2TEE-pnos::Luc*

4.9 Block 2 can drive expression without a minimal promoter

The phylogenetic shadowing showed that the 3' end of Block 2 comprises a number of core promoter elements, such as a TATA Box and a CT rich repeat (see chapter 2). Moreover, a weak expression of RNA in this region was detected by real time PCR (data not shown). Together this provides evidence that Block 2 might be capable of driving transcription. In order to test if Block 2 can confer a transcriptional start site, stably transformed *A. thaliana* lines with Block 2 as well as Block 2 with the three mutated EEs fused to the luciferase open reading without the *pnos* minimal promoter were established.

Surprisingly, these two constructs gave a very clear expression pattern comparable to the respective *pnos* minimal promoter constructs (Fig. 4.10). Moreover, the peak in relative expression in the Block 2 lines exactly matched the timing of the *GI::Luc* control, whereas the B2TEE::Luc lines showed a broader peak with lower amplitude and an earlier peak in expression (Fig. 4.10). No expression was detected for the B1 and the B 3 fragments fused to luciferase (data not shown).

Therefore Block 2 contains a transcriptional start site capable of driving a *GI*-like pattern of *Luc* expression even in the absence of a minimal promoter.

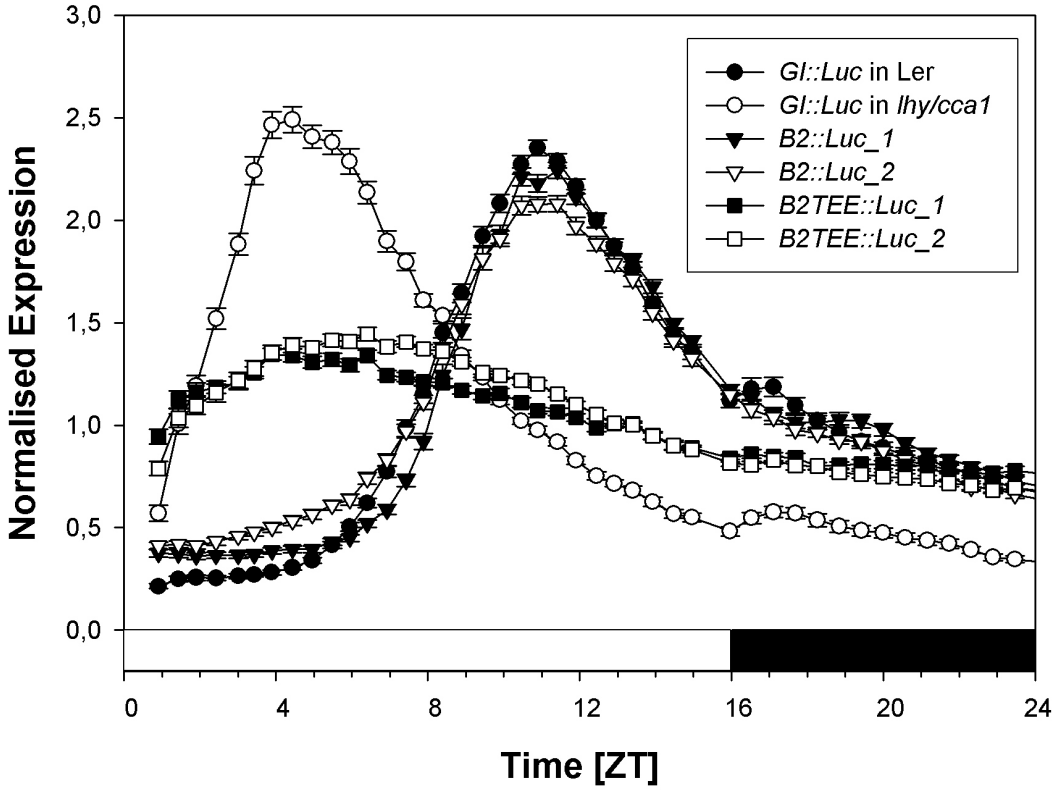


Fig. 4.9 Block 2 can drive luciferase expression without the pnos minimal promoter
Each 2 independent lines of the, *B2::Luc* and the *B2TEE::Luc* constructs without the pnos minimal promoter were grown under LD 16 conditions and compared to the reference line of *GI::Luc*. Luciferase expression was normalised to the average expression of the measured time frame. Error Bars=Standard Error

4.10 Trichostatin A delays the phase of GI expression

Epigenetic regulation provides an additional layer of transcriptional regulation by modifying the tails of histones and thus making the chromatin generally more or less accessible to transcription factors (general review). In the plant circadian clock it was shown that the chromatin state of the *TOC1* promoter is regulated in a circadian rhythm and that the proper timing of *TOC1* expression relies on the acetylation and de-acetylation of histones within its promoter (Perales 2008).

In order to test if histone acetylation and de-acetylation also play a role for *GI* transcriptional regulation, seedlings harbouring the *GI::Luc* transgene were grown for 7d and then transferred to medium containing different concentrations of Trichostatin A (TSA). TSA is a potent inhibitor of histone de-acetylases (Pikaard and Chang, 2005), thus leading to hyperacetylation of histones, which generates a more open chromatin state and therefore enhances transcription.

Exposing seedlings to increasing concentrations of TSA caused a shift in *GI::Luc* expression towards the evening with a maximum phase shift of almost 3h compared to control plants (Fig.4.11a). To investigate a general clock effect of TSA, the average period of TSA-treated plants was calculated under the different conditions. This analysis revealed no significant differences between the differently treated seedlings (Fig. 4.11b).

Taken together, TSA caused a delay in *GI* phase, suggesting that chromatin state is important for *GI* transcriptional regulation, although it remains to be tested whether this is a direct effect on the *GI* promoter or an indirect effect through a *GI* regulator.

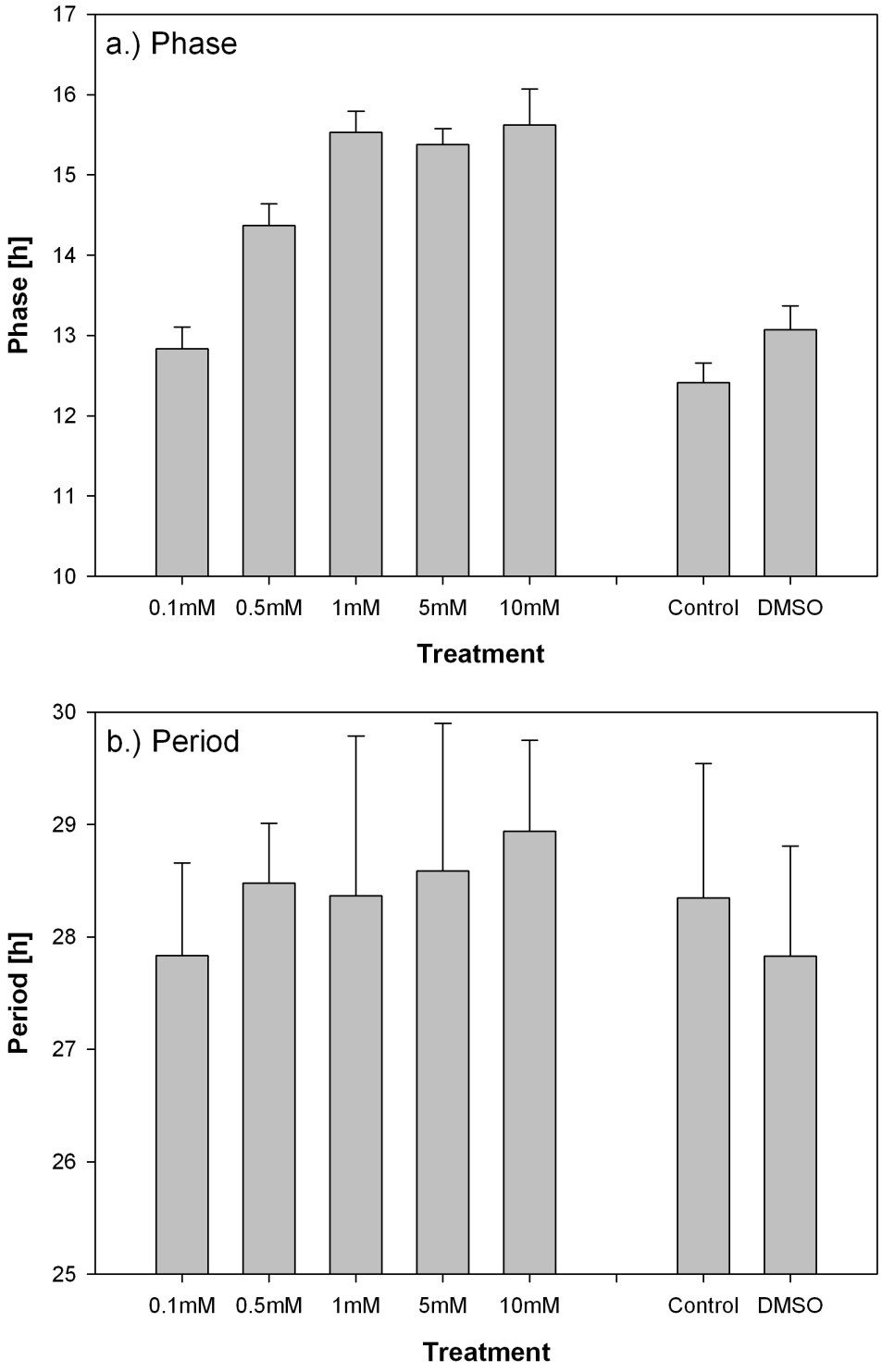


Fig. 4.10 Trichostatin A affects the Phase of *GI::Luc* expression
At least 12 seedlings with the GI::Luc construct were grown under LD 16 conditions for 7d. Subsequently, plants were transferred to medium containing the indicated concentration of Trichostatin A. Phase was calculated the following day in DD. Values represent mean values +/- SE.

4.11 Discussion

In order to test conserved *cis*-regulatory modules and elements that were predicted by the phylogenetic shadowing, stable transgenic *Arabidopsis* lines carrying the luciferase reporter system were used.

A 2,5kb *GI* promoter construct was used as the basis for all *GI* promoter studies. Previous experiments with this 2,5kb construct of *A.thaliana* demonstrated that it confers the same diurnal expression pattern of luciferase expression as *GI* RNA abundance measured by qPCR in previous reports (Fowler et al., 1999; Park et al., 1999). Moreover, this 2,5kb *GI* promoter fragment (Cremer et al., unpublished) is used in the laboratory for a number of different experimental approaches, such as chemical genetic screens (Toth, Nougalli et al, unpublished) and in the search for natural variation within the *Arabidopsis* circadian clock (de Montaigu et al, unpublished). The intergenic region between *GI* and its upstream gene, *PAB3*, has a length of 3,6kb and was used for the phylogenetic analysis. Conserved block 4/5 is not present in the 2,5kb fragment, whereas Block 1, Block 2 and Block 3 are present. Nevertheless a previously reported transgenic luciferase construct with the full 3,6kb promoter gave the same expression pattern as the 2,5kb construct (Onai et al., 2004).
Taken together, the 2,5kb fragment reflects *GI* transcriptional regulation and is therefore used as the basis for the *GI* promoter analysis within this study. Moreover, the 2,5kb fragment fused to the *GI* cDNA fully complements the *gi* mutation (see chapter 5).

A phylogenetic shadowing approach is based on the assumption that the compared orthologous sequences retain the same function, e.g. the same expression pattern on orthologous genes. The *AaGI::Luc* lines in *A. thaliana* give strong evidence that this is the case for the *GI* promoters of the Brassicaceae. Moreover, as the *AaGI* promoter is heterologously expressed in *A. thaliana* and confers the same expression, this indicates that not only the promoters, but also the upstream regulators of *GI* are conserved between *A. thaliana* and *A. alpina*. As all plant species investigated so far, including distantly related species such as Arabidopsis and rice, show evening expression of *GI* (Hayama et al., 2003) it is not surprising that two species within the same family display the same expression pattern. As *AtGI* and *AaGI* promoters have the same activity, consequently the definition of the five conserved blocks within the *GI* promoters of eight Brassicaceae was based on the conservation between *A. thaliana* and *A. alpina*.

Analysis of conserved Block 1, Block 2 and Block 3 revealed that Block 2 is capable of conferring a *GI*-like expression pattern on luciferase, whereas Block 1 and Block 3 are only capable of driving a very weak luciferase expression pattern, which most likely reflects the basal expression pattern conferred by the pnos minimal promoter (Puente et al., 1996). Block 2 is therefore likely to be important for *GI* regulation. However, the weaker amplitude of the *B2-pnos::Luc* constructs compared

to *GI-pnos::Luc* demonstrates that other parts of the promoter are important for high level *GI* expression as well. Block 1 and Block 3 are the most likely candidates for such enhancers, even if they are not capable of conferring robust *GI* expression on their own. Moreover, Block 2 is capable of responding to different light spectra, demonstrating that this integral part of *GI* regulation is encoded within Block 2. The phylogenetic shadowing analysis also showed that Block 2 comprises a large number of previously described clock-associated *cis*-regulatory elements. Among these were 3 conserved EEs, which were promising candidates for generating *GI* evening expression.

The mutations in *TEE::Luc* and the *B2TEE-pnos::Luc* constructs have related effects. However, deletion of the 3 EEs has a strongly enhanced effect if carried out only in the context of Block 2. This demonstrates that EEs are active within the *GI* promoter, but that within the *GI* promoter outside Block 2 there is some redundancy. Such a redundancy could be conferred by other EE-like elements, for instance by a CBS that is located in the proximal part of the *GI* promoter.

The data presented here suggest that mutation of EEs in the *GI* promoter has two major effects: It shifts the peak in expression of *GI* towards the morning and it impairs the gated response to a light pulse. In contrast, mutating an EE in fragments from other clock-related genes such as PRR9 or CCR2 caused complete arhythmicity in such transgenic plants (Harmer and Kay, 2005). However, these fragments were only 20-30bp long and then fused to a *pnos* minimal promoter. This shows that EEs can confer rhythmicity, but it does not give further insights into the interaction of EEs with surrounding DNA.

Mutagenesis of individual EEs within the full length *GI* promoter was expected to progressively shift the peak in expression towards the morning and finally result in an expression pattern as in the background of the *lhy/cca1* double mutation. Even though *Luc* expression is shifted towards the morning in the *TEE::Luc* plants, this effect is less pronounced as in a *lhy/cca1* mutant. One possibility for this pattern is that the *GI* promoter still contains redundant elements to the mutagenised EEs in Block 2 and LHY/CCA1 can still repress *GI* during the day, possibly due to binding of a CBS in the proximal part of the promoter.

However, data from the *B2-pnos::Luc* and *B2TEE-pnos* constructs argue that other factors recognise the EEs in *lhy/cca1* double mutants. If LHY/CCA1 were unique repressors of *GI* via EEs in Block 2, no difference between the mutated and the non-mutated constructs would be discovered in *lhy/cca1* double mutants. However, rhythmicity is completely disrupted in the B2TEE-pnos lines compared to the *B2-pnos* lines in the double mutant. The pattern of the two different constructs in the *lhy/cca1* double mutant is very similar to the pattern observed in wildtype plants, suggesting that an EE-binding factor that is not LHY/CCA1 mediates the rhythmic expression of *GI*.

The *lhy/cca1* double mutation causes transcriptional miss-programming of many clock controlled genes (Alabadi et al., 2001; Carre et al., 2001; Schaffer et al., 2001; Alabadi et al., 2002; Mizoguchi et al., 2002a). Therefore the morning expression of *GI* in the *lhy/cca1* double mutant might be caused by indirect effects, such as altered expression of unknown transcription factors that regulate GI. Moreover, it was shown recently that LHY associates with the *GI* promoter (Lau et al., 2011), suggesting that at least LHY directly regulates *GI*.

In summary, data obtained during this study suggests that LHY/CCA1 regulates *GI* probably by a combination of direct and indirect mechanisms.

One interesting feature of Block 2 within the *GI* promoter is that it can confer expression in the absence of a minimal promoter, indicating that it encodes a transcription start site within its sequence. Nevertheless, Block 2 is 500bp upstream of the established transcriptional start site of *GI*. Core promoter elements were identified by the phylogenetic analysis at the 3' end of Block 2 and are likely to be the cause of this ability to initiate transcription. It is so far unclear what might be the biological function of such a second transcription start. Compared to the *pnos*-constructs, the expression pattern generated by the B2 constructs is more similar to the *GI* expression as the *B2-pnos* constructs, possibly due to interference of two different transcriptional start sites.

Experiments with the histone deacetylase inhibitor TSA demonstrated an effect on the peak time of *GI* transcription. However, TSA generally affects acetylation of histones and therefore might cause a general transcriptional reprogramming. Therefore it is not clear if this effect on *GI* phase is direct or indirect. For the transcriptional regulation of the evening gene *TOC1* a similar phase shift upon TSA treatment was reported (Perales and Mas, 2007). Moreover, it was shown that the chromatin state at the *TOC1* promoter is directly regulated by the circadian clock. Therefore one can propose that similar mechanisms might confer the phase shift of *GI,* another evening expressed gene. Moreover, the absence of period changes upon TSA treatment suggests that the plant circadian clock is not generally changed in TSA treated plants, thus also supporting the hypothesis of a direct effect on *GI* transcriptional regulation.

Taken together, the luciferase data highlight the importance of Block 2 and the EEs within the *GI* promoter of *A. thaliana*. Three different *GI* promoter constructs confer a variety of different expression patterns of *GI* and therefore provide an effective tool to test the effect of missexpression of *GI* conferred by these promoters in transgenic plants.

Functional characterisation of different *GI* expression patterns

5.1 Perspective

One method to analyse promoters in different organisms is the use of marker gene fusions in which promoter fragments are to a reporter gene such as luciferase (e.g. this study, (Spensley et al., 2009)) or Beta-glucuronidase (Adrian et al., 2010). Although this approach provides important insights into sequences required for promoter activity, it does not necessarily test the biological relevance of these promoter regions or *cis*-regulatory elements.
Here I follow different approaches in order to directly study the effects of *cis*-regulatory modules and elements within the *GI* promoter on the ability of *GI* to confer its biological function in plants. The approaches used include the analysis of T-DNA insertion mutants, the analysis of natural variation within the *GI* promoter using a broad range of natural accessions as well as the generation and analysis of transgenic *gi* complementation lines. Combining the data obtained by these approaches with the luciferase results described in chapter 4 will further elucidate the functionality of *cis*-regulatory modules and elements within the *GI* promoter and relate this to the biological function of the gene.

5.2 Analysis of plants with a T-DNA insertion within the GI promoter

A.thaliana libraries of T-DNA insertion mutants are frequently used for reverse genetics (Alonso et al., 2003). Commonly such lines are exploited to study the function of a gene that harbours a T-DNA insertion and therefore is likely to be mutant for that gene (Alonso et al., 2003). However, for the analysis of promoter functions T-DNA insertion lines might also be a promising tool. The part of the promoter that is upstream of the T-DNA insertion may be so far separated from the rest of the gene that it no longer influences transcription.

In order to further elucidate the function of the *GI* promoter, seven T-DNA insertions located at different positions across the *GI* promoter were identified in the SALK or the SAIL T-DNA collections. Lines were assigned numbers from 1 to 7 according to their location within the *GI* promoter (Fig. 5.1) and homozygous plants were selected for all seven lines. Subsequently flowering time of these T-DNA insertion mutants was analysed under long day conditions (LD 16).
This analysis revealed that lines 1 to 3 flowered comparably to the wildtype control, whereas lines 4 to 7 showed a subtle late flowering phenotype (Fig. 5.1). This late flowering was most pronounced in

line 5 that flowered with approximately 8 leaves more compared to the wildtype. Nevertheless, all T-DNA insertion mutants flowered clearly earlier than the *gi-2* mutant, indicating that *GI* is expressed and functional in all these lines. Lines 1 and 2 harbour the T-DNA insertion upstream of the conserved Block2, lines 3 and 4 within Block 2 and lines 6 and 7 have the T-DNA insertion between Block 2 and conserved Block 1. Line 5, which shows the most pronounced late flowering phenotype, harbours the T-DNA at the end of Block 2, thus probably uncoupling this conserved region from the rest of the promoter. Taken together, these data show that T-DNA insertions within the *GI* promoter affect flowering time in *A. thaliana*, presumably due to reductions in or changes in the spatial or temporal pattern of *GI* mRNA.

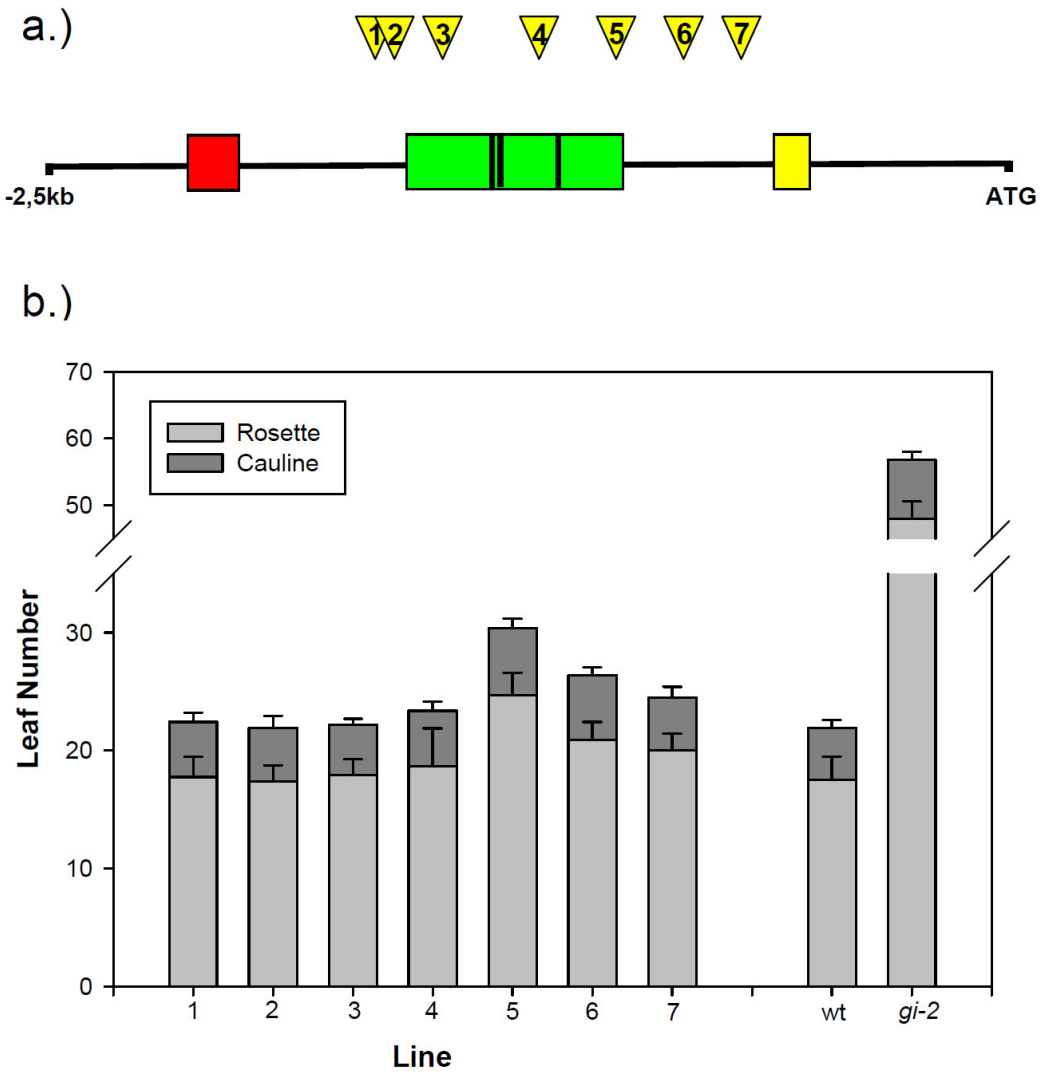


Fig.5. 1 Flowering time of lines carrying T-DNA insertions within the *GI* promoter
a.) Location of T-DNA insertion within the *GI* promoter. Yellow triangles indicate number and insertion site of T-DNA within the *GI* promoter. Colored boxes indicate location of conserved blocks as described in chapter 1; Bars within Block 2 indicate location of EEs. Locations of the T-DNA insertions are given in the Materials and Methods.
b.) Flowering time of 7 T-DNA insertion lines under long day conditions (LD 16). A second independent experiment gave similar results.

5.3 Natural variation within the GI promoter of Arabidopsis

Many plants from the same species colonise different geographic regions and are adapted to these local habitats. This adaption is reflected by differences in the DNA sequence between different local accessions. In *A. thaliana* this natural variation within different accessions has been used in various reverse genetics approaches (de Meaux and Koornneef, 2008).

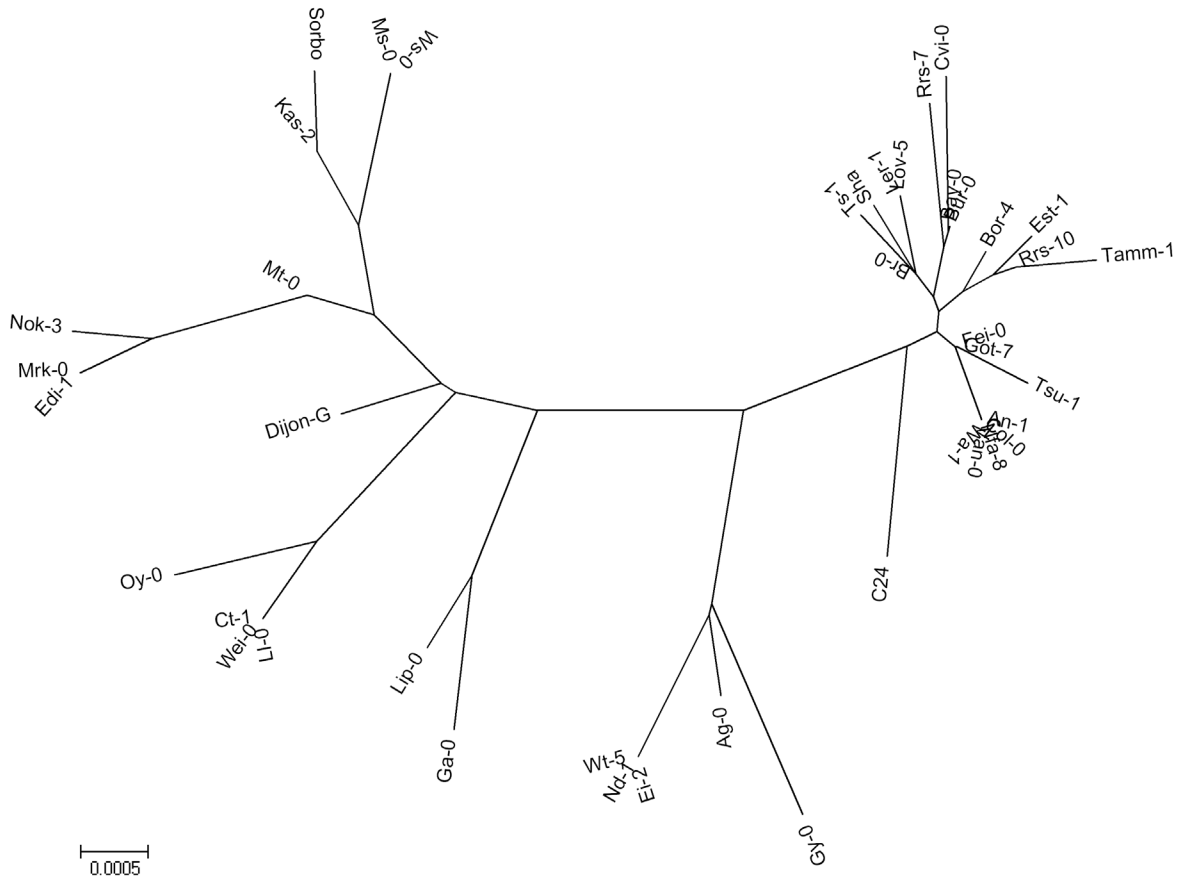


Fig.5. 2 Phylogenetic tree of the 1,8kb *GI* promoter of 42 different *A. thaliana* accessions; Distance of 0,0005 corresponds to 1 SNP between different promoters

In order to identify SNPs in *cis*-regulatory elements within the *GI* promoter, a 1,7kb upstream region measured from the translational start site (which contains conserved Block 2 almost completely) of 42 different *Arabidopsis* accessions was compared (Fig. 5.2). These 42 sequences were obtained from different databases and by sequencing (see Material and Methods). Approximately half of the *GI* promoters from this dataset varied in a range of 3 to 4 SNPs, whereas the other half contained up to 20 SNPs. Interestingly, the majority of these SNPs occurred in the regions between conserved blocks (data not shown). No variation was detected in any EE or CBS within this set of *GI* promoters. Inspection of other putatively important *cis*-regulatory elements within the *GI* promoter revealed that there were 2 different accessions that harbour the same SNP within ABREL 3. These accessions are

Edi-1 (from Edinburgh, Great Britain) and Mrk-0 (from Märkt am Rhein, Germany), where the SNP changes ABREL3 from CACGT to CAC**A**T, which might be an interesting subject of future studies.

5.4 Natural variation within the GI promoter of Capsella rubella

During the phylogenetic analysis (chapter 2), *GI* promoter sequences from two different *Capsella rubella* accessions were obtained. One is from the 'Monte Gargano' accession, which was collected from the Monte Gargano region in western Italy and is the strain that is being sequenced by an international consortium (Barbara Neuffer, personal communication). The second sequence is from an accession that was obtained from IPK in Gatersleben. The true origin of this accession is unclear, because it was obtained from a botanical garden in Portugal, but there is no record of its initial collection site. Therefore this accession is referred to as 'IPK'. Another *C. rubella* accesession which is used by the international community is 'Circus Maximus', which was collected in Rome, Italy. So far no *GI* promoter fragment was obtained from this accession.

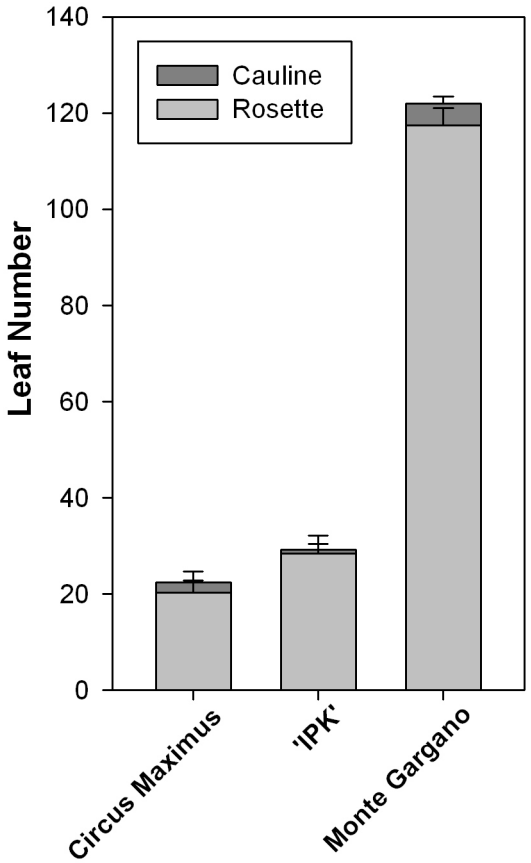


Fig. 5.3 Flowering time of three different *C. rubella* accessions grown under long day conditions (LD 16)

Interestingly, the 'IPK' and the 'Monte Gargano' accession have relatively diverse *GI* promoters, which include an SNP in EE2 of the IPK accession. As EE1 is mutated in *Capsella* (chapter 2), the *GI* promoter of the IPK accession only harbours one full length EE (EE3).

Luciferase data from chapter 4 suggest that the loss of EE1 and EE2 changes timing of the peak in expression of *GI* and therefore might change the flowering behaviour of these two different *Capsella* accessions. Therefore, flowering time of the three aforementioned accessions was determined under LD conditions. This analysis revealed that the 'IPK' and the 'Circus Maximus' accessions flowered clearly earlier than the 'Monte Gargano' accession. 'Circus Maximus' flowered at approximately 20 leaves, 'IPK' at approximately 30 leaves and 'Monte Gargano' at approximately 120 leaves under long day conditions and without vernalisation (Fig. 5.3). This shows that there is tremendous variation in flowering time between different accessions of *C. rubella.* Analysis of *GI* mRNA in these accessions will determine whether its timing of expression is affected by polymorphisms in the promoter region.

5.5 Analysis of transgenic gi complementation lines

Analysis of T-DNA insertions within the *GI* promoter and the search for natural variation within the *GI* promoter of different *Arabidopsis* accessions did not reveal which *cis*- regulatory modules or elements are crucial for the effect of *GI* on flowering time.

To precisely answer these questions a set of *GI* promoter fragments was fused to the cDNA of *GI* and introduced into the *gi-2* mutant via *Agrobacterium*-mediated transformation. In order to compare these transgenic plants with the previously obtained results from the luciferase reporter lines, the same *GI* promoter fragments were used in both experiments: These were the full length *GI* promoter, the full length *GI* promoter with all three EEs mutated, conserved Block2 and conserved Block2 with all three EEs mutated (Tab. 4.1).

Homozygous lines were established for the four constructs and different experiments were carried out in order to characterise these transgenic plants. A focus was set on *gi* - related phenotypes and therefore flowering time under different photoperiods, length of the hypocotyl and the response to freezing stress were analysed in the transgenic plants.

Name	Background	Description
GI::GI	*gi-2*	2,5kb *GI* promoter fused to the *GI* cDNA
TEE	*gi-2*	2,5kb *GI* promoter fused to the *GI* cDNA, 3 EEs mutated
B2	*gi-2*	conserved block 2 (700bp) fused to the *GI* cDNA
B2 TEE	*gi-2*	conserved block 2 (700bp) fused to the *GI* cDNA, 3 EEs mutated

Tab. 10 Overview of transgenic *gi* complementation lines

5.5.1 Flowering time of GI promoter fusion containing transgenic lines under different day lengths

In order to analyse the flowering phenotype of transgenic *gi* complementation lines, flowering time was first analysed under long day (16h light / 8h darkness) and under short day conditions (8h light / 16h darkness). Under both conditions, the *GI::GI* plants flowered like wildtype, demonstrating that the construct is fully functional und thus can complement the *gi* mutation (Fig 5.4).
Under long day conditions, 8 independent T-DNA insertion lines (4 shown in Fig. 5.4) with the TEE construct flowered comparably to wildtype plants. The same was true for two independent lines carrying the B2 construct or the B2 TEE construct. This shows that all three different types of construct can fully rescue the *gi* mutant flowering phenotype under long day conditions similarly to the *GI::GI* construct.

Under short day (SD 8) conditions, all *gi* complementation lines flowered as late as wt plants. The *gi-2* mutants only flowered slightly later than wt and *GI::GI* plants, whereas *35S::GI* plants flowered clearly earlier (Fig. 5.4). These data indicate that none of the constructs caused a *GI* overexpression phenotype under SDs.

In order to further investigate the response to different photoperiods in the *gi* complementation lines, the flowering times of all transgenic lines were determined under the three intermediate photoperiods of 14h light / 10h darkness, 12h light / 12h darkness and 10h light / 14h darkness. At least two independent transgenic lines were analysed for each genotype and compared to *35S::GI*, *gi-2* and *GI::GI* plants. *GI::GI* and wt plants (not shown) were almost equally late flowering under SD 8 and SD10 (Fig. 5.5). Lengthening photoperiod led to pronounced acceleration of flowering with the greatest acceleration in flowering time between SD 10 and LD 12.
In contrast, *35S::GI* plants flowered early, largely irrespective of the photoperiod. The *gi-2* mutants flowered late in all photoperiods. However, longer days slightly accelerate flowering even in *gi-2*, especially under LD 16 conditions. Flowering time of the three *gi* complementation lines TEE, B2 and B2 TEE plants was very similar to *GI::GI* control plants under all different photoperiods. Slight differences were detected under the SD 10 condition between these lines: TEE and B2 TEE plants flowered slightly later as *GI::GI*, but almost as late as *gi* plants under these conditions (Fig. 5.5).

Taken together, no strong differences in flowering time were detected between the 3 lines carrying derivates of the *GI* promoter and those carrying the full wt promoter under 5 different photoperiods.

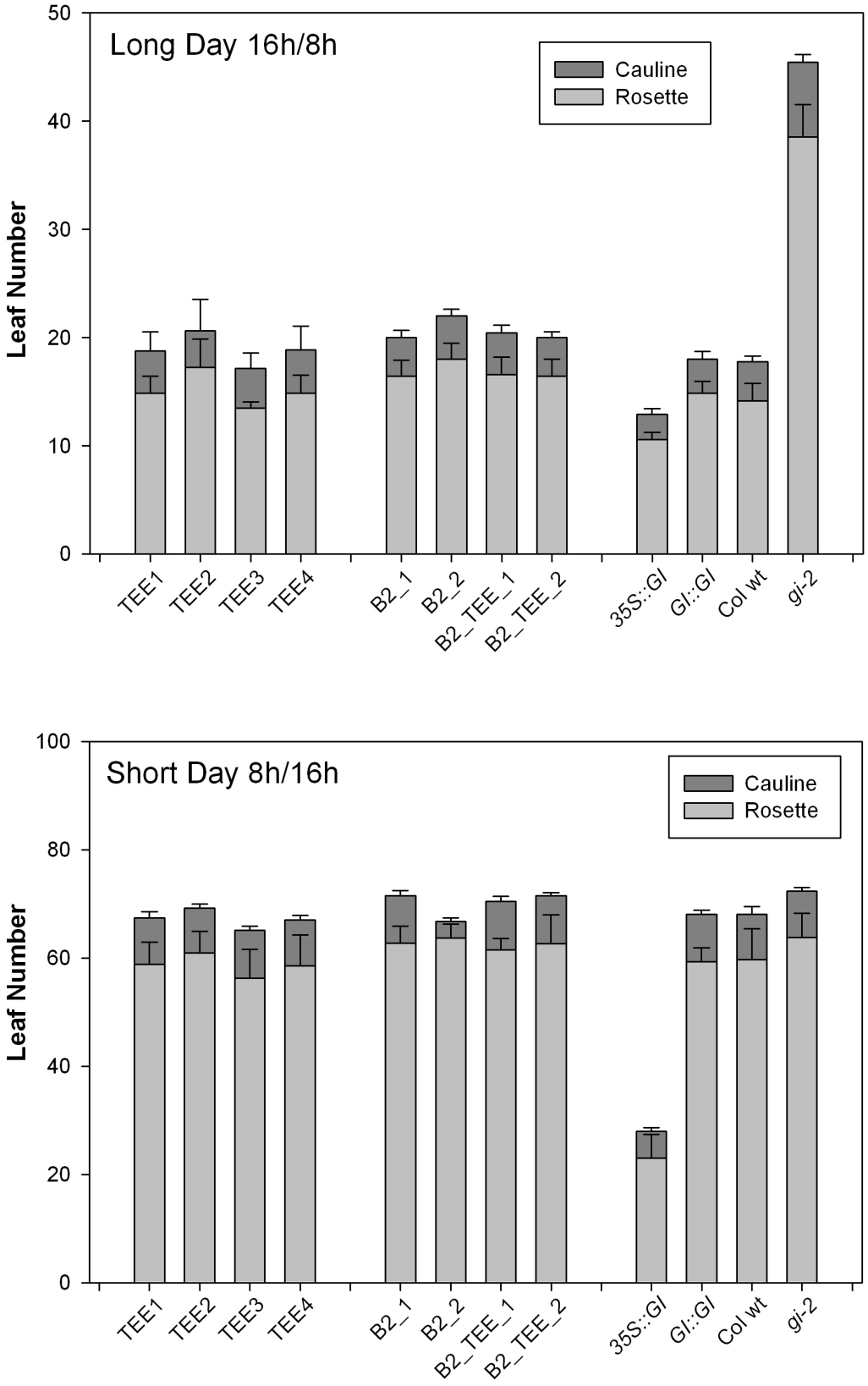


Fig. 5.4 Flowering times of different *gi* complementation lines
a.) long day conditions (LD 16) b.) short day conditions (SD 8);
Representative data of two independent experiments with same results

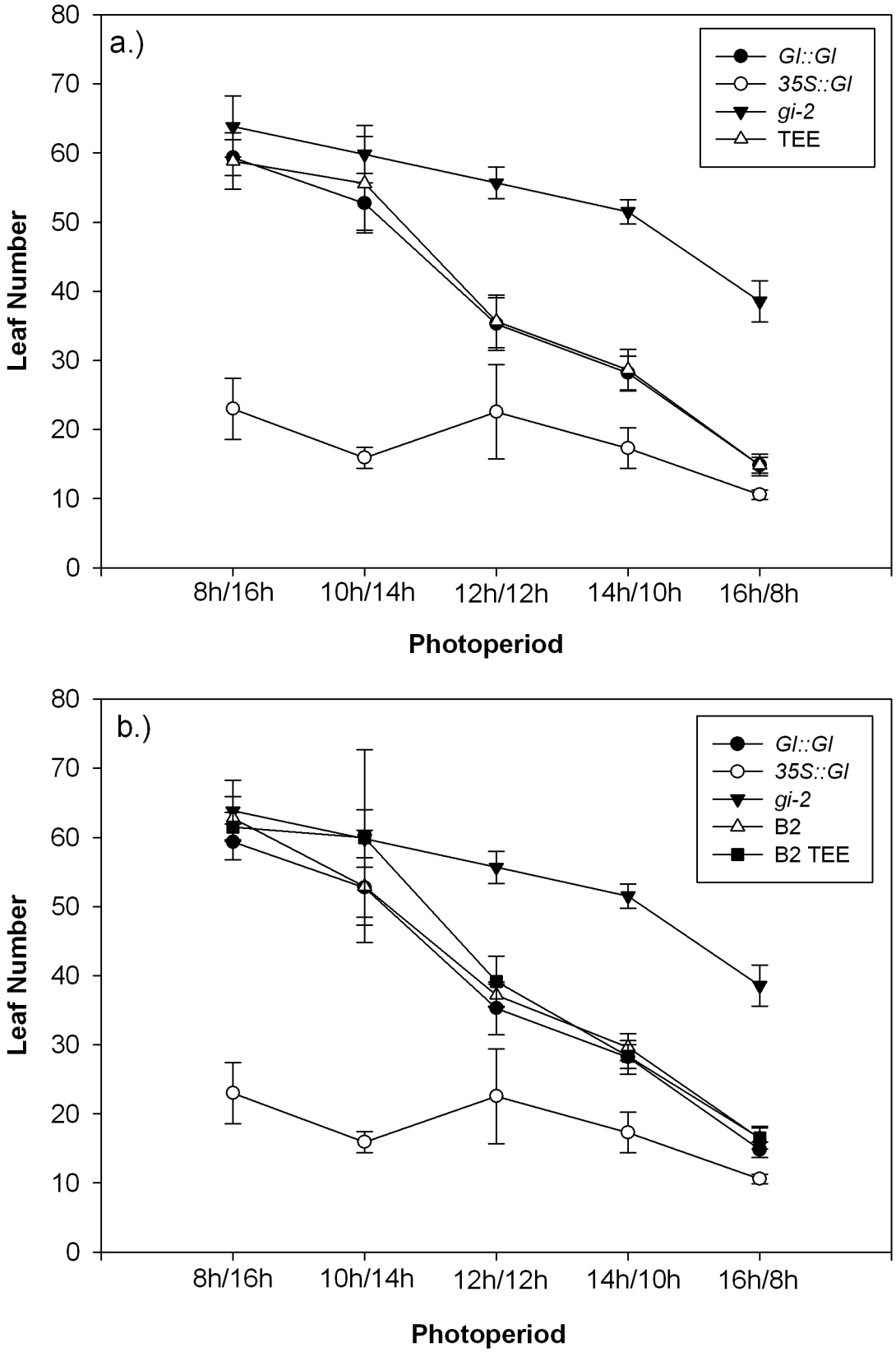


Fig. 5.5 Flowering times of different *gi* complementation lines under different photoperiods.
a.) Flowering time of one representative TEE line compared to *GI::GI*, 35S::GI and *gi-2*
b.) Flowering time of one representative B2 and B2 TEE line compared to *GI::GI*, 35S::GI and *gi-2*

5.5.2 Flowering time under non-optimal conditions

Motifs within promoters exhibit a high degree of redundancy, which has been observed in many organisms. In *Drosophila*, for instance, it was shown that so called shadow enhancers confer fitness advantages under non-optimal conditions (Frankel et al., 2010). These sequences confer a similar expression pattern as primary enhancers, but their deletion did not lead to a visible phenotype under standard laboratory conditions, but showed a strong phenotype under extreme temperature conditions (Frankel et al., 2010). As the *gi* complementation lines did not display a clear phenotype under different photoperiods and as GI is proposed to play a role in temperature compensation to the circadian clock (Gould et al., 2006), the set of *gi* complementation lines was grown under different temperature regimes.

The 'Heat Stress' experiments were performed in a greenhouse without where the temperatures ranged between 16 and 38°C degrees, depending on the outside temperature. Under this condition, all genotypes flowered earlier than under the previously described LD conditions. However, no clear difference in flowering time was detected between any of the transgenic *gi* complementation lines and wildtype plants, even though the B2 and the B2 TEE lines all flowered slightly later under this condition (Fig. 5.6).

The 'Cold Stress' experiment was done in a cold chamber at a temperature of 4°C, in combination with high light intensity (approximately 250 µmol $m^{-2}s^{-1}$) under long day conditions (16h/ 8h). Due to the strong radiation of the vapour discharge lamps that were used temperatures at the plant level were well above the surrounding air temperatures, thus resulting in day temperatures of approximately 12°C and night temperatures of approximately 4°C.

Under this 'Cold Stress' condition, the two B2 as well as the two B2TEE lines did flowered later than the respective wt and *GI::GI* control plants (Fig. 5.6). The TEE lines flowered slightly later than wt plants, whereas SAIL 5 plants showed a clear *gi* like phenotype and flowered almost as late as *gi-2* plants under this cold stress condition (Fig.5.6). Interestingly, *35S::GI* plants flowered almost as late as wildtype plants under this condition and were almost indistinguishable in size from wildtype (not shown).

In Summary, the 'Cold Stress' condition seems to enhance the phenotype if *GI* expression is strongly changed or abolished (as in B2, B2TEE, SAIL 5 and *gi-2*), whereas no strong differences in phenotype were observed under the 'Heat Stress' condition.

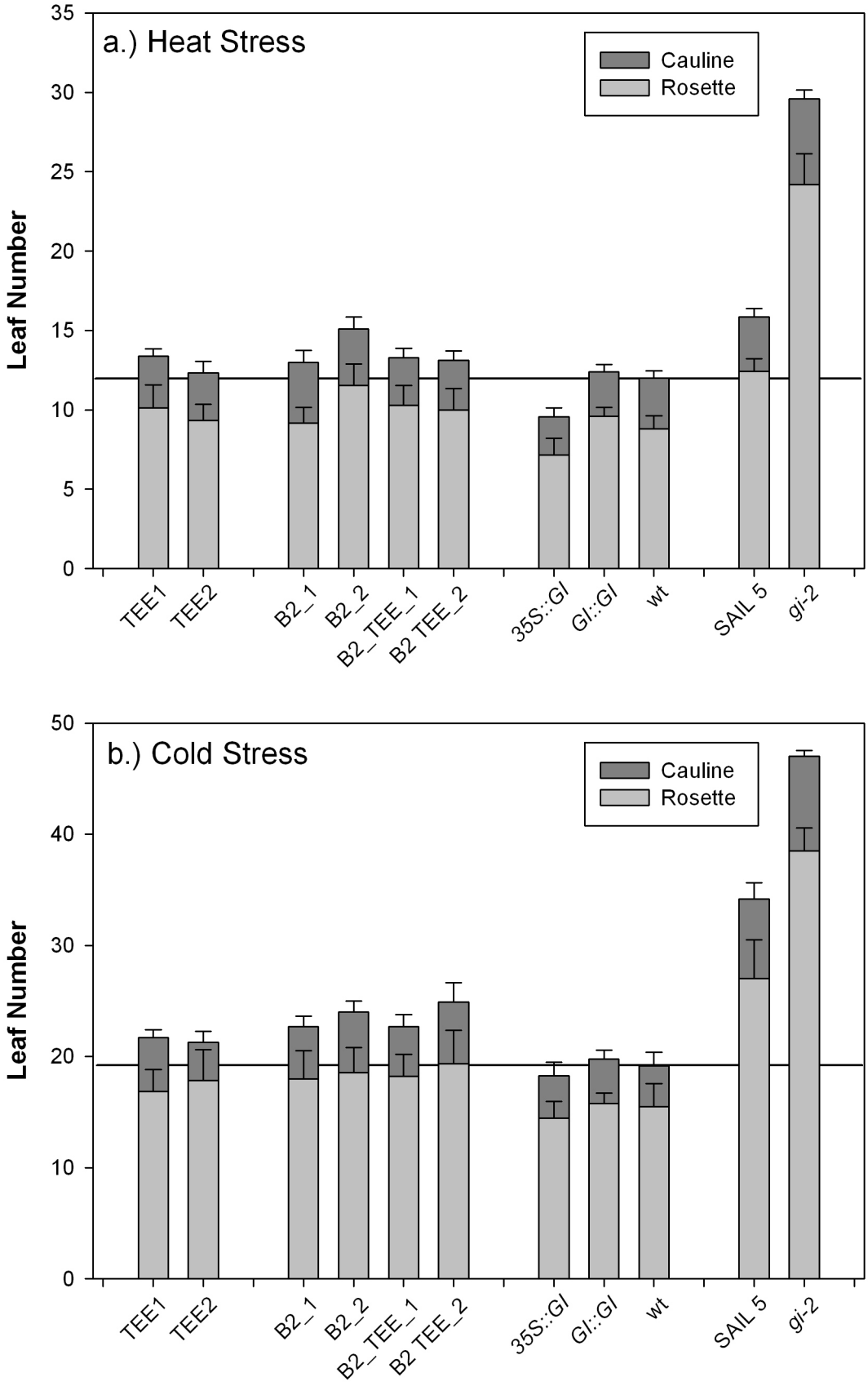


Fig.5. 6 Flowering time of different *gi* complementation lines under different stress conditions a.) Heat stress condition b.) Cold stress condition; horizontal black line indicates wt value.

5.5.3 Hypocotyl growth

One of the well-defined phenotypes of the *gi* mutant is its long hypocotyl under red light conditions (Huq et al., 2000a; Martin-Tryon et al., 2007). In order to analyse whether Block 2 and the EEs contribute to these phyB-mediated processes, *gi* complementation lines were grown under different light regimes and the hypocotyls of 1 week old seedlings were analysed.

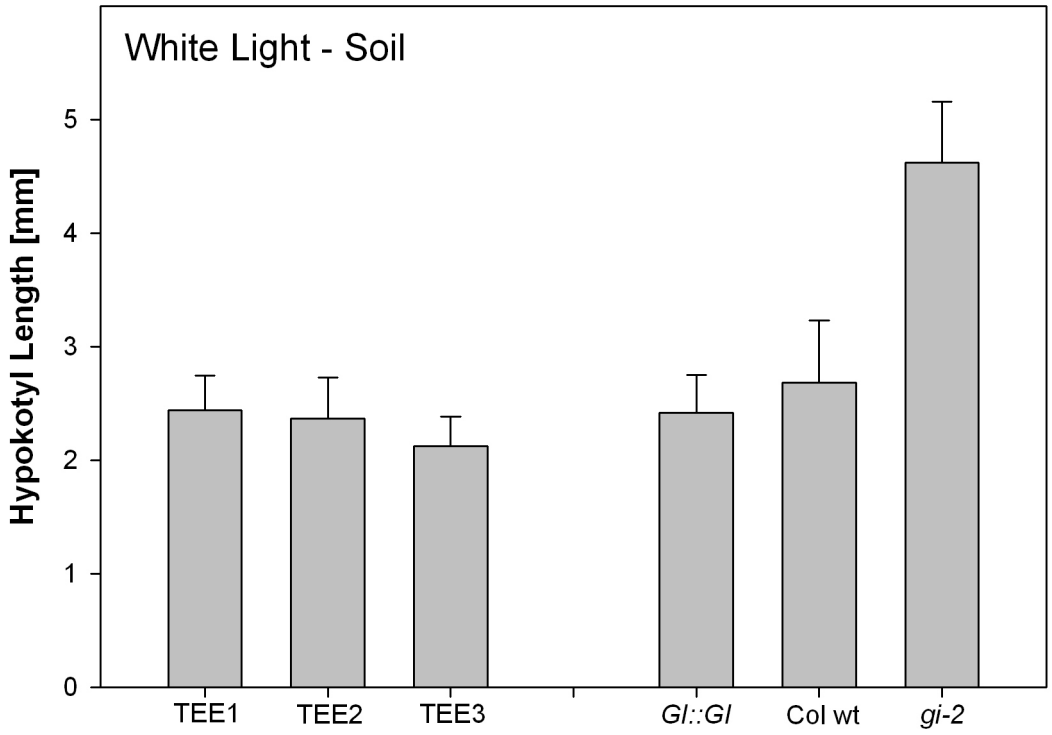


Fig. 5.7 Hypocotyl length of soil-grown TEE lines under white light; Representative data of two independent experiments with same results

First, three independent TEE lines were grown in SDs under cool white light on soil and their hypocotyls compared to controls after 1 week. This analysis revealed no difference between wildtype plants and the transgenic lines, whereas *gi-2* plants had a clearly elongated hypocotyl (Fig. 5.7). Subsequently all four different *gi* complementation lines were grown under red, blue and white light as well as in darkness on agar. As under the previous condition, no difference was detected between transgenic lines and wildtype plants under all conditions (Fig. 5.8). In accordance with previous publications, the hypocotyl of *gi-2* plants was elongated under red and under white light (Fig.5.8).

In summary, no hypocotyl phenotype was discovered for any of the *gi* complementation lines under any of the different light regimes, demonstrating that Block 2 is sufficient to mediate the short hypocotyl phenotype of wildtype plants and that EEs play no role in mediating this response.

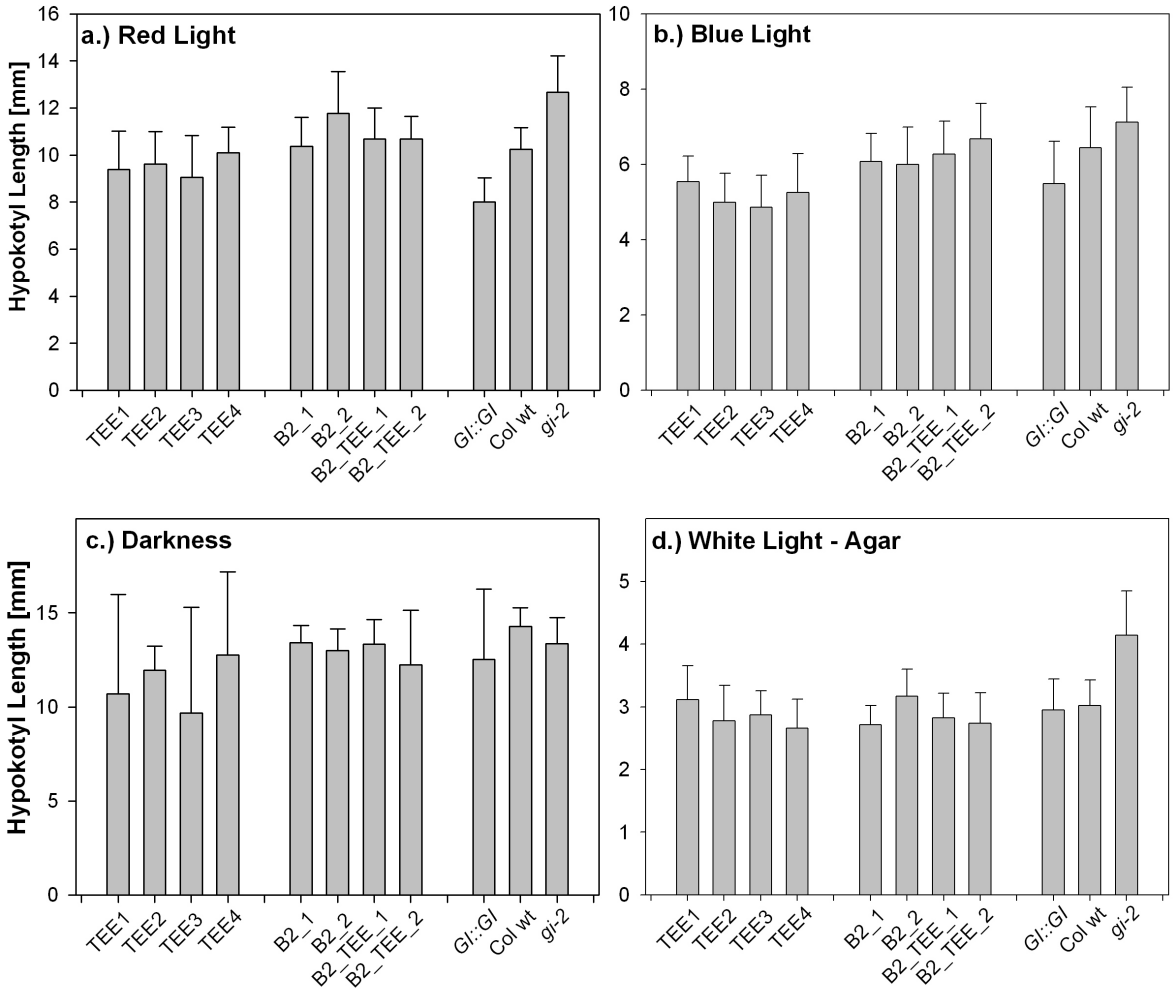


Fig. 5.8 Hypocotyl length of different *gi* complementation lines under different light conditions a.) red light b.) blue light c.) darkness d.) white light

5.5.4 Tolerance to freezing stress

GI transcription is up-regulated under cold temperatures in various species (Fowler and Thomashow, 2002; Paltiel et al., 2006; Hong et al., 2010) and Arabidopsis *gi* mutants are more sensitive to freezing temperatures compared to wildtype plants (Cao et al., 2005). Moreover, EEs play a crucial role in upregulating transcription of the *COL1* and *COR27* genes in response to cold (Mikkelsen and Thomashow, 2009). Finally, the bioinformatic analysis in chapter 3 revealed a strong over-representation for cold-stress related GO terms within EE-enriched promoters. Collectively this suggests that the EEs within the *GI* promoter might play a role in regulating the cold-induced up-regulation of *GI*.

The *gi-3* mutant was described to be more susceptible to freezing stress compared to wt plants (Cao et al., 2005). In order to confirm these results, freezing stress tolerance of *gi* mutants in the Col (*gi-2*) and Ler (*gi-3*) backgrounds was compared to the respective overexpressor and wildtype plants. This was done both with and without an acclimation period before the cold stress treatment.

Therefore, plants were grown 3 weeks in long day conditions and then exposed to -9°C for 5h. Acclimated plants were placed to 4°C 1d before the cold treatment. After a recovery time of 5 days survival of all plants was scored.

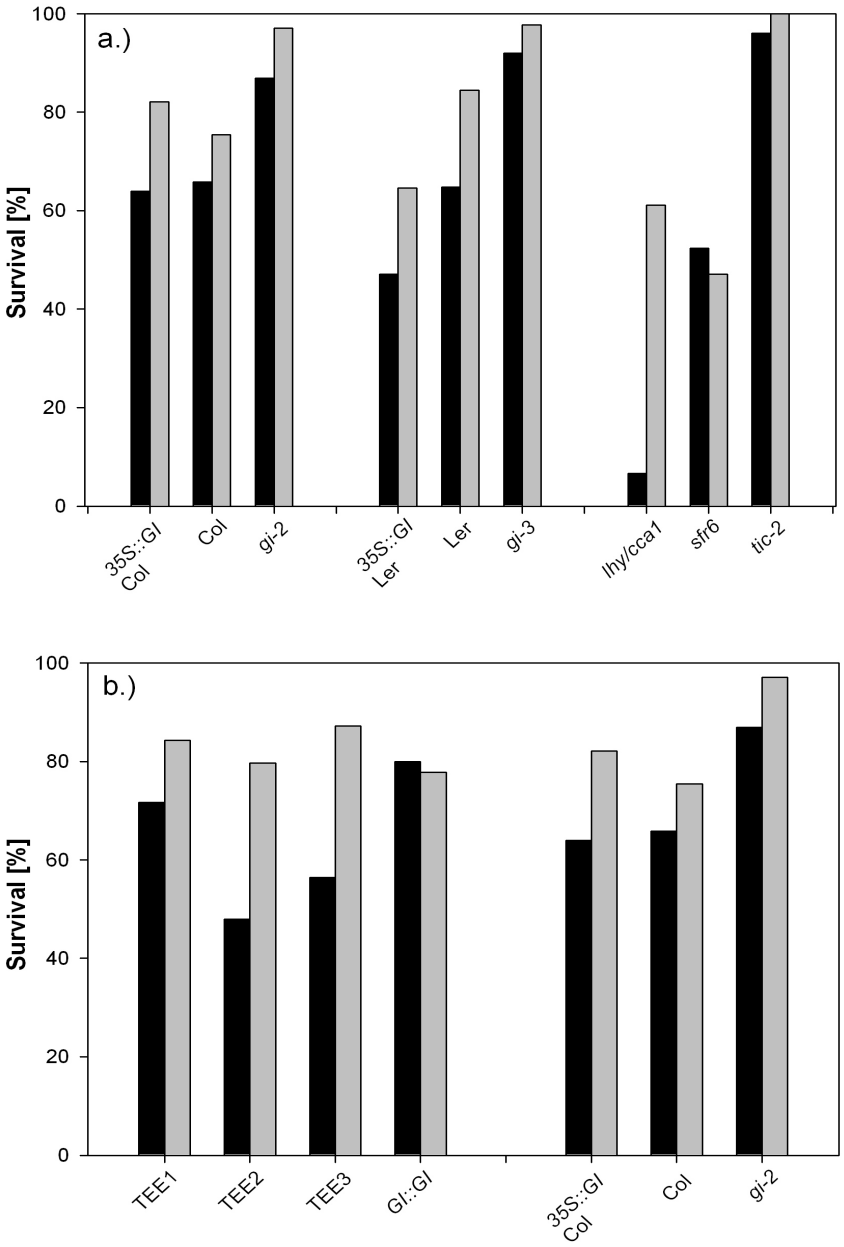


Fig.5. 9 Tolerance to freezing stress of different Arabidopsis genotypes.
Black bars = without acclimation; grey bars = with 1d acclimation at 4°C; Representative data of two independent experiments with similar results.

This analysis revealed that for both backgrounds *gi* mutants were more tolerant to the freezing treatment than the wildtype plants (Fig. 5.9). In the Ler background, *35S::GI* plants were more sensitive to the freezing treatment, whereas there was no difference between over-expressor and wildtype in the Col background. These data show that *gi* mutants are more tolerant to freezing stress than wildtype plants, which is the opposite of the published results (Cao et al., 2005). In all these lines, acclimation increases freezing tolerance, indicating that acclimation is not affected by loss or miss-expression of GI in *gi* mutants or *gi* complementation lines.

As additional controls, more genotypes were exposed to the same freezing treatment. For the *lhy/cca1* and the *sfr6* mutants, a decreased tolerance to freezing temperatures has been described before. TIC is implicated to play a role in many different metabolic and stress-related processes, probably also in freezing tolerance (Sanchez et al., unpublished). The *lhy/cca1* double mutants were more sensitive to freezing, especially when not being acclimated. *sfr6* mutants were slightly less tolerant than the wildtype regardless of an acclimation period, whereas the *tic-2* mutant showed enhanced tolerance to the freezing treatment.

Subsequently, the TEE, the B2 and the B2 TEE transgenic plants were tested for their tolerance to a freezing treatment. Whereas B2 and TEE plants had a similar freezing tolerance as wt plants, B2 TEE and SAIL5 plants were more tolerant to the freezing stress than wt plants in this experiment (Fig. 5.10). This suggests that the B2TEE construct can not fully complement the effect of the *gi* mutation on enhanced freezing tolerance.

Taken together, this analysis shows that *gi* mutants are more tolerant to freezing stress conditions as wt plants. Moreover this analysis suggests that B2 TEE and SAIL_5 plants have an intermediate freezing tolerance phenotype between wt and *gi-2*, suggesting that the EEs are involved in conferring sensitivity to freezing.

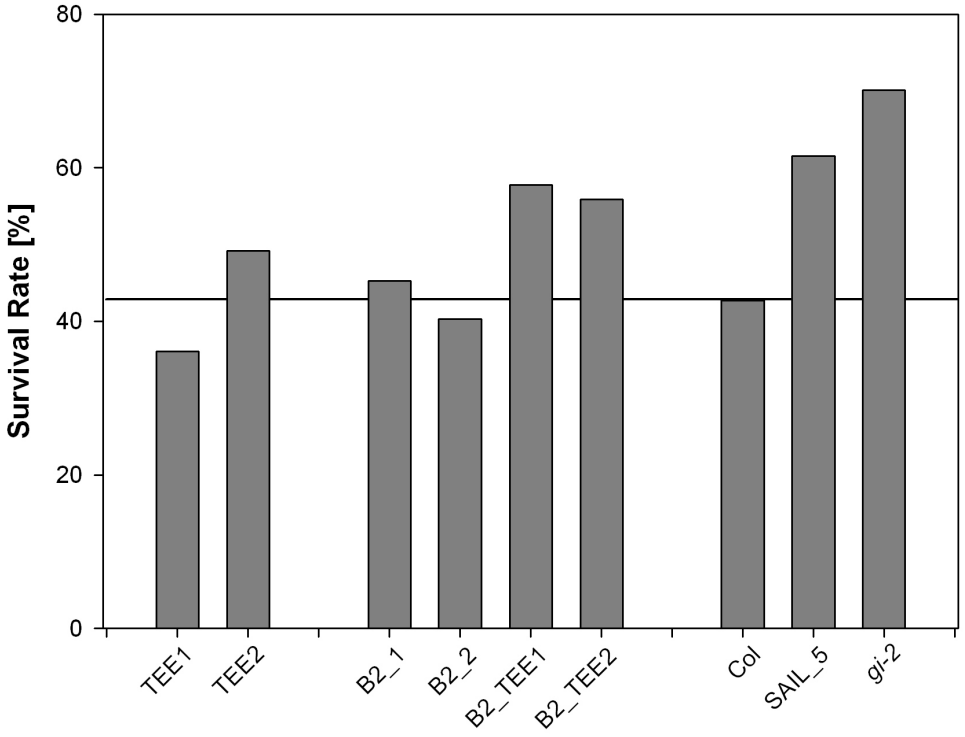


Fig. 5.10 Tolerance to freezing stress of different Arabidopsis genotypes;
Horizontal black line indicates wt value; Representative data of two independent experiments with similar results

5.5.5 Summary of gi complementation lines

Taken together, 4 different *gi* complementation constructs were created and a number of *gi* specific traits were analysed in these plants. Under standard laboratory conditions, none of the *gi* complementation lines showed a visible phenotype for flowering time or for the length of the hypocotyl compared to wt plants. By contrast under cold stress conditions, differences were observed in flowering time and tolerance to freezing stress between wt plants and the *gi* complementation lines.

5.6 Discussion

In order to study directly the contributions of *cis*-regulatory elements within the *GI* promoter on the plant phenotype, several experimental approaches were taken.

Arabidopsis lines harbouring a T-DNA insertion within the *GI* promoter only displayed a weak flowering phenotype compared to *gi* mutants. SAIL 5, a line that harbours the T-DNA insertion directly downstream of conserved Block 2, displayed the strongest phenotype. This indicates that the function of this block might be disrupted in these plants and therefore highlights the importance of conserved Block 2 in the role of GI in flowering time.
The other T-DNA lines were from the SALK collection and therefore it cannot be completely excluded that differences between the transgenes in the T-DNA collections of SAIL and SALK could influence their effect on *GI* expression. Measuring the *GI* mRNA abundance in SAIL 5 and the other SALK T-DNA insertion lines will show if *GI* expression is impaired in these plants.

In a comparable study (Adrian et al., unpublished) T-DNA insertions within the promoter of *FT* did not show any obvious flowering phenotype. Together this indicates that the insertion of a T-DNA does not necessarily disrupt promoter function, either because the upstream DNA is still functional despite the insertion or that at least motifs within the promoters of *FT* and *GI* are redundant so that they can compensate for the effect of an inserted T-DNA.

The analysis of natural variation in 42 Arabidopsis accessions did not identify SNPs within EEs in those accessions, indicating that there is little intra-species variatiation in these motifs. This is consistant with their strong conservation between Brassicaceae as shown in chapter 2. By contrast, SNPs were detected within the ABRELs of some *Arabidopsis* accessions and within EE2 of *Capsella rubella.* This might suggest that that this ABREL can confer a differential expression of *GI*. In *C. rubella* a connection between variation of EEs within the *GI* promoter and GI mRNA expression has not yet been tested. Interestingly the different accessions from geographically relatively close locations (approximately 250km between Rome and the Monte Gargano region) showed large differences in flowering time, but their relationship to *GI* expression remains to be tested.

Even tough no SNPs in EEs have been found within the *GI* promoters of 42 Arabidopsis accessions within this study, the approach of searching desired mutations within sequenced strains of different species will be a promising tool in the future. The revolution in sequencing technologies has led to an enormous increase in available whole genome sequences, a process that certainly will accelerate in the future (Heard et al., 2010). This is especially true for the re-sequencing of different accessions of the

same species, as demonstrated by the ongoing 1001 genome initiative for *A. thaliana* (Weigel and Mott, 2009), thus making it possible to screen a multitude of different accessions to find SNPs at a desired position in the future. As long as direct DNA-targeting techniques such as homologous recombination are not available for *A. thaliana*, such an analysis of natural occurring variation within different accessions will provide a great resource for studying the effect of *cis*-regulatory and other point mutations directly in plants.

As the analysis of T-DNA insertion mutants and natural variation within the *GI* promoter only provided subtle insights into functionality, a set of transgenic *gi* complementation lines was generated. This approach allowed to specifically pose questions about the contribution of conserved Block 2 and the EE to *GI* promoter function. Moreover, it directly allowed comparison with the luciferase reporter-lines (chapter 5), which were generated with the same *GI* promoter fragments as fused to luciferase.

Surprisingly, the *gi* mutation was either fully or partially complemented by all 4 transgenic constructs. This demonstrated the importance of conserved Block 2, which was capable of fully complementing the hypocotyl and flowering phenotypes to carry out these functions of *gi-2* under standard laboratory conditions, showing that all necessary transcribed information for *GI* for is encoded within that 700bp region. This is further supported by the SAIL 5 line, where Block 2 is most probably uncoupled from the proximal part of the *GI* promoter, thus causing an intermediate *gi* mutant phenotype in flowering time and for freezing tolerance. However, under cold stress conditions, Block 2 cannot fully rescue the *gi-2* mutation. This indicates that under this condition, other promoter regions are required for the full response. Such an effect has been described previously in other organisms: In *Drosphila*, so called shadow enhancers are necessary for the correct expression pattern of the developmental gene *shavenbaby* (Frankel et al., 2010) under extreme temperature conditions, whereas no effect was observed under standard laboratory conditions (Frankel et al., 2010).
The full complementation of the *gi* mutation by the 2,5kb *GI::GI* construct demonstrates that such shadow enhancers within the *GI* promoter must be located within these 2,5kb. The most likely candidates for shadow enhancers within the *GI* promoter are conserved Blocks 1 and 3. As shown in the previous chapter, these blocks cannot drive luciferase expression on their own, but their high degree of conservation suggests an important role in *GI* regulation. This function could be the propagation of *GI* expression under cold temperatures. Such cold temperature conditions with strong diurnal temperature differences between day and night in combination with a high light intensity is a condition that *Arabidopsis* plants are exposed to regularly in their natural habitat (Wilczek et al., 2009). Therefore it is likely that the shadow enhancers within the *GI* promoter would have a more important function under natural environmental conditions.

Various *gi* mutant alleles display a long hypocotyl phenotype under white and red light, but not under far-red light, thus suggesting a role for *GI* downstream of PHYB (Huq et al., 2000a). Whereas a clearly elongated hypocotyl of *gi-2* mutants was reproduced under white light conditions, all *gi* complementation lines displayed a hypocotyl phenotype similar to wildtype plants. Under different light conditions, only a slightly elongated hypocotyl of *gi-2* was observed, compared to more extreme phenotypes under red (Huq et al., 2000a) and blue (Martin-Tryon et al., 2007) light. This is most probably due to different light intensities in the different studies: whereas approximately 50 µmol $m^{-2}s^{-1}$ of red and blue light were used in the present study, the strongest effects between different *gi* mutants and wildtype plants were reported between 0.1 and 1 µmol $m^{-2}s^{-1}$ (Huq et al., 2000a).However, under none of the light conditions was there an obvious hypocotyl difference between any of the *gi* complementation lines and wildtype plants. This indicates that EEs within the *GI* promoter are not involved in the GI-dependent regulation of hypocotyl growth in *Arabidopsis*. However, the full rescue of the hypocotyl phenotype by the B2 lines suggests that other *cis*-regulatory elements within Block 2 might be involved in this process. Candidate elements would be the two conserved HUD boxes within Block 2. HUD boxes are bHLH-binding *cis*-regulatory elements that have been proposed to play a role in PHYB-mediated responses, thus making them good candidates for mediating hypocotyl growth-related processes within the *GI* promoter (Michael et al., 2008a).

The freezing stress experiments showed that two different *gi* mutant alleles both in the Ler and in the Col background are more tolerant to freezing temperatures and that the overexpression of *GI* – at least in the case of the Ler background – decreases the tolerance to freezing stress. This conclusion is the opposite of what has been published before. Cao et al. found that *gi-3* plants are less resistant to freezing temperatures (Cao et al., 2005). These studies, however, were carried out with 10 day old seedlings on agar plates, compared to the 3 week old plants on soil used to obtain the data shown in this study. Therefore one possibility is that the age of the plants or the growth conditions might be responsible for this observed difference. However, experiments performed in this study with 10 day old seedlings grown on soil showed an enhanced freezing tolerance of *gi-2* as well (data not shown), indicating that the age of plants is not responsible for this discrepancy. I also tested plants grown on agar plates and obtained very inconsistent results that were not reproducible (data not shown) for most genotypes (including wt and *gi-3*). Therefore, in my hands freezing tolerance experiments with agar-grown seedlings are not a sufficient method to determine freezing tolerance in plants and such inconsistencies might provide an explanation for the conclusions by Cao and colleagues.

It was demonstrated that *gi* mutants have an altered carbohydrate metabolism (Eimert et al., 1995). As the carbohydrate status of *Arabidopsis* is crucial for its freezing tolerance (Yano et al., 2005; Nagele et al., 2011), the altered starch degradation pattern in *gi-2* might explain its enhanced freezing tolerance. Moreover, *gi-2* mutants constitutively over-express *COR* genes (Fornara et al. unpublished), which

provides a further explanation for the enhanced freezing tolerance phenotype of *gi* mutants. Analysis of freezing tolerance from other circadian clock mutants gave interesting insights into the cross-talk between the circadian clock and freezing tolerance: *lhy/cca1* double mutants were less tolerant to freezing temperatures in my experiments, a result which has been published recently (Dong et al., 2011). A role for *tic* mutants in freezing tolerance has also been proposed (Alfredo Sanchez, unpublished). As TIC, similar to GI, is involved both in the circadian clock and in metabolism (Sanchez, unpublished), it will be interesting to see if these two proteins act in the same or in different pathways.

The TEE lines and the B2 lines did not display differences in freezing tolerance compared to wt plants, whereas the B2 TEE lines showed an intermediate freezing tolerance phenotype between wt and *gi-2* plants. This indicates that EEs are important for the sensitivity of B2 plants to freezing. Moreover, it demonstrates redundancy within the *GI* promoter, because the mutation of EEs in the context of Block 2 has a strong effect, but this is not true for the full length promoter, suggesting that other elements within the 2,5kb might be capable of compensating for their function. One possible candidate is a CBS located in conserved Block 1. However, the B2 TEE lines are still less resistant to freezing temperatures than *gi-2*, suggesting that further *cis*-regulatory elements within Block 2 are involved in this process. Like the B2 TEE lines, SAIL 5 displays an intermediate freezing tolerance phenotype. This might be caused by the insertion of the T-DNA close to EE3 within the *GI* promoter, which could disrupt the function of this EE and probably of further *cis*-regulatory elements within Block 2.

In summary, different approaches were taken to directly study *GI* promoter function on the plant phenotype. Collectively these approaches demonstrate the importance of conserved Block 2 within the *GI* promoter, as suggested by the phylogenetic analysis (chapter 2) and by the luciferase experiments (chapter 4). Moreover, an enhanced freezing tolerance phenotype of the *gi* mutant that was not previously described was analysed and further assays suggest that this phenotype is partially dependent on the conserved EEs within the *GI* promoter.

General Discussion and Outlook

6.1. Phylogenetic shadowing and genome-wide analysis

The phylogenetic shadowing approach as presented here was successful in identifying conserved *cis*-regulatory elements and modules within the *GI* promoter. As shown by luciferase-reporter lines and *gi* complementation lines, Block 2 promotes patterns of *Luc* expression similar to the *GI* promoter and can complement the *gi* mutation under standard long day conditions. Also conserved *cis*-regulatory elements such as the EEs predicted by the phylogenetic analysis were important in Block 2.

Such an extensive phylogenetic study combined with the streamlined analysis of reporter and complementation constructs as well as genome-wide analysis for identified *cis*-regulatory elements has not been carried out before for a plant promoter. Nevertheless other studies have combined phylogenetic analysis with the use of transgenic reporter or complementation constructs. However, in several of these analyses the phylogenetic analysis was conducted after the experimental work, for instance for the promoters of *LHY* (Spensley et al., 2009) or *FT* (Adrian et al., 2010). Other studies have used specific fragments directly based on phylogenetic analysis, but use only two different species in comparison to *A. thaliana* (Lee et al., 2005).

The phylogenetic shadowing of the *GI* promoter might be further improved by comparison of the *GI* promoters from more distant species in order to provide a better resolution for highly conserved elements. However, various approaches to amplify the *GI* promoters of species from the order Brassicales such as *Cleome spinosa* or *Tropaeolum majus* (Schranz and Mitchell-Olds, 2006) with the degenerated primers employed for other species were not successful. It is not known if the synteny of *GI* and its upstream gene *PAB3* is still conserved in these species, which would be necessary for a PCR-based amplification of the *GI* promoter with these primers to work. The *GI* promoters of more distant species within the Angiosperms, such as *Populus trichocarpa* or *Medicago truncatula* could not be aligned to any *GI* promoter from the Brassicaceae (data not shown).
The genus *Aethionema* represents the most basal Brassicaceae (Franzke et al., 2011) and therefore might provide a good evolutionary distance for further comparative analyses. A recently obtained BAC library of *Aethionema* might facilitate the identification of the orthologous *GI* promoter. Such an *Aethionema* sequence would most probably further decrease the conserved regions within the *GI* promoter and thus might allow better predictions about functional *cis*-regulatory elements.

However, the main potential improvement of future phylogenetic analysis as presented in this thesis will be the reduction of work and time consuming steps. Primarily this tackles the PCR based

amplification of orthologous promoters. Recent and future advances in sequencing entire genomes of different Brassicaceae already provide the necessary sequence information for such an analysis. Analysis of the *PRR9* promoter demonstrates that an informative phylogenetic analysis is possible by using *A. lyrata, C. rubella* and *A. alpina*, three species whose full genome sequences will be available soon.

Moreover, the comparisons of the *A. thaliana* promoters from *FKF1* and *CCR2* to their orthologs from *A. alpina* show that these two species have an appropriate evolutionary distance to compare transcriptional features. The second step that can be simplified is the phylogenetic pipeline itself that involved a number of manual steps as done for the *GI* promoter.

Future tools might automatically identify the orthologous regions within a sequence of interest, extract conserved *cis*-regulatory modules based on pairwise alignments and finally identify potentitial *cis*-regulatory elements based on databases with known elements and based on their conservation.

Such tools would greatly facilitate phylogenetic analyses and make them applicable to understanding transcriptional regulation within a wide range of biological processes in plants.

As for the phylogenetic analysis, the statistical analysis of the co-occurrence of different *cis*-regulatory elements gave some interesting insights into combinatorial *cis*-regulatory networks. However, one general problem of genome-wide analysis is that it provides statistical over-representations and does not give any direct information about the functional significance of elements. A particular difficulty is that *cis*-regulatory elements are relatively short sequences and therefore are expected to occur frequently in genomes by chance, as demonstrated by the background levels in Fig 3.8 and 3.9. One option to overcome this problem is the experimental validation of putative important *cis*-elements (as conducted in this study for the *GI* promoter), but this is time-consuming and therefore not applicable on a genome-wide level.

Phylogenetic shadowing approaches are another powerful tool to predict functionally important sequences. On the level of comparative genomics for transcriptional regulation, the upcoming draft genome sequence of *A.alpina* is of particular interest. The evolutionary distance of this species from *A.thaliana* allows efficient comparison of promoters, which is not only demonstrated here for the promoters of *GI, FKF1, CCR2* and *PRR9,* but also by others for the promoters of *FT* (Adrian et al., 2010)*, CO* (Simon et al., unpublished) or *LFY* (Wagner et al., unpublished). Thus a genome-wide comparison of *A. thaliana* promoters to the respective orthologs from *A.alpina* will provide an enormous amount of information on conserved promoter regions and *cis*-elements that have a high probability of being functional within their respective promoters. Analysis of several EE-enriched promoters within this thesis already demonstrates the power of this method.

Combining genome-wide phylogenetic shadowing together with statistical analysis of *cis*-regulatory elements is likely to greatly reduce the background of non-functional *cis*-elements identified *in silico* and reveal important insights into the transcriptional network of *A. thaliana*. Such methods are

applicable for clock- and stress-mediated transcriptional responses, but also any other transcriptional network such as hormonal pathways or the transcriptional response to pathogen stress. Therefore extensive phylogenetic and statistical analysis of such networks will contribute to cracking the general transcriptional code in plants.

Understanding this *cis*-regulatory code is key for understanding transcriptional regulation of general biological processes not only in plants, but all organisms. Already in 1975 differences in transcriptional regulation and not in protein coding genes were proposed to be mainly responsible for the differences between humans and chimpanzees (King and Wilson, 1975). In support of this idea, the sequencing of whole genomes and subsequent extensive comparisons between the human and the mouse or chimpanzee genomes indeed showed surprisingly few differences in protein coding genes (Waterston et al., 2002; Waterston et al., 2005). An elegant study by Wilson et al. provides more direct evidence for this hypothesis. The heterologous expression of human chromosome 21 in mouse liver cells resulted in the same mRNA expression pattern as observed in human liver cells, whereas the orthologous mouse chromosome in the mouse cells was expressed differently (Wilson et al., 2008). Therefore Wilson et al. concluded that not the cellular environment, but the sequence itself mainly determines the expression patterns (Wilson et al., 2008). *Cis*-regulatory changes have also been shown to have profound effects in plants as well. In the promoter of the *Flaveria* PHOSPOENOLPYROVATE CARBOXYLASE (PEPC), for instance, point mutations are responsible for its mesophyll-specific expression pattern and thus determine one of the key features of C3 / C4 photosynthesis (Akyildiz et al., 2007).

6.2 Combinatorial regulation of GI

Multiple studies in different organisms highlighted the importance of *cis*-regulatory modules for the regulation of complex transcriptional patterns (Nguyen and Xu, 1998; Zinzen et al., 2009). The data obtained from the phylogenetic shadowing approach combined with the luciferase-reporter lines and the *gi* complementation constructs all imply that conserved Block 2 represents such a *cis*-regulatory module within the *GI* promoter. Therefore current and ongoing work on the *GI* promoter mainly focuses on Block 2. The phylogenetic shadowing analysis correctly predicted the functionality of the conserved EEs within Block 2, suggesting that other conserved *cis*-elements within Block 2 are likely to be functional as well.

ABRELs in combination with EEs are responsible for the cold-induced up-regulation of *COL1* and *COR27* (Mikkelsen and Thomashow, 2009). Moreover, the statistical analysis of co-occurrence of promoter elements performed here demonstrated that ABRELs occur at high frequency certain

distances from EEs, suggesting a combinatorial interaction between these elements in multiple promoters. As the phylogenetic analysis revealed three absolutely conserved ABRELs within Block 2 upstream of the EEs, this implies a combinatorial action of EEs and ABRELs within Block 2. Freezing stress experiments with *B2TEE::cDNA* lines suggested that these lines are more resistant to freezing temperatures and have an intermediate phenotype between *B2::cDNA* lines and *gi*-2 mutants (Fig.5.10). This implies that the mutation of the EEs in the context of Block 2 is partly responsible for the phenotype of *gi* and other elements are necessary to confer the full response. The most likely candidates for these additional elements are the ABRELs. In order to test these hypotheses the ABRELs are currently being mutagenised and the effects of these mutations will then be studied separately and in concert with the EE mutations.

Further candidates for site-directed mutagenesis within Block 2 are the two HUD Boxes as well as the LBS. HUD Boxes are closely related to G Boxes and are implicated in PHYB-mediated processes (Michael et al., 2008a). Analysis of the luciferase reporter lines demonstrated that Block 2 is still induced by light even in the absence of functional EEs. Moreover, the hypocotyl measurements of *gi* complementation lines demonstrate that all constructs, including the *B2TEE::cDNA* lines, fully rescue the *gi* mutant phenotype. Collectively this suggests that other *cis*-regulatory elements than EEs confer wavelength dependent light-sensitivity to the *GI* promoter and the HUD boxes are likely candidates for that.

Finally, the LBS is another promising candidate to test its functionality, as recent reports suggest that LUX in a complex with ELF3 and ELF4 binds the *GI* promoter (Herrero et al., unpublished).

In order to further dissect conserved Block 2, three subfragments A, B and C were cloned, fused to the 105bp fragment of the pnos minimal promoter inserted upstream of the luciferase open reading frame. These three fragments were defined in a way that fragment A contains the 3 ABRELs, fragment B the 3 EEs and fragment C the 2 HUD Boxes (Fig.6.1). Mutagenesis of the EEs and the ABRELs will make it possible to specifically study the contribution of these different *cis*-regulatory elements to *GI* expression on their own and in concert.

Ultimately this set of constructs will allow study of the possible combinatorial interactions between EEs and other *cis*-regulatory elements within Block 2, such as the ABRELs and the HUD-Boxes. Therefore these transgenic plants will provide a good tool to further dissect the *GI* promoter and its features on a molecular level by studying gating responses or induction by different light spectra. Moreover, these mutagenised *GI* promoter fragments might be useful to identify novel upstream regulators of *GI* in the future by Yeast-1-Hybrid (Y1H) assays or EMSA gelshift assays (discussed below).

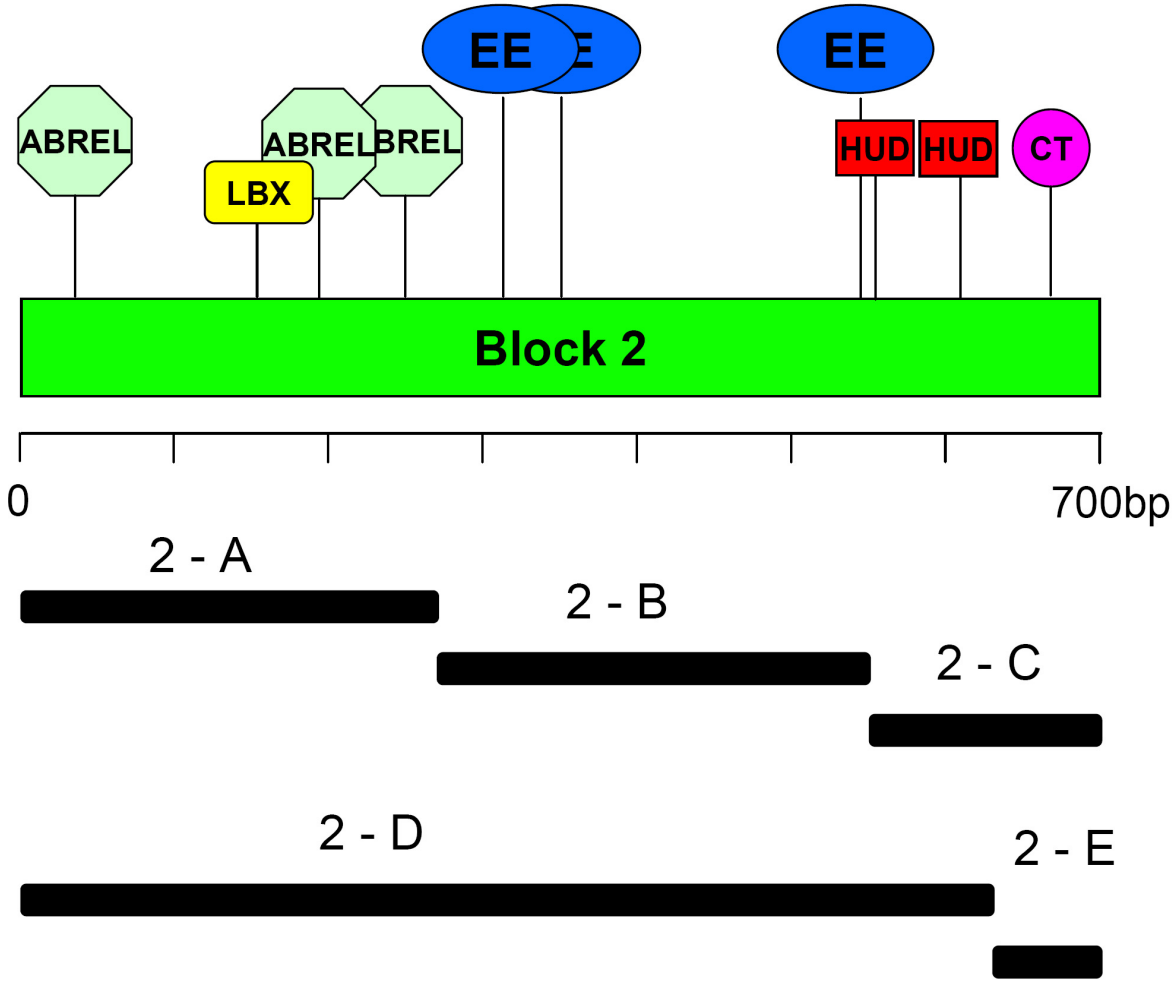


Fig. 6.1 Overview of sub fragments of Block 2
Five sub fragments of Block 2 were fused to the pnos minimal promoter and subsequently cloned upstream of the luciferase open reading frame.

ABREL = ABA Response Element-like; LBX = LUX Binding Site; EE = Evening Element; HUD = Hormone Up at Dawn Box; CT = CT repeat

One classical approach in promoter analysis is the reduction in length from the distal 5'end towards the 3' end proximal to the transcriptional start. In order to further analyse the contribution of Block 2, two different constructs that either include (1,8kb *GI* promoter) or exclude Block 2 (1.1kb *GI* promoter) were constructed and fused to luciferase. Luciferase expression from multiple independent T1 plants harbouring these two different transgenes shows a pronounced decrease in luciferase expression in the lines missing Block 2 (data not shown), thus further supporting the importance of Block 2.

One surprising finding is that Block 2 confers luciferase expression and complements the *gi*- mutant phenotype if fused to a *GI* cDNA without a transcriptional start site provided by a minimal promoter. An explanation might be provided by the phylogenetic shadowing analysis, which revealed a number of conserved core promoter motifs within the last 80bp at the 3' end of Block 2. In order to test the functionality of these elements, a fragment of Block 2 without the last 80bp was fused to luciferase with and without the pnos minimal promoter (Fig. 6.1, 2-D). The resulting transgenic plants will show whether the core promoter elements in these 80bp of Block 2 can drive transcription and are therefore capable of driving the robust expression of the constructs without a minimal promoter.

However, what is the biological function of this striking pattern? Does this region provide another transcriptional start site for *GI*? This seems unlikely, as mapping the transcription start site of *GI* by different groups resulted in identification of the annotated start site closer to the translational start (Terryn et al., 1997; Fowler et al., 1999; Park et al., 1999). However, it would be interesting to map the TSS in the *B2pnos::Luc* and the *B2::Luc* plants in order to reveal where luciferase transcription initiates.

Another possible explanation is provided by the annotation of a protein-coding gene (*At22767*) within Block 2 of the *GI* promoter, suggesting that this potential transcriptional start might be associated with *At22767*. The prediction of this putative 44 amino acid protein without any known function is based on a study by Hanada and colleagues (2007). In this study low level expression measured by tiling arrays was correlated to computationally calculated open reading frames on a genome-wide level (Hanada et al., 2007). This analysis revealed more than 3000 putative, novel protein-coding genes within the *A. thaliana* genome, including *At22767* (Hanada et al., 2007).
However, comparisons of the putative protein coding region in the eight different Brassicaceae revealed that the open reading frame is only present in *A. thaliana*, whereas the entire region is highly conserved in all eight species. This demonstrates that the conservation of this promoter region is evolutionary older as the putative open reading frame. Nevertheless, novel genes can arise from intergenic regions, as shown for the *Polymorphic derived intron-containing* (*Poldi*) gene in mice (Heinen et al., 2009). *At22767* may be such a novel gene in *A. thaliana*. However, a line carrying a T-DNA insertion located within putative *At22767* (which is line SALK 4 from the flowering time experiment in Chapter 5, see Methods) did not show any obvious phenotype under standard laboratory conditions, demonstrating that putative *At22767* at least does not have an important biological function under these conditions.

A third possibility is that the transcription starting in Block 2 has a regulatory function for *GI* transcription itself. It was demonstrated for different organisms that a large part of the non protein coding sequence is transcribed (Kim et al., 2005; David et al., 2006b; Li et al., 2006; Heard et al.,

2010). Such pervasive transcription was shown to play a role in the regulation of transcription of protein-coding genes. In a study by Hirota et al. it was demonstrated that stepwise transcription within the upstream region of the *fructose*-1,6-*bis-phosphatase* ($fbp1^+$) gene in fission yeast is responsible for loosening the chromatin in that region and is necessary for proper transcription of the protein coding gene (Hirota et al., 2008). Similar processes might act at the *GI* promoter. Indeed the experiments conducted with Trich A during this thesis suggest that histone modifications play a regulatory role in determining the precise timing of *GI* expression, even though it is unclear whether this effect is direct or indirect.

Moreover, the co-occurence of conserved sequence and the prediction of nucleosome occupancy (Segal et al., 2006) imply that this pattern is significant. If this prediction is true, conserved regions of the *GI* promoter are densely occupied by nucleosomes. Generally it is assumed that regions that are occupied by histones are less accessible to transcription factors and are therefore less likely to be important for regulation (Bryant et al., 2008; Segal and Widom, 2009). However, it was shown that this is not necessarily the case and regions of high nucleosome occupancy often correlate with important regulatory promoter regions (Bryant et al., 2008; Oren et al., 2010; Tillo et al., 2010).

More research is required to elucidate the epigenetic impact on *GI* transcriptional regulation. ChIP experiments using histone-specific antibodies could reveal the occupation of histones at the *GI* locus and therefore would clarify if the clustering of nucleosomes indeed co-occurs with conserved regions as predicted bioinformatically. Moreover, further experiments with Trich A would give insights into possible contributions of histone modifications of *GI* specific features such as the gated response to a light pulse.

6.3. Upstream regulators of GI

One goal of this project was to elucidate *trans*-acting factors that regulate *GI* specific evening expression. At the beginning of this study, nothing was known about such direct regulators of *GI*. However, based on mathematical modelling (Locke et al., 2006a) and genetic data (Mizoguchi et al., 2002a; Mizoguchi et al., 2005) it was proposed that the Myb transcription factors LHY and CCA1 repress *GI* transcription in the morning, probably due to direct binding to the EEs in the *GI* promoter (Cremer et al., unpublished; (Mizoguchi et al., 2005)).

In order to test this hypothesis, Chromatin Imunoprecipitation (ChIP) experiments were conducted with polyclonal antibodies raised against LHY and CCA1 protein. Transgenic *B2TEE::Luc* plants were used for these ChIP experiments and two sets of primers that could distinguish between the endogenous *GI* promoter with the full EEs and the transgenic *GI* promoter fragment with the mutated EEs were used. However, no reproducible binding of LHY or CCA1 to *GI* promoter regions was

detected during these ChIP experiments (data not shown). Therefore future ChIP experiments will focus on available tagged versions of LHY and CCA1 that are expressed under their native promoters and fully complement the respective mutant plants (Yakir et al., 2011). Recently it was demonstrated by ChIP assays that LHY indeed associates with the *GI* promoter in the region of EE1 and EE2 (Lau et al., 2011). Consistent with LHY protein abundance, this binding was stronger at ZT0 compared to ZT12 (Lau et al., 2011).

Nevertheless different *GI* promoter constructs fused to luciferase in the background of the *lhy/cca1* double mutant used in this study demonstrated that other factors than LHY /CCA regulate *GI* in an EE dependent way. The identity of these factors is unknown. However, in a recent study various EE-binding transcription factors from several different families were identified (Rawat et al., 2011). Most candidates were Myb-like transcription factors such as the REVEILLE proteins, which are closely related to LHY and CCA1. Most of these RVEs show circadian-clock regulated expression in the morning, such as RVE1, RVE3 and RVE8 (Rawat et al., 2011). This makes them possible candidates for regulating *GI* in an EE-dependent manner in the morning.

Different strategies could be applied in order to test possible binding of RVEs to the *GI* promoter. These strategies would include ChIP experiments using tagged RVE lines or *in vitro* EMSA gelshift assays using recombinant RVE protein. In order to test specifically binding to the EEs within the *GI* promoter, the *GI* promoter fragment with mutated EEs would provide a further useful tool.

Despite the importance of EEs within the *GI* promoter, Block 2 contains a number of other highly conserved *cis*-regulatory elements. Therefore the *GI* promoter could also be bound by a number of different unknown transcription factors such as the bZIP and the bHLH family transcription factors predicted to bind to ABREL or HUD motifs.

Yeast-1-Hybrid screens would be a classical approach to identify candidate transcription factors that directly regulate *GI*. The focus in such Y1H assays would be the three subfragments of conserved Block 2, which would provide an ideal basis for such an approach. This analysis would most probably yield a number of different candidates, which would then be further tested by genetic and biochemical experiments.

An alternative strategy to identify novel direct regulators of *GI* would be EMSA gelshift experiments with crude plant extracts. Putative *GI* promoter binding transcription factors from protein-DNA complexes on the EMSA gel could be purified by DNA affinity chromatography and subsequently analysed via mass spectrometry. Performing such EMSA experiments with extracts harvested at different times of the day might identify different transcription factors that bind the *GI* promoter specifically in the morning or in the evening. Moreover, the generated EE and ABREL mutagenised

fragments can be used as controls in order to identify transcription factors that bind specifically to the EEs or the ABRELs within the *GI* promoter. However, one general disadvantage of EMSA gelshift assays is that it the DNA-Protein complex has to survive the electrophoresis step on a gel, which is not always the case (Carey, 2008).

One putative upstream regulator of *GI* suggested by previous work and current experiments was *SFR6* (Knight et al., 1999). The *sfr6* mutant was shown to be late-flowering due to downregulation of *GI* mRNA (Knight et al., 2008). Moreover, it was shown that EEs were the most overrepresented motifs in the promoters of genes downregulated in *sfr6*, thus suggesting SFR6 as an upstream regulator of *GI* (Knight et al., 2008). However, establishing transgenic lines harbouring different *GI* promoter luciferase constructs and creating *sfr6/gi* double mutants was not successful during this thesis. Moreover, the later cloning of SFR6 showed that it is a protein of unknown function without transcription factor activity (Knight et al., 2009). Furthermore the bioinformatics analysis in Chapter 3 has demonstrated an over-representation of EEs in promoters of cold-regulated genes. As *sfr6* regulates cold related pathways, this likely explains the over-representation of EEs in the promoters of its targets.
Taken together, SFR6 regulates photoperiodic flowering upstream of *GI* via unknown mechanisms that likely do not involve direct regulation of *GI*.

Although no direct interaction between a transcription factor and the *GI* promoter could be demonstrated within this study, it nevertheless provided novel insights into the likely nature of such transcription factors. Moreover, material that was generated during this thesis will be helpful to identify novel direct regulators of *GI* in the future.

6.4. Implications of different GI expression patterns

One intriguing conclusion from the comparison of luciferase reporter lines and *gi* complementation lines is that the earlier expression of *GI* from mutated promoters has no effect on flowering time under standard laboratory conditions. However, these conclusions are based on the assumption that the luciferase expression in the transgenic plants reflects the *GI* mRNA and GI protein abundance in the *gi* complementation lines.

All expression data is based on the analyses performed with the luciferase reporter system (Millar et al., 1992; Millar et al., 1995). This system has been proven to be extremely reliable for *in vivo* experiments within different species (Hooper et al., 1990; Contag and Bachmann, 2002) and gives comparable results to RNA abundance measured by RT-PCR in different studies (Mizoguchi et al.,

2005). Moreover, the luciferase expression conferred by *GI::Luc* gives the same expression pattern as *GI* mRNA abundance determined by RT-PCR at 4h intervals (Cremer et al., unpublished)(Mizoguchi et al., 2005).

Nevertheless it cannot be excluded that *GI* mRNA in the transgenic *gi* complementation lines is not precisely the same as luciferase expression in the transgenic reporter lines. One possibility would be that RNA degradation or post transcriptional processes are differentially regulated in the transgenic *gi* complementation lines. In order to test this hypothesis, timecourse RT-PCR experiments will be conducted with the transgenic *gi* complementation lines in order to compare the *GI* mRNA abundance with the luciferase expression.

Moreover, this analysis would also allow quantifying the mRNA abundance of *CO* and *FT* and thus addressing whether the photoperiodic flowering time pathway is normally activated in these plants or not. However, both luciferase expression data and mRNA abundance still do not necessarily reflect GI protein abundance in transgenic plants. The peak in GI protein follows the peak in mRNA of *GI* (David et al., 2006a). However, plants constitutively over-expressing *GI* under the control of a *35S* promoter still maintain the evening abundance of GI protein (David et al., 2006a). This clearly demonstrates that post-transcriptional and / or posttranslational mechanisms control GI protein abundance in *A. thaliana* and thus create one layer of redundancy to maintain an evening-specific abundance of GI. As all tested *gi* complementation constructs display a robust level of *GI* expression (based on luciferase data), it is possible that such a post-translational mechanism also maintains the GI evening abundance in the *gi* complementation lines and therefore no flowering phenotype is detected under standard laboratory conditions.

One way to test this hypothesis would be the detection of GI protein in the *gi*-complementation lines with a specific polyclonal antibody raised against *A. thaliana* GI. However, such an antibody is not available despite many attempts to raise one (Kishore Panigrahi, personal communication). An alternative strategy would be the detection of GI protein via tagged protein versions. However, this would require the generation of transgenic plants with a tagged version of GI which is expressed under the control of the 4 promoter constructs analysed in the background of a *gi* mutant.

In addition to controlling flowering GI is an important component of the plant circadian clock, so that the EEs within the *GI* promoter may also contribute to plant circadian clock function. In the different *gi* complementation lines the clock might be changed due to the mis-expression of *GI*. In order to test this, the different *gi* complementation lines were crossed to *GI::Luc* and *CCR2::Luc* marker lines and luciferase expression of the resulting plants will be compared. Another strategy would be the analysis of mRNA abundance of different clock genes such as *TOC1* or *LHY/CCA1* under LL conditions. As the functions of *GI* are genetically separable between the regulation of flowering and the circadian

clock (Mizoguchi et al., 2005; Martin-Tryon et al., 2007), the mutation of EEs might have a more pronounced effect on the clock than on the flowering traits assayed here.

One problem that is generally poorly addressed in plants is the underlying tissue specificity of transcriptional regulation. Mostly transcriptional regulation is studied in plants by analysis of entire seedlings and therefore the contribution of different tissues or even different cell types are not represented. For *GI* its expression in the vasculature is crucial to fulfil its function in propagating flowering time (Sawa and Kay, 2011). As the plant circadian clock is proposed to be cell autonomous (Dunlap et al., 2003), it is likely that *GI* is required in every plant cell for its clock function. How far different regions of the *GI* promoter regulate tissue-specific expression is unclear. Promoter fragment GUS lines would be one method to study possible tissue specificities within different *GI* promoter regions.

A recent study has shown that the circadian clock works differently between the shoot and the root and implies that this difference is due to non-binding of LHY/CCA1 to the EEs in *TOC1* and *GI* (James et al., 2008). This shows that EE-mediated regulation in the *GI* promoter is at least partially tissue specific.

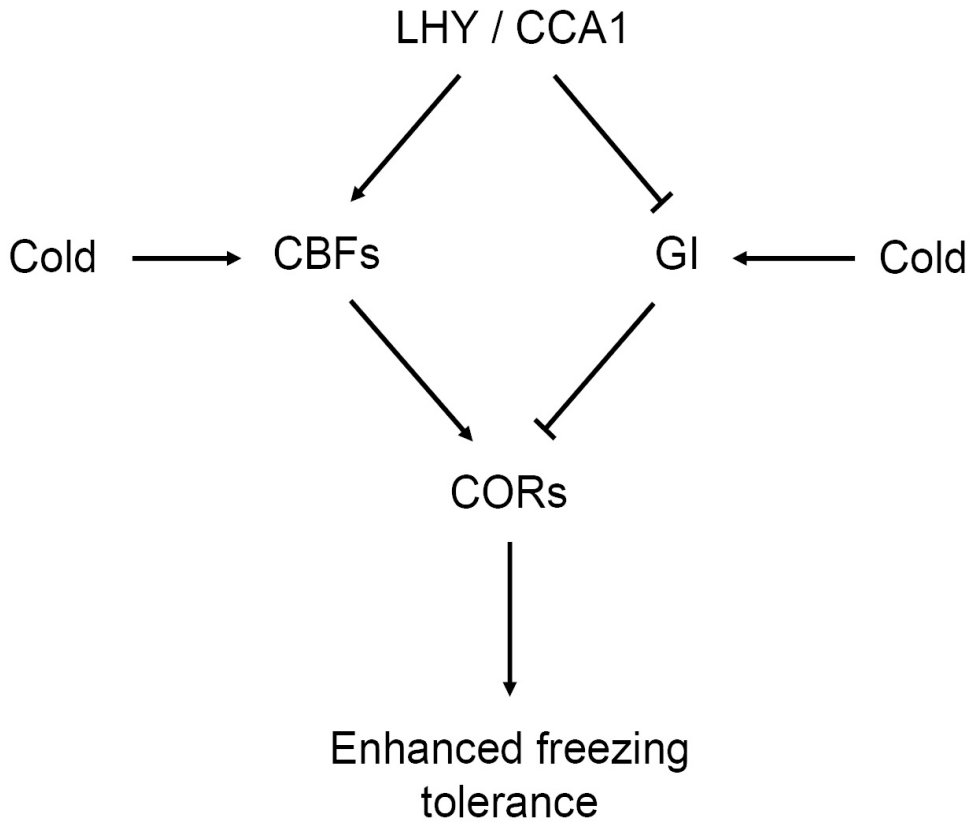


Fig. 6.2 Overview of the proposed network that controls freezing tolerance in *A. thaliana*
Arrows indicate positive regulation; blunt arrows indicate negative regulation.

A role for GI in freezing stress has been proposed before, but the position of *GI* within the previously described network of cold-induced genes remained unclear so far (Cao et al., 2005; Cao et al., 2007). *GI* and the three CBFs share a number of transcriptional features. Both the three CBFs as well as *GI* have a peak in expression in the evening and are strongly transcriptionally induced after the plant is exposed to cold temperatures (Fowler and Thomashow, 2002; Mizoguchi et al., 2002a; Fowler et al., 2005; Dong et al., 2011). Moreover, both have several EEs within their promoters and LHY binds the region of these EEs in both promoters (Dong et al., 2011; Lau et al., 2011). However, the *lhy/cca1* mutation causes different effects on the regulation of both genes. The peak in *GI* mRNA expression is shifted towards the morning in the *lhy/cca1* background, but its expression remains rhythmic (Mizoguchi et al., 2005). In contrast, *CBF1* and *CBF3* expression are completely arrhythmic in a *lhy/cca* mutant, whereas *CBF2* displays impaired rhythmicity (Dong et al., 2011). Therefore LHY/CCA1-mediated effects seem to be opposite between *CBFs* and *GI* and consequently on *COR* mRNA. Moreover, *CBFs* induces the expression of *COR* genes (Gilmour et al., 1998), whereas *GI* represses them (Fornara et al., unpublished).

However, when wildtype plants are subjected to cold temperatures, *COR* genes are clearly upregulated, indicating that the CBF pathway overcomes repression by the *GI* pathway (Baker et al., 1994; Gilmour et al., 1998). However, under natural conditions this apparently contrary system might have a regulatory function that sharpens the correct expression pattern of COR genes. Further experiments are required to untangle this network and to reveal the functions of EEs within the promoters of *GI* and the three CBFs for the response to cold temperatures. This will include the analysis of *GI* mRNA abundance after cold treatments in the different *gi* complementation lines as well as electrolyte leakage assays, which will provide a more quantitative output of freezing tolerance in the different lines.

The data obtained within this thesis demonstrate that results from promoter-reporter lines require careful interpretation. Profound changes in transcriptional expression of *GI* did not cause a visible phenotype under standard growth conditions, suggesting that such effects might be also true for other genes. On the other hand, not detecting a phenotype does not exclude different transcriptional regulation of underlying genes. Therefore it appears to be necessary to study as many different aspects of transcriptional regulation and its impact in order to form a complete picture. This especially includes the exposure of plants to non standard conditions, as phenotypes might be visible only under extreme conditions.

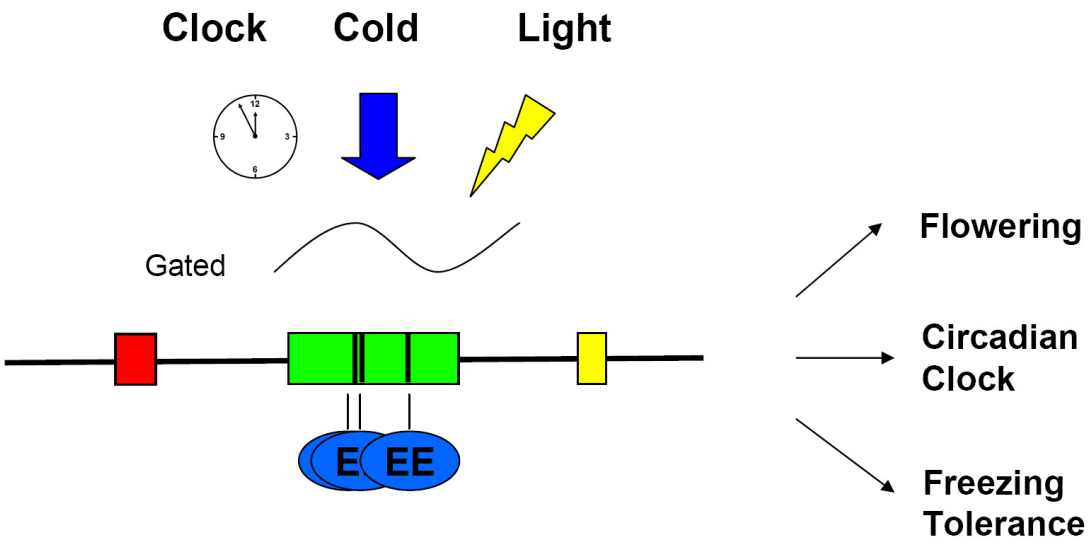


Fig. 6.3 Model of multiple inputs that regulate *GI* via Block 2
The circadian clock, cold temperatures and light regulate the precise timing of *GI* expression via Block 2.

Plants have evolved enormous plasticity that allows them to cope with many different environmental factors. However, this 'external plasticity' seems to be reflected in a number of different, but interconnected regulatory mechanisms. In the case of *GI*, redundancy within the *GI* promoter, post-translational regulation and probably epigenetic regulation together display an 'internal plasticity' that seems to manifest the evening abundance of GI. Similar effects are observed for the flowering time gene *CO* and other regulatory proteins such as *PIFs*. These results emphasise that responsiveness to environmental stimuli is conferred by mechanisms acting on different regulatory layers of gene expression.

Final Conclusions and Perspective

The aim of this study was to elucidate *cis*- and *trans* acting factors that determine the precise expression pattern of *GI* and to elucidate its functions for the plant. Therefore the *GI* promoter was studied intensively by combining phylogenetic, genetic and biochemical methods. The combination of these different approaches demonstrated that a highly conserved 700bp block within the *GI* promoter is crucial for many aspects of *GI* regulation. Moreover, conserved EEs within this block determine the precise timing of *GI* expression dependent on the photoperiod. However, this work also shows that the *GI* promoter is highly redundant and the mis-expression of *GI* can be buffered under many conditions by unknown mechanisms. Having shown the importance of EEs within the *GI* promoter, all EEs were mapped on a genome-wide level and interactions with other circadian-clock related *cis*-regulatory elements were determined. This analysis revealed some striking patterns between such elements and gave insights into the general transcriptional code in plants.

Future work will aim in two different directions. First, conserved Block 2 will be further dissected in order to study combinatorial interactions of different *cis*-regulatory elements within the *GI* promoter. Due to the high conservation of its expression pattern and to the multiple cues the promoter can respond to, the *GI* promoter provides a unique model in plants to further study biological processes such as gating or the induction by light at a molecular level. Moreover, this work is lays the basis to further study known and novel direct regulators of *GI*. The second direction will aim at the genome-wide level in order to further decipher the *cis*-regulatory code in plants. A combination of statistical analysis and the evolutionary conservation of promoter regions will give interesting insights into the general transcriptional code embedded in *cis*-regulatory elements.

Almost 50 years after the identification of the *gi* mutant and intensive studies of its pleiotropic phenotype this study demonstrates that these pleiotropic functions of GI in light signalling, circadian clock, freezing tolerance and the regulation of flowering time are also reflected within its promoter.
Taken together, this work has not only contributed to the understanding of the complex transcriptional regulation of *GI* and its functions in the plant, but also gave novel insights into the regulation of co-expressed genes and the general transcriptional code in plants.

Materials and Methods

8.1 Isolation of GI promoters

All sequences from *Arabidopsis thaliana* and *Arabidopsis lyrata* were obtained from TAIR or Phytozome, respectively. *Capsella rubella* sequences were assembled from raw sequence reads available at NCBI (http://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi).
The sequence reads of *Capsella rubella* were assembled using Newbler (version 2.1) resulting in a genome assembly of 132 Mb (~22x coverage) with an estimated gene space coverage of 98% (Ver Loren van Themaat, unpublished).
All *Arabis alpina* sequence was identified from the *A.alpina* sequencing project at the MPIPZ Cologne (Nordström, Albani, Castaings and Coupland, unpublished). The *Brassica rapa* sequence was obtained from the Brassica Genome Project (brassica.info).
Sequences from *Turritis glabra, Diplotaxis erucoides* and *Sinapis alba* were amplified from genomic DNA using degenerated primers (Bürstel and Cremer, unpublished). Seeds from *Turritis glabra, Diplotaxis erucoides* and *Sinapis alba* were obtained from the IPK Gatersleben.

All *GI* promoter sequences used in this thesis are attached in the Supplementary Information.

8.2 Phylogenetic analysis

All pairwise alignments were done with Shuffle LAGAN using default settings (Brudno et al., 2003b). VISTA Plots were made with the VISTA Browser (Mayor et al., 2000), with a calculation window of 100 base pairs and a consensus identity of 70%. For the multiple VISTA plot, the *A.alpina* sequence was fixed as reference and the number of perfect matches from the other 7 sequences at each position was counted. The data was smoothened by averaging over a sliding window of 100 bp (done by Karl Nordström, unpublished).
Multiple sequence alignments were done with DIALIGN (Morgenstern, 2004) or CLUSTALW (Larkin et al., 2007) using default parameters. Conserved *cis*-regulatory elements were visualized with WEBLOGO (Crooks et al., 2004). Conserved Blocks were defined based on conservation between *A.thaliana* and *A.alpina*. Regions were a stretch of 100bp or longer showed at least 70% conservation was considered as conserved Block. Sequences of conserved Blocks 1-5 from *A.thaliana* are attached in the Supplementary Information.

8.3 Analysis of co-expressed genes

Analysis of genes with at least 3 EEs in their promoters was done with PHASER (Michael et al., 2008b), FatiGO (Al-Shahrour et al., 2004; Al-Shahrour et al., 2005; Al-Shahrour et al., 2007) or data available from the eFP browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi). Further information is provided in the respective figures.

8.4 Co-occurrence of cis-regulatory elements

All calculations were performed by Karl Nordström. The TAIR9 data set containing 3000 bp upstream of each translation start was downloaded from the TAIR website (www.arabidopsis.org) and used for all calculations. Matches for each motif were identified on both strands of all promoter regions. To compare two groups of motifs, the absolute distances between all valid pairs in each promoter were calculated. Furthermore, the total number of motifs for each motif and promoter was counted.

A random background was generated by bootstrapping. The number of motifs for each group was sampled from the true distribution of counts, then the given number of positions was sampled from the true distribution of positions. If the strand was considered, the sampled positions were randomized to either strand with equal probability. This process was repeated 1,000,000 or 2,000,000 times, depending on whether the strand was considered or not. For each random promoter, the absolute distances between all valid pairs were calculated.

For distances below 500bp, counts for bins with the size of two base pairs were generated for both the true distribution and the randomized background. For each bin, the probability is to find at least as many motifs at the given distances as found in the true distribution. This was done with a binomial distribution. The probability of success in a single draw equaled the percentage of distances in the current bin in the randomized background. The number of draws was the total number of distances in the true distribution. The calculations were made in the statistical software R and each probability was adjusted with the internal method p.adjust. Values below 0,05 (EE comparison) or 0,01 (bHLH-bZIP-LBS comparison) were considered significantly, respectively.

8.5 Plasmid constructions

Desired fragments were amplified with primer pairs containing GATEWAY recombination sites using a proofreading polymerase (Expand High Fidelity Tag, Invitrogen, Carlsbad, CA). Primer combinations and fragments are described in the primer tables (Supplementory Materials).

Mutagenic primers were used for site directed mutagenesis. In a first PCR reaction, two fragments were amplified with flanking primers MB38 and MB39 and the respective mutagenic primers. In a second PCR reaction these two fragments were combined with the flanking primers MB38 and MB39, but without the mutagenic primers. Primer combinations and mutated elements are described in the primer tables.

Amplified and mutated promoter fragments were recombined with entry vector pDONR 207 (Invitrogen, Carlsbad, CA) in a BP reaction according to the user's manual, but with 50% of the recommended reaction volume overnight (Invitrogen , Carlsbad, CA). *E. coli* strain DH5 alpha was transformed with 1µl of this BP reaction by heat shock and transformed bacteria were selected on Gentamycin-containing LB media. Plasmids from liquid grown cultures were isolated with the NucleoSpin MiniPrep Kit (Macherey-Nagel, Düren, Deutschland) according to the user manual and inserts were fully sequenced.
Subsequently plasmids were recombined with the desired binary vector in a LR reaction (Invitrogen , Carlsbad, CA) according to the user's manual, but with 50% of the recommended reaction volume overnight. *E. coli* strain DH5 alpha was transformed with 1µl of this LR reaction by heat shock and transformed bacteria were selected on Kanamycin-containing LB media.
Binary vectors from liquid grown cultures were isolated with the NucleoSpin MiniPrep Kit (Macherey-Nagel, Düren, Germany) according to the user manual.

These binary vectors were introduced into *Agrobacterium tumefaciens* strain GV3101 containing the helper plasmid pSOUP via electroporation. Transformed bacteria were selected on LB medium containing the appropriate antibiotics and presence of the plasmid was verified by colony PCR.

8.6 Construction of the pLucGW_pnos vector

A 105bp fragment of the *NOS* promoter was amplified from vector pLucGW (Cremer et al., unpublished) using primers containing Hind 3 restriction sites (MB14 and MB15). The amplified fragment was digested with Hind3 and cloned into the Hind3 restriction site of the pLucGW vector

upstream of the luciferase open reading frame, thus creating pLucGW_pnos (done by Samson Simon, unpublished). The insert was verified by sequencing.

8.7 Generation of transgenic plants

Binary Plasmids were transformed into Arabidopsis plants by the floral dip method (Clough and Bent, 1998) or by a simplified floral dip method (Davis et al., 2009)
T1 plants carrying the plasmid were selected on soil based on their resistance to BASTA (Bayer, Leverkusen, Germany). T2 were identified based on segregation analysis on ½ strength Murashige and Skoog medium supplemented with 1% of sucrose and containing 12µg/ml Phosphinotricin (PPT). Lines that showed a segregation ratio between 1:2 and 1:4 (based on 40-60 seedlings) were considered for further analysis. At least two independent homozygous lines were established in the T3 generation for all constructs that are described in this thesis.

8.8 Flowering Time Experiments

All plants were grown in separate 9*9 cm square pots. If not otherwise stated, plants were grown under cool white light (approx. 80 µE) without additional incandescent lambs and at a temperature of 22°C during the light phase and 18°C or 20°C during the dark phase. The photoperiod is indicated in the figures of each experiment.
Different lines were randomised across the growth space and rosette and cauline leaves were counted from a minimum of 10 plants from each line. Error bars represent Standard Errors.

8.9 Hypocotyl measurements

Seeds were surface sterilised, stratified for 3d at 4°C and seedlings were either grown on soil or on ½ strength Murashige and Skoog medium supplemented with 1% of sucrose for 7d under the indicated light condition. Dark controls were placed for 8h in red light to induce germination. For hypocotyl measurements, at least 20 seedlings from each genotype and each condition were transferred to agar plates. A picture was taken together with a size standard and hypocotyl length was determined with ImageJ (http://rsb.info.nih.gov/ij/).

8.10 Freezing tolerance assays

Seeds were stratified for 3d at 4°C and each 4 seeds were transferred to the corners of 7*7 cm pots. At least 50 plants from each genotype were grown under long day conditions (LD 16) for 21d and then treated with a temperature of -9°C for 5h between ZT 3 and ZT 8 and then transferred back to the intitial conditions. Acclimated plants were transferred to 4°C for 1d before the freezing treatment. Survival of plants was scored 5d after the freezing treatment.

8.11 Isolation of GI promoters

GI promoters from Lip-0 and Dijon-G were amplified from genomic DNA (primers MB 16 / MB 17) and sequenced.
The sequence of the following promoters were obtained from the 1001 genomes initiative (http://www.1001genomes.org)
Bay-0, Bor-4, Br-0, Bur-0, C24, Col-0, Cvi-0, Est-1, Fei-0, Got-7, Ler-1, Lov-5, Nfa-8, Rrs-10, Rrs-7, Sha, Tamm-2, Ts-1, Tsu-1, Van-0

A 1,7kb fragment of the *GI* promoter of the following accessions was obtained from Genbank based on a study on natural variation in flowering (Flowers et al., 2009).
Ag-0 An-1, Bay-0, Br-0, C24, Ct-1, Cvi-0, Edi-1, Ei-2, Ga-0, Gy-0, Kas-2, Ll-0, Mrk-0, Ms-0, Mt-0, Nd-1, Nok-3, Oy-0, Sorbo, Wa-1, Wei-0, Ws-0, Wt-5

All sequences were combined and duplicates were removed, which resulted in a final dataset of 42 different 1,7kb GI promoter accessions. The phylogenetic tree was built with MEGA5 (Tamura et al., 2011).

8.12 T-DNA insertion lines

T-DNA insertion lines that carry a T-DNA insertion were identified by TAIR (www.arabidopsis.org) and homozygous lines were established based on genotyping with primers MB40 – MB59. Lines were assigned names from 1 to 7 according to table below. Location of the T-DNA insertion was determined based on flanking sequence done by TAIR (respective Genbank accession number of these sequences is given in the table below).

ID	SALK Line	Accession No / Sequence
SALK 1	SALK_005476.42.35	ED606671
SALK 2	SALK_005474.30.35	ED606669
SALK 3	SALK_005475.53.85	ED606670
SALK 4	SALK_057774.34.10	CC179053
SAIL 5	SAIL_147_D11	CL470741.1
SALK 6	SALK_114233.29.05	BZ379925
SALK 7	SALK_086534.47.80	BH855163

Tab. 11 Overview about SALK and SAIL lines used for flowering time experiments.

8.13 Luminescence measurements

Seeds were surface sterilised, stratified for 3d at 4°C and grown for 7d under cool white light (approximately 70µE) on ½ strength Murashige and Skoog medium supplemented with 1% of sucrose. Seedlings were then manually transferred into 96 well opaque microtiter plates with each 200µl of ½ strength Murashige and Skoog medium supplemented with 1% of sucrose. 20 µl of 1 mM D-luciferin was added for each plant and luminescence was measured from the next day in a TopCount (PerkinElmer Inc; Waltham, Massachusetts, USA). Each plate was manually transferred every 30Min from cool white light (approximately 70µE) to the TopCount to measure luminescence.

For Trichostatin A experiments, plants were grown for 7d under LD 16 conditions and were then transferred into 96 well opaque microtiter plates with each 200µl of ½ strength Murashige and Skoog medium supplemented with 1% of sucrose and the indicated concentration of Trichostatin A.

For all measurements, expression of 12-36 seedlings was normalised to the average expression during the time measured with Microsoft Excel (Microsoft, Redmond, WA). Error bars represent Standard Errors. Peak times were calculated with the Excel Macro BRASS. Period lengths of free running cycles were estimated from at least 72 h of luminescence measurements starting 12h after transfer into DD using BRASS (Plautz et al., 1997).

Bibliography

Adrian, J., Farrona, S., Reimer, J.J., Albani, M.C., Coupland, G., and Turck, F. (2010). cis-Regulatory Elements and Chromatin State Coordinately Control Temporal and Spatial Expression of FLOWERING LOCUS T in Arabidopsis. Plant Cell **22,** 1425-1440.

Akyildiz, M., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., and Westhoff, P. (2007). Evolution and function of a cis-regulatory module for mesophyll-specific gene expression in the C4 dicot Flaveria trinervia. Plant Cell **19,** 3391-3402.

Al-Shahrour, F., Diaz-Uriarte, R., and Dopazo, J. (2004). FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics **20,** 578-580.

Al-Shahrour, F., Minguez, P., Vaquerizas, J.M., Conde, L., and Dopazo, J. (2005). BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. Nucleic Acids Res **33,** W460-464.

Al-Shahrour, F., Minguez, P., Tarraga, J., Medina, I., Alloza, E., Montaner, D., and Dopazo, J. (2007). FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. Nucleic Acids Res **35,** W91-96.

Alabadi, D., Yanovsky, M.J., Mas, P., Harmer, S.L., and Kay, S.A. (2002). Critical role for CCA1 and LHY in maintaining circadian rhythmicity in Arabidopsis. Current Biology **12,** 757-761.

Alabadi, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Mas, P., and Kay, S.A. (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. Science **293,** 880-883.

Alberts, B., Johnson, A., Walter, P., Lewis, J., Raff, M., and Roberts, K. (2007). Molecular Biology of the Cell. (Taylor & Francis).

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.

Androsthenes. Androsthenes of Thasos: Theophrast, hist.pl. IV 7,8.

Bailey, C.D., Koch, M.A., Mayer, M., Mummenhoff, K., O'Kane, S.L., Jr., Warwick, S.I., Windham, M.D., and Al-Shehbaz, I.A. (2006). Toward a global phylogeny of the Brassicaceae. Mol Biol Evol **23,** 2142-2160.

Baker, S.S., Wilhelm, K.S., and Thomashow, M.F. (1994). THE 5'-REGION OF ARABIDOPSIS-THALIANA COR15A HAS CIS-ACTING ELEMENTS THAT CONFER COLD-REGULATED, DROUGHT-REGULATED AND ABA-REGULATED GENE-EXPRESSION. Plant Molecular Biology **24,** 701-713.

Barbaric, S., Gregory, P.D., Munsterkotter, H., Korica, T., and Horz, W. (1999). Regulation of the yeast PHO5 and PHO8 genes: A network of regulatory proteins, transcription factors and chromatin. Food Technology and Biotechnology **37,** 1-8.

Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., and Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. Proc Natl Acad Sci U S A **107,** 18724-18728.

Bernard, V., Brunaud, V., and Lecharny, A. (2010). TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. Bmc Genomics **11**.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science **302,** 1956-1960.

Birnbaum, K., Jung, J.W., Wang, J.Y., Lambert, G.M., Hirst, J.A., Galbraith, D.W., and Benfey, P.N. (2005). Cell type-specific expression profiting in plants via cell sorting of protoplasts from fluorescent reporter lines. Nat. Methods **2,** 615-619.

Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K.D., Ovcharenko, I., Pachter, L., and Rubin, E.M. (2003). Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science **299,** 1391-1394.

Boss, P.K., Bastow, R.M., Mylne, J.S., and Dean, C. (2004). Multiple pathways in the decision to flower: Enabling, promoting, and resetting. Plant Cell **16,** S18-S31.

Bresinsky, A., Körner, C.C., Kadereit, J.W., and Neuhaus, G. (2008). Strasburger - Lehrbuch der Botanik. (Spektrum der Wissenschaft mbH).

Bretzl, H. (1903). Botanische Forschungen des Alexanderzuges. (B. G. Teubner).

Brock, M.T., Tiffin, P., and Weinig, C. (2007). Sequence diversity and haplotype associations with phenotypic responses to crowding: GIGANTEA affects fruit set in Arabidopsis thaliana. Molecular Ecology **16,** 3050-3062.

Brudno, M., Poliakov, A., Minovitsky, S., Ratnere, I., and Dubchak, I. (2007). Multiple whole genome alignments and novel biomedical applications at the VISTA portal. Nucleic Acids Research **35,** W669-W674.

Brudno, M., Malde, S., Poliakov, A., Do, C.B., Couronne, O., Dubchak, I., and Batzoglou, S. (2003a). Glocal alignment: finding rearrangements during alignment. Bioinformatics **19,** i54-i62.

Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Green, E.D., Sidow, A., and Batzoglou, S. (2003b). LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. **13,** 721-731.

Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Green, E.D., Sidow, A., Batzoglou, S., Progra, N.C.S., and (2003c). LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. **13,** 721-731.

Bryant, G.O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., and Ptashne, M. (2008). Activator Control of Nucleosome Occupancy in Activation and Repression of Transcription. Plos Biology **6,** 2928-2939.

Bünning, E. (1936). Die endogene Tagesrhythmik als Grundlage der photoperiodischen Reaktion. Ber. Deut. Bot. Ges. **54:590-607**.

Cao, S.Q., Ye, M., and Jiang, S.T. (2005). Involvement of GIGANTEA gene in the regulation of the cold stress response in Arabidopsis. Plant Cell Reports **24,** 683-690.

Cao, S.Q., Song, Y.Q., and Su, L. (2007). Freezing sensitivity in the gigantea mutant of Arabidopsis is associated with sugar deficiency. Biologia Plantarum **51,** 359-362.

Carey, M.P., C.L.; Smale, S.T. (2008). Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques. (Cold Spring Harbor Laboratory Press).

Carre, I.A., Kim, J.Y., Song, H.R., and Taylor, B. (2001). Multiple functions of the LHY transcription factor in the Arabidopsis circadian clock. Perspective in Comparative Endocrinology: Unity and Diversity**,** 103-108.

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B.H., Hong, X., Agarwal, M., and Zhu, J.K. (2003). ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes Dev **17,** 1043-1054.

Chory, J., Michael, T.P., Mockler, T.C., Breton, G., McEntee, C., Byer, A., Trout, J.D., Hazen, S.P., Shen, R.K., Priest, H.D., Sullivan, C.M., Givan, S.A., Yanovsky, M., Hong, F.X., and Kay, S.A. (2008). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. Plos Genetics **4**.

Clauss, M.J., and Koch, M.A. (2006). Poorly known relatives of Arabidopsis thaliana. Trends Plant Sci **11,** 449-459.

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

Contag, C.H., and Bachmann, M.H. (2002). Advances in in vivo bioluminescence imaging of gene expression. Annu Rev Biomed Eng **4,** 235-260.

Cook, P.R. (2010). A model for all genomes: the role of transcription factories. J Mol Biol **395,** 1-10.

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.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome Research **14,** 1188-1190.

David, K.M., Armbruster, U., Tama, N., and Putterill, J. (2006a). Arabidopsis GIGANTEA protein is post-transcriptionally regulated by light and dark. Febs Letters **580,** 1193-1197.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006b). A high-resolution map of transcription in the yeast genome. Proceedings of the National Academy of Sciences of the United States of America **103,** 5320-5325.

Davis, A.M., Hall, A., Millar, A.J., Darrah, C., and Davis, S.J. (2009). Protocol: Streamlined sub-protocols for floral-dip transformation and selection of transformants in Arabidopsis thaliana. Plant Methods **5,** 3.

de Meaux, J., and Koornneef, M. (2008). The cause and consequences of natural variation: the genome era takes off! Current Opinion in Plant Biology **11,** 99-102.

Dixon, L.E., Knox, K., Kozma-Bognar, L., Southern, M.M., Pokhilko, A., and Millar, A.J. (2011). Temporal Repression of Core Circadian Genes Is Mediated through EARLY FLOWERING 3 in Arabidopsis. Current Biology **21,** 120-125.

Dodd, A.N., Salathia, N., Hall, A., Kevei, E., Toth, R., Nagy, F., Hibberd, J.M., Millar, A.J., and Webb, A.A.R. (2005). Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. Science **309,** 630-633.

Dong, M.A., Farre, E.M., and Thomashow, M.F. (2011). CIRCADIAN CLOCK-ASSOCIATED 1 and LATE ELONGATED HYPOCOTYL regulate expression of the C-REPEAT BINDING FACTOR (CBF) pathway in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **108,** 7241-7246.

Dunford, R.P., Griffiths, S., Christodoulou, V., and Laurie, D.A. (2005). Characterisation of a barley (Hordeum vulgare L.) homologue of the Arabidopsis flowering time regulator GIGANTEA. Theoretical and Applied Genetics **110,** 925-931.

Dunlap, J.C., Loros, J.J., and Decoursey, P.J. (2003). Chronobiology. (Sinauer Associates;).

Dunn, M.A., White, A.J., Vural, S., and Hughes, M.A. (1998). Identification of promoter elements in a low-temperature-responsive gene (blt4.9) from barley (Hordeum vulgare L.). Plant Molecular Biology **38,** 551-564.

Dutrow, N., Nix, D.A., Holt, D., Milash, B., Dalley, B., Westbroek, E., Parnell, T.J., and Cairns, B.R. (2008). Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping. Nat Genet **40,** 977-986.

Edwards, J., Martin, A.P., Andriunas, F., Offler, C.E., Patrick, J.W., and McCurdy, D.W. (2010a). GIGANTEA is a component of a regulatory pathway determining wall ingrowth deposition in phloem parenchyma transfer cells of Arabidopsis thaliana. Plant J **63,** 651-661.

Edwards, K.D., Akman, O.E., Knox, K., Lumsden, P.J., Thomson, A.W., Brown, P.E., Pokhilko, A., Kozma-Bognar, L., Nagy, F., Rand, D.A., and Millar, A.J. (2010b). Quantitative analysis of regulatory flexibility under changing environmental conditions. Mol Syst Biol **6,** 424.

Eimert, K., Wang, S.M., Lue, W.L., and Chen, J.C. (1995). Monogenic Recessive Mutations Causing Both Late Floral Initiation and Excess Starch Accumulation in Arabidopsis. Plant Cell **7,** 1703-1712.

Farre, E.M., Harmer, S.L., Harmon, F.G., Yanovsky, M.J., and Kay, S.A. (2005). Overlapping and distinct roles of PRR7 and PRR9 in the Arabidopsis circadian clock. Current Biology **15,** 47-54.

Flowers, J.M., Hanzawa, Y., Hall, M.C., Moore, R.C., and Purugganan, M.D. (2009). Population Genomics of the Arabidopsis thaliana Flowering Time Gene Network. Mol. Biol. Evol. **26,** 2475-2486.

Fornara, F., de Montaigu, A., and Coupland, G. (2010). SnapShot: Control of flowering in Arabidopsis. Cell **141,** 550, 550 e551-552.

Fornara, F., Panigrahi, K.C., Gissot, L., Sauerbrunn, N., Ruhl, M., Jarillo, J.A., and Coupland, G. (2009). Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. Dev Cell **17,** 75-86.

Fowler, S., and Thomashow, M.F. (2002). Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell **14,** 1675-1690.

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 Journal **18,** 4679-4688.

Fowler, S.G., Cook, D., and Thomashow, M.E. (2005). Low temperature induction of Arabidopsis CBF1, 2, and 3 is gated by the circadian clock. Plant Physiology **137,** 961-968.

Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature **466,** 490-493.

Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A., and Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. Trends in Plant Science **16,** 108-116.

Fraser, P., Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., Eskiw, C.H., Luo, Y.Q., Wei, C.L., Ruan, Y.J., and Bieker, J.J. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nature Genetics **42,** 53-U71.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., Alves, P., Chateigner, A., Perry, M., Morris, M., Auerbach, R.K., Feng, X., Leng, J., Vielle, A., Niu, W.,

Rhrissorrakrai, K., Agarwal, A., Alexander, R.P., Barber, G., Brdlik, C.M., Brennan, J., Brouillet, J.J., Carr, A., Cheung, M.S., Clawson, H., Contrino, S., Dannenberg, L.O., Dernburg, A.F., Desai, A., Dick, L., Dose, A.C., Du, J., Egelhofer, T., Ercan, S., Euskirchen, G., Ewing, B., Feingold, E.A., Gassmann, R., Good, P.J., Green, P., Gullier, F., Gutwein, M., Guyer, M.S., Habegger, L., Han, T., Henikoff, J.G., Henz, S.R., Hinrichs, A., Holster, H., Hyman, T., Iniguez, A.L., Janette, J., Jensen, M., Kato, M., Kent, W.J., Kephart, E., Khivansara, V., Khurana, E., Kim, J.K., Kolasinska-Zwierz, P., Lai, E.C., Latorre, I., Leahey, A., Lewis, S., Lloyd, P., Lochovsky, L., Lowdon, R.F., Lubling, Y., Lyne, R., MacCoss, M., Mackowiak, S.D., Mangone, M., McKay, S., Mecenas, D., Merrihew, G., Miller, D.M., 3rd, Muroyama, A., Murray, J.I., Ooi, S.L., Pham, H., Phippen, T., Preston, E.A., Rajewsky, N., Ratsch, G., Rosenbaum, H., Rozowsky, J., Rutherford, K., Ruzanov, P., Sarov, M., Sasidharan, R., Sboner, A., Scheid, P., Segal, E., Shin, H., Shou, C., Slack, F.J., Slightam, C., Smith, R., Spencer, W.C., Stinson, E.O., Taing, S., Takasaki, T., Vafeados, D., Voronina, K., Wang, G., Washington, N.L., Whittle, C.M., Wu, B., Yan, K.K., Zeller, G., Zha, Z., Zhong, M., Zhou, X., Ahringer, J., Strome, S., Gunsalus, K.C., Micklem, G., Liu, X.S., Reinke, V., Kim, S.K., Hillier, L.W., Henikoff, S., Piano, F., Snyder, M., Stein, L., Lieb, J.D., and Waterston, R.H. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science **330,** 1775-1787.

Gilmour, S.J., Sebolt, A.M., Salazar, M.P., Everard, J.D., and Thomashow, M.F. (2000). Overexpression of the Arabidopsis CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. Plant Physiology **124,** 1854-1865.

Gilmour, S.J., Zarka, D.G., Stockinger, E.J., Salazar, M.P., Houghton, J.M., and Thomashow, M.F. (1998). Low temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. Plant J **16,** 433-442.

Giuliano, G., Pichersky, E., Malik, V.S., Timko, M.P., Scolnik, P.A., and Cashmore, A.R. (1988). An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. Proc Natl Acad Sci U S A **85,** 7089-7093.

Gong, W., He, K., Covington, M., Dinesh-Kumar, S.P., Snyder, M., Harmer, S.L., Zhu, Y.X., and Deng, X.W. (2008). The development of protein microarrays and their applications in DNA-protein and protein-protein interaction analyses of Arabidopsis transcription factors. Molecular Plant **1,** 27-41.

Gould, P.D., Locke, J.C.W., Larue, C., Southern, M.M., Davis, S.J., Hanano, S., Moyle, R., Milich, R., Putterill, J., Millar, A.J., and Hall, A. (2006). The molecular basis of temperature compensation in the Arabidopsis circadian clock. Plant Cell **18,** 1177-1187.

Hanada, K., Zhang, X., Borevitz, J.O., Li, W.H., and Shiu, S.H. (2007). A large number of novel coding small open reading frames in the intergenic regions of the Arabidopsis thaliana genome are transcribed and/or under purifying selection. Genome Res. **17,** 632-640.

Hanikenne, M., Talke, I.N., Haydon, M.J., Lanz, C., Nolte, A., Motte, P., Kroymann, J., Weigel, D., and Kramer, U. (2008). Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. Nature **453,** 391-395.

Harmer, S.L., and Kay, S.A. (2005). Positive and negative factors confer phase-specific circadian regulation of transcription in Arabidopsis. Plant Cell **17,** 1926-1940.

Harmer, S.L., Hogenesch, J.B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A., and Kay, S.A. (2000a). Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science **290,** 2110-2113.

Harmer, S.L., Hogenesch, L.B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A., and Kay, S.A. (2000b). Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science **290,** 2110-2113.

Hay, A., and Tsiantis, M. (2006). The genetic basis for differences in leaf form between Arabidopsis thaliana and its wild relative Cardamine hirsuta. Nat Genet **38,** 942-947.

Hayama, R., Izawa, T., and Shimamoto, K. (2002). Isolation of rice genes possibly involved in the photoperiodic control of flowering by a fluorescent differential display method. Plant and Cell Physiology **43,** 494-504.

Hayama, R., Yokoi, S., Tamaki, S., Yano, M., and Shimamoto, K. (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature **422,** 719-722.

Heard, E., Tishkoff, S., Todd, J.A., Vidal, M., Wagner, G.P., Wang, J., Weigel, D., and Young, R. (2010). Ten years of genetics and genomics: what have we achieved and where are we heading? Nature Reviews Genetics **11,** 723-733.

Hecht, V., Knowles, C.L., Schoor, J.K.V., Liew, L.C., Jones, S.E., Lambert, M.J.M., and Weller, J.L. (2007). Pea LATE BLOOMER1 is a GIGANTEA ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. Plant Physiology **144,** 648-661.

Heinen, T.J.A.J., Staubach, F., Haming, D., and Tautz, D. (2009). Emergence of a New Gene from an Intergenic Region. Current Biology **19,** 1527-1531.

Hintz, M., Bartholmes, C., Nutt, P., Ziermann, J., Hameister, S., Neuffer, B., and Theissen, G. (2006). Catching a 'hopeful monster': shepherd's purse (Capsella bursa-pastoris) as a model system to study the evolution of flower development. J Exp Bot **57,** 3531-3542.

Hirota, K., Miyoshi, T., Kugou, K., Hoffman, C.S., Shibata, T., and Ohta, K. (2008). Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature **456,** 130-134.

Hogenesch, J.B., and Ueda, H.R. (2011). Understanding systems-level properties: timely stories from the study of clocks. Nature Reviews Genetics **12,** 407-416.

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.

Hong, S.-Y., Lee, S., Seo, P.J., Yang, M.-S., and Park, C.-M. (2010). Identification and molecular characterization of a Brachypodium distachyon GIGANTEA gene: functional conservation in monocot and dicot plants. Plant Molecular Biology **72,** 485-497.

Hooper, C.E., Ansorge, R.E., Browne, H.M., and Tomkins, P. (1990). CCD imaging of luciferase gene expression in single mammalian cells. J Biolumin Chemilumin **5,** 123-130.

Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J.D., Ossowski, S., Ottilar, R.P., Salamov, A.A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M.E., Bergelson, J., Carrington, J.C., Gaut, B.S., Schmutz, J., Mayer, K.F., Van de Peer, Y., Grigoriev, I.V., Nordborg, M., Weigel, D., and Guo, Y.L. (2011). The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat Genet **43,** 476-481.

Huq, E., Tepperman, J.M., and Quail, P.H. (2000a). GIGANTEA is a nuclear protein involved in phytochrome signaling in Arabidopsis (vol 97, pg 9789, 2000). Proceedings of the National Academy of Sciences of the United States of America **97,** 11673-11673.

Huq, E., Tepperman, J.M., and Quail, P.H. (2000b). GIGANTEA is a nuclear protein involved in phytochrome signaling in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **97,** 9789-9794.

Hurka, H., and Neuffer, B. (1997). Evolutionary processes in the genus Capsella (Brassicaceae). Plant Systematics and Evolution **206,** 295-316.

Imaizumi, T. (2010). Arabidopsis circadian clock and photoperiodism: time to think about location. Current Opinion in Plant Biology **13,** 83-89.

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.

Jacob, F., and Monod, J. (1961). GENETIC REGULATORY MECHANISMS IN SYNTHESIS OF PROTEINS. Journal of Molecular Biology **3,** 318-&.

James, A.B., Monreal, J.A., Nimmo, G.A., Kelly, C.L., Herzyk, P., Jenkins, G.I., and Nimmo, H.G. (2008). The circadian clock in Arabidopsis roots is a simplified slave version of the clock in shoots. Science **322,** 1832-1835.

Jung, C., and Müller, A.E. (2009). Flowering time control and applications in plant breeding. Trends in Plant Science **14,** 563-573.

Jung, J.H., Seo, Y.H., Seo, P.J., Reyes, J.L., Yun, J., Chua, N.H., and Park, C.M. (2007). The GIGANTEA-regulated MicroRNA172 mediates photoperiodic flowering independent of CONSTANS in Arabidopsis. Plant Cell **19,** 2736-2748.

Kaul, S., Koo, H.L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L.J., Feldblyum, T., Nierman, W., Benito, M.I., Lin, X.Y., Town, C.D., Venter, J.C., Fraser, C.M., Tabata, S., Nakamura, Y., Kaneko, T., Sato, S., Asamizu, E., Kato, T., Kotani, H., Sasamoto, S., Ecker, J.R., Theologis, A., Federspiel, N.A., Palm, C.J., Osborne, B.I., Shinn, P., Conway, A.B., Vysotskaia, V.S., Dewar, K., Conn, L., Lenz, C.A., Kim, C.J., Hansen, N.F., Liu, S.X., Buehler, E., Altafi, H., Sakano, H., Dunn, P., Lam, B., Pham, P.K., Chao, Q., Nguyen, M., Yu, G.X., Chen, H.M., Southwick, A., Lee, J.M., Miranda, M., Toriumi, M.J., Davis, R.W., Wambutt, R., Murphy, G., Dusterhoft, A., Stiekema, W., Pohl, T., Entian, K.D., Terryn, N., Volckaert, G., Salanoubat, M., Choisne, N., Rieger, M., Ansorge, W., Unseld, M., Fartmann, B., Valle, G., Artiguenave, F., Weissenbach, J., Quetier, F., Wilson, R.K., de la Bastide, M., Sekhon, M., Huang, E., Spiegel, L., Gnoj, L., Pepin, K., Murray, J., Johnson, D., Habermann, K., Dedhia, N., Parnell, L., Preston, R., Hillier, L., Chen, E., Marra, M., Martienssen, R., McCombie, W.R., Mayer, K., White, O., Bevan, M., Lemcke, K., Creasy, T.H., Bielke, C., Haas, B., Haase, D., Maiti, R., Rudd, S., Peterson, J., Schoof, H., Frishman, D., Morgenstern, B., Zaccaria, P., Ermolaeva, M., Pertea, M., Quackenbush, J., Volfovsky, N., Wu, D.Y., Lowe, T.M., Salzberg, S.L., Mewes, H.W., Rounsley, S., Bush, D., Subramaniam, S., Levin, I., Norris, S., Schmidt, R., Acarkan, A., Bancroft, I., Brennicke, A., Eisen, J.A., Bureau, T., Legault, B.A., Le, Q.H., Agrawal, N., Yu, Z., Copenhaver, G.P., Luo, S., Pikaard, C.S., Preuss, D., Paulsen, I.T., Sussman, M., Britt, A.B., Selinger, D.A., Pandey, R., Mount, D.W., Chandler, V.L., Jorgensen, R.A., Pikaard, C., Juergens, G., Meyerowitz, E.M., Dangl, J., Jones, J.D.G., Chen, M., Chory, J., Somerville, M.C., and Ar Gen, I. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408,** 796-815.

Kay, S.A., Helfer, A., Nusinow, D.A., Chow, B.Y., Gehrke, A.R., and Bulyk, M.L. (2011). LUX ARRHYTHMO Encodes a Nighttime Repressor of Circadian Gene Expression in the Arabidopsis Core Clock. Current Biology **21,** 126-133.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423,** 241-254.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J., and Harter, K. (2007). The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. Plant J **50,** 347-363.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C.X., Singer, M.A., Richmond, T.A., Wu, Y.N., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature **436,** 876-880.

Kim, W.Y., Fujiwara, S., Suh, S.S., Kim, J., Kim, Y., Han, L.Q., David, K., Putterill, J., Nam, H.G., and Somers, D.E. (2007). ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. Nature **449,** 356-+.

King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science **188,** 107-116.

Knight, H., Thomson, A.J.W., and McWatters, H.G. (2008). Sensitive to freezing6 integrates cellular and environmental inputs to the plant circadian clock. Plant Physiology **148,** 293-303.

Knight, H., Veale, E.L., Warren, G.J., and Knight, M.R. (1999). The sfr6 mutation in arabidopsis suppresses low-temperature induction of genes dependent on the CRT DRE sequence motif. Plant Cell **11,** 875-886.

Knight, H., Mugford, S.G., Uelker, B., Gao, D., Thorlby, G., and Knight, M.R. (2009). Identification of SFR6, a key component in cold acclimation acting post-translationally on CBF function. Plant Journal **58,** 97-108.

Koch, M., Haubold, B., and Mitchell-Olds, T. (2001). Molecular systematics of the Brassicaceae: evidence from coding plastidic matK and nuclear Chs sequences. American Journal of Botany **88,** 534-544.

Koch, M.A., Kiefer, C., Ehrich, D., Vogel, J., Brochmann, C., and Mummenhoff, K. (2006). Three times out of Asia Minor: the phylogeography of Arabis alpina L. (Brassicaceae). Mol Ecol **15,** 825-839.

Koornneef, M., Hanhart, C.J., and van der Veen, J.H. (1991). A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana. Mol Gen Genet **229,** 57-66.

Kurepa, J., Smalle, J., Van Montagu, M., and Inze, D. (1998). Oxidative stress tolerance and longevity in Arabidopsis: the late-flowering mutant gigantea is tolerant to paraquat. Plant Journal **14,** 759-764.

Lam, F.H., Steger, D.J., and O'Shea, E.K. (2008). Chromatin decouples promoter threshold from dynamic range. Nature **453,** 246-250.

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.

Lau, O.S., Huang, X., Charron, J.B., Lee, J.H., Li, G., and Deng, X.W. (2011). Interaction of Arabidopsis DET1 with CCA1 and LHY in Mediating Transcriptional Repression in the Plant Circadian Clock. Mol Cell **43,** 703-712.

Lee, J.Y., Baum, S.F., Alvarez, J., Patel, A., Chitwood, D.H., and Bowman, J.L. (2005). Activation of CRABS CLAW in the nectaries and carpels of Arabidopsis. Plant Cell **17,** 25-36.

Li, G., Siddiqui, H., Teng, Y., Lin, R., Wan, X.Y., Li, J., Lau, O.S., Ouyang, X., Dai, M., Wan, J., Devlin, P.F., Deng, X.W., and Wang, H. (2011). Coordinated

transcriptional regulation underlying the circadian clock in Arabidopsis. Nat Cell Biol **13,** 616-622.

Li, L., Wang, X.F., Stolc, V., Li, X.Y., Zhang, D.F., Su, N., Tongprasit, W., Li, S.G., Cheng, Z.K., Wang, J., and Deng, X.W. (2006). Genome-wide transcription analyses in rice using tiling microarrays. Nature Genetics **38,** 124-129.

Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K., and Hake, S. (1994). A Knotted1-Like Homeobox Gene in Arabidopsis Is Expressed in the Vegetative Meristem and Dramatically Alters Leaf Morphology When Overexpressed in Transgenic Plants. Plant Cell **6,** 1859-1876.

Locke, J.C., Kozma-Bognar, L., Gould, P.D., Feher, B., Kevei, E., Nagy, F., Turner, M.S., Hall, A., and Millar, A.J. (2006a). Experimental validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana. Mol Syst Biol **2,** 59.

Locke, J.C.W., Southern, M.M., Kozma-Bognar, L., Hibberd, V., Brown, P.E., Turner, M.S., and Millar, A.J. (2005). Extension of a genetic network model by iterative experimentation and mathematical analysis. Molecular Systems Biology.

Locke, J.C.W., Kozma-Bognar, L., Gould, P.D., Feher, B., Kevei, E., Nagy, F., Turner, M.S., Hall, A., and Millar, A.J. (2006b). Experimental validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana. Molecular Systems Biology.

Lu, S.X., Knowles, S.M., Andronis, C., Ong, M.S., and Tobin, E.M. (2009). CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL Function Synergistically in the Circadian Clock of Arabidopsis. Plant Physiology **150,** 834-843.

Martin-Tryon, E.L., Kreps, J.A., and Harmer, S.L. (2007). GIGANTEA acts in blue light signaling and has biochemically separable roles in circadian clock and flowering time regulation. Plant Physiology **143,** 473-486.

Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., and Dubchak, I. (2000). VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics **16,** 1046-1047.

McClung, C.R. (2006). Plant circadian rhythms. Plant Cell **18,** 792-803.

Michael, T.P., and McClung, C.R. (2002). Phase-specific circadian clock regulatory elements in Arabidopsis. Plant Physiol **130,** 627-638.

Michael, T.P., and McClung, C.R. (2003). Enhancer trapping reveals widespread circadian clock transcriptional control in Arabidopsis. Plant Physiology **132,** 629-639.

Michael, T.P., Breton, G., Hazen, S.P., Priest, H., Mockler, T.C., Kay, S.A., and Chory, J. (2008a). A morning-specific phytohormone gene expression program underlying rhythmic plant growth. Plos Biology **6,** 1887-1898.

Michael, T.P., Mockler, T.C., Breton, G., McEntee, C., Byer, A., Trout, J.D., Hazen, S.P., Shen, R.K., Priest, H.D., Sullivan, C.M., Givan, S.A., Yanovsky, M., Hong, F.X., Kay, S.A., and Chory, J. (2008b). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. Plos Genetics **4**.

Mikkelsen, M.D., and Thomashow, M.F. (2009). A role for circadian evening elements in cold-regulated gene expression in Arabidopsis. Plant Journal **60,** 328-339.

Millar, A.J., Short, S.R., Chua, N.H., and Kay, S.A. (1992). A Novel Circadian Phenotype Based on Firefly Luciferase Expression in Transgenic Plants. Plant Cell **4,** 1075-1087.

Millar, A.J., Carre, I.A., Strayer, C.A., Chua, N.H., and Kay, S.A. (1995). Circadian Clock Mutants in Arabidopsis Identified by Luciferase Imaging. Science **267,** 1161-1163.

Mizoguchi, T., Wheatley, K., Hanzawa, Y., Wright, L., Mizoguchi, M., Song, H.R., Carre, I.A., and Coupland, G. (2002a). LHY and CCA1 are partially redundant

genes required to maintain circadian rhythms in Arabidopsis. Developmental Cell **2,** 629-641.

Mizoguchi, T., Wheatley, K., Hanzawa, Y., Wright, L., Mizoguchi, M., Song, H.R., Carre, I.A., and Coupland, G. (2002b). LHY and CCA1 are partially redundant genes required to maintain circadian rhythms in Arabidopsis. Dev Cell **2,** 629-641.

Mizoguchi, T., Wright, L., Fujiwara, S., Cremer, F., Lee, K., Onouchi, H., Mouradov, A., Fowler, S., Kamada, H., Putterill, J., and Coupland, G. (2005). Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in Arabidopsis. Plant Cell **17,** 2255-2270.

Mockler, T.C., Michael, T.P., Priest, H.D., Shen, R., Sullivan, C.M., Givan, S.A., McEntee, C., Kay, S.A., and Chory, J. (2007). The Diurnal project: Diurnal and circadian expression profiling, model-based pattern matching, and promoter analysis. Cold Spring Harbor Symposia on Quantitative Biology **72,** 353-363.

Morgenstern, B. (2004). DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. Nucleic Acids Research **32,** W33-W36.

Mun, J.H., Kwon, S.J., Seol, Y.J., Kim, J.A., Jin, M., Kim, J.S., Lim, M.H., Lee, S.I., Hong, J.K., Park, T.H., Lee, S.C., Kim, B.J., Seo, M.S., Baek, S., Lee, M.J., Shin, J.Y., Hahn, J.H., Hwang, Y.J., Lim, K.B., Park, J.Y., Lee, J., Yang, T.J., Yu, H.J., Choi, I.Y., Choi, B.S., Choi, S.R., Ramchiary, N., Lim, Y.P., Fraser, F., Drou, N., Soumpourou, E., Trick, M., Bancroft, I., Sharpe, A.G., Parkin, I.A., Batley, J., Edwards, D., and Park, B.S. (2010). Sequence and structure of Brassica rapa chromosome A3. Genome Biol **11,** R94.

Nagele, T., Kandel, B.A., Frana, S., Meissner, M., and Heyer, A.G. (2011). A systems biology approach for the analysis of carbohydrate dynamics during acclimation to low temperature in Arabidopsis thaliana. Febs Journal **278,** 506-518.

Negre, N., Brown, C.D., Ma, L.J., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R., Li, Z.R., Ishii, H., Spokony, R.F., Chen, J., Hwang, L., Cheng, C., Auburn, R.P., Davis, M.B., Domanus, M., Shah, P.K., Morrison, C.A., Zieba, J., Suchy, S., Senderowicz, L., Victorsen, A., Bild, N.A., Grundstad, A.J., Hanley, D., MacAlpine, D.M., Mannervik, M., Venken, K., Bellen, H., White, R., Gerstein, M., Russell, S., Grossman, R.L., Ren, B., Posakony, J.W., Kellis, M., and White, K.P. (2011). A cis-regulatory map of the Drosophila genome. Nature **471,** 527-531.

Nelson, C.E., Hersh, B.M., and Carroll, S.B. (2004). The regulatory content of intergenic DNA shapes genome architecture. Genome Biology **5**.

Nelson, D.C., Lasswell, J., Rogg, L.E., Cohen, M.A., and Bartel, B. (2000). FKF1, a clock-controlled gene that regulates the transition to flowering in Arabidopsis. Cell **101,** 331-340.

Nguyen, H.T., and Xu, X. (1998). Drosophila mef2 expression during mesoderm development is controlled by a complex array of cis-acting regulatory modules. Dev Biol **204,** 550-566.

O'Neill, J.S., van Ooijen, G., Le Bihan, T., and Millar, A.J. (2011). Circadian Clock Parameter Measurement: Characterization of Clock Transcription Factors Using Surface Plasmon Resonance. Journal of Biological Rhythms **26,** 91-98.

Oliverio, K.A., Crepy, M., Martin-Tryon, E.L., Milich, R., Harmer, S.L., Putterill, J., Yanovsky, M.J., and Casal, J.J. (2007). GIGANTEA regulates phytochrome A-mediated photomorphogenesis independently of its role in the circadian clock. Plant Physiology **144,** 495-502.

Onai, K., Okamoto, K., Nishimoto, H., Morioka, C., Hirano, M., Kami-ike, N., and Ishiura, M. (2004). Large-scale screening of Arabidopsis circadian clock mutants by

a high-throughput real-time bioluminescence monitoring system. Plant Journal **40,** 1-11.

Oren, M., Nili, E.L., Field, Y., Lubling, Y., Widom, J., and Segal, E. (2010). p53 binds preferentially to genomic regions with high DNA-encoded nucleosome occupancy. Genome Res. **20,** 1361-1368.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., and Fraser, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet **36,** 1065-1071.

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

Park, D.H., Somers, D.E., Kim, Y.S., Choy, Y.H., Lim, H.K., Soh, M.S., Kim, H.J., Kay, S.A., and Nam, H.G. (1999). Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. Science **285,** 1579-1582.

Pauli, S., Rothnie, H.M., Gang, C., He, X.Y., and Hohn, T. (2004). The cauliflower mosaic virus 35S promoter extends into the transcribed region. Journal of Virology **78,** 12120-12128.

Perales, M., and Mas, P. (2007). A functional link between rhythmic changes in chromatin structure and the Arabidopsis biological clock. Plant Cell **19,** 2111-2123.

Pikaard, C.S., and Chang, S. (2005). Transcript profiling in Arabidopsis reveals complex responses to global inhibition of DNA methylation and histone deacetylation. Journal of Biological Chemistry **280,** 796-804.

Pittendrigh, C. (1966). CIRCADIAN OSCILLATION IN DROSOPHILA PSEUDOOBSCURA PUPAE - A MODEL FOR PHOTOPERIODIC CLOCK. Zeitschrift Fur Pflanzenphysiologie **54,** 275-&.

Plautz, J.D., Straume, M., Stanewsky, R., Jamison, C.F., Brandes, C., Dowse, H.B., Hall, J.C., and Kay, S.A. (1997). Quantitative analysis of Drosophila period gene transcription in living animals. J Biol Rhythms **12,** 204-217.

Pruneda-Paz, J.L. (2009). A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock (vol 325, pg 1481, 2009). Science **326,** 366-366.

Puente, P., Wei, N., and Deng, X.W. (1996). Combinatorial interplay of promoter elements constitutes the minimal determinants for light and developmental control of gene expression in Arabidopsis. EMBO J **15,** 3732-3743.

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.

Quail, P.H. (2000). Phytochrome-interacting factors. Semin Cell Dev Biol **11,** 457-466.

Quail, P.H., Boylan, M.T., Parks, B.M., Short, T.W., Xu, Y., and Wagner, D. (1995). Phytochromes: photosensory perception and signal transduction. Science **268,** 675-680.

Rawat, R., Takahashi, N., Hsu, P.Y., Jones, M.A., Schwartz, J., Salemi, M.R., Phinney, B.S., and Harmer, S.L. (2011). REVEILLE8 and PSEUDO-REPONSE REGULATOR5 form a negative feedback loop within the Arabidopsis circadian clock. PLoS Genet **7,** e1001350.

Rawat, R., Schwartz, J., Jones, M.A., Sairanen, I., Cheng, Y., Andersson, C.R., Zhao, Y., Ljung, K., and Harmer, S.L. (2009). REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. Proc Natl Acad Sci U S A **106,** 16883-16888.

Redei, G.P. (1962). Supervital Mutants of Arabidopsis. Genetics **47,** 443-460.

Sawa, M., and Kay, S.A. (2011). GIGANTEA directly activates Flowering Locus T in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America **108,** 11698-11703.

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

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M., and Wisman, E. (2001). Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis. Plant Cell **13,** 113-123.

Schranz, M.E., and Mitchell-Olds, T. (2006). Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. Plant Cell **18,** 1152-1165.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends in Genetics **25,** 335-343.

Segal, E., Fondufe-Mittendorf, Y., Chen, L.Y., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P.Z., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature **442,** 772-778.

Shirsat, A., Wilford, N., Croy, R., and Boulter, D. (1989). Sequences Responsible for the Tissue Specific Promoter Activity of a Pea Legumin Gene in Tobacco. Molecular & General Genetics **215,** 326-331.

Simpson, S.D., Nakashima, K., Narusaka, Y., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Two different novel cis-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. Plant J **33,** 259-270.

Solano, R., Nieto, C., Avila, J., Canas, L., Diaz, I., and Paz-Ares, J. (1995). Dual DNA binding specificity of a petal epidermis-specific MYB transcription factor (MYB.Ph3) from Petunia hybrida. EMBO J **14,** 1773-1784.

Spensley, M., Kim, J.Y., Picot, E., Reid, J., Ott, S., Helliwell, C., and Carre, I.A. (2009). Evolutionarily conserved regulatory motifs in the promoter of the Arabidopsis clock gene LATE ELONGATED HYPOCOTYL. Plant Cell **21,** 2606-2623.

Stadler, R., and Sauer, N. (1996). The Arabidopsis thaliana AtSUC2 gene is specifically expressed in companion cells. Bot. Acta **109,** 299-306.

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.

Suzuki, M., Ketterling, M.G., and McCarty, D.R. (2005). Quantitative Statistical Analysis of cis-Regulatory Sequences in ABA/VP1- and CBF/DREB1-Regulated Genes of Arabidopsis. Plant Physiology **139,** 437-447.

Tagle, D.A., Koop, B.F., Goodman, M., Slightom, J.L., Hess, D.L., and Jones, R.T. (1988). Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. J Mol Biol **203,** 439-455.

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.

Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J.R., Wickramasinghe, P., Lee, M., Fu, Z., and Noma, K. (2010). Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. Nucleic Acids Res **38,** 8164-8177.

Terryn, N., Neyt, P., DeClercq, R., DeKeyser, A., VandenDaele, H., Ardiles, W., Dehais, P., Rouze, P., Gielen, J., Villarroel, R., and VanMontagu, M. (1997). Sequence

analysis of a 24-kb contiguous genomic region at the Arabidopsis thaliana PFL locus on chromosome 1. Febs Letters **416,** 156-160.

Tillo, D., Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Field, Y., Lieb, J.D., Widom, J., Segal, E., and Hughes, T.R. (2010). High Nucleosome Occupancy Is Encoded at Human Regulatory Sequences. PLoS One **5**.

Tompa, M., and Chen, X.Y. (2010). Comparative assessment of methods for aligning multiple genome sequences. Nature Biotechnology **28,** 567-U553.

Tseng, T.S., Salome, P.A., McClung, C.R., and Olszewski, N.E. (2004). SPINDLY and GIGANTEA interact and act in Arabidopsis thaliana pathways involved in light responses, flowering, and rhythms in cotyledon movements. Plant Cell **16,** 1550-1563.

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 de Peer, Y., Fawcett, J.A., Proost, S., Sterck, L., and Vandepoele, K. (2009). The flowering world: a tale of duplications. Trends in Plant Science **14,** 680-688.

Walther, D., Brunnemann, R., and Selbig, J. (2007). The regulatory code for transcriptional response diversity and its relation to genome structural properties in A. thaliana. PLoS Genet **3,** e11.

Wang, R.H., Farrona, S., Vincent, C., Joecker, A., Schoof, H., Turck, F., Alonso-Blanco, C., Coupland, G., and Albani, M.C. (2009). PEP1 regulates perennial flowering in Arabis alpina. Nature **459,** 423-U138.

Waterston, R.H., Mikkelsen, T.S., Hillier, L.W., Eichler, E.E., Zody, M.C., Jaffe, D.B., Yang, S.P., Enard, W., Hellmann, I., Lindblad-Toh, K., Altheide, T.K., Archidiacono, N., Bork, P., Butler, J., Chang, J.L., Cheng, Z., Chinwalla, A.T., deJong, P., Delehaunty, K.D., Fronick, C.C., Fulton, L.L., Gilad, Y., Glusman, G., Gnerre, S., Graves, T.A., Hayakawa, T., Hayden, K.E., Huang, X.Q., Ji, H.K., Kent, W.J., King, M.C., Kulbokas, E.J., Lee, M.K., Liu, G., Lopez-Otin, C., Makova, K.D., Man, O., Mardis, E.R., Mauceli, E., Miner, T.L., Nash, W.E., Nelson, J.O., Paabo, S., Patterson, N.J., Pohl, C.S., Pollard, K.S., Prufer, K., Puente, X.S., Reich, D., Rocchi, M., Rosenbloom, K., Ruvolo, M., Richter, D.J., Schaffner, S.F., Smit, A.F.A., Smith, S.M., Suyama, M., Taylor, J., Torrents, D., Tuzun, E., Varki, A., Velasco, G., Ventura, M., Wallis, J.W., Wendl, M.C., Wilson, R.K., Lander, E.S., and Consortium, C.S.A. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. Nature **437,** 69-87.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S.E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M.R., Brown, D.G., Brown, S.D., Bult, C., Burton, J., Butler, J., Campbell, R.D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A.T., Church, D.M., Clamp, M., Clee, C., Collins, F.S., Cook, L.L., Copley, R.R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K.D., Deri, J., Dermitzakis, E.T., Dewey, C., Dickens, N.J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D.M., Eddy, S.R., Elnitski, L., Emes, R.D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G.A., Flicek, P., Foley, K., Frankel, W.N., Fulton, L.A., Fulton, R.S., Furey, T.S., Gage, D., Gibbs, R.A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T.A., Green, E.D., Gregory, S., Guigo, R., Guyer, M., Hardison, R.C., Haussler, D., Hayashizaki, Y., Hillier, L.W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D.B., Johnson, L.S., Jones, M., Jones, T.A., Joy, A., Kamal, M., Karlsson, E.K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W.J., Kirby, A., Kolbe, D.L., Korf, I., Kucherlapati, R.S.,

Kulbokas, E.J., Kulp, D., Landers, T., Leger, J.P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D.R., Mardis, E.R., Matthews, L., Mauceli, E., Mayer, J.H., McCarthy, M., McCombie, W.R., McLaren, S., McLay, K., McPherson, J.D., Meldrim, J., Meredith, B., Mesirov, J.P., Miller, W., Miner, T.L., Mongin, E., Montgomery, K.T., Morgan, M., Mott, R., Mullikin, J.C., Muzny, D.M., Nash, W.E., Nelson, J.O., Nhan, M.N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M.J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K.H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C.S., Poliakov, A., Ponce, T.C., Ponting, C.P., Potter, S., Quail, M., Reymond, A., Roe, B.A., Roskin, K.M., Rubin, E.M., Rust, A.G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M.S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J.B., Slater, G., Smit, A., Smith, D.R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J.P., Von Niederhausern, A.C., Wade, C.M., Wall, M., Weber, R.J., Weiss, R.B., Wendl, M.C., West, A.P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R.K., Winter, E., Worley, K.C., Wyman, D., Yang, S., Yang, S.P., Zdobnov, E.M., Zody, M.C., and Lander, E.S. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature **420,** 520-562.

Weigel, D., and Mott, R. (2009). The 1001 genomes project for Arabidopsis thaliana. Genome Biol **10,** 107.

Wilczek, A.M., Roe, J.L., Knapp, M.C., Cooper, M.D., Lopez-Gallego, C., Martin, L.J., Muir, C.D., Sim, S., Walker, A., Anderson, J., Egan, J.F., Moyers, B.T., Petipas, R., Giakountis, A., Charbit, E., Coupland, G., Welch, S.M., and Schmitt, J. (2009). Effects of genetic perturbation on seasonal life history plasticity. Science **323,** 930-934.

Wilson, M.D., Barbosa-Morais, N.L., Schmidt, D., Conboy, C.M., Vanes, L., Tybulewicz, V.L.J., Fisher, E.M.C., Tavare, S., and Odom, D.T. (2008). Species-specific transcription in mice carrying human chromosome 21. Science **322,** 434-438.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PLoS One **2,** e718.

Yakir, E., Hilman, D., Kron, I., Hassidim, M., Melamed-Book, N., and Green, R.M. (2009). Posttranslational regulation of CIRCADIAN CLOCK ASSOCIATED1 in the circadian oscillator of Arabidopsis. Plant Physiol **150,** 844-857.

Yakir, E., Hassidim, M., Melamed-Book, N., Hilman, D., Kron, I., and Green, R.M. (2011). Cell Autonomous and Cell-Type Specific Circadian Rhythms in Arabidopsis. Plant J.

Yanagisawa, S., and Schmidt, R.J. (1999). Diversity and similarity among recognition sequences of Dof transcription factors. Plant Journal **17,** 209-214.

Yano, R., Nakamura, M., Yoneyama, T., and Nishida, I. (2005). Starch-related alpha-glucan/water dikinase is involved in the cold-induced development of freezing tolerance in arabidopsis. Plant Physiology **138,** 837-846.

Ziermann, J., Ritz, M.S., Hameister, S., Abel, C., Hoffmann, M.H., Neuffer, B., and Theissen, G. (2009). Floral visitation and reproductive traits of Stamenoid petals, a naturally occurring floral homeotic variant of Capsella bursa-pastoris (Brassicaceae). Planta **230,** 1239-1249.

Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M., and Furlong, E.E. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature **462,** 65-70.

Abbreviations

ABA	absisic acid
ABREL	ABA Response Element-like Element
AG	*AGAMOUS*
AP2	*APETALA2*
B1	Block 1
B2	Block 2
B3	Block 3
bHLH	basic helix loop helix
bp	base pair
bZIP	leucine zipper
CAT3	*CATALASE3*
CBF	*C-REPEAT BINDING FACTOR*
CBS	CCA1 Binding Site
CCA1	*CIRCADIAN CLOCK ASSOCIATED 1*
CCR2	*CLOCK AND COLD REGULATED 2*
CDF	*CYCLING OF DOF FACTOR*
cDNA	complementary DNA
CHE	*CCA1 HIKING EXPEDITION*
ChIP	Chromatin Imunoprecipitation
CO	*CONSTANS*
Col	Columbia
COR	*COLD REGULATED*
CRC	*CRABS CLAW*
CT-BMY	*CHLOROPLAST BETA AMYLASE*
DD	constant darkness
EE	Evening Element
ELF3	*EARLY FLOWERING3*
EMSA	electromobility shift assay
FKF1	*FLAVIN-BINDING; KELCH-DOMAIN; F-BOX PROTEIN 1*
FT	*FLOWERING LOCUS T*

GA	gibberellin
GI	*GIGANTEA*
Go term	gene ontology term
HOT	high occupancy target
HUD Box	Hormone Up at Down Box
ICE1	INDUCER OF CBF EXPRESSION1
LD	Long Day
Ler	Landsberg erecta
LFY	*LEAFY*
LHY	*LATE ELONGATED HYPOCOTYL*
LL	continuous light
LUX	*LUX ARHYTHMO*
ME	Morning Element
mRNA	messanger RNA
NOS	*NOPALINE SYNTHASE*
ORF	open reading frame
PAB3	*POLYADENYLATE BINDING PROTEIN 3*
PBX	Protein-Box
PCR	polymerase chain reaction
PEPC	*PHOSPOENOLPYROVATE CARBOXYLASE*
PHY	*PHYTOCHROME*
PIF3	*PHYTOCHROME INTERACTING FACTOR 3*
Pol II	RNA polymerase II
PRR7	*PSEUDO RESPONSE REGULATOR7*
RVE	*REVEILLE*
SBX	Starch-Box
SD	short day
SEE	short EE
SEX1	*STARCH EXCESS 1*
sfr6	*SENSITIVE TO FREEZING 6*
SNP	single nucleotide polymorphism
SOC1	*SUPRESSOR OF OVEREXPRESSION OF CONSTANS1*
SPY	*SPINDLY*

TBX	Telobox
TCP	TEOSINTE BRANCHED1, CYCLOIDEA and PCF
TEE	triple Evening Element mutation
TIC	*TIME FOR COFFEE*
TOC1	*TIMING OF CAB EXPRESSION* 1
TOE	*TARGET OF EAT1*
TSA	Trichostatin A
TSS	transcriptional start site
UTRs	untranslated region
wt	wildtype
Y1H	Yeast-1-Hybrid
ZTL	*ZEITLUPE*

Primer Tables

Cloning Primers			
ID	Stock Name	Sequence	Comment
MB1	CB1F-FOR	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGTGGGATGATGATGGTTTATATGG	to amplify Conserved Block 1(Trans-start) in Gi-Promoter of A.thaliana, with attB-Site
MB2	CB1R-REV	GGGGACCACTTTGTACAAGAAAGCTGGGTAAAGAATACATACACAGAAGAAGAGATTG	to amplify Conserved Block 1 in Gi-Promoter of A.thaliana, with attB-Site
MB3	CB2F-FOR	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGCCCAGTTACATAAC	to amplify Conserved Block 2 in Gi-Promoter of A.thaliana, with attB-Site
MB4	CB2R-REV	GGGGACCACTTTGTACAAGAAAGCTGGGTATTGGTAAAGAATTTGT	to amplify Conserved Block 2 in Gi-Promoter of A.thaliana, with attB-Site
MB5	CB3F-FOR	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTGATGTCACGGTCCTTCTGT	to amplify Conserved Block 3 in Gi-Promoter of A.thaliana, with attB-Site
MB6	CB3R-REV	GGGGACCACTTTGTACAAGAAAGCTGGGTTAAGTTTATAGTCCAACG	to amplify Conserved Block 3 in Gi-Promoter of A.thaliana, with attB-Site
MB7	174m-R2-80_R_GW	GGGGACCACTTTGTACAAGAAAGCTGGGTATTTAATTTAAGTATATAACATCTCTCATGTGATTG	Primer to amplify Subfragment D of Block2 +GW
MB8	175m-R2-A_R_GW	GGGGACCACTTTGTACAAGAAAGCTGGGTCACATAAACAAGTAGCACAATAAAATGATGAC	Primer to amplify Subfragment A of Block2 +GW
MB9	176m-R2-B_F_GW	GGGGACAAGTTTGTACAAAAAAGCAGGCTGTACAGTAGAGACAAGTGGTAAGATTTT	Primer to amplify Subfragment B of Block2 +GW
MB10	177m-R2-B_R_GW	GGGGACCACTTTGTACAAGAAAGCTGGGTCACAGATATTTTGAGGGAGCTAAG	Primer to amplify Subfragment B of Block2 +GW
MB11	178m-R2-C-F-GW	GGGGACAAGTTTGTACAAAAAAGCAGGCTGACATATGTCACTAGCTTTTTTTTTGTG	Primer to amplify Subfragment C of Block2 +GW
MB12	179m-80-R2-GW	GGGGACAAGTTTGTACAAAAAAGCAGGCTAACTCAATAATACAATCATGACCACC	Primer to amplify Subfragment E of Block2 +GW
MB13	180m-R2-B_R_GW_EEMUT	GGGGACCACTTTGTACAAGAAAGCTGGGTCACAGACACTCTGAGGGGAGCTAAG	Primer to amplify Subfragment B of Block2 +GW, EE3 mutated
MB14	83mc-pnos-101-Hind-F	AGACTCGTAAGCTTTTGCGCGTTCAAAAGTCG	to amplify pnos minimal promoter -101 +4, +Hind 3, + 8bp random
MB15	84mc-pnos-101-Hind-R	ACGAGTCTAAGCTTACTCTAATTGGATACCGAGGGGAA	to amplify pnos minimal promoter -101 +4, +Hind 3, + 8bp random
MB16	63at-at_GI-P_36_FOR	CTACCACCGACTGAAATTTGC	to amplify 3.6kb of Arabidopsis GI _Promoter, for Lip-0, Dijon-G
MB17	64at-at_GI-P_36_REV	CATCCAGGAACCGAAACTAAA	to amplify 3.6kb of Arabidopsis GI _Promoter, for Lip-0, Dijon-G
MB18	65at-GI-P_36_FOR_GW	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTACCACCGACTGAAATTTGC	to amplify 3.6kb of Arabidopsis GI _Promoter, for Lip-0, Dijon-G, +GW site
MB19	66at-GI-P_36_REV_GW	GGGGACCACTTTGTACAAGAAAGCTGGGTCATCCAGGAACCGAAACTAAA	to amplify 3.6kb of Arabidopsis GI _Promoter, for Lip-0, Dijon-G, +GW site
MB20	143m-FKF1_big_FOR	TGACAAGTTTTTGGCTGTGTT	to amplify FKF1 promoter in Arabidopsis
MB21	144m-FKF1_small_FOR	AATCGCGTTTCACAAAATATCT	to amplify FKF1 promoter in Arabidopsis
MB22	145m-FKF1_REV	CTCGCTTACAGAAATATCTTCTTC	to amplify FKF1 promoter in Arabidopsis
MB23	146m-FKF1_big_GW_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGACAAGTTTTTGGCTGTGTT	to amplify FKF1 promoter in Arabidopsis, +GW
MB24	147m-FKF1_small_GW_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTAATCGCGTTTCACAAAATATCT	to amplify FKF1 promoter in Arabidopsis, +GW
MB25	148m-FKF1_GW_REV	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCGCTTACAGAAATATCTTCTTC	to amplify FKF1 promoter in Arabidopsis, +GW

Mutagenic Primers			
ID	Stock Name	Sequence	Comment
MB26	giEE1mutFW	GTACAGTAGAGACAAGTGGTAAGATTTTAGAGTGTCTTAAAAGATGTGTGAGGCCC	to mutagenise EE1 in GI promoter
MB27	giEE1mutREV	GGGCCTCACACATCTTTTAAGACACTCTAAAATCTTACCACTTGTCTCTACTGTAC	to mutagenise EE1 in GI promoter
MB28	giEE2mutFW	TATCTTAAAAGATGTGTGAGGCCCATTTAGACACTCTCTCAACGGGCTGTTC	to mutagenise EE2 in GI promoter
MB29	giEE2mutREV	GAACAGCCCGTTGAGAGAGTGTCTAAATGGGCCTCACACATCTTTTAAGATA	to mutagenise EE2 in GI promoter
MB30	giEE3mutFW	GCATCTGAACTTAGCTCCCCTCAGAGTGTCTGTGGACATATGTCACT	to mutagenise EE3 in GI promoter
MB31	giEE3mutREV	AGTGACATATGTCCACAGACACTCTGAGGGGAGCTAAGTTCAGATGC	to mutagenise EE3 in GI promoter
MB32	135m-A1_FOR_Mut	AATTTAGCCCAGTTACATAACTCCTCGAAAGGTTTGAGCCCATCACATCA	to mutagenise ABREL 1 in GI promoter
MB33	136m-A2_FOR_Mut	ATATTGCTCACATCTTATTGCGCCTCGACTCTTATTTGGCGAGTGATCAAC	to mutagenise ABREL 2 in GI promoter
MB34	137m-A3_FOR_Mut	CTCTTGCTCCACTCTGATTGGTACTCGACATCATTTTATTGTGCTACTTG	to mutagenise ABREL 3 in GI promoter
MB35	138m-A1_REV_Mut	TGATGTGATGGGCTCAAACCTTTCGAGGAGTTATGTAACTGGGCTAAATT	to mutagenise ABREL 1 in GI promoter
MB36	139m-A2_REV_Mut	GTTGATCACTCGCCAAATAAGAGTCGAGGCGCAATAAGATGTGAGCAATAT	to mutagenise ABREL 2 in GI promoter
MB37	140m-A3_REV_Mut	CAAGTAGCACAATAAAATGATGTCGAGTACCAATCAGAGTGGAGCAAGAG	to mutagenise ABREL 3 in GI promoter
MB38	PGI-1F-MB-GW	GGGGACAAGTTTGTACAAAAAAGCAGGCTACCAGCATATCTCTAATCAG	to amplify 2,5kb of GI Promoter + GW Sites
MB39	PGI-1R-MB-GW	GGGGACCACTTTGTACAAGAAAGCTGGGTGGAACCGAAACTAAACCCCAAC	to amplify 2,5kb of GI Promoter + GW Sites

Genotyping Primers			
ID	Stock Name	Sequence	
MB40	28LP-SALK_005476.42	AAGGTTTGGTGGTCATGATTG	Genotyping Primer for SALK_005476.42
MB41	29RP-SALK_005476.42	GCCTACAATTTGAGAATACGCC	Genotyping Primer for SALK_005476.42
MB42	30LP-SALK_057774.34	TTTCCCTAATTGGTAAATCATAAACC	Genotyping Primer for SALK_057774.34
MB43	31RP-SALK_057774.34	TCACCAAGAAAAATGTTTGCC	Genotyping Primer for SALK_057774.34
MB44	32LP-SALK_057774.25	ATCTGCAATGGTCATTGAAGG	Genotyping Primer for SALK_057774.25
MB45	33RP-SALK_057774.25	GAAATTTTTCGATGCGATGAG	Genotyping Primer for SALK_057774.25
MB46	34LP-SALK_086534.47	TTAACATTCGCCAGCAAGAAC	Genotyping Primer for SALK_086534.47
MB47	35RP-SALK_086534.47	AGCATCTGAACTTAGCTCCCC	Genotyping Primer for SALK_086534.47
MB48	36LP-SALK_005474.30	TCACCAAGAAAAATGTTTGCC	Genotyping Primer for SALK_005474.30
MB49	37RP-SALK_005474.30	GCCTACAATTTGAGAATACGCC	Genotyping Primer for SALK_005474.30
MB50	38LP-SALK_005475.53	GGTTTAAGATTTCCCGTCTGC	Genotyping Primer for SALK_005475.53
MB51	39RP-SALK_005475.53	CAAGAGCCATCCAGACAGAAG	Genotyping Primer for SALK_005475.53
MB52	40LP-SALK_114233.29	CAAGAGCCATCCAGACAGAAG	Genotyping Primer for SALK_114233.29
MB53	41RP-SALK_114233.29	CTCGGTTTAAGATTTCCCGTC	Genotyping Primer for SALK_114233.29
MB54	42-LBb1	GCGTGGACCGCTTGCTGCAACT	check SALK-Lines for homozygous plants, LBb1
MB55	43-LBb1.3	ATTTTGCCGATTTCGGAAC	check SALK-Lines for homozygous plants, new LBb1.3
MB56	44-LB1_4SAIL	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC	check SAIL-Lines for homozygous plants
MB57	45-LB3_4Sail	TAGCATCTGAATTTCATAACCAATCTCGATACAC	check SAIL-Lines for homozygous plants
MB58	46LP-SAIL_147_D11	TGAGAGATCGACGAATATCGG	Genotyping Primer for SAIL_147_D11
MB59	47RP-SAIL_147_D11	CTCTTGCTCCACTCTGATTGG	Genotyping Primer for SAIL_147_D11

Side project - Tissue specific cell sorting

The most important processes during the initiation of flowering in response to day length occur in the vascular tissue. *FT*, for example, is almost exclusively expressed in the vasculature (Corbesier et al., 2007). Thus it would be of enormous interest to study the transcriptome of the vascular tissue, i.e. phloem companion cells. Therefore, phloem companion cells have to be separated efficiently from the surrounding tissue. I was aiming to separate these cells via FACS (Fluorescence Activated Cell Sorting), which has been shown to work efficiently to sort different root tissues (Birnbaum et al., 2003; Birnbaum et al., 2005). Therefore, the target cells have to be labeled with a specific fluorescence marker. To do so, I took a construct that fuses YFP (YELLOW FLUORESCENT PROTEIN) to Histone 2b under the control of the SUC2 (SUCROSE-PROTON SYMPORTER 2) promoter, an overexpressing promoter that is exclusively active in phloem companion cells (Stadler and Sauer, 1996). The same fusion construct was also made under the control of the KNAT2 (KNOTTED-LIKE FROM ARABIDOPSIS THALIANA 2) promoter, whose expression is restricted to the meristem (Lincoln et al., 1994). Both constructs were kindly provided by Daniel Schubert's group in Düsseldorf (Lafos et al., unpublished)

A. thaliana plants were transformed with these constructs and homozygous T3 plants were analysed for YFP expression under a CLSM (Confocal Laser Scanning Microscope).

10 selected lines with the SUC2 construct showed fluorescence in the vasculature, though with a differing intensity. The three strongest expressing lines were selected. Additionally, one KNAT2-line with pronounced fluorescence in the meristem was selected. Protoplasting of seedlings with both constructs were successful and fluorescent a limited amount of protoplasts was detected under a CLSM. However, the future challenge will be to collect a sufficient number of fluorescent cells to conduct a subsequent transcriptome analysis.

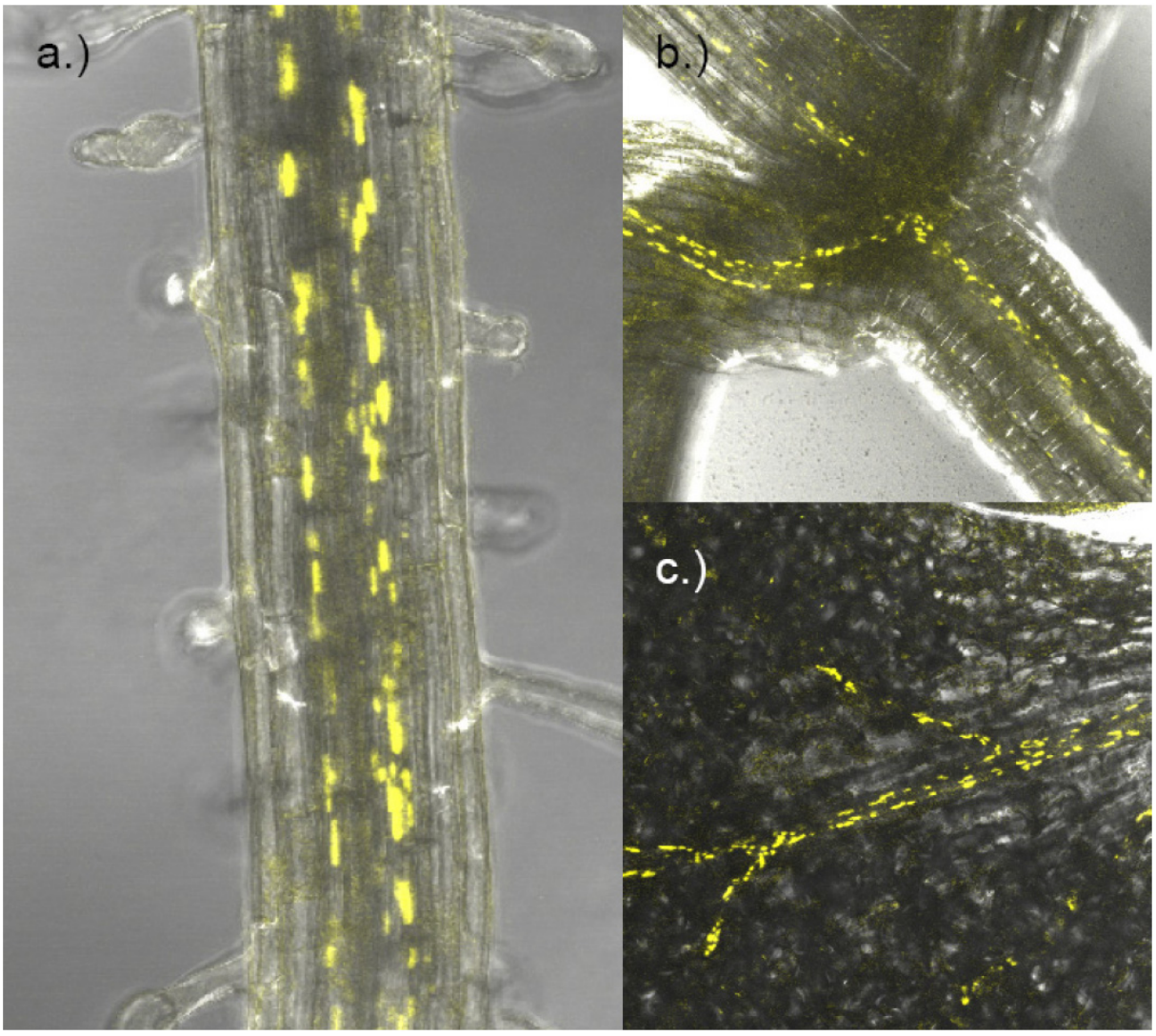
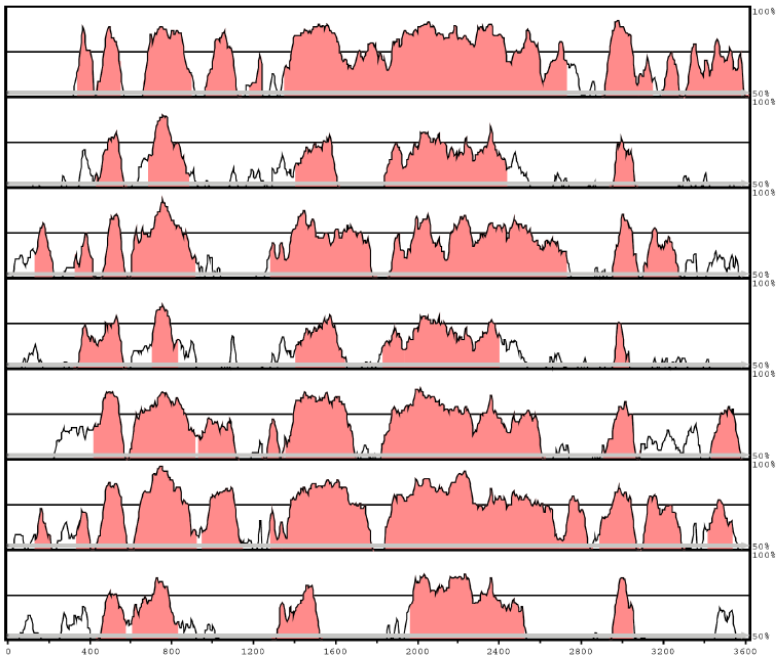


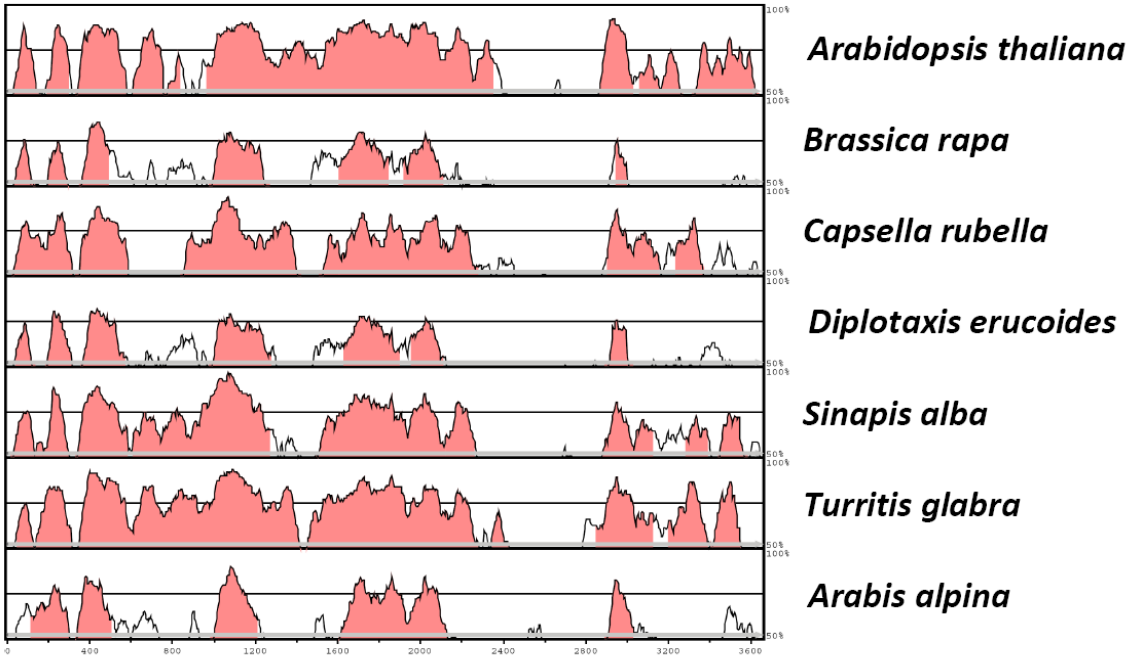
Fig. S1 CLSM pictures of homozygous T3 plants containing the SUC2::H2b:YFP transgene; 10d old seedlings were analysed; a.) root b.) rosette c.) leaf vains;

Supplementary Information

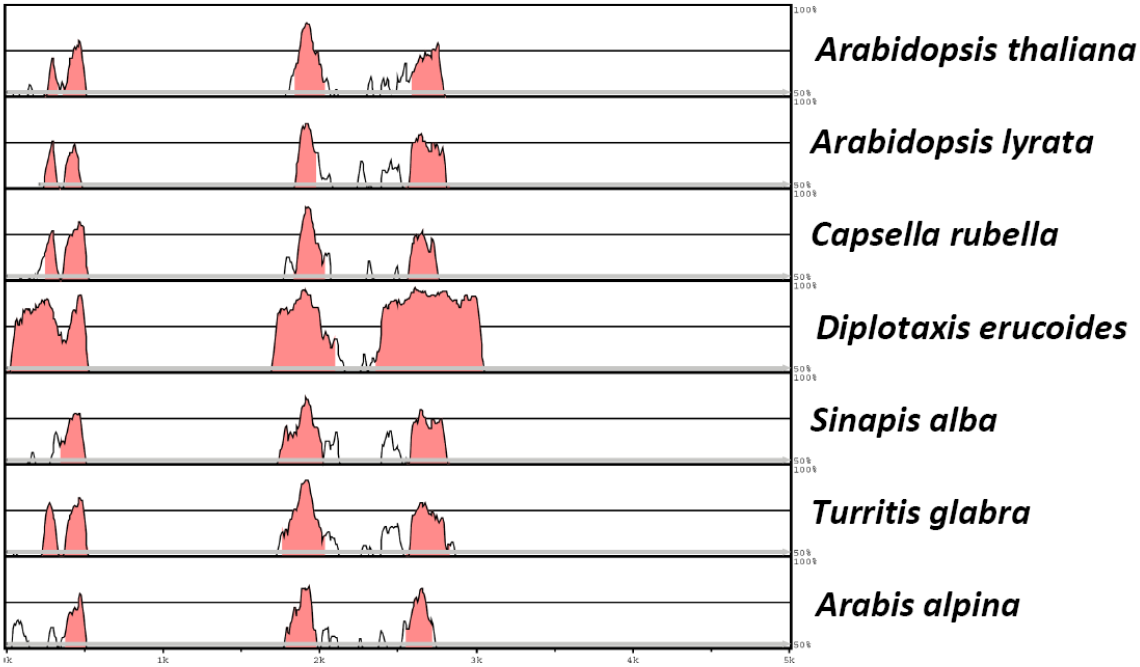
Additional VISTA plots

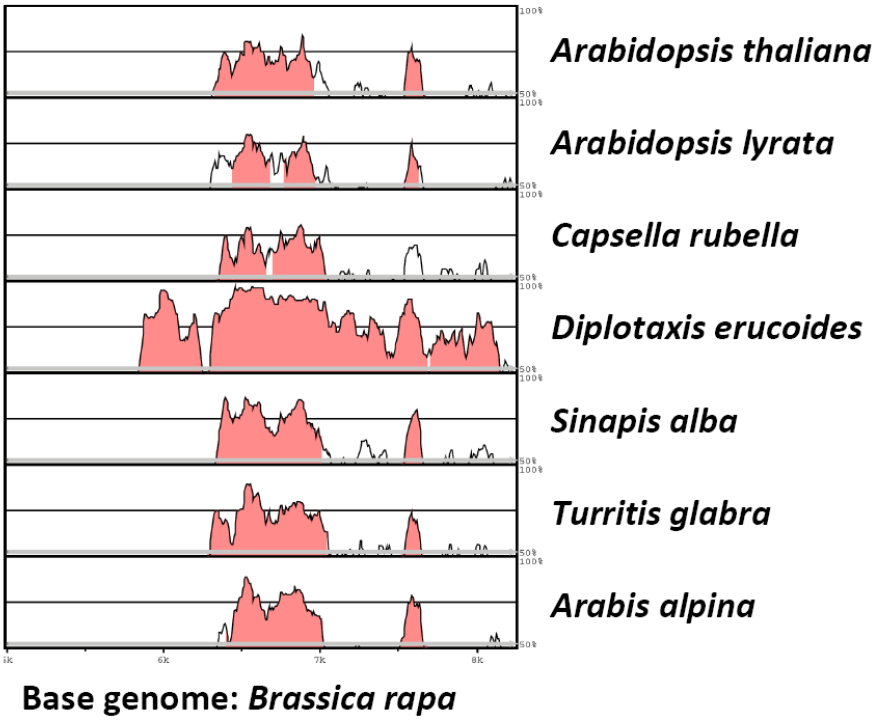


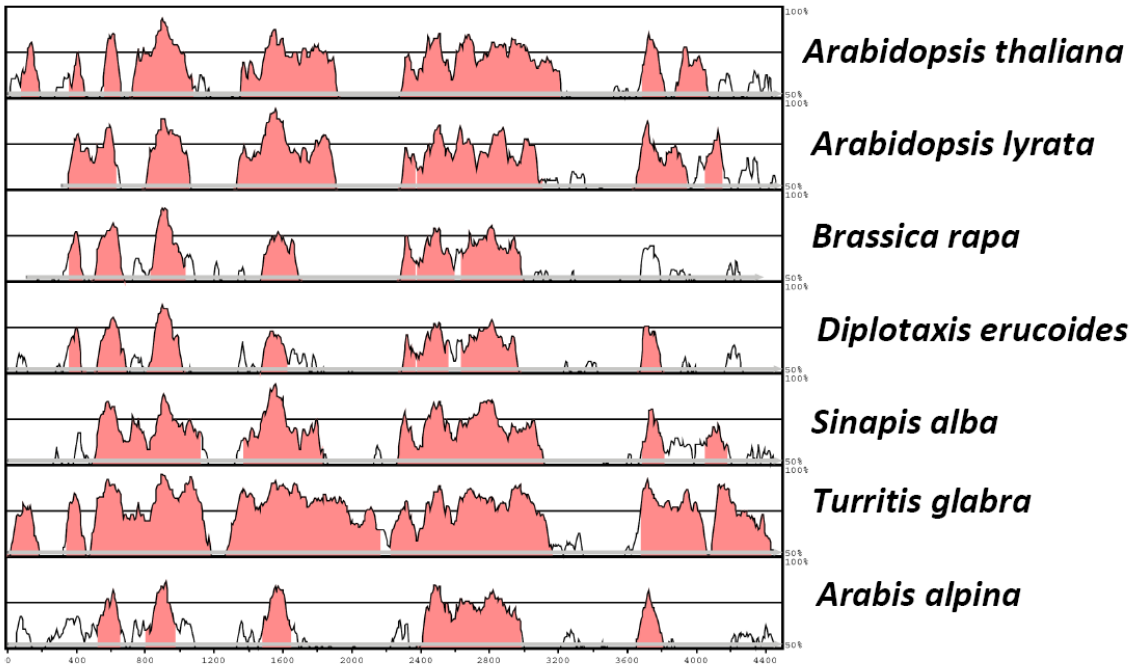
Base genome: *Arabidopsis thaliana*



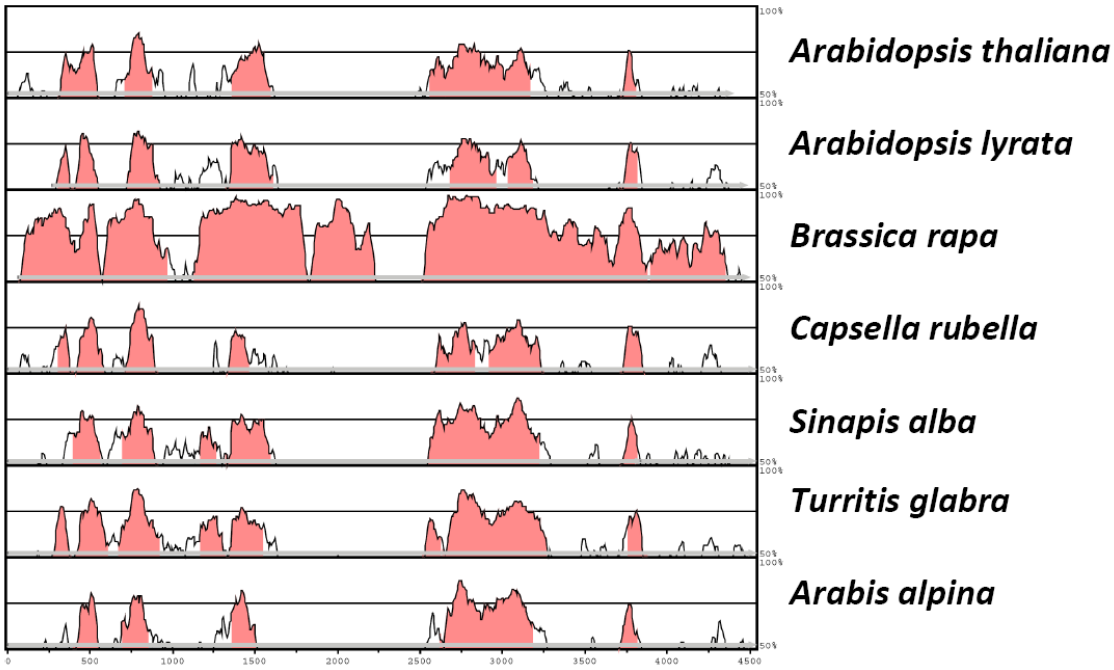
Base genome: *Arabidopsis lyrata*


Arabidopsis thaliana
Arabidopsis lyrata
Capsella rubella
Diplotaxis erucoides
Sinapis alba
Turritis glabra
Arabis alpina
Base genome: Brassica rapa

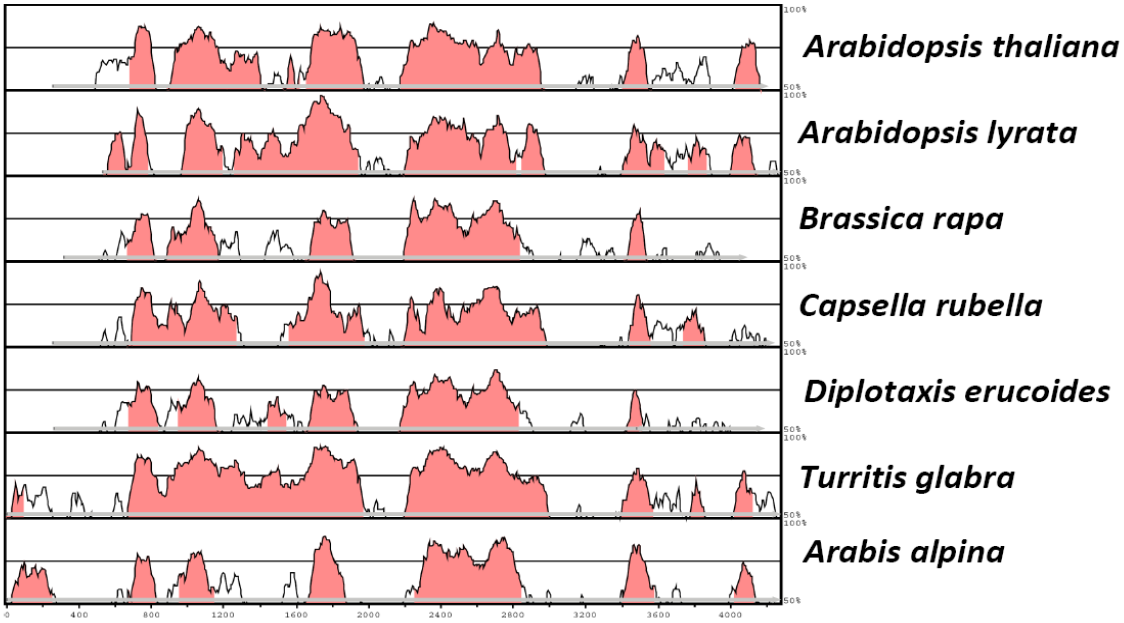

Arabidopsis thaliana
Arabidopsis lyrata
Capsella rubella
Diplotaxis erucoides
Sinapis alba
Turritis glabra
Arabis alpina
Base genome: Brassica rapa



Base genome: *Capsella rubella*



Base genome: *Diplotaxis erucoides*



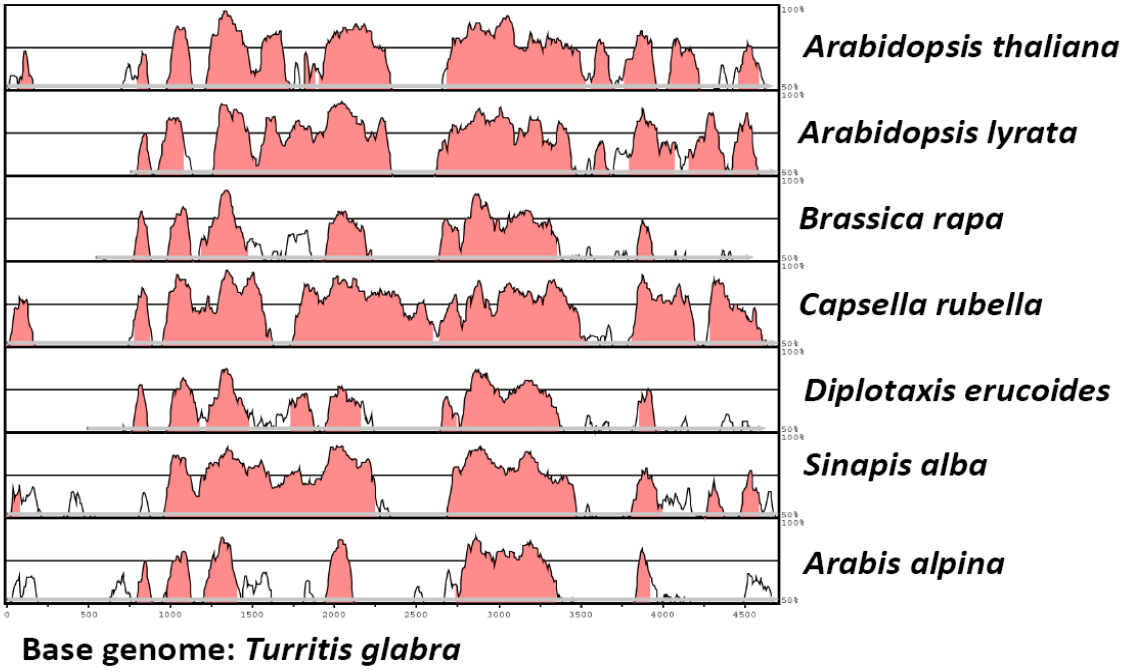


Fig. S2 **Pairwise alignments of eight Brassicaceae *GI* promoters. Alignments performed with CHAOS / Shuffle-LAGAN and visualised with VISTA Browser. Sliding window=100bp; Conservation=70%; Red colour indicates regions where a sliding window of 100bp shows at least 70% of conservation. Base genome is indicated in respective panel.**

Additional figures

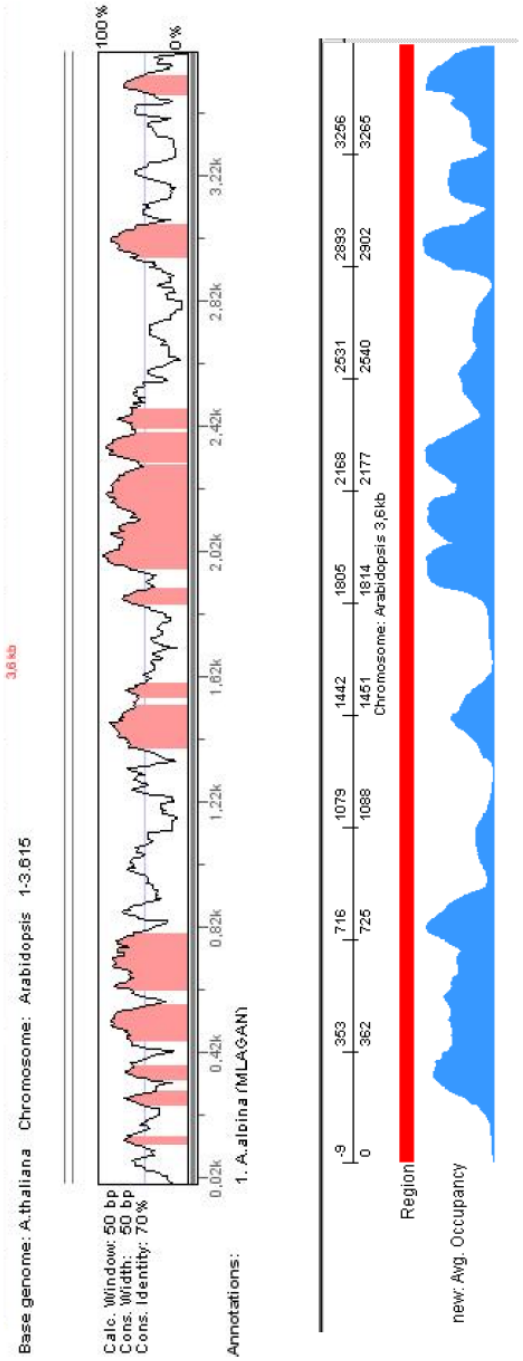


Fig. S3 Co-occurance of conservation and putative nucleosome positioning in the *GI* promoter.
Nucleome prediction was calculated with
http://genie.weizmann.ac.il/software/nucleo_prediction.html (Segal et al., 2006)

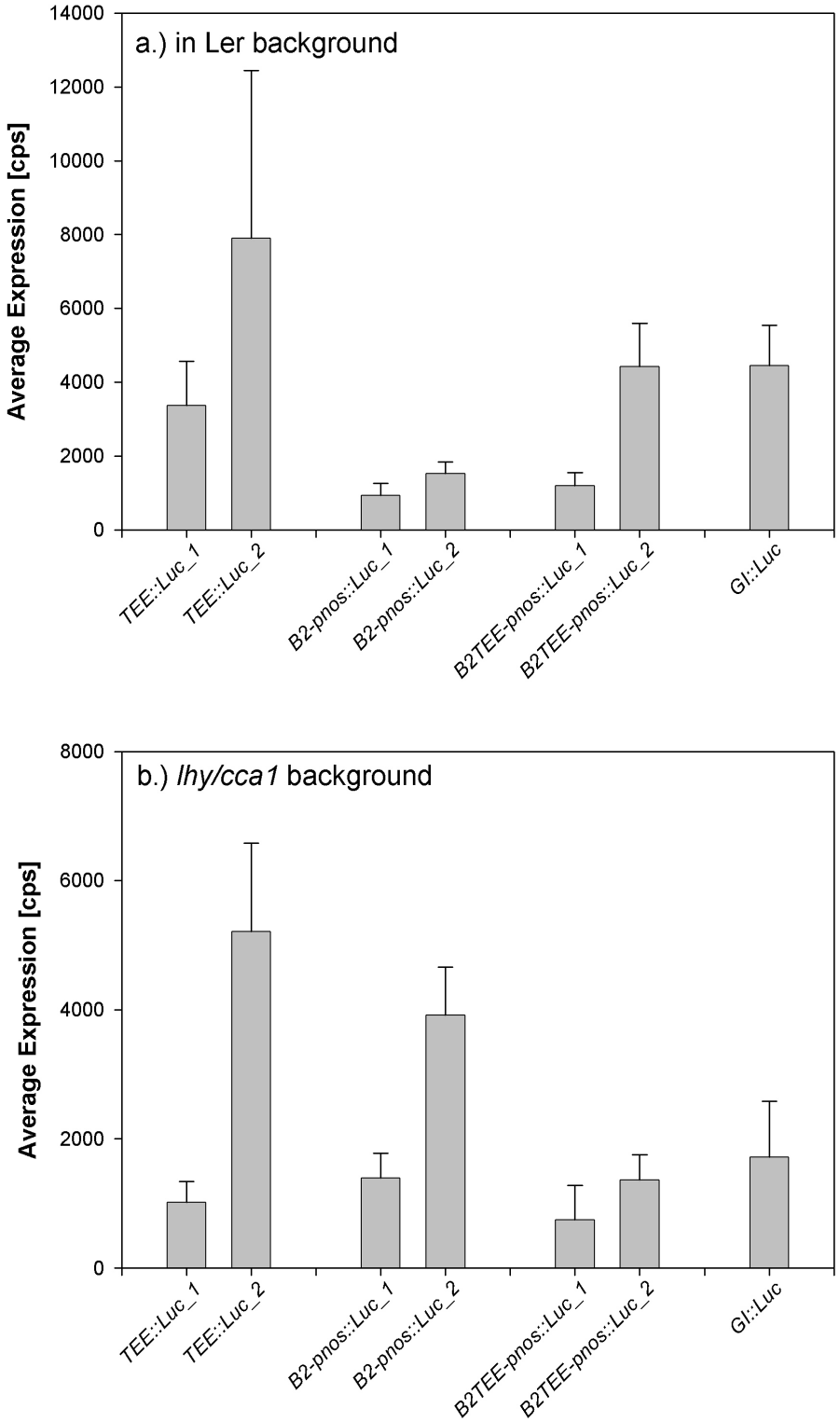


Fig. S4 All transgenic *GI* constructs confer robust *GI* expression
a.) Average luciferase expression per seedling of each 2 independent lines with the *TEE::Luc, B2-pnos::Luc* and *B2TEE-pnos::Luc* construct compared to the reference line of *AtGI::Luc* over 24h grown under LD 16; Errorbars = Standard Deviation
a.) All constructs in the background of Ler wt b.) All constructs in the background of the *lhy/cca1* double mutation

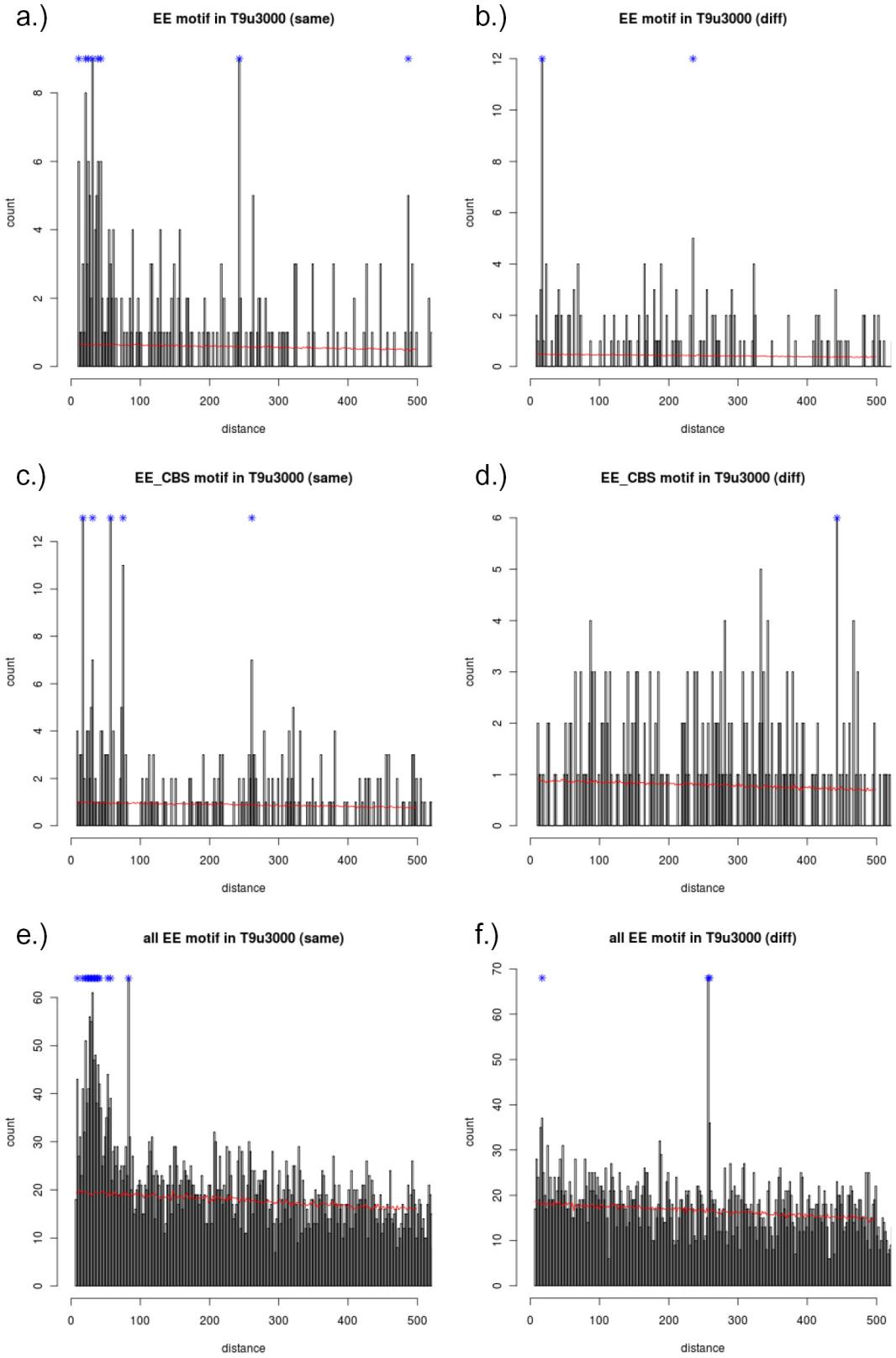


Fig.S5 Clustering of EEs within promoters
EEs (AAAATATCT) and CBS (AAAAAATCT) were mapped in relation to neighbouring EEs; a.) /b.) only EEs, same and different orientation; c.) / d.) only CBS, same and different orientation e.) / f.): All EEs, SEEs and CBS together, same and different direction (e.) and f.) are same data as in Fig. 3.8 a.) and b.); Asterisks indicate statistical significance p<0,05

List of 71 genes whose promoters contain at least 3 EEs

ID	# EE	# ABREL	Description
AT1G07040	5	3	unknown protein
AT1G10760	3	11	SEX1 (STARCH EXCESS 1); alpha-glucan, water dikinase
AT1G21945	3	12	transposable element gene
AT1G22770	3	4	GI (GIGANTEA)
AT1G28060	3	5	small nuclear ribonucleoprotein family protein
AT1G42650	3	6	transposable element gene
AT1G44100	3	0	AAP5; amino acid transmembrane transporter
AT1G45332	3	8	mitochondrial elongation factor, putative
AT1G45474	3	5	LHCA5; pigment binding
AT1G48330	3	4	unknown protein
AT1G54410	3	1	dehydrin family protein
AT1G60270	3	4	pseudogene, glycosyl hydrolase family 1
AT1G65360	3	0	AGL23 (AGAMOUS-LIKE 23); transcription factor
AT1G68050	3	6	FKF1 (FLAVIN-BINDING, KELCH REPEAT, F BOX 1)
AT1G70650	3	12	zinc finger (Ran-binding) family protein
AT1G71710	3	2	inositol polyphosphate 5-phosphatase, putative
AT1G75790	3	14	sks18 (SKU5 Similar 18); copper ion binding / pectinesterase
AT2G15880	3	6	leucine-rich repeat family protein / extensin family protein
AT2G21660	4	4	CCR2 (COLD, CIRCADIAN RHYTHM, AND RNA BINDING 2)
AT2G21680	4	5	FUNCTIONS IN: molecular_function unknown
AT2G25190	3	2	unknown protein
AT2G33830	3	1	dormancy/auxin associated family protein
AT2G34840	3	2	coatomer protein epsilon subunit family protein
AT3G05790	3	4	LON4 (LON PROTEASE 4); ATP binding
AT3G05800	3	5	transcription factor
AT3G07650	3	2	COL9 (CONSTANS-LIKE 9); transcription factor
AT3G14270	3	3	phosphatidylinositol-4-phosphate 5-kinase family protein
AT3G15450	3	3	unknown protein
AT3G15830	3	2	phosphatidic acid phosphatase-related / PAP2-related
AT3G20800	3	1	rcd1-like cell differentiation protein, putative
AT3G42050	3	2	vacuolar ATP synthase subunit H family protein
AT3G45190	3	1	SIT4 phosphatase-associated family protein
AT3G51400	3	5	unknown protein
AT3G61570	3	7	GDAP1 (GRIP-RELATED ARF-BINDING DOMAIN-CONTAINING ARABIDOPSIS PROTEIN 1)
AT3G61580	3	6	delta-8 sphingolipid desaturase (SLD1)
AT4G03530	3	2	transposable element gene
AT4G16860	4	2	RPP4 (recognition of peronospora parasitica 4)
AT4G16890	3	1	SNC1 (SUPPRESSOR OF NPR1-1, CONSTITUTIVE 1)
AT4G17090	3	1	CT-BMY (CHLOROPLAST BETA-AMYLASE); beta-amylase
AT4G25470	4	3	CBF2 (C-REPEAT/DRE BINDING FACTOR 2)
AT4G25480	3	10	DREB1A (DEHYDRATION RESPONSE ELEMENT B1A); CBF3
AT4G25490	3	6	CBF1 (C-REPEAT/DRE BINDING FACTOR 1)r
AT4G25500	3	10	ATRSP35; RNA binding / nucleic acid binding
AT4G26530	3	2	fructose-bisphosphate aldolase, putative
AT4G29190	3	2	zinc finger (CCCH-type) family protein
AT4G30190	3	1	AHA2; ATPase/ hydrogen-exporting ATPase
AT4G31360	3	2	selenium binding
AT4G31410	4	0	unknown protein

AT4G31875	3	2	unknown protein
AT4G33980	4	7	unknown protein
AT5G01050	3	2	laccase family protein / diphenol oxidase family protein
AT5G03940	3	6	CPSRP54 (CHLOROPLAST SIGNAL RECOGNITION PARTICLE 54 KDA SUBUNIT)
AT5G05400	3	6	disease resistance protein (CC-NBS-LRR class), putative
AT5G09450	3	5	pentatricopeptide (PPR) repeat-containing protein
AT5G12270	4	9	oxidoreductase, 2OG-Fe(II) oxygenase family protein
AT5G15980	3	7	pentatricopeptide (PPR) repeat-containing protein
AT5G23550	3	2	FUNCTIONS IN: molecular_function unknown
AT5G23570	3	0	SGS3 (SUPPRESSOR OF GENE SILENCING 3)
AT5G24470	5	2	APRR5 (ARABIDOPSIS PSEUDO-RESPONSE REGULATOR 5)
AT5G39910	3	4	glycoside hydrolase family 28 protein
AT5G41830	3	2	F-box family protein-related
AT5G47020	4	2	glycine-rich protein
AT5G48250	3	3	zinc finger (B-box type) family protein
AT5G50130	3	2	short-chain dehydrogenase/reductase (SDR) family protein
AT5G51510	3	7	unknown protein
AT5G56260	4	3	dimethylmenaquinone methyltransferase family protein
AT5G56270	3	3	WRKY2; transcription factor
AT5G56490	3	4	FAD-binding domain-containing protein
AT5G59810	3	6	SBT5.4; identical protein binding / serine-type endopeptidase
AT5G59820	3	5	RHL41 (RESPONSIVE TO HIGH LIGHT 41)
AT5G64400	3	1	FUNCTIONS IN: molecular_function unknown

Suppl. Tab. 1 List of 71 genes whose promoters contain at least 3 EEs.

Sequences

GI promoter sequences of 8 Brassicaceae as used for the phylogenetic analysis

>Arabidopsis_thaliana
ctaccaccgactgaaatttgcaatgcgtttgctcaactctgttttcttctttttagttttgatatcgactccttttctttttcttttttgggttttagtcttcctttactattttagcagttttttttgtttacttagt
ttatggttttgtgtctttaacttagtttatggtgtttttctttttctttggacaatgatgatttcttaataatctgagtagacttcttcttccgaatccccaaataatccatctctacttgaaatttaagttgt
tcggtattcatggtcagcaacaagggacaatccttctattatggtagcagccttctattatgaatttattcgaatgtgtctactcataatttgtggtttgtatttcaacttgtagttgatatcaattata
tgcacacagacttttttcttttttgaggggttggaagcaagcgattcaatgattgtaccactcatttcataatagtgtttgatatgacgttaatttgaatggtttatgttgatcgaaaatactctcagtc
acaacaagaatctaattcgattaatttatttgaaaaacatatgttggcatggtacatgaaccagctcgttcatatgaaaatattttaaaaagtcaacatcagtgctaatttcaaattgaccataca
ttagcacacaagaagactaacaaaaaaaatgagtgaataaagaagagagtgagtggggttgatacatgtcaccgtccaaaatgtgtctgaccttctcattggagctacaagagccga
ccctttgcctcatcatttgttgacttcactctccttttacatatttaaattataacttgctcatatttccattcattgtgttcggttctagattgttccatccacttttcttttttctcttacttttgaaatttagaa
atatatgaatacttgacaaatatttgaatattgtaactgatttgtttgttcgtaatacttgacaaactttttgaatatacaccttacatatattttgagttgagataaaaagggaattatgttagtttat
ggattatattatgtaaatacacatctaactatgatcggatgaaaactaaaccagcatatctctaatcagtaatatttactaaatttatatatcattaaagatttttgaatttagcaaaaaaagaa
gaagatatcattaaagatttttaaatttaaagtataacataatgaataaatcaagtgttgagcttagtttcgcgttgcaaagatgttcatttgatttttttttttttttccatttgatttttcttatactaa
attgattgatgtcacggtccttctgtgaaaaaagtatattgtactgtataaatttataatagtcatacttgaagataactaatatcattttctaaaggttttgttgacaatgaaatgggtaatgatga
acatctcgtttgctaaaccacaaaagaaataattcgttggactataacttaaacttataatatatcaactgaaatgcctacaatttgagaatacgccttttcaaaaagttttagaatttttataaaaa
tttctccaccataatattactttaataaaagcttgttattaaaatgatatattaacaaaaaaaaaaagaggtaggcaaagtagcttttattaaaattaatgttttccctaattggtaaatcataaacca
aaaaattgaggtaattcttaatatatttactatgattagaatgatacattaatgtaacattttatcatcttcattaataatagtaagttctcctcaaaattacaatttgatatcaaaaaatatatttgtatt
acaagcaattactgaaaaattaaaattacaatttagcccagttacataactccacgtaaggtttgagcccatcacatcacatggtacatgaccaaagcccagttatgttaccatggagtattc
ttcagacaagagccatccagacagaagtggaccacacaatcacgaatcgtatggagatcattacaatatattgctcacatcttattgcgccacgtctcttatttggcgagtgatcaacgtcc
acagatctcttgctccactctgattggtacacgtcatcattttattgtgctacttgtttatgtggtacagtagagacaagtggtaagattttaaaatatcttaaaagatgtgtgaggcccatttaga
tattttctcaacgggctgttcctcaaaagctgcccgtttatctacaaccatcgaacggctcgaaactattctgggctagctgcccgaaatctcttaaaggaatttgttgagtgggcttgaatca
acggtggataattcgaaggtgtacgtaacatctcgtagcatctgaacttagctcccctcaaaatatctgtggacatatgtcactagctttttttttgtgtaatgaaactattaattaattaggcaat
cacatgagagatgttatatacttaaattaataactcaataatacaatcatgaccaccaaaccttgaaataaaataatctctctgattacttgtacaaattctttaccaatatatgccatactgtgaa
tgtaatggtttatttttaaaacagccaaatattattgatataattatatgtttaagtagattttaaacttagctcatggttaaaaatagacagaattttggagtaaatctgagtttacaaaatttatttatt
aggattaaaattaattacttaaattggcaaacatttttcttggtgattgtaacatacaatatacgaatttgaattcgcattgtgattccaaaacaacactaacataaactaccagtaaaatttttttaa
aataaaatttcatatatatatgcttaaaaaatgtaacaaaaaatatggtaaatttttttaaccatggtatgggtggagatgtatgtgggatgatgatggttatatggtaatggcgcataaaggtggt
ggcaaaggcaaggaaatatcgatgacacgtaagcagacgggaaatcttaaaccgagtggagaagcctccaaatctctttttctctctctctcctaaggccaccacaatctcttcttctg
tgtatgtattcttttttttctccgatattcgtcgatctctcaaaaatatcctcaaagccaaaaaaaagcaagagtagagataaccaaccaacaaactcataggaagacatatttactttcaggtct
ttcttctactacctttttcgtggtttttggttgattgattctcagatgagatttttcccgaatcatttgatgtctttttttttaaataagttcttggtaaatcactgtttctactagcatttttgttttctcag
cttagatttgttgtttgatttgacatcttatctgattgagattttgttcttctgaattgttgttacagggtttagctgtttgattcagcttcgatttagtgtacagtgtgttgattagtataaaaaggattta
aaagaatctgggagatgtgtatagtgattgttcttgctggcgaatgttaacactacgtcgtattgattctcgattgatagtagaagaaaggtgttagttagattgttcgttcatctagttggggtt
tagtttcggttcctggatg

>Arabidopsis_lyrata
cagtacaagccttctattatgaaattattcgaatgtgtcatactcaaactttgtggtttgtattcaatatgtagttgatatcaattatatgcacacaagacttctttttttttttttctttttttttgaggggt
tgaatataagagactctttatgttttataattggaagcaagcaattcaatgattgtatttcataataaatgtttgatatgacgttaatttgaatggtttatgttgatcgaatatactttcagtcacaac
aagaatctaaatacaattactaaactatatatcagctgagttctgaccatactttataagaagatttaaaaaaatgattgaataaagaagagagtgagtggggttgatacgtgtcaccgtcca
caatgtgtctgactctgacccttctcattggagctatactagccgaccctttgcctcatcatttgttgacttcactcacctttgcatatttttttaaatgaacttgctaatgtttccattcattgttcgat
tctagattgttccattcactttctttttttctctttgacgtaattaagaatatactaacatagtaaatatttgacaaaatttttaggatttaggttttacatctatcttgagttgagataaaaatggaattat
gataggttagggattattttatgtaaatacccacatctaaactatgacggatgacacactaaaccaacatatcttatttactaaatttagatatgatcgttaaagatttttaaatttaaagtataatat
aatgaataaatcaaatttttgagcttaatttggtttgaggattttctaaatcccgcgttgcaaagatattcgattatatttttttgtctataaaattgattgatgtcacggctcttttgtcaaaaaagtatatt
gtataatgtcaaaaaaaaaaaaggtatataaatttataatagagtaatacttgaagaaaaataatagcattttcaaaggttgtattgacaatgaaaggggtaatgaatatctcgcttgctaaac
cacaaaagaatcaattcattggactataacttaaacatataatatatctaccgaaatgtctacaatttgtgaatgcgacttttcaaaaagttttagaatttttataaaatttttctccactatagtattac
tttaataaaagattgctaataaaatgatatcttaattcacaacaaaaaagtaggcaaagtagctttattagaaaactaaatggttttttttttttgttaatggtaaatcataaacctaaatttaagct
aattgcaaaacgcgtttactatgatcagcatgatacatcaatgtaacaatttaacgtcttgattagtatctaattagtaaattatcctcaaaattacaatttgacataaaaaaatatatttgttttaca
agcaatcacagtgaaaaatagattgttttctcttaaattacaattatttttttagcccagttacataactgcacgtatggtttgagactttgagcccatcacattacatggtacatgaccaaagccc
agttatgttaccatattattcagacaagagccgccatccagacagaagtggaccacacaatcaagaatcgtatggagagcattacaatatattgctcgcatcttattgtgccacgtctcttg
gtaggcgagagatcaacgtccacagatctcttgctccactctggttggtacacgtcatcatatttatgttctacttgtttatgtagacgttcatgtagacaagtgggaagatttaaaaatatctt
aaagcaagtgtgaggcccatttagatatttctcaacgggctgttcctcaaaagctgcccgttgatcaaacggctccaaactgttctgggcaagctgcccgaaatcttataaaggaaattgt
ggagtgggactgaatcaacggtggagaattcgagggtgtacgtaacaactcgtagtatctgaacttagcttctctcaaaatatctgtggacatatgtcactagttttgtttgtgaaagtatta
actaattaggcaatcacatgagaggttgatatagtaaaatacaatcatgaccacaaaacttttaaataaataatctctctgattacttctgcaaattctttaccaatatttgccatactgtgaatct
agagttttatttaaaaacaactactatgaatttttgatatgatttcatgtttaaggagatttagacagaattttggagtaaatctgagtttacaaaatatattattaggattaaaaactaattacttaa
attggtaaatatatttgttggtaatttgaattcgaattgtaatccaaaaacaacactaaaacaaactactagactattttaaaaattagaattcttatgactgagtgatgaatcatatatacttcttct
gtttcttttaggaaatttttgtttgtttcattatatttgacttttccaagttgctagacaatttgaatatagtataaaattgttgaattaatttcatgatgcattaaataatagtttttcaatttttgtgtgtttt
agtaaaaaaaaaaaatcttatataaagaaatggagggagtattagatgcttaaaaaataaagacaaatatgataaaattgtaattatagtatgtatggttaacacaagaatttgattcgactaa
aattcaaaaacaacaaactactaaactatttttaaaagtagaattgttttcattgagtgataaaccatatatgaattgtttaaaaattaagaaaaatatgataaaatttttaaccatatacatatatgtg
aggatgatgatgatggtcatatggtaatggagcataaaggtcgtggcaaggctagggaatatcgatgacacgtaagcggacgggaaatcttaaaccgagtggagaagcctccaaatc

tctttctctctctctccaaaggccaaagccaccacaatctctcttgtgtgtatgtattgcttcagaatccttttttgatattcttcttcttttttctccgattttcttccatctctcaaaaaaatatcctcaa
agctaaaagaacagcaagactaggtgatgataatcaacaatcttaggtctttctttttattatgatttatttttttccgtggtttttgcttgattgatttgtcagatgagttttttttaccgaatcatttggtg
tttggtgaatttttttgttgattgattctccttgatttgcttaatttctccagttttgtgttgtgtgtgttggttttcgtcactgtttctacttgcatttttttttttctcagcttagatttcatttttttttttgttatcac
tgattgagattttgttcttctgaattgtaacaggaaggagcgaataacaaagacatagcgttttaattcagcttcgatttagtgtacagtgtgttggtagtataaaaggagttaaatctgggtga
tgtgtatagagtttgttcttgccggcgacgacgtattgattctcgactggtagaagaaaggtgttagttagttagattgtgtaaatcaatcattctagtgagggattagtttcggtttctggttttta
cttcttctggatg

>Brassica_rapa
tctatttttttagttcaactttagagtatttcttttagtatttttattaatttttttaacttaatttatgattttgaaacttttaacttaatgtggttggtccattttatttagagtctttaacgctacatgaaattt
aattgtttatcatttggtagtgattgtcagcaataggaagccactcaatcatatatggttactgtttagtgacatgatggtgcgaatccaaacgtgggtgttttctgtaaatgaattgattcgaat
gtgtcctacttgaaatttgtggttctgcagtcaacatgcttggtttcagttatatacacacaaacgtctttttaaaacatttatggttttagaatcggaaattaatgagcaaaactgaatgattatac
cattttcatttcataaaaaaaatggtttgatatgacgttgattataatggtttatgttgatcgaatattctccgaacgagagcaagaatctaattcattttagtttattcggaaaacacatgtttgcata
taatactagggtcggcccgccctacgggcgggatatacttttacttgtgatcattttattatttttgtatgattttatagttagtgttttaagttttcatttgcgtgaaatatgtatgatacactacaaaa
attcatattatatcacataaatagtatcagaaatataaatgataaataaattatgatgtattctagggttctgatttgagaataatgtcacttatgtattatttttaaattgatggtcatgtttttaggcg
tttagtaagatttttttttccagttaaattgatatgtgaatgcatgattgagtttgttttgtctattatccatattttgtattaatgaactttagttattgggttattagtggactttaacttgattgggtacat
caaaattaaacgaaaggttaactgaactaaaaaagaaaaaaaaagttcttttattgttttttctctcagccgactaaacctaattttcttatttctttcactatctcatccgacccgtattcttttttcat
cacctcaccttcgtcttcacctcctcaccgtatcttcttcatcacctaatctccctttttcatctcatccatcttcttcgtttaaccacctcaccttcttctttaccacctgaccccatcttcttcttcacc
acctaatctccctttttcatctaattatgcttcgtcgttcactttcttcttttctgttttacttgataaattaggatttgatttacagttcatcaacataaagtgacagtatttttatctgtttggttcaattatat
agatatttttcagtctgaaaaaaaaacagaaaacagtttaacttaatcgatttttttccagatttacctaattcgcaggcttagatgtaatctttaggagtttttgacggaaaaaagtttgtttggatt
aaaaaaatgcttggattacagatgtgatttgtctcagttttttggttggagaaaaggttggatggtggctcagaaaataatttttcaataactcatttaataaaaacggaagcccactcatgaa
gcccaaaaaaaaacaaattaaaaagatgaaataaccagagttgaacacgtgtcatactaagaagaattgacttagtgacgtggatgcacaggagtgaagcgaacatttttttatatatata
gattagttttgctgagacttcttgtcagatttttgttttttattaaattatgagatggtacatgcaattagtgaataaacaatctctgttcatgtgaagtagtatattattaaaagtcaacatcagtgctc
tatttttcaaattgactatacttataaaatgattgaacaaagaagagatggtgagtgggtttgatacgtgtcaccggccaaaatgtggctgccccttctcattggagcttcactagccgacctt
tgcctcatcatttgttgacttcactcgccttttttacatatttttactcaaaactcatttctttcggtttctagattgttccacccacttttggttctttctttgaaactcaagttcattatatcaacgtagtg
aatgaatttatttttcgtgtgtttgcaaaatttttaatcgaatttagatctttttcatttatttggtgttttaagtcctaatatgcaaagaaacactagaacgagttagttctggtggtaaatagactgtgctt
gtaactcccatcacccatattcaatttgctgtaagaaagactatttataatgtttggactctggtaactatttatgtaaacaccacatctaaattatgacataatgacactaaaacaaggtaatat
cataagtaatggtggtgtttgaaaaaaaatctaatcagtattggtgaattctacatttagatatatatcataaaagatcaaagtataacctaatgagcgaatcaagtttgagtttagttttattaac
gttgcgttttgaaatctacatcggttggaaaaaatttcgattctcttatattatattgattactgtcacggttctccgatgaaaaaagtatgtaacataataatagagtaacacttgaagacgaat
aaaaaagactgttgacaatgtaaggtgtaatgaatatctggctcagtaaaccacaatagaatcgattcattggactataacttaaattaaatctaaagaaatgtctaccgtctgagagcttac
gacttttcaaaaagctttttgaagtttcgtaaaaagttttctccaccatagtatcactttaatagaatatggttttaaacatgatataataacaaaaagtagacacgctagctttattagacctga
atactctttagattaacaatttatttttcacaaaggtaaatattaaacacaaattgaagccaagttcaatgaataaagcttttgttgccactcttaaaaagattgcaatgaccttcacaagaaaca
cagtctgataatcaataataatgaatatttatttgtgtacaagtttggttattacaacaatcatacaactgctttctgagtcgttataagtcaactgctagaaaatatagaaagtattacactgaga
ggcagtttcaagtttaatattacactataaggtagtttgtgctactacggtagtttttctcaaccaagcttattttctattcataaactaactaaataaatttgtaaacaaaagaaagtgggttccatt
tacttaacatatgatttctatagaatttagtcaccgttcttcctagcaaacaaaagaaagtggaccccatatatcttcttcttcgtttctttttctttagttcttcatttttttgtaaaatgattgttgttttttt
cagaacattcttttaataaaatatgtttataattagtatcttcattattcaatctatctttgatatatatcaagctaatgttattatacactgcaaatcacggtcacaaaagctccaccgccacagtc
tcctttctttccacgagacttatctttgtggtgtttcaagtccggtgggagcttcaacgtcataccctatgctacatctttgccgtgatcttagcctctctccatgacctccggcggactatca
agcgaagccgctgttgcaccgagatgttgatcgccattgttgagctttgagaagaagtttctttaattttcagttttggtccttgtaattttaaatgtttccattgttgtctcaaaacatctagatttg
catataagtcatgtttaattagcccaaaactttcaaatgtgaactttagaccttgtcaaatgcaaataaacaaaaataaacaaatatatgcaccaaatacctctaaatacatactaaaattattta
catgaggtaaatgcaaatgttaaaacctactaaaattatgagttaccaatgaacaacatatttacataagttgagtaggtcatattgactcatttttgtatcatattgaccatacaaataacgaa
acttgcaatttgtgtacaaaaatgatctggtgttcatgtattattgtacaaatcaataagtgtacatgcgaactattgtacgcatggatgaaaaatatttgtgtacatagtatagtgtacacatca
aatagcttttaaagttacaatttttttaccgtataaatattgtaatttgcaaatgtgcctaaaacatttatcaggtgtgcatttgaatccttgcaaataaatctaaaatactgtctggtgtccatatatta
ttgtacaaattaattgatgaacatgtgcattattgtatatgtggatagaataaaatttatgtacgtattgtcgggatttttctattgtttatgtatggccacatgtacgctgaattttcaatatccatgt
gtacgaaagatcatgtgtacacctaaatccggtaagagactgagttaccaaatcgatagaaagtgatagagagtatcacagttagtactgatgctggttgtgagttatccaaatcatgctct
catcatgttttagtagaagagaccgaaaaatccaaaccttgtcttatctccgacgatgtttactcaaaagacaacttctgaaagaggttaaaacaccataatcaaaaccataacacatatgc
atgtgtgcatgtagctccttatgaacttaatattgtgaatttatgtttcatcttcttcttcatctaaatttacttctctttcttataagaataatttttagaaatgtatacatgtagatgtgtatgcacgtgta
catgtagattttataaaaatggaaacatagaaattgtgtaatcattttgtgtcggtgtccacgaatttaggttgtacacatgcacattaaattcatatgttgtccacagttatgatatttacatgtac
actaattctacatatatgtggtcatgttttaaggttaataatttgttttaggtttaagtgcttaggatttattataaggggttattgagttgtaaatctgagttcttttgttcatatatatgtactataatgt
ttttgcgaatgtgtacaattaagttctttaatacaattaagttttccggcaattagctgaatgtgtacatttaataaattatttatgcgagagtgtacaagtaagcgagatggtaaggcgcttgtttt
tattcttagactctgtttctattaacgttttcatcatctaagttttacttaatttttttaatgagtggtaatttcgtaataacttcatcttcttgttagggcttcttgagtttcctgcaaatatatctcagagc
cgccattgatttttatcatgcaatcttcattatatttttgtgtgttttaactctagaacttataaatattttacagtttaactgattaggacccaaaacccagaaaaaacaagttttaatttcccataact
aaattccccatatccaaatctatggatagttaattccgacgtctggacacgtgattcatcttaggaatggtaatgagttttaaaaatctaactatttgacatgaattctagcggaactgcgtgat
atatgatatacaggttgattgaaagaagaaaaccgatgagaaaaaaatataaagatctgagatttatgttttggagaggctgagatacaatgagttttttctaaatcaattttctgttataaaag
aaagaaaaaacgagatcaatggagaagaaattcaggaaagagaagatgataggtcccactaattagaaaagtaactttgtttggtcttagttaagggcactaaagtaaacaaccataata
aactaccctagtagctgaatctgtctttggatgtaataaaatcttagaaactgccctatactgtaaaatactcgaaaatatataaacatgcaagaacagagagaaatagtcaatgtcatttag
aacatgggtgtggtgttatcttagaggttagatcgatacatctagcatcccttctgatatcacatcattagaacatatatgtgcttcatatgttgaaacaatgataggttccataaattattggga
agagcaaatgatgtgatgttcgtttgtttatttttatgcgttaatgcttacggggttcatttacatcaaccatcgaactattaattatcaaaaaaataagtaactacacataatgttcaaagagtcta
tagcaaatcatcaatatcagccattcacaaatttgaggtataagaaaatttaaattatttatttttatttctaatattcaatataattaagtattatatttatttttttcaaattttaaaatgttatattaacttt
aaaagtagttacagcgcttaataattaaaaaaaaactattttcttattaaaatggcaaatattttcagcccagttacaagtctccacgtatggtttagagcccatcgcatacatatagaagagc
atgaccatgaaaagcccacagttatgttccatgttcagacaccaacctgccagccatccaccagagtgtggaccccacacccacgaatcgtatggtagatcactacaatatattgctggc
atcctgttgcgccacgtcgtcgttaggaggaagatcaacgtccatatatctcttgcgcctctctcattggtacacatcgttttgttatatttcaacttgtttatgtggtagagtggacaagtggg
aagaagaagataaatatctaaaaagagtgtgggccttatgtgtagtcttgagatattttctcaacgggctgtctccacgaaaagctgccccgttaatcaatacacaccatcagacggctcg
aatctcttcaggtcaagctgcccgaaaactttgccgaggaatttgtgtgggactcactcaccactgaaatcaacggcggaaatttcgaaggtgtacgtaacactgcgtagcatctgaatat
agctcctctgaaaatatctagggacacatgtcacttagctttgtaagaaaagcaattaactaatcacatgacagagcttatagttaaactaatataaaaaggaggaaataatacaatcatgacc
acctttaaataatgtaattaatggggaagagtttatttaaccgaactcaccataccatgaactagagtttaaaaaaaataaagtaaacaaaaataagcttgtttattatctatacacaattttactt
gtatttttttgttttcagaaacttttctttgtccaaagtaaaagaaacatttctttttaatttagaatatacccaaattttgaagtaaaacaaaagtaaaatcagaatttatagtatatattataactaaa
agttaaatacaaaattagttgaaattaactgaatttgtaatattttaacagtcaaaaatctatttggaatataaaattaaattagttaatttacttagaataagaaaaaaaacaaaggtaaaaaatt
aaaaaaaaaacaaaggggatgatgcaatgtaatgtatgtgagaggggaatagtaatgtttgatgatggccatggggacatggaggagggaagggtgagatctaaaaaatgagagag

cataaagaaaggggtcgtggtggcaaaggcagaagaagtatcggtacgccacgtcaggttcggggaaatcttaaaccgagtggagaaagcctccaaaatctctttctctcttcctcttc
ctcttccatttccacaaagacacaatctctcgtcttctcctctctctcttttctgtaataataataataataatcctcctctccagaaaatcacttctctgatatttctttctttctttcctcctaaatctttt
ctctctctttcaccactgatattcttcccctcatctatttttttttagctctcaaataaaacaaaatcgtaaaagggaaccaaaccagagcaatctcatagaagattccttctgacaagtccgtgttt
ggttttctatggttggttcttcagattatatgagatcttgtagttaaattgtgacgaattctttgtacttgcattttctgtcagcaaagatctgttctttatcactgtgtccatttcaaataacacaatctt
gtcttttcttatatctgattgagatttttcttcttcttcttgattgtgttgtatctacaggaaggagcctatatcagttttaaccagcccttgtataagagatattccctttttttttatagaaagaactcga
tttactgtattagagtagatcaagaaccgttacgacccattgattctcgattggttagaaaaaagggtttgtaaaccattctctctagatg

>Capsella_rubella_MG

TAAAATTCGCAATGCGTTTGCTGAACTTTGTTTCTTTTCTTTAGTTTCCTTGTGGACTCGTTTTCTTTTTTCTTTTT
TTATAGCATTTTTCTAAACTTAGTTTATGGTTTTTGTGTCTTTTAACTTAGTTTATGAAGTTTTTTTTATCTGAGC
AATGAAGATTTTTTTAAATAATCTGCGGAAACAAAGCATTGTCTCAATTTCCGACCTCCAAAATCTTTATACCT
AGAAAAATAAAATGAATGGACATGTAGCGACATACTTAAAATTTAGGTTGTTCATGGTCAGCAAAAAGAGACA
AAAGCCCTTCTATCATGGTAGTAGTACAGTACTATACTGATGTGAGTGTGTACAGTTATGAATTGATTCGAATG
TGTCCTACATCAAAATTTGTGGTTTTTATTCGACATGGTTCGTTTCAATTATATGCACACATGATTACTCCTTTTT
CTTTTTTGGATGATATGTAGAATGAATACAAGAGTCTCTTTATGTTTTTATAGTCGGAAACAAAGCCATTCAAT
TTCAATGATTGTAACATTCATTTCATTAGTATATGTTTTGATATGACGTTGAATCAGAATGGTTTATGTTGATCG
GAATATACTCTCAGTCACAACAAGAATCTAATTATTCAATTAGTTTATTCGGAAACACACATGGTGGCATATAA
CTATACTAGCTGAGAGATATTATAAAAAAGTTTGGTACATGAACCATCTCTCTCTGGTCATGTAAAAAGATATT
ATGAAAAGTCAACATTAGTGCTAATTTCAAATTGACCATGCTCTAACGCACAAGAAGATTTGTTTATATAATAA
AGATGCTTGAATAAAGAAAGAGAAGAGTGTGAGTGGGGGTTGATACGTGTCACCGTCCAAAATGCGTCTGACC
CTTCTCATTGGAGCTACACTAGCCGACCCTTTTGCCTCATCATTTGTTGACTTCAGTCACCTTTTTTTACATAATT
TTAATTAGCTTACTTAATAATGTTTCCATTCATCGTCTGATTCTAGATGGTTCCATCCACTTTCATTTTTCTCTTT
GACATTTACTTTCAAACGTAGAAATCATATTAACGTAGTATATGATTTGTTTGTTTTTAGAATTTGACAAAAAA
ACAAAATTGAATTTGATCATTGTATGTATCTTTTTTGAGTTGAGATGATAGGTTAGTATGTTAATATCACAACTA
AATTTTAATGACACTTATTTTTTTTTACTTTTTTATCTGAATAGTAAATAAAAACAATCTAATCCATAAATAAATA
TATCTGCCTCAACCGCCATTAACCATTAAATCTCTAAATATGTTTGAGGTTCCGTTTTCTAAACTTATGCGTTTG
CAAAGATGATGTCCGATTTGATTTTTCTTGTAACAAATTGATTGATGTCACGGTCATTCTGTGAATAAAAGTAT
CTTGTGTCAAAGAAGGAAAAGAAAGTATATTGTATAAATTTTATAGTAGAGTTAATAACTTGAAGAAAAATAA
TGGCATTTTATAAAGGTTTTGTTTGACAATGAAAGGGGTAATGAATATCTAGCTTGCTAAACCACAAAAGAATC
AATTCATTGGACTATAACTTAAACATATAATATATTTATAGCTAAGCGAAATGTCTACAATTTGAGAATACGAC
TTTTTCGAAAAAGTTAATAGAACTCTTATATAAAATTTCTCCACCATAATGTTACTTTATTAAAGCTTAAGTATT
AAATTGATATATATATTTACAATAATATAAAAAAATGTAGGCAAAGTAGCTTTAAATAAAATTATTTGTCTTTT
CGCTAATGGTAAAACATAAACCAAAATTTGAGCTAATTCTGAAACGCGTTTTACCATGATCAGCATAATACATC
AATGTTATAACACTTTCATCATCTTGATTAATAATATCTAATCAGTAAGTTCTCCTTAAAACACTTTAATAAGCT
AAAGTTGTTTGTTAATTTTCTCTTCATCTCTTCTCTCAAGATATCTACTGGAACAAATAAATAATGTATATCGTG
TAACACAGTAATAATTACTAATATATTTTTATTATATATACAAATAATTAATGAGTATAAAAAGTAAAAATTGC
TATCATTTAGAAAACTTTATATTTGTTTCACTGTTAATTCTAGAACAAAAAAAATGCGGTATTGTTTATTATAAT
TAGATCGTGAAAAATATGATAGTTAAAATGTAAATTTATATTTATTAAAAAAATAATTATAGTATAAACATAA
AATAATTTATTTTAATTATTGTAATTGAGTTATTTTCAGTCCAGTTATGCAGCTCCACGTATGGTTTGAGCCCAT
GACATTAACATGGTTACATGACCATGAAAAAGGCCCAGTTATGTTCACCATATTCAGAGTTTCAGACACAGAC
ACCAGCCAAGCGATCCAGACAGAACACACAAGTGGACAGCACCAATCACGAATCGTATGGAGATCTGAATAC
AATATATTGCTCGCATCTTATTGCGCCACGTCTGTTTGTTTTTTAAGGCGCAAGATCAACGTCCATAGATCTCTT
GCTCCACTCTGATTGGTACACGTCACCTCGGTTATGTGGTATAGAGTAGTAGACAAGTGGGAGAAAGATTAAA
TTAATATCTTAAAGGGAAATTGTGAGGCCCATTTAGATATTTTCTCAACGGGCTGATCTCAAAAGCTGCCCGGT
TTATTTTACAGCCATCCGACGGCTCAAAACTGTTCCGGGAAAGCTGCCCGAAACTTTTTAAAGGAATTTTATAG
GTGGCACTCACTGATGAGTGAATCGACGGCGGAAAAATTCGAAGGTGTACCCAACATCTCGTAGCATCTGAAC
AAAGCTCCTCCCAAAATATCTGTGGACACATGTCATGTCACTAGTTTTTGTTTGGGCGAAAATATTAATTAAGT
AGGCAATCACATGAGAGGGTTATATAGTTAAATTAATAAAAATAGGAGAATAATACAATCATGACCATACCAA
ACCTTTAAAAATAAATAATCTCTGCTTACTTCAACAAATTCTTTACCAAATATTTACCATACACACTGAATCTGT
ATAGTTTTTATTTTAAAACAACCACATTTCCGGGGTAATTTTATGATGTAAGGAGATTTTTAGAAACTTTACTTG
TGGCTTAATAGATCAATTTTTGGAGTAAATCAAGTTTACAAACAAAAACTTGTTTATTCAGATTTAGAAAACTT
AATATTTAAATATTTCCTTGTTAACAACATGCAATATACATATTTGAATAAAATTTGACAAGAATACATATTTG
AATAAAATTTTGATTCGGAAACAACAATACAAGCACAAACCATTATACTGTTTAAGAAATAAAAATTACCGGA
GTTGGGTTTAAGTTAAATAAATGAGGGATAAATAATCCTTTGTTTCAAATTAAAAATGATTACAACGGGAAAA
AATTAGTCTTTGAACTAAACTATTTTTTATTTACTGTGTTATAATTTCGTAGTGGTAAAGATTCAAACAAAAATT
GAAAAAAACAGGTTTGATGAATTTATCCATTTTAACTGTGGTAACCATAGTTGGTAAATTTTTGTTTTTAACCGT
AGTATGAGATGGACATGTATGTTAATGGTGAGGAGGGTGATGGTATGTGGCTATGTGCCATAGTGCCATGGTA
CCAATAGGCGCATAAACAAAAAAAAAGGGTACGTGGCAACGCTAGGAAAATATCGATGACACGTCAGCCGGA
CGGGAAATCTTAAACCGAGTGGAGAAGCCTCTCCAAATCTCTTCCTCTTTCTCTCTCTCTCTTCCTCTCAAAA
GGCCACCACCACAATCTCAATCTCTTCTCTGTATTTTTTCTTGCTTCAGAATCCTTTTTTTCTTTCTCTCTTGATAT
TTTTTCCTCATCTTTTTCTCCGATCTCTCAGAAAAATATCCTCATAGCTCAAAAAGCAAAGAGTAGAGAGAGAT
AAATCAACAAATCTCATAGGAAGATTCCTTACAGGTCTTTCTTCTACTTTACTTGTGTGTGTCTTTTTCTTTCTTC
TTTTGGGATTGATTGATTCTCAGATGAGTTTTTTTCTTTCTGAATCATTTGGTAATTTTTTTGTTTGATTTGCTTA
ATTTTTCCCAGTTTGGTGTTTGTGTGGTGTTCGTTCTCGTTACTGTTTCTACTTGCATTTTTATTTTCTTCCTCAGC
TTAGATTTGTTGTTTGATTTTTCAAATTTCTTTGTATCACATCTCGTATTGTTTTATCTGATTGAGAATTTTTGTTT
TCCCTCTACAGGAAGGAGCGAATATATAAGAGAGACACAGACACACACGGATTTAGCGTAATAACTCAGCTTC

GATTGACTGTACAGTGTTTTGGTTAGTTTAAAATCTTAGGAGGGTAGTGATTGTTCTGAATCTCGCCGGGGAAT
GTTAAAACTAGGACGTATTGAAAGGTGTTGGAGATCAAGGGTGTAAAGAGATTCTTCATTTTTAGTTTCGGGTT
TCTGGTATATTTTTATTCTTCTGAGATG

>Diplotaxis_erucoides
tacacctgatagagtccccagatgcactcaagtcttaaatttctgaagctttggatgttggctagtttaacgttggagtatttttttagttttcttagtaatttttatttaattttaaacttagcctatgtt
tttgaaacttttaactcaatggggttggtccattttattaatatttaacgctacatgaaatttaagctgttatcatttggtattgccattgtcagcaactggaagccactcaatcatagtatctgcact
gttgagtgacatgatgatgcgaatccaaacgtgggagtttgctgtaatgaattgattcgaatgtgtcctacttcacatttgtggtttggagtcagcatgcttgatttcaataacatatacccaca
aacttatttttttgtggttttaaaattggaaacgaacaactcaatgattataccattttataacaaatggtttttgatatgacgttgattacaatggtttatgttgatcgaatattctctgaacaagag
caagaatctaactcattttagtttattcggaaaactcattttggcatataataatggttttgctgagacttcatcagatttttgttttgattaaaatgtgagatgtcacatgcaatgagtgaaggaac
aatctctgttcatgtgaaatttattattagtagtcaacatcagtgctctaatttttcaaattgactcgttataaaatatgattgaacaaagaacagatggtgagtgggtttgatacgtggcaccgg
ccaaaatgtggctgccccttctcattggagcttcactagccgacttttgtcttatcatttgttgacttcaccaacctttttacatatttttaatgaaattgaaaactcatttctttcggtttctagattg
ttccatccacttgtttctttttgtattaacctagtgaatgaacatatttctgtttgtgtttgtatatttttaccgggatttagatactttcatttatttggtgttgaaaacaagatgaattataattgaaaatg
gatcctaatatgcaaggaagctattttatgtaaacagttaaacaccacacctaaactacgacctaattacagtaaaactgcaaaacaaggtaatatataatcagtaatggtgaattctacattt
agatattataaaagattaaagtatgacctaattagttaatcaagtttgagtttagttttgttgaagttgcgttttgaaacctacctcggttggaaaaatgttcgattctcttatattaaattgattaatg
tcacagttctcctgtgaataaagtatgtaacataataaataatagagtaacacttgaagacgaataaaaaagactgttgacaatgcaaggtgtaatgaatatctcgctcagtaaaccacaat
agaatcgattcattggacaaaaacttaaattaaatcaaaagaaatgtctaccgtgtgagagcttatgacttttcaaaaagcttttgaagttttgtaaaaagttttctccaccatcgtatcacttta
aatagaatatggttttcaacatgatataataacaaaaaagtagacacgctagctttatcagacctgaatactctttagattaacaatcttttaattcagaaaggtaaatagtaaaccaaaattga
agctaagttcaatgaataaagctttgtatcgccactcttaataagtttgcaatgaccttcacaagaaacacaatctgataatcaataataatgaatatttgtgtacaagtttggttattagaaca
atcatgcaactgctttgtgagtcgttataagtaactaggtcaactgcgctaggaaatacataaacatgcaagaacatagagaaatagtcaatgtcatttagaacatggatgtggtgtgtcat
cttagaggttagttcggcaatcctgcagcaattatcatcccttctgatatcacatcattagaacatatacgtgcttcatatgttgaaacaatgatagcttccataaattattgggaaacgcaaat
gatgtgatgttcgtttgtttatttttatacgttagtgcttatgaaattcatttacatcaaccattgagctattaattatatataaaaaggtatcacacataacgttcaagtcatcaatttcaacaatcata
aatttgaggtataagaaaatttacattatttatttttcttttctaatgttcaatataattaaagatagtattttaaaatttttcagaataaactatactctctccgtttcatattaagtgtcaccgtaaagaaa
attttcgttgcaaaataaatgtcgttttagtatttcaatgcaacatttattaattttatttttcagattatttttctattggttgaaatatgagtagatgtatgagaaatgatgtttttatattgaaaataaa
tataaaattaaatgatttattaatttatgtgcagaagtttaaaatgacacttataatgaaacagagtgataatgagtatggaagtagttacagtgctttataattaaaaataaatagatattttatatt
aaaatggaaataattttcagcccagttacaagtctccacgtatagagcccatcccatatatatgcatgaccatgaaaagcccacagttatgttctatgttcagacaccaacctgccagccat
ccaccagagtgtggacccacacaccacgaatcgtatggtagatcgctacgatatattgctggcatcctgttgcgccacgtcttcgttaggaggaagatcaacgcccatatatctcttgc
gcctctctcattggtacacatcgttttgttatatattcaacttgtttatgtggtggagtggacaagtgggaagaagatgataaatatctaaaaagagtgtgggtcacagtttaggcttgagatat
tttcacaacgggctgtctccacgaaaagctgcctccttaaccgataccaccatcagacggctcgaatctcttcgggtcaagctgcccgaaaaactttgttgaggaatttgtgtgggactc
acccaccactgagatcaacggaggaaatttcgaaggtgcacgtaacacatcgtagcatctgaacgtagctccttcgaaaatatctaggacacatgtcactttgctttgtaagaaaagta
attaactaatcacatgacagagtttagtgttattaaactaataaaaaggaggtaataatacaatcatgacccacctgtaaataacagtatgttattaacggggaagagtttatttaaccgatct
caccataccatgataccataaattaggttttaatatttaaataaacaacagataagcttatttatttatacacaattttgatggtaaatttttgttttgcagaactttttttttgtcgaaagaaaaataaa
attttcttttaatttagaatatagcagattttgaagtataccaaataaacaatcagattttataatatgttatatatgaataaaaattataacaattaaatattaaattagttaaaatgaactgaatttg
taatttttttacagtaaaaatgtcaatttggagtttaaattagttaatttactgaggacaaaacaaaacaatggtaaattaaaaaaaccaaggggggaatgcatgtatgtgatgatgcaaatggg
catggacatggaggaggacgagatctaaaaacgagagcataaagaaagggtcgtggtggcaaaggcagagaagaatcggtagtacgccacgtcagggtcgggaaatcttaaacc
gagtggagaagcctccaaaatctctttctcgcttcctctcatttttacaaagtacacacccaatctctctctctctctttctttttatctcctcctctttgtgtgataataatcctccagaatcat
ctctgatatttctttccacctaaatctttttctctttccccactactgattattcatccctcaccgagtttctttagctctcacaaaaaatatccagaaaaaaatagacagagacagaactaaagcaga
gcaaactcatagaggacgattccatctgacaggttcgtgtttggtttcttctcttcttcttcgatatggttgattcttcagattatatgagatcttgtagtaaaactcctgtcgaattctcctttaattta
gttttttttttctacttgcatttctctcaccaaagatctgttctttctcgctctgttcatttcaaaagaccacatcttgtctttgtctttttgtctgattgagattctttcttctttgtatctacaggaaggagc
caatatcagttttaccagccttgttataagagtgattccttttttagagaaaataactctatagtttacagttaattctccagcttcgatctactactatgctgtgtgttccgtagtggagtagatca
aggactattgattctcgattgatagaaaaaaaagttggagctttttatttttggtttactggtggtggttaaggttgtaaccaattctctctagatg

> Sinapis_alba
tgaaatctgcaatgcgttaatcgagctcaatttttatttttcttttaatttatttctagttttgttatggattatttctttactctttttagcatttttttttttcttttttaacttagtttatgttttttatgtctttttaatttaa
tgaggttgatcattctagttagagtttttgtccggacgatgatgatcttaatttctaataatctgccattatacgattagtctctctagaaatctgaatgaactttgtcgtctacgtatcttgggttaa
gaagcaacattagtaatggtcattgatctcattaaaggtaaattcgaccgtctcttctgtcccgcattgagggtctttaccatgaaattaacttttcctcacgtgttatgtcagattgtcagaatat
gattgaatgatatgaccaaaccatgtaacttagtagtttgtattagtgaggtatacgtgacaactatacatattgttgacataagtatatttatgattggatggtatatgttcttgaaaaatttaagt
tgttcagtagtcattgtcagcaaaaacatgccttctattatggtagcagtacagtgagtgtgtcatactcaaatttgtggtttatattcatatcatggttgatttcaattatgtgcacacagactcat
ctttttttttttctctctatttttataattggaaacaagcaattcaatgataataacattcatttcataataaatgtttgatatgacgttaatttgaatggcttatgttgatcgaatataccctcaatcccaa
ctagaatctagttcacttaacttattcggaaaacacataggtggcatgtagtgtgtaaactatatatgtattagctgagacttattattataggagagagtacatgaatgaaccatctctgttcat
gtgaaagatattataaagagtcaacatcagtgctaatttcaactatactttagcaccttaagaagattaaaaaacga
ttgaataaataagtgggtaagtggggttgatacgtgtcaccgtccaaaatgtggctgacgacccttctgattggagctatagccgacccttgcctcatcatttgttgacttaagtcaccttta
catattttaattaacttgccatcgctcaactcattcttcggttctagattgttccatacattttcgttttctcttttacatttacttttgaatttcgaaacgtaaaatatattaacgtagtaaatgagtagtt
cttttgtaatatttgaccaaattttgaatttagatccttacatctattatgagttgaaattaaaataaatcataatacgtaagtaatttattgtatgtaacaccacatctacgattctaaactatgatcc
gatgacactaaaacaacacatctccaagcagtaatattaactacgtatataacttaagattttcaaatctaaagtataacttaatgaataaaccaaatttttgagcttaattttgttgaggttgcgtt
ttatgttttgcgttggttcctttgcaaaaatgttcgatttgaatttttgttttcgtttcttctattaattgattgatgtcacggtccttctatgagaaattattattgtataatgtcagaaaaagaaaaaa
gaaagtatattgtataatagagtaatacttgaagaaaaataatagcatcgtaaaggttttattgacaatgaaaggggtaatgaatatctcgcttgctaaaccacaaaagaatcaattcattggt
ctataacttaaacatataatatatctaccaaaatgtctaccatttgagaaaacgactttaaaaaaaagttttcgagctttataaaattttccccaccatagtattactttaataaaagcttgttattaa
aatgataaaaaaaaaagaaagtaggcaaagtagctttattataattaatgtttttctctaatggtaaaccataaaccaaaaattaatggttt
tcgctaaagcttctaatttttgttttcactgccatgtataacgcgtttactatgatcaacatgatacatcttaacatcaatgttacacttcatcatctagatatataacttttataatttgtttcagtgtta
gttcttatttgttacatataatttgacgacttgagaaatatatgcattatataatctgatttttgttttttaatactaaaatacgattattttcagcccagtaacatgtctccacgtatggtttgagcccat
agcatacacatggaacatgaccatgaaaagcccagttatgtgtccatctccagacaccagccatcgacagaagtggaccacacaatcacgaatcgtatggtgattactacaatataatg
ctcgcatcttattgcgccacgtctgttgtttggcggaagatcaacgtccatatatcttatgctccattatgattggtacacgtcatcttgttatgttctacttgtttatgtggtggcgtagacaagt
gggaagaagaaaatatctttaaaggaagccttggacccatttagatattttgaacgggctggttcttcaaaagctgcccgttgatctatacccatcatcaagacgggtctatactcttccgg
gcaaagctgcccgaagtttttaaaagaattttgtaggtgggactcactcaacactgaatcaacggtggaaaaattcgagggtgtacgtaacacctcgtagcatctgaacgtagctcctcc

caaaatatcttcggacacatgtcactagtttcgtaaactattaacttaaatcagagaatcacatgagagaccatatagtttaaattaattaaaaggaaattatgcaatcatggaccaacaaact
tttaaataatgatctgtaaggttagctatacaaaattctttaccaatattaccattctctgaatctagagttttattttaaaacaattacaatttttctgctaattttatgttt
gagtctttgaggaggtaagcagtgcaattcaaatagtaacaaaaatacatatatatatatatatatatagatattttgacactattacgatactgttgcagtgcagttttggagtgcggtatggtt
accatttagagcctagttttggagtaaattaagttaacaagtacactagattttgtgatatacttttaccgtcaaaatctatttttaatattaggattaaaaattaaattacttaattattgagaacaa
atatttgtaaacagtattttaaagagatgagtcatatactgtattaccaaatatgaatttcttgatttgacacaaaagtcttataaattcaaaatcaaacagttttaacacgaggaataagcaaag
agatgagtaataacatgtatgtgtaatgtgtatggtgatgtgatgttatgttatgggatgatgaaatgtataatggcgcataaaggtcaggaaatatcgatgccacgtcaaacggaaatctta
aaccgagtggagaagcctccaaatctctttctctctctctccttcctcctcctcaaaggccacaatctcctctgtgtgtatgcttcacaatcttttctgattttatttcttcttcacctatattcttc
catccctcaaaagtatccacaaagctaaaagctagacaaagggaagagagatataatcagattcctgacaggtgtctttctttcttctctactttgtcatgtgtctttttgttgaattctcagatga
gatttttccgattcatttggtaaaaccttgaggatagatttgcttaatttctcaaatttggtgtttagttttgtttactgtttctaacttgcatttttttctcagcttagatttgttgtctttctctgtgtttaat
aacttttctgatttgaggttttttttttcttcttctgaattgttgtatccacaggaaggatcttatacttttggtcttcgctatagagataaa
agccctcaagtgttaaccaacaacaacacacaggatttagcgtcttttcttcagccctcggattttactgtacagtgtgttggttgattaaaggagttatttctgggggatgctgttgtgtatag
tgattgttcttgctggcgaatgttaaactacgacctattgattctcgattggtagaaaggtgtttggagatcagactcttgttgcctcttttaaaaaaaagattcttcttttttggtcttgttagattgg
tgtaatcaaatcactctagtgggttagttttggtttctggtattgattcttctggatg

>Turritis_glabra
tgaaatctgcaatgcgtttgctgaactctacttttttctagttttgttatggaatctttctttactctttttagcagttttaaatttagcttatgattcttgtgtctttaacttagtttatggagtttttgtctgga
cgatgatgatttcttaattatctgcgattagccgactagtctttttagaaatctaagtaaatttcctcttggaattcctaaccgtactgtcctaaaatcttatgcctaaaaatatatagggtgtagag
ttatagagggaataacgcggcaatcccaaatggatccttaatcacaaatttttacctcgtcggatgaatcaacatgaccacgctattcctgagtgaggatcttaccagtgtctaatcatgaaa
ttaactttccctcacaaatgtgttatctattctagttaggtctgaatatggttgggtggtgtggccatgacggatatatagcgacttagtgatttgtgataggggtgtttgtgcaacaccacccg
ggttaagtccacgtataatgcatgttattgacgagaaaaacaaatatatgatcgaatgatctatgttatgatgtttgtttctttatgacctgaatttcagatgctgtatgttttggaagaaaaagtt
aatagtttcaacaacaacaaaaaaaaaatgaactgaaaagaccgaactgacttgctgaattcccaaatgcccaactctacttgaaatttggttgttcggtagtcttggtcaacaacagtaat
atactgacgggaatccatatgtgagtgtgtatagtaatgaatttattcgaatgtgtcctacataaaatttgtggtttgtattggacatggttgatttcaattatatgcacgcactctcttttgtggttt
gtattcaacatagttgatttcaattatatgcacacaactcttttcttggggtgttgaatataagaggctctttatatttttataattggaagcaagccct
tcaatgattgtaacattcatttcactatacatgttttgatatgacgttaatcagaatggtttaagttgatcgaatatactctcaatcacaacaagaatctacaattcaattgatttattcggaaaaca
catggtggcatataatattatatactagctgagagttcttactagatttattttgagatggtacacgaaccatctctggtcatgtaaaagatagtataaaaagtcaacatcagtgctaatttca
aatttagcgcacaagaagatttaaaaaaatgattgaataaagaagagagtgagtggggttgatacgtgtcaccgtccaaaatgtgtctgacccttctcattggagctacactagccgacc
ctttgcctcatcatttgttgacttaagtcgccttttacatattttaattaactttgctaatgtttccattcattgttcgattttctagattgttccatccacttttcttttctctttgacatttacttttgaaacgt
agaaacatattaacgtagtatagggatttgtttgtttttagtatttgacaaaatttttgaatttagatccttatatctattttgagtagagataaaaatgaaattatgttaggttagggagtatttaatgtta
ataccacatctaaattatgatcgggtgacacaaaaacaacatatctctaatcagtaatctttaatacatttatatatcattaaaaagattcacaaatttgaagtataacatattgaataaatcaagt
tttgagcttaaatttgtttgaggttgcgttttctaaatttttgcgttgcaaagatgttcgatttgatttttcttatatagttgattgatgtcacggtccttctgtgaaaaaagtatcttgtgttaatgtcaa
agaagaaaaataaattatattgtacaaatttatataataatagagtaatacttgaagaaaaaataatagcattttaaaggttatgttgaca
atgaaaaaggggtaatgaatatctcgcttgcgaaaccacaaaagaatcaattcattggactataacttaaatatataatatatgtagcgaaatgtctacaatttgagaatacgacttttcaaaa
aagttatagaactcttataaaatttttctccaccataatcttacttttataaaagcttagtattaaaatgatatatttacaataaaaaaaagtaggcaaagtagctttattaaaattaatgtctttcgct
aatggtaaaacataaaccaaaatttgagctaattgcgaaacgcgtgaaatgatcagcatgatacatcaatgtaacacttcatcatcttgattaatatctaatcagtaagttctcctcaaaacac
tttttataagctatttgtaagctaatgttttctccatctcttctctgtggatgtctattgaaacaaagaaatattgcataacgtgtaaacatacatagtaattacctttttaatatataaaatttattaatta
gtataaaacaaaataaaaaattgctataatttataaaactctataatttgttcagcgttaatctataactaaaaaatatatgcagtatgatttattatatttagacagaaaaaatatacattattattta
ttttaaaagcaattacagtgctaaataataattaaccaaatgattgttttctattattacagttattttcagcccagttacataactccacgtatggtttgagcccatcacattaacatggtgcatac
catagcccagtgatgtaaccatattcagacaccagccatccagacagaagtagactccgcaaacacgaatcgtatggagatcactacaatatattgctcgcatcttattgcgccacgtct
gttgttgggaggaagatcaacgtccatagatctcttgctccactctgattggtacacgtcatcttggttatgttttacttgtttatgtggtagagtggccaa
gtgggaggattataaaatatcttaaagcaagtgtgaggcccattttagatattttctcaacgggctgttcctcaaaagctgcccgttaatctacagccatcgaacggttcgaaactgttctgg
gcaagctgcccgaaacttttaagggaatttgtgggtgggactcactcgtaactgaatcgacggtggaagatttgagggtgtacgtaacacctcatagcatctgaacgtatctcctcccaa
aatatctgtggacacacgtcactagttctttgtgaaactattaagttattaggcaatcacatgagaggggttatattgttaaataaataaaaaggaaataatacaatcatgaccacaccaaac
ctttaaaaatatataatctctgattacctcaacaaattctttaccaaatattgccatattgtgaatctatagttctatttttaaaacaaccacaatttttggggttaattttatgttttaagaagagtttaa
aactcaactcatgactaaaaaaaaaaattatatattttggagtaaatcgaatttacagtcagatttgttctacttttacaggtaaaattgtttattaaaaactaattatgtaaatcagtaaacatttttt
tgttaataacatgtaatatacgaatatgaattcgaattgtgattcaaaaacaacactagaacaaaccactaaactatttaaaaaacaaaactagaattgtgtataattgagctgatgaatcaga
taactttaaaaattaagaaaaaaaaggtaaaatttttaaccaatagtatatggaatggacgtgtaagtgattgtgaggatgatggtatgtaccaatgggggcataaaaaaggtcgtggcaa
agctaggaaatatcgatgacacgtcagccggacgggaaatcttaaaccgagtggagaagcctctccaaatccctttctctctctcctcttcccctcaaaggccaccacaatcaaa
tctcttctgtgtatgttttgcttcagaatccttttttttctttctttctgataatttgttttcatatttttctccgatactttcgatctctcaaaaaatatcctcaaagctaaaagcaaagagtagagatata
atcaacaatctcataggaagattccttgcaggtctttcttcttctactacttgtgtatctttttttttcgggttgattgattctcagattagttttttccgaatcacttggtgcttttattttttgaatttctt
ggtgaatttttgttcatggattttccttgatttacttaatttctccagttttggtgttgtgtgtgtgtttgttctcgtcactgtttctacttgcatttttttttttctttctcagcttagatttgttgtttgatttca
atttttttgtatcacatctttgtcccgttttatctgattgagattattttatttacaggaaggagcgaacatctaagagacacacggatttagagttataattcagcttcgattaactgtacagtgtgt
tggtagtataaaaggagtttaaatcttgggagggtgatgtgtatagagtttgttgttcttgctggcgaatgttagacgacgacgtattgaaaggtgttggagatcaaggctagtgcacttttgt
tgcctcgattcttctgtgggtgttagattgtgttatcaatcattagtttcggtttctggtaattaattcttctgggatg

> Arabis_alpina
AAATTCTCTAAAATCTCATTTTGCTAGTTGTTTTTTGTTTGCTTGATTTCTAGTTTTGTTATGGAGTCTTTC
TTTCTTTATTTACTATCATTAGTTGTTTAATTTTTTTAACTTAGTTTATGGTTTTTGAGTCTTTGATTTTAT
GTAGTTCGCGTTATTTAGAGTTTTTGTTCCGAAAGATTATGATCTCTTCTTAATAATTTTTTATTAGCCGAT
TAGCATTTTTTAGAAATCTGAGTGAAATACTTGAATTTCAGATTTTATATGCATTTGATACAAAAAAGGGAT
AATAGTTTCTGATTTTTTTTATAAGAAAAATCTTTATTGAAACTGGATTTATGAATGACCAAATGCCCACAA
CTCTACTTGAAACTGAAGTTGTTTGTAGTATCAACAACAAGAAGCTTTGTATTGTGGTTGCAGTAGAGTAAG
ATATTCACAACAGTAATGAATTTATTCGGATGTGTCCTACAAATAATTTGGTGCTTGTACAATTCATGGTTA
ATTTCTGTTATACCACAGATACATTTTTGTTTTGTTTTGTTTTGTGTCTGTTGAATTTGCTAACATAGAATT
ATGTTTTTATAATTGGAAATAAGTAATTCAATGATTATAACATTCATTTCGTGATATGACGTTGTTTAAAAT
GGTTTATGTTGATGGAATATACTCTCAATTACAAGAAGAATCTAGTTCAGTTAATTTATTCGGAAATCACAT
TACACATAAAATAATAACGTGGGATGATACATGATTGATGTTTGTGGAAGATATTATGAAAAGCCAACATCA

GTGCTTCTAATTTGGAATTGACCATACTTTTAATTAGCAAATTAAAAAAATGATTGAATTAAAGAAGAGAGT
GAGTGGGGTTGATACTTGTCAGTGTCACCGTCCAAAATGTGGCTGACCCTTTCTCATTGGTGCGACACTAGC
CGACCTTTGCCTGCCTCATCATTTGTTGACTTTGGTCACCTTTTTTAAACTATTTTCTTGTCGTCATTTTA
ATCAAACTTTTAAGCAATGTTCATTGTTTGGTTCTAGATTGTTCCATTTCATTTTATGTTTCTCTCTCTCTT
TGACATATTAGTTTTAAACTTTACATTATAGTATTCAGTGTATTTTTTTAGTATTTTGAATTTTGATCCTTA
CATCTATGTTGAGTTGAAATTGAAATGGAATACGTTACGGGATAATTTATGCAAACATTAGATGATCCGATG
GTATAAGAAAACAACATTTCTTAATTTTATTTTATAAACCATTTAAAATTTGATGGTAATATATTGACAACA
TTAGTACAAATCAGTGAATCACTACGTTATATATCATCGTCTAAACATAACTATATTCATCTATATTCAGTA
GTATTTGCCACATTTGTAATTCATTGTTATATTAAAAGTATGGTCCAATGAGCCACTCAAGTTTGCTCCTAT
TTTTTCTATAAAAAAAAAGTTTGCTCCTAATTTTGTTAAATATTTTTTTTTTTAAATAAATTTTAGAAGTTC
CAAATTTTTAATCTATGAGTTCTTCCGGTATCTAAAATCAATTGTCGATAATAGTATGTTTATAACGTAGAT
CATCATGTGTATAAAAGATTTAATAAAAAGAAAATTATGGGTATATCTTTCTTTTGAGAAAAAAATTATAGG
TGCTGAATTAGTGGCTTTAAATATTTTCAAAAGCCAAGAGATTATAGTTCATGGTGTATTATATAGAATAAA
AAAATGGAAGCAAAATATGTTTGGATCCACTATACTGTATTAAAGTAAAAAACTTGGTATCGAATATTTTAA
TTTTTTCCAGCATAAATTTGTTCAATAGATAAAAGTGGTGGGTAGATTTATTCTAATATGGTAAATCTAAAA
AAATGATTTACTTGATCCAACTTTATTATTCTATTATAACATACGAGTCACACCTGGTTATTTTTATTTATT
TATGTAATTTCGTTTTAAATCTCTACGTTGCAAAATTAGTGGTCAATCGGAATTTGTATTTCTTTCTTATAT
TAAATTGATTGATGTCACGGATCTGCCATCTGCCGTGAAAAAAGTTTATTGTATAAACTGAGTAATAATACA
TGGAGAAAAATAATACTATCAATTAAAAGTTGAAAATGAAAGGGGTAATGAATATCTCGCTCACTATACCAC
AATAGAATCAATTCATTGGAGTATAACTTAAACATATAATATATCCAACGAAATGTCTACCGTTTGAGAATA
CCACTTTAAAAAAAAATATTAAAAATTCTTCACTATAGTTACACTTTAATAAAAAAAAAAATTGGTATTACA
AAATAGTAAATTGTATATCTACAAATTTATGGAAATTTTTTTAAAATGTCACAAACGTTAAAAATCTAATGT
AAACGTATTTCTATTAATTTGCATATTTATTAATTATTATATTATTTGGAGATCGAATCTCATATATTTTTT
GGTAAACCGTTTAAAAAGTATTTTTTTTTTAAACGTCTGAATTTTTTTTTTTACAAATTAGGAAACTATTAC
ATCGACGATTCTACGATGAGCATGTCTTCCGTCTTGATGTACACCACACCATACGCTTTTACCCTACTCCAA
ATTGAAGATCTCCTGCAAACTTGTCTCCATTCATCTGCAAATATAATAAACTGCATATTCAAATATTCGAAC
CCCAAATATTTTGGTGTAAAATTTTTATAGCCCTTAGTTAATCACTACGTCAACGGAGCTTCTAAGCTCCAC
TAATTTTTGGGTGTTTATAAGAGATTTAAAAATAGAATGTTTTCCACTAAATTAGAAATGATTATAAGCCCA
ATTACGTACGAATCTCCACGTATGGTCTGAGCCCATCCATACATGGGAGGCTCAAGGCTGGGAGCATGCATG
ACCATGAAAAAGCCCATTCATCTTCAGACTCCAGCCATCTAAAGAGTGGACCCCACACACCACGAATCGTAT
GGAGATCATTACAATAAATTGCTCGCATCTTATTGCGCCACGTCTTCCTTTAGGCGGAAGATCAACGTCCAT
ATATCTCTTGCTCTACTTTCATTGGTCCACGTCATCGTGCTACTTGTTTATGTGGTAGACTAGACAAGTGGG
AAAATTATAAAATATCTTAAAAGGAGCATGGGACCCATTTAGATATTTTTCTCAACGGGCTGTGTCCTCAAA
AGCTGCCCGTTAATCCATATCCATCAAACGGCTTGAAACTCTTCCGGGTAGAAGCCCGAAAACTTTGTAAAG
GAATTTGTGTGTAGGACTCACTTATCACTGAATCAACGGTCGAGAGTTGAAGGTGTACGTAACATCTCGTAG
CATCTGAACGTGGCTCCTCACAAAATATCTTTGGACACACATGTCACTAGTTCTCGTAAACTAATTAACTAA
TCAGGCAATCACATGAGAGCTTGTAGTTAAATAAATTAATAAATGGAAATAATACAATCATTCATGACCAAA
CAAAGTAACAAACCTTTAAAAATTCCGTGATTAGTTTATTCATATTCTTTCCCAATACTACCGAATTTTTGT
GTAAACCAGAATTTAAAATTAGGTTGTGTGATATTTTTTCTTTTTTTTTTGTCATAGGTTTTGTGATATTTCT
GTGGCTAAATTCTGTAATGAAAAATTGAATTATTTAACTATCAAAAATAAATCTGTTTATATATATATATGG
CATATAGTCTAACACATTGTGCAAATTGATTTGATCGTGCAATAGTAATTTTAAATCCCATATAGTCTAATA
CTTTGTGTAATCTAATGGTAAATACCACTAGATGTCGTCGTATGGTAGTGGTTAAGCGCAAATGTTGGTTTT
GTTGTGCTCCTATTTCAAATCCCCGGTGGGAAGGATGTTTTACGCCATGTTATAAGATATTAGCAAAAAATA
ATCCCGAGCTTTGGGAAAGATTAAGGCCAACTCCTCCTAGTTATCAAAAAAAAAAACATAATGGTAAATAAT
TTGTTAAAAAACATAATTTGTTAAAAAAAACGTAAAATTTTACTTTTGTAAAATTCAATTTTCAGGCACAGA
CAATTATATTATGAGATGAATAATAGAACTTATTGATAAATATGTTTTTCCGCTAAGCATGATCTAATATCG
AAACTAAATCAACTCTACTAGACTTAGAATATTAGTATACGTCTTTGGAGTCGTTGCTAGTGGCGAACCTAT
ATCACAAGTAGATGATACATTTGTCCATCATAATTTATTATTTTTCCCTTGAGTTTTCTTTAAGCAATTAGT
TGAATACCTCCGTTTGGATCACTAAAAAATTACTACTAATATATATATGAATAACATTTTAAGAAGTTAAAA
CATAGTAATTTTAACCATACATTATATAAAAAAAAACATATCAATCCACTTATTTGCCTGTATGCGAGTGA
GAGGATGATGATGGTATGACCCATGGATGGTTGATGGATGATGGATGATGGATGATGGATGATGGATGGTAT
GAGCCGGGGGCATAAAGGTCGTGGCAAGTGCAGAAAATATCGATGCCACGTCAGCGGAAATCTTAAACCGAG
TGGAGAAGCCTCCAAATCTCTTTCTCTTCTCTCTCTCTCCCAACGCCACAATTTCACGACTTGTATCCTCTC
CACAATCATTTTTTTTCTCTTTTTTTCACTCGCCGACTTTTATCTCTCCAAAATATCCAACCACATACAAGA
GATTATTCTTCATACAGGTCTTGTTCTTGTTCTTGTTCTTGTTTTGGTAGATTCTCAGATAAGATT
TGTTGTGTTGATTTCACATCTCGCTGATTTGAGAGTTTTTTTCTTATCTACAGGAAGGGAGACGAATATCAG
TTTTACAATCCATCTACACTAGTTTTGATAGAGTGATTCCTTTTTAGGAACCAGATTCGATTTTACTGTACA
GAGTGTTGTTGGTAGATCAAACGACTAATCTGAGGTTTTATTGTATAGTGTTGTGTTCTTGACGGCAAATGT
TAAACTACGACCCGTATTGATTCCAGATCAAAGGGTTTAGGGATTTTTTTCTTTTGTTGGTGTTAGTCTGTA
ACCATTTCTAGTTAGTTTAGTTAGTTTCTGTTTTTTATTTGGATG

Sequence of 5 conserved blocks in the GI promoter of A.thaliana

>Block_1
TGTGGGATGATGATGGTTATATGGTAATGGCGCATAAAGGTGGTGGCAAAGGCAAGGAAATATCGATGACAC
GTAAGCAGACGGGAAATCTTAAACCGAGTGGAGAAGCCTCCAAATCTCTTTTTTCTCTCTCTCTCTCCTAAGGC
CACCACAATCTCTTCTTC

>Block_2
TAAAATTACAATTTAGCCCAGTTACATAACTCCACGTAAGGTTTGAGCCCATCACATCACATGGTACATGACCA
AAGCCCAGTTATGTTACCATGGAGTATTCTTCAGACAAGAGCCATCCAGACAGAAGTGGACCACACAATCACG
AATCGTATGGAGATCATTACAATATATTGCTCACATCTTATTGCGCCACGTCTCTTATTTGGCGAGTGATCAAC
GTCCACAGATCTCTTGCTCCACTCTGATTGGTACACGTCATCATTTTATTGTGCTACTTGTTTATGTGGTACAGT
AGAGACAAGTGGTAAGATTTTAAAATATCTTAAAAGATGTGTGAGGCCCATTTAGATATTTTCTCAACGGGCTG
TTCCTCAAAAGCTGCCCGTTTATCTACAACCATCGAACGGCTCGAAACTATTCTGGGCTAGCTGCCCGAAATCT
CTTAAAGGAATTTGTTGAGTGGGCTTGAATCAACGGTGGATAATTCGAAGGTGTACGTAACATCTCGTAGCATC
TGAACTTAGCTCCCCTCAAAATATCTGTGGACATATGTCACTAGCTTTTTTTTTGTGTAATGAAACTATTAATTA
ATTAGGCAATCACATGAGAGATGTTATATACTTAAATTAATAACTCAATAATACAATCATGACCACCAAACCTT
GAAATAAAATAATCTCTCTGATTACTTGTACAAATTCTTTACCAATATATG

>Block_3
GAGCTTAGTTTCGCGTTGCAAAGATGTTCATTTGATTTTTTTTTTTTTTTTTCCATTTGATTTTTCTTATACTAAAT
TGATTGATGTCACGGTCCTTCTGTGAAAAAAGTATATTGTACTGTATAAATTTATAATAGTCATACTTGAAGAT
AACTAATATCATTTTCTAAAGGTTTTGTTGACAATGAAATGGGTAATGATGAACATCTCGTTTGCTAAACCACA
AAAGAAATAATTCGTTGGACTATAACTTAAACTTATAATATATCAACTGAAATGCCTACAATTTGAGAATACGC
CTTTTCAAAAAGTTTTAGAATTTTATAAAAATTTCTCCACCATAATATTACTTTAATAAAAGCTTGTTATTAAAA
TGATATATTAACAAAAAAAAAAAGAGGTAGGCAAAGTAGCTTTTATTAAAATTAATGTTTTCCCTAATTGGTAA
ATCATAAACCAAAAAATTGAGGTAATTCTTAATATATT

>Block_4
ATATGAAAATATTTTAAAAAGTCAACATCAGTGCTAATTTCAAATTGACCATACATTAGCACACAAGAAGACT
AACAAAAAAAAATGAGTGAATAAAGAAGAGAGTGAGTGGGGTTGATACATGTCACCGTCCAAAATGTGTCTGA
CCCTTCTCATTGGAGCTACAAGAGCCGACCCTTTGCCTCATCATTTGTTGACTTCACTCTCCTTTTACATATTTA
AATTATAACTTGCTCATATTTCCATTCATTGTGTTCGGTTCTAGATTGTTCCATCCACT

>Block_5
TTGGAAGCAAGCGATTCAATGATTGTACCACTCATTTCATAATAGTGTTTGATATGACGTTAATTTGAATGGTT
TATGTTGATCGAAAATACTCTCAGTCACAACAAGAATCTAATTCGATTAATTTATTTGAAAAACATATGTTGGC
A

Sequence of conserved Block 2 from 7 Brassicaceae

>Arabidopsis_lyrata
taaattacaattatttttagcccagttacataactgcacgtatggtttgagactttgagcccatcacattacatggtacatgaccaaagcccagttatgttaccatattattcagacaagagcc
gccatccagacagaagtggaccacacaatcaagaatcgtatggagagcattacaatatattgctcgcatcttattgtgccacgtctcttggtaggcgagagatcaacgtccacagatctct
tgctccactctggttggtacacgtcatcatatttatgttctacttgtttatgtagacgttcatgtagacaagtgggaagatttaaaaatatcttaaagcaagtgtgaggcccatttagatatttctc
aacgggctgttcctcaaaagctgcccgttgatcaaacggctccaaactgttctgggcaagctgcccgaaatcttataaaggaaattgtggagtgggactgaatcaacggtggagaattc
gagggtgtacgtaacaactcgtagtatctgaacttagcttctctcaaaatatctgtggacatatgtcactagtttttgtttgtgaaagtattaactaattaggcaatcacatgagaggttgatata
gtaaaatacaatcatgaccacaaaactttaaataaataatctctctgattacttctgcaaattctttaccaatatttg

>Brassica rapa
TAAAATGGCAAATATTTTCAGCCCAGTTACAAGTCTCCACGTATGGTTTAGAGCCCATCGCATACATATAGAAG
AGCATGACCATGAAAAGCCCACAGTTATGTTCCATGTTCAGACACCAACCTGCCAGCCATCCACCAGAGTGTG
GACCCCACACCCACGAATCGTATGGTAGATCACTACAATATATTGCTGGCATCCTGTTGCGCCACGTCGTCGTT
AGGAGGAAGATCAACGTCCATATATCTCTTGCGCCTCTCTCATTGGTACACATCGTTTTGTTATATTTCAACTTG
TTTATGTGGTAGAGTGGACAAGTGGGAAGAAGAAGATAAATATCTAAAAAGAGTGTGGGCCTTATGTGTAGTC
TTGAGATATTTTCTCAACGGGCTGTCTCCACGAAAAGCTGCCCCGTTAATCAATACACACCATCAGACGGCTCG

AATCTCTTCAGGTCAAGCTGCCCGAAAACTTTGCCGAGGAATTTGTGTGGGACTCACTCACCACTGAAATCAAC
GGCGGAAATTTCGAAGGTGTACGTAACACTGCGTAGCATCTGAATATAGCTCCTCTGAAAATATCTAGGGACA
CATGTCACTTAGCTTTGTAAGAAAAGCAATTAACTAATCACATGACAGAGCTTATAGTTAAACTAATAAAAAG
GAGGAAATAATACAATCATGACCACCTTTAAATAATGTAAGGAAATAATACAATCATGACCACCTTTAAATAA
TGTA

>Capsella rubella
AATTATTGTAATTGAGTTATTTTCAGTCCAGTTATGCAGCTCCACGTATGGTTTGAGCCCATGACATTAACATG
GTTACATGACCATGAAAAAGGCCCAGTTATGTTCACCATATTCAGAGTTTCAGACACAGACACCAGCCAAGCG
ATCCAGACAGAACACACAAGTGGACAGCACCAATCACGAATCGTATGGAGATCTGAATACAATATATTGCTCG
CATCTTATTGCGCCACGTCTGTTTGTTTTTTAAGGCGCAAGATCAACGTCCATAGATCTCTTGCTCCACTCTGAT
TGGTACACGTCACCTCGGTTATGTGGTATAGAGTAGTAGACAAGTGGGAGAAAGATTAAATTAATATCTTAAA
GGGAAATTGTGAGGCCCATTTAGATATTTTCTCAACGGGCTGATCTCAAAAGCTGCCCGGTTTATTTTACAGCC
ATCCGACGGCTCAAAACTGTTCCGGGAAAGCTGCCCGAAACTTTTTAAAGGAATTTTATAGGTGGCACTCACTG
ATGAGTGAATCGACGGCGGAAAAATTCGAAGGTGTACCCAACATCTCGTAGCATCTGAACAAAGCTCCTCCCA
AAATATCTGTGGACACATGTCATGTCACTAGTTTTTGTTTGGGCGAAAATATTAATTAAGTAGGCAATCACATG
AGAGGGTTATATAGTTAAATTAATAAAAATAGGAGAATAATACAATCATGACCATACCAAACCTTTAAAAATA
AATAATCTCTGCTTACTTCAACAAATTCTTTACCAAATATTT

>Diplotaxis erucoides
TAAAATGGAAATAATTTTCAGCCCAGTTACAAGTCTCCACGTATAGAGCCCATCCCATATATATGCATGACCAT
GAAAAGCCCACAGTTATGTTCTATGTTCAGACACCAACCTGCCAGCCATCCACCAGAGTGTGGACCCCACACA
CCACGAATCGTATGGTAGATCGCTACGATATATTGCTGGCATCCTGTTGCGCCACGTCTTCGTTAGGAGGAAGA
TCAACGCCCATATATCTCTTGCGCCTCTCTCATTGGTACACATCGTTTTGTTATATATTCAACTTGTTTATGTGGT
GGAGTGGACAAGTGGGAAGAAGATGATAAATATCTAAAAAGAGTGTGGGTCACAGTTTAGGCTTGAGATATTT
TCACAACGGGCTGTCTCCACGAAAAGCTGCCTCCTTAACCGATACCCACCATCAGACGGCTCGAATCTCTTCGG
GTCAAGCTGCCCGAAAAACTTTGTTGAGGAATTTGTGTGGGACTCACCCACCACTCAGATCAACGGAGGAAAT
TTCGAAGGTGCACGTAACACATCGTAGCATCTGAACGTAGCTCCTTCGAAAATATCTAGGGACACATGTCACTT
TGCTTTGTAAGAAAAGTAATTAACTAATCACATGACAGAGTTTAGTGTTATTAAACTAATAAAAAGGAGGTAA
TAATACAATCATGACCCACCTGTAAATAACAGTATGTTATTAACGGGGAAGAGTTTATTTAACCGATCTCA

>Sinapis alba
TAATACTAAAATACGATTATTTTCAGCCCAGTAACATGTCTCCACGTATGGTTTGAGCCCATAGCATACACATG
GAACATGACCATGAAAAGCCCAGTTATGTGTCCATCTCCAGACACCAGCCATCGACAGAAGTGGACCACACAA
TCACGAATCGTATGGTGATTACTACAATATAATGCTCGCATCTTATTGCGCCACGTCTGTTGTTTGGCGGAAGA
TCAACGTCCATATATCTTATGCTCCATTATGATTGGTACACGTCATCTTGTTATGTTCTACTTGTTTATGTGGTG
GCGTAGACAAGTGGGAAGAAGAAAATATCTTTAAAGGAAGCCTTGGACCCATTTAGATATTTTGAACGGGCTG
GTTCTTCAAAAGCTGCCCGTTGATCTATACCCATCATCAAGACGGGTCTATACTCTTCCGGGCAAAGCTGCCCG
AAGTTTTTTAAAAGAATTTTGTAGGTGGGACTCACTCAACACTGAATCAACGGTGGAAAAATTCGAGGGTGTA
CGTAACACCTCGTAGCATCTGAACGTAGCTCCTCCCAAAATATCTTCGGACACATGTCACTAGTTTCGTAAACT
ATTAACTTAAATCAGAGAATCACATGAGAGACCATATAGTTTAAATTAATTAAAAGGAAATTATGCAATCATG
GACCAACAAACTTTTAAATAATGATCTGTAAGGTTAGCTATACAAAATTCTTTACCAATATTAC

>Turritis glabra
TATTATTACAGTTATTTTCAGCCCAGTTACATAACTCCACGTATGGTTTGAGCCCATCACATTAACATGGTGCAT
ACCATAGCCCAGTGATGTAACCATATTCAGACACCAGCCATCCAGACAGAAGTAGACTCCGCAAACACGAATC
GTATGGAGATCACTACAATATATTGCTCGCATCTTATTGCGCCACGTCTGTTGTTGGGAGGAAGATCAACGTCC
ATAGATCTCTTGCTCCACTCTGATTGGTACACGTCATCTTGGTTATGTTTTACTTGTTTATGTGGTAGAGTGGCC
AAGTGGGAGGATTATAAAATATCTTAAAGCAAGTGTGAGGCCCATTTTAGATATTTTCTCAACGGGCTGTTCCT
CAAAAGCTGCCCGTTAATCTACAGCCATCGAACGGTTCGAAACTGTTCTGGGCAAGCTGCCCGAAACTTTTTAA
GGGAATTTGTGGGTGGGACTCACTCGTAACTGAATCGACGGTGGAAGATTTGAGGGTGTACGTAACACCTCAT
AGCATCTGAACGTATCTCCTCCCAAAATATCTGTGGACACACGTCACTAGTTCTTTGTGAAACTATTAAGTTAT
TAGGCAATCACATGAGAGGGGTTATATTGTTAAATAAATAAAAAGGAAATAATACAATCATGACCACACCAAA
CCTTTAAAAATATATAATCTCTGATTACCTCAACAAATTCTTTACCAAATATTG

>Arabis alpina
CTAAATTAGAAATGATTATAAGCCCAATTACGTACGAATCTCCACGTATGGTCTGAGCCCATCCATACATGGGA
GGCTCAAGGCTGGGAGCATGCATGACCATGAAAAAGCCCATTCATCTTCAGACTCCAGCCATCTAAAGAGTGG
ACCCCACACACCACGAATCGTATGGAGATCATTACAATAAATTGCTCGCATCTTATTGCGCCACGTCTTCCTTT
AGGCGGAAGATCAACGTCCATATATCTCTTGCTCTACTTTCATTGGTCCACGTCATCGTGCTACTTGTTTATGTG
GTAGACTAGACAAGTGGGAAAATTATAAAATATCTTAAAAGGAGCATGGGACCCATTTAGATATTTTTCTCAA
CGGGCTGTGTCCTCAAAAGCTGCCCGTTAATCCATATCCATCAAACGGCTTGAAACTCTTCCGGGTAGAAGCCC
GAAAACTTTGTAAAGGAATTTGTGTGTAGGACTCACTTATCACTGAATCAACGGTCGAGAGTTGAAGGTGTAC
GTAACATCTCGTAGCATCTGAACGTGGCTCCTCACAAAATATCTTTGGACACACATGTCACTAGTTCTCGTAAA
CTAATTAACTAATCAGGCAATCACATGAGAGCTTGTAGTTAAATAAATTAATAAATGGAAATAATACAATCAT
TCATGACCAAACAAAGTAACAAACCTTTAAAAATTCCGTGATTAGTTTATTCATATTCTTTCCCAATACTAC

Erklärung

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

Die von mir vorgelegte Dissertation ist von Prof. Dr. George Coupland betreut worden.

Köln, November 2012 _____
Markus Berns

Teilpublikationen

Eva Herrero, Elsebeth Kolmos, Nora Bujdoso, Ye Yuan, Mengmeng Wang, **Markus C. Berns**, Heike Uhlworm, George Coupland, Reena Saini, Mariusz Jaskolski, Alex Webb, Jorge Gonçalves & Seth J. Davis

EARLY FLOWERING4 recruitment of EARLY FLOWERING3 in the nucleus sustains the Arabidopsis circadian clock

Plant Cell, 2012

Acknowledgements

I want to thank:

Prof. George Coupland for his wise supervision, many fruitful discussions and for the freedom to develop the project into the direction that it took.

Prof. Ute Höcker for being the second examiner and her continuous interest in my work from the very beginning.

Prof. Martin Hülskamp for being the head of my committee.

Dr. Imre Somssich for fruitful discussions during my PhD, for being part of my thesis committee and for being the 'Beisitzer' in my defence.

Frederic Cremer for having started work on the *GI* promoter and for introducing me into lab-life and this project.

Karl Nordström for many lines of computer code to decipher biological codes.

Maret Linda Kalda for photographs and Elmon Schmelzer for help with microscopy.

The MPIPZ gardeners for taking care of my plants, especially Monika Bürgel and Heike Eichler.

Martin Hartke for help in the lab and the student helpers, especially Lisa Hallau and Jula Werres, for cleaning seeds and transferring plants for TopCount experiments.

All the collegues that became friends, particularly:

Alfredo Sanchez, Amaury de Montaigu, Christiane Kiefer, Emma Miatton, Eva Herrero, Fernando Andres Lalaguna, Isabella Nougalli, Jessika Adrian, Julia Engelhorn, Julia Reimer, Julieta Mateos, Karl Nordström, Katharina Heidrich, Kerstin Richau, Kishore Panigrahi, Liron Krebs, Loren Castaings, Maida Romera Branchat, Maria Albani, Mark Rühl, Nadine Ilk, Nico Tintor, Nico Villacorta, Nora Peine, Omer Sarid-Krebs, Reka Toth, Samson Simon, Sara Bergonzi, Sara Farrona, Seth Davis, Simon von Berlepsch, Stefan Dieterle, Stefano Torti, Vipul Patel, Yang Bai and Xunli Lu.

Many thanks for stimulating discussions, fantastic travels and having good times during hard training, hard rocking and hard working!

Special thanks to 'my' year of IMPRS students, especially to Sara and Julia for sharing the PhD game from the beginning to the end.

The entire staff of the MPIPZ for creating such an international, scientifically stimulating and vibrant atmosphere.

My family and friends, especially my parents and my sister for their continuous love and support.

Lebenslauf

Persönliche Daten

Name:	Markus Christian Berns
Geburtsort:	Moers, Deutschland
Geburtsdatum:	30.05.1981
Nationalität:	deutsch
Adresse:	Konradstrasse 4 50937 Köln

Ausbildung:

10/2007 -	Doktorarbeit am Max-Planck-Institut für Pflanzenzüchtungsforschung, Köln; Betreuer: Prof. George Coupland Thema: Transcriptional regulation of the *Arabidopsis thaliana* flowering-time gene *GIGANTEA*
5/2007	Diplom in Biologie, Gesamtnote: 1 (ausgezeichnet) Thema der Diplomarbeit: Growth dynamics of plants under the influence of different light conditions; (Betreuer: Prof. Ulrich Schurr und Dr. Achim Walter, Forschungszentrum Jülich)
10/2002 – 05/2007	Studium der Biologie an der Heinrich-Heine-Universität Düsseldorf, Schwerpunkte: Pflanzen –genetik, -ökophysiologie und –evolution; Meeresbiologie
2001	Abitur, Julius-Stursberg Gymnasium in Neukirchen-Vluyn, Deutschland

Teilnahme an internationalen Konferenzen:

‘Control of flowering time and applications for plant breeding’ in Salzau, Deutschland (2008), Poster Präsentation

‘Arabidopsis – Konferenz’ in Edinburgh, Schottland (2009), Poster Präsentation

‘FESPB – Konferenz’ in Valencia, Spanien (2010), Mündliche Präsentation

‘Evolution of plant developmental regulatory systems’ in Roscoff, Frankreich (2011), Poster Präsentation

‘Botanikertagung’ in Berlin, Deutschland (2011), Mündliche Präsentation

Teilnahme an internationalen Retreats und Workshops:

1st International PhD retreat in plant sciences in Wageningen, Niederlande (2008)

PhDnet meeting in Jena, Deutschland (2009)

2nd International PhD retreat in plant sciences in Köln, Deutschland (2010)

2nd VIB PhD student symposium in Leuven, Belgien (2010)

‚Bioinformatics’ in Bertinoro, Italien (2008), Organisator: Prof. Werner Mewes

‚Biocrystallography with synchrotron radiation’ in Poznan, Polen (2010), Organisator: Prof. Mariusz Jaskolski

Besuchte Kurse

Scientific writing (Gadi Rothenberg und Christopher Lowe), Presenting professionally (Andrea Roos), Getting funded (Ruth Willmott), Patenting (Prof. Peter Schreier)

Weitere Aktivitäten

Organisator des ‘2nd International PhD retreat in plant sciences’ in Köln 2010

Doktorandenvertreter des Coupland-Departments 2009/2010

Betreuer beim Girls Day (2009 – 2010) und bei der Kinderuni (2008 – 2010) am MPIPZ

Mitarbeit bei der Wissenschaftsscheune des MPIPZ (seit 2010)