Evolution of *CONSTANS* Transcriptional Regulation and Protein Function in the Brassicaceae Family

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Samson Simon

aus Köln

Köln, November 2013

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

Berichterstatter:	Prof. Dr. George Coupland
	Prof. Dr. Ute Höcker
Prüfungsvorsitzender:	Prof. Dr. Martin Hülskamp
Tag der Disputation:	16. Januar 2014

ABSTRACT

Plants respond to seasonal changes in day length to control developmental transitions such as flowering and bud dormancy. In the model plant *Arabidopsis thaliana* (*A. thaliana*), CONSTANS (CO) activates the expression of *FLOWERING LOCUS T* (*FT*) specifically under long days, and FT protein then moves from the leaf to the shoot apical meristem where it promotes the transition to flowering. *CO* transcription is precisely regulated so that its mRNA is repressed in the morning and rises in the evening of a long day allowing activation of *FT*. Therefore, the diurnal pattern of *CO* transcription is an essential component of the regulation of flowering time in *A. thaliana*. The two closest homologs of *CO* in *A. thaliana, CONSTANS-like 1* (*COL1*) and *COL2* have no function in photoperiodic flowering response and show different mRNA patterns. All three genes evolved from a common ancestor after the latest whole genome duplication (*At-α*) in the lineage leading to *A. thaliana* that is believed to have occurred ~43 mya.

The transcriptional patterns of *CO* and *COL2* are conserved in the basal Brassicaceae species *Aethionema arabicum*. Also the *CO* homolog of this species but not its *COL2* homolog can complement the *co* mutation in *A. thaliana*. As the *Aethionema* lineage represents the earliest split in the evolution of the Brassicaceae these findings indicate that *CO* function was already present at an early stage of Brassicaceae evolution and its activity is likely conserved throughout the family.

The promoter of *CO* contains conserved sequence blocks which are shared in orthologs examined in ten Brassicaceaen species. DOF binding sites, E-Boxes and TCP binding sites are the most abundant *cis*-elements identified in ten *CO* promoters. Furthermore, DOF and TCP binding sites are strongly enriched in the promoters of *CO* genes compared to *COL* gene promoters. DOF binding sites are reported to enable binding of repressive CDF Transcription factors and were among the most highly conserved motifs in the promoters of *CO* orthologs. E-Boxes are putative binding sites for FLOWERING bHLH (FBH) TFs that function as transcriptional activators of *CO*. No role has so far been assigned for TCP transcription factors in regulation of *CO* transcription. Promoter reporter studies in *A. thaliana* revealed that the *CO* promoter contains functional regulatory sequences up to 2.6kb from the TS and that a short conserved regulatory motif that comprises DOF binding sites and is present both distally and proximally is sufficient to recreate mRNA patterns highly similar to endogenous *CO*. A model for the evolution of *CO* promoter is proposed in which downstream conserved DOF binding sites appeared distally upstream to regulate *CO* after the gene duplication leading to its diversification from *COL1*.

Tarenaya hassleriana, a member of the Cleomaceae family which is sister to the Brassicaceae, has only one ortholog of *CO* (*ThCOL*) which shares one last common ancestral gene with *CO*, *COL1* and *COL2*. ThCOL protein sequence is most similar to COL2 homologs from the Brassicaceae family.

Additionally mRNA expression pattern and lower ability to promote early flowering in *A. thaliana* are common characteristics of this gene with *COL1* and *COL2* but not with *CO*. Hence, the transcriptional regulation of *CO* as well as an increase in capacity to promote flowering when compared to its close homologs *COL1* and *COL2* evolved in the Brassicaceae family and represents a novelty that arose in this family.

My evolutionary studies indicate that *CO* evolved by gene duplication at the root of the Brassicaceae family. This gene then evolved a function in photoperiodic flowering through changes to its transcriptional regulation and protein function. Such a history contradicts the current view that *CO* has an ancient history predating the divergence of monocotyledonous and dicotyledonous angiosperms, but rather suggests that the role of *CO* homologous genes in conferring photoperiodic flowering might have evolved independently in different families of higher plants.

ZUSAMMENFASSUNG

Pflanzen reagieren auf saisonale Änderung der Tageslänge um Entwicklungsschritte wie das Blühen oder die Knospenruhe zu kontrollieren. In der Modellpflanze *Arabidopsis thaliana* (*A. thaliana*) aktiviert CONSTANS (CO) die Expression von *FLOWERING LOCUS T* (*FT*) am Ende eines Langtages. Daraufhin bewegt sich FT Protein vom Blatt zum Sprossapikalmeristem um dort den Wechsel zur reproduktiven Entwicklung einzuleiten. Die Transkription von *CO* ist präzise reguliert damit dessen mRNA am Morgen unterdrückt wird und erst am Abend eines Langtages ansteigt, was die Aktivierung von *FT* erlaubt. Das von der Tageszeit regulierte Profil der *CO* Transkription ist somit eine essentielle Komponente der Blühzeitregulation in *A. thaliana*. Die zwei ähnlichsten Homologe von *CO* in *A. thaliana*, *CONSTANS-like 1* (*COL1)* und *COL2* haben keine Funktion in der photoperiodischen Regulation der Blühzeit und weisen abweichende mRNA Profile auf. Alle drei Gene stammen von einem letzten gemeinsamen Vorgänger ab und sind nach der letzten Genomduplikation (*At-α*) die zur *A. thaliana* Linie geführt hat, welche vor ca. 43 Millionen Jahren stattfand, entstanden.

Die Transkriptionsprofile von *CO* und *COL2* sind in der basalen Brassicaceae Spezies *Aethionema arabicum* konserviert. Das *CO* Homolog aber nicht das *COL2* Homolog von *Aethionema arabicum* ist in der Lage die *co* Mutation in *A. thaliana* zu komplementieren. Weil die *Aethionema* Linie sich als erstes von den Brassicaceae abgetrennt hat weisen diese Ergebnisse darauf hin, dass die *CO* Funktion schon früh in der Brassicaeae Familie entstanden ist und das dessen Aktivität wahrscheinlich in der kompletten Familie konserviert ist.

Der Promoter von *CO* enthält konservierte Blöcke, die in zehn verschiedenen Brassicaceae Spezies vorkommen. DOF Bindestellen, E-Boxen und TCP Bindestellen sind die am häufigsten vorkommenden *cis*-Bindestellen die in den *CO* Promotoren identifiziert werden konnten. Weiterhin sind DOF und TCP Bindestellen überrepräsentiert in den Promotoren von *CO* Genen im Vergleich zu denen von *COL* Genen. DOF Bindestellen werden von unterdrückenden CDF Transkriptionsfaktoren (TFs) gebunden und stellen die am höchsten konservierten Motive dar. E-Boxen sind mögliche Bindestellen für FLOWERING bHLH (FBH) TFs welche als transkriptionelle Aktivatoren von *CO* gelten. Für TCP Transkriptionsfaktoren wurde noch keine Funktion in der Regulation von *CO* festgestellt. Promoter Reporter Untersuchungen in *A. thaliana* haben gezeigt, dass *CO* Promoter regulatorische Sequenzen bis zu 2.6kb vom Transkriptionsstart enthält und das kurze konservierte regulatorische Motive, die DOF Bindestellen enthalten und sowohl distal als auch proximal vorkommen, ausreichend sind um mRNA Expressionsmuster ähnlich dem von endogenem *CO* zu vermitteln. Ein Model für die Evolution von dem *CO* Promoter wurde vorgeschlagen in dem nachliegende konservierte DOF Bindestellen nach der Genduplikation, die zur Diversifikation von *COL1* geführt hat, distal vorliegend auftauchen um *CO* zu regulieren.

Tarenaya hassleriana, ein Mitglied der Familie der Cleomaceae, einer Schwesterfamilie der Brassicaceae, hat nur ein Ortholog von *CO* (*ThCOL*) welches einen letzten gemeinsamen Vorgänger mit *CO*, *COL1* und *COL2* hat. ThCOL Proteinsequenz ist der von COL2 Homologen der Brassicaceae Familie am Ähnlichsten. Zusätzlich ist das mRNA Expressionsprofil und die geringere Fähigkeit frühzeitiges Blühen in *A. thaliana* zu aktivieren gemeinsame Charakteristiken mit *COL1* und *COL2* aber nicht mit *CO*. Also entstanden sowohl die transkriptionelle Regulation von *CO* also auch die erhöhte Fähigkeit Blühen auszulösen in der Familie der Brassicaceae und stellen somit Neuheiten dar.

Meine evolutionären Untersuchungen weisen darauf hin, dass *CO* am Ursprung der Brassicaceae Familie durch Genduplikation entstanden ist. Dieses Gen hat dann eine Funktion in der photoperiodischen Blühregulation durch Änderungen in der transkriptionellen Regulation sowie der Proteinfunktion evolviert. Diese Darstellung wiederspricht der jetzigen Auffassung nach der *CO* vor der Entstehung von monokotyledonen und dikotyledonen Angiospermen schon vorhanden war, sondern deutet an, dass die Rolle von *CO* als Regulator vom photoperiodischem Blühen unabhängig in verschiedenen Familien der höheren Pflanzen entstanden ist.

TABLE OF CONTENTS

ABSTRACT ..... **I**

ZUSAMMENFASSUNG ..... **III**

1. INTRODUCTION **1**

1.1. Preamble 1

1.2. Flowering time regulation by day length 1

1.3. Transcriptional gene regulation 2

1.3.1. Plant transcriptional gene regulation and the role of *cis*-elements 3

1.3.2. Phylogenetic shadowing identifies *cis*-regulatory modules in plants 3

1.3.3. Transcriptional regulation of photoperiodic flowering in *A. thaliana* 4

1.4. CO protein functions as a transcriptional activator of *FLOWERING LOCUS T* 5

1.4.1. CO protein structure and function in transcriptional regulation 5

1.4.2. The influence of photoreceptors on CO protein stability 6

1.4.3. The *CONSTANS-like* gene family shows little redundancy to CO 6

1.4.4. Conservation of *CO* function in other species 7

1.5. Regulation of *CO* transcription by external coincidence 8

1.5.1. *CO* mRNA levels are diurnally regulated 9

1.5.2. *FBHs* activate *CO* transcription 9

1.5.3. *CDFs* are transcriptional repressors of *CO* 11

1.5.4. FKF1 and GI co-ordinately degrade CDFs and relieve repression of *CO* 11

1.5.5. Additional regulators implicated in *CO* transcriptional regulation 12

1.5.6. Involvement of Photoreceptors in *CO* mRNA regulation 12

1.5.7. Conservation of *CO* mRNA regulation in other species 12

1.6 The sister families Brassicaceae and Cleomaceae as research tools 13

2. AIM OF THE STUDY **17**

3. MATERIAL AND METHODS **19**

4. Phylogenetic relationship and functional diversification of *CO* Homologs in the Brassicaceae and Cleomaceae **29**

4.1. *A. arabicum* and *T. hassleriana* are long day plants 29

4.2. Subgroup Ia *COL* genes are conserved in the Brassicaceae and *T. hassleriana* contains only one such gene that is closely related to *AtCOL2* 30

4.3. Expression patterns of *A. thaliana*, *A. arabicum* and *T. hassleriana* group Ia *COL* genes are conserved 33

4.4. Subgroup Ia *COL* genes from *A. thaliana*, *A. arabicum* and *T. hassleriana* are able to complement the *co-10* mutant 36

4.5. Amino acid sequence changes may explain functional differences between CO, COL1, COL2 and ThCOL 39

4.6. Discussion Chapter 4 40

5. Phylogenetic Shadowing of *CONSTANS* promoter and functional test **45**

5.1. *CO* promoters are present in all Brassicaceaen species investigated 45

5.2. The promoter of *CO* is conserved in ten Brassicaceaen Species and features a unique *cis*-element composition 47

5.3. Strategy for functional analysis of the *A. thaliana CO* promoter 54
5.4. Analysis of *A. thaliana CO* regulation by promoter deletion constructs identifies a minimal promoter 54
5.5. Dissection of *CO* promoter by the gain of function approach reveals separable functions for subelements in shaping mRNA expression patterns 59
5.6. Analysis of conserved regulatory motifs by gain of function approach supported findings from phylogenetic shadowing and *cis*-element analysis 64
5.7. Tandem duplication leading to emergence of *CO* caused appearance of a regulatory DOF binding site cluster in the distal *CO* promoter 69
5.8. Discussion Chapter 5 71
5.8.1. *CO* promoter conservation in the Brassicaceae 71
5.8.2. The composition of *cis*-elements in the *CO* promoter 72
5.8.3. Functional analysis of the *CO* promoter in *A. thaliana* 73
5.8.4. The model for *CO* promoter evolution 74
6. Regulation of *CO* transcription by CDFs is a novelty that arose in the Brassicaceae 77
6.1. CDF1 and CDF2 binding to the *A. thaliana CO* promoter could not be detected *in planta*.. 77
6.2. CDF transcription factors regulate *CO* but not *COL1* and *COL2* in *A. thaliana* 79
6.3. Light downregulates *CO* mRNA in the morning in the absence of CDF Transcription factors 80
6.4. Light shapes the expression patterns of *AtCOL1* and *AtCOL2* as well as *ThCOL* in the morning 81
6.5. The mRNA of *AtCO, AtCOL1* and *AtCOL2* is highly unstable 84
6.6. Red and blue light are partially responsible for downregulating *CO* mRNA in the morning .. in the absence of CDF Transcription factors 85
6.7. Discussion Chapter 6 87
6.7.1. ChIP of CDF1 and CDF2 to *CO* and *FT* locus did not detect binding 87
6.7.2. Regulation of mRNA levels by CDFs is not conserved between subgroup Ia *COL* genes 88
6.7.3. Light downregulation of subgroup Ia *COL* genes is conserved and depends on light quality 89
7. CONCLUSIONS AND PERSPECTIVES 91
7.1. Insights into transcriptional regulation of *CO* 91
7.2. A model for *CO* promoter evolution 92
7.3. *CO* as a model for promoter evolution 92
7.4. Interplay of *CO* Function and Transcription 93
7.5. The last common ancestral gene of *CO* 93
7.6. Function of *CO* orthologs beyond the Brassicales 94
LITERATURE 95
APPENDIX 109
ABBREVIATIONS 129
ACKNOWLEDGEMENTS 133
ERKLÄRUNG 135
LEBENSLAUF 137

1. INTRODUCTION

1.1. Preamble

Plants are sessile organisms that are highly adapted to their environment to ensure optimal survival and reproduction in each habitat. Plants sense day length and contain an internal timekeeper, called the circadian clock. This information is used for timing of developmental processes, of which the reprogramming from vegetative to reproductive phase, generally termed “flowering”, is a key step. Photoperiodic flowering response, which enables flowering under favourable and inhibits it under inappropriate day length conditions was first described almost a century ago (Garner and Allard, 1920). Long-day conditions typical for spring and summer accelerate flowering in the model organism *Arabidopsis thaliana*. Mutants impaired in the photoperiodic flowering pathway in *Arabidopsis thaliana* were identified as early as 1962. Interestingly out of four “supervital” mutants investigated three exhibited mutations in genes of the photoperiodic flowering pathway, namely *GIGANTEA* (*GI*) and *CONSTANS* (*CO*) (Rédei, 1962). It was only much later that a higher number of late-flowering mutants was identified and their genetic relationship defined (Koornneef et al., 1991). The three key regulators of photoperiodic flowering in *Arabidopsis thaliana*, *GI*, *CO* and *FLOWERING LOCUS T* (*FT*) were thereby placed in the same pathway. This set the stage for investigating the mechanism by which *Arabidopsis thaliana* is able to discriminate between day lengths and integrate this information into a signal that triggers flowering. A central component in the generation of flowering response is *CO*, which integrates internal signals from the circadian clock and the external signal light both at the transcriptional and post-transcriptional levels to activate *FT* exclusively under inductive LD conditions.

1.2. Flowering time regulation by day length

The switch from vegetative to reproductive development in plants, referred to as floral induction or flowering, is a crucial step in the life of angiosperms as it initiates reproduction. Plants have evolved many regulatory mechanisms to trigger flowering only under favourable conditions. The signals that are integrated to regulate flowering time include internal and external factors. Flowering can be delayed for many years in an age dependant way by an extended juvenile phase for example in trees (Bohlenius et al., 2006). Seasonal regulation is important for many plants and is sensed by integrating different signals. Perennial plants for example are able to sense winter by responding to cold temperatures to avoid flowering in autumn but allow it in the following spring (Wang et al., 2009b). Another important signal for discriminating seasons is the measurement of day length (Garner and Allard, 1920).

Flowering can be initiated in short-day plants like *Oryza sativa* (Rice) when the days shorten in autumn indicating that the hot and often dry summer is over (Hayama et al., 2003). Day neutral plants like *Solanum lycopersicum* (Tomato) do not flower in response to day length (Lifschitz and Eshed,

2006). This behaviour coincides with these species often originating from regions where day length does not substantially change in one year. Long day (LD) plants on the other hand initiate flowering in the spring or summer when a certain day length is exceeded which indicates favourable conditions in temperate climates. The plant model organism *Arabidopsis thaliana* (*A. thaliana*) is a LD plant (Koornneef et al., 1991). Flowering is accelerated by LD conditions resulting in the plant producing fewer leaves before flowering and initiating flowering faster than under SD conditions. *A. thaliana* is a facultative long day plant meaning it eventually flowers independently of an inductive photoperiodic signal under SD. A plant with an obligate requirement for a certain photoperiod on the other hand does not initiate flowering under non-inductive conditions. *A. thaliana* is a model for inductive photoperiodic flowering and interactions between genes in the photoperiodic flowering pathway are the molecular cause of flowering under inductive conditions (Koornneef et al., 1991). The gene products of flowering time regulators are part of complex transcriptional networks that will be discussed below. New insights in the evolution and regulation of an important hub in this network, the gene *CONSTANS*, will be presented in this study.

1.3. Transcriptional gene regulation

Transcription of DNA into RNA is the first step in the central dogma of molecular biology (Crick, 1970). The regulation of transcription provides an important regulatory step enabling a cell to control the amount of protein that is produced from a gene at any given time. Gene regulation allows multicellular organisms to make differential use of the same genome to create different cell types. Transcriptional regulation is a complex mechanism in prokaryotes but becomes even more complex in eukaryotic organisms (Alberts et al., 2008). In general transcription is facilitated by the recruitment of a large protein complex, called RNA polymerase II (Pol II), which transcribes the template DNA strand of genes into messenger RNA (mRNA). Transcription was found to occur combined in so called transcription factories (Cook, 2010; Osborne et al., 2004). The regulation of the recruitment of Pol II to the transcriptional start (TS) site in a gene is facilitated by transcription factors (TFs). TFs are proteins which are able to bind DNA and regulate transcription (Alberts et al., 2008). They can be part of protein complexes and act both as activators or repressors depending on the molecular context (Beckerman and Prives, 2010). Specific DNA sequences, so called *cis*-elements are recognized by TFs. Gene regulatory regions are generally located 5´ of the coding region but can also be found in introns, coding and downstream regions. The proximal regulatory part is called the promoter of a gene, while more distal regulatory regions are called enhancers. These enhancer regions can be kilo base pairs (kb) away from the coding region (Alberts et al., 2008). Another important layer of transcriptional regulation involves epigenetic mechanisms. These are changes in the genome that do not affect the DNA sequence but regulate gene expression. The effects range from modification of single loci to large parts of chromosomes. The most prominent epigenetic changes occur as posttranscriptional modification of histones (e.g. acetylation) and methylation of DNA sequence (Alberts et al., 2008). In contrast to the central paradigm of molecular biology it was found more

recently that transcription is not restricted to protein coding genes but also occurs in intergenic space. In rice for example ~58% of all transcripts identified are located in intergenic space (Li et al., 2006). Studies in yeast suggest an involvement of non-coding RNAs close to the TS to be involved in mechanisms of chromatin remodelling and POL II recruitment and thus transcriptional initiation (Dutrow et al., 2008; Hirota et al., 2008).

1.3.1. Plant transcriptional gene regulation and the role of *cis*-elements

About 5.9% (1533) of all *A. thaliana* genes code for transcription factors (Riechmann et al., 2000), highlighting the importance of transcriptional regulation in plants. Regulation of transcription is organized in complex transcriptional networks, that are additionally often interconnected (Huang et al., 2012; Kaufmann et al., 2010). Transcription factors associate to DNA via so called *cis*-elements located in the regulatory region or promoter of a gene (Alberts et al., 2008). C*is*-elements are generally short sequence stretches. Different families of TFs have preferences for certain core *cis*-elements, while the binding sites for individual members of a given TF family can vary. In *A. thaliana* the basic helix-loop-helix (bHLH) TF family for example has 147 members which can be categorized by their DNA binding preference (Toledo-Ortiz, 2003). Most bHLH (109) were predicted to bind the E-Box (CANNTG), while the others either bind varying sequences (11) or have no predicted DNA binding capacity (27). Of the E-Box binding proteins, 89 bHLH have a preference for the G-Box (CACGTG), a specific version of the E-Box. The preferred *cis*-element of a TF can be identified by different methods both *in vitro* using methods such as protein binding microarrays (PBM, Helfer et al., 2011) and *in vivo* with utilizing Chromatin Immunoprecipitation followed by next generation sequencing (ChIP-seq, Kaufmann et al., 2010). Nonetheless transcription factor binding sites do not occur independently but are usually organized in so called *cis*-regulatory modules (CRMs) to collectively regulate gene expression both in plants and animals (Michael et al., 2008; Nguyen and Xu, 1998; Zinzen et al., 2009). In animals CRMs were defined to be ~300 base pairs (bp) long and carry ten or more TF binding sites for at least four TFs (Levine and Davidson, 2005).

1.3.2. Phylogenetic shadowing identifies *cis*-regulatory modules in plants

Evolution tends to conserve functionally important DNA sequences, while changes in less significant regions are tolerated. This makes sequence comparisons of orthologous genes or proteins among species a powerful tool to identify functionally important regions both in Proteins and DNA. Conservation is in general lower in non-coding regions compared to protein coding regions which complicates the identification of regulatory regions of a gene (Haudry et al., 2013). However, conserved *cis*-elements in promoters leave a “phylogenetic footprint”, meaning a short conserved and functionally important motif that can be identified in sequence alignments of regulatory regions from homologous genes (Gumucio et al., 1992; Tagle et al., 1988). This method was advanced to identify whole *cis*-regulatory modules, which contain many TF binding sites, by comparing non-coding regions of closely related species from the same family in a method coined phylogenetic shadowing

(Boffelli et al., 2003). In plants phylogenetic footprinting and shadowing were used to identify regulatory regions in genes such as the floral homeotic gene *AGAMOUS* (*AG*, Hong et al., 2003), the clock component *LONG ELONGATED HYPOCOTYL* (*LHY*) (Spensley et al., 2009) and the circadian clock-associated *PSEUDO RESPONSE REGULATOR 9* (*PRR9*, Herrero et al., 2012). Recently phylogenetic shadowing was applied on a whole genome level for the first time in plants. Genome sequences of nine species from the Brassicaceae family were employed to develop an atlas of conserved non-coding regions which serves as a resource for extensive identification of *cis*-regulatory regions in intergenic space (Haudry et al., 2013).

1.3.3. Transcriptional regulation of photoperiodic flowering in *A. thaliana*

Regulation of flowering initiation in *A. thaliana* is organized in a signalling cascade starting from perception of the inductive signal in the leaves to the induction of structural changes in the meristem which is reprogrammed to produce flowers. Most genes involved in this day length dependent cascade mediate transcriptional changes of their downstream targets and either code for TFs or proteins involved in the regulation of TFs (Andrés and Coupland, 2012). The LD pathway in *A. thaliana* is a transcriptional network that evolved to sense changes in day length and triggers the switch to reproductive development in response to these inductive conditions. Triggering of flowering starts with *CO* mRNA levels rising in the phloem companion cells at the end of a long day, a mechanism mediated by the circadian clock and light (Suarez-Lopez et al., 2001; Yanovsky and Kay, 2002). CO protein is subsequently stabilized by light and acts as a transcriptional activator of the floral integrator *FT* (Adrian et al., 2010; Tiwari et al., 2010; Valverde et al., 2004; Wigge et al., 2005). *FT* and its close homolog *TWIN SISTER OF FT* (*TSF*) are part of the mobile signal "florigen" and move from the phloem to the shoot apical meristem (SAM) (Michaels, 2005; Yamaguchi, 2005; Corbesier et al., 2007; Jaeger and Wigge, 2007; Mathieu et al., 2007), where they interact with the bZIP transcription factor FD (Abe et al., 2005; Wigge et al., 2005). This complex was shown to include 14-3-3 proteins in rice (Taoka et al., 2011). *FT*, *TSF* and *FD* are believed to contribute to transcriptional changes in the SAM including activation of the two MADS Box genes *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS* 1 (*SOC1*) (Borner et al., 2000; Lee, 2000; Samach et al., 2000) and *FRUITFULL* (*FUL*) as well as the floral meristem identity gene *APETALA 1* (*AP1*) (Schmid et al., 2003; Teper-Bamnolker and Samach, 2005; Wigge et al., 2005). SOC1 in turn interacts with another MADS Box protein called AGL24 to activate the transcription of *LEAFY* (*LFY*), a meristem identity gene involved in the first steps of flower development (Lee et al., 2008). Another class of TFs that are both mediators of flowering response and direct downstream genes of the FT-FD complex and SOC1 in *A. thaliana* are *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) genes, namely *SPL3*, *SPL4* and *SPL5* (Jung et al., 2012). These SPL factors directly bind to and activate transcription of *SOC1*, *AP1*, *LFY* and *FUL* which leads to the formation of flowers (Wang et al., 2009a; Yamaguchi et al., 2009).

Thus the transcriptional activation of *FT* in the leaves by CO triggers a downstream transcriptional cascade that eventually leads to flowering response (Andrés and Coupland, 2012). Misregulated *CO* mRNA and protein levels led to disruption of accurate flowering response in *A. thaliana* as shown by experiments on mutant and overexpression lines (Onouchi et al., 2000; Rédei, 1962). It is thus crucial for reproductive success of *A. thaliana* that *CO* as a key sensory gene in the LD pathway is regulated tightly at both mRNA levels as well as protein abundance.

1.4. CO protein functions as a transcriptional activator of *FLOWERING LOCUS T*

A. thaliana plants with mutations in the *CO* gene exhibit late flowering in LD while flowering time is unaffected in SDs (Koornneef et al., 1991; Putterill et al., 1995; Rédei, 1962). Flowering time of *co-10* null mutant is indistinguishable in LD and SD conditions (Laubinger et al., 2006). Hence loss of *CO* gene function disables the ability of *A. thaliana* to discriminate day length leading to a “constant” flowering time in all conditions. Other phenotypes were not assigned to *co* mutants. The cause of late flowering in *co* mutants is the reduction of expression of “florigen” gene *FT* under inductive long-day conditions (Samach et al., 2000). Overexpression of *CO* from the ubiquitous *Cauliflower Mosaic Virus* (CaMV) *35S* promoter as well as the leaf vasculature specific *SUCROSE TRANSPORTER2* (*SUC2*) promoter leads to high *FT* mRNA levels and early flowering regardless of the photoperiod (An et al., 2004; Imlau et al., 1999; Suarez-Lopez et al., 2001). To ensure that flowering is only triggered under inductive conditions CO protein is regulated at the posttranscriptional level. One key mechanism is the regulation of CO protein abundance internally by the circadian clock and by the external signal light. This process, referred to as the external coincidence model, ensures that CO protein abundance leading to direct activation of *FT* transcription occurs at the correct time of the day under inductive conditions (Sawa et al., 2007; Song et al., 2012; Valverde et al., 2004).

1.4.1. CO protein structure and function in transcriptional regulation

CO protein consists of two zinc binding B-Box motifs and a CO, CO-like, *TIMING OF CAB2 EXPRESSION 1* (*TOC1*) (CCT) domain (Putterill et al., 1995; Robson et al., 2001). The B-Box tandem motifs show homology to zinc finger motifs in animal transcription factors. In animals B-Box motifs co-occur with coiled-coil and RING finger domains (tripartite motif) and are involved in protein-protein interactions but do not confer DNA binding properties (Torok and Etkin, 2001). The B-Box family in *A. thaliana* is comprised of 32 members containing either one or two B-Box motifs. Mutations in either B-Box were shown to impair CO protein function (Fig. 1, Robson et al., 2001). Experiments in yeast fusing the CO B-Box domain to a GAL4 DNA binding domain indicate that this motif can mediate transcriptional activation, which might be its function in CO as being a part of a DNA binding complex in *A. thaliana* (Ben-Naim et al., 2006).

The CCT domain is plant specific and has a nuclear localization signal (Robson et al., 2001). Mutation in this domain impairs CO protein function (Fig. 1). The CCT domain is not only found in CO and

CO-like proteins but also in TOC1, a PSEUDEO-RESPONSE REGULATOR (PRR) protein (Strayer et al., 2000). The CCT domain of TOC1 and other PRR proteins, which associate with chromatin *in vivo*, is sufficient for DNA binding *in vitro* (Gendron et al., 2012; Nakamichi et al., 2010). Similarly, direct association of CO to *FT* locus was identified and the CCT domain of CO enables it to bind DNA *in vitro* via a *cis*-regulatory element present in the *FT* promoter (Song et al., 2012; Tiwari et al., 2010). Additionally the CCT domain allows physical interaction of CO with members of the trimeric CCAAT-binding complex (CBF) also known as Nuclear Factor Y (NF-Y), which was claimed to be important for CO-mediated triggering of flowering (Kumimoto et al., 2010; Ben-Naim et al., 2006; Wenkel et al., 2006). This interaction of CO with the NF-Y complex and the presence of CCAAT-Boxes in *FT* promoter could in this scenario potentiate transcriptional activation (Laloum et al., 2013; Tiwari et al., 2010).

1.4.2. The influence of photoreceptors on CO protein stability

The regulation of CO protein abundance is crucial for photoperiodic flowering control in *A. thaliana* (Valverde et al., 2004). Interactions with different proteins control the stability of CO throughout the day to ensure that CO protein is exclusively stable at the end of a light period in LD conditions. The red light photoreceptor phytochrome B (phyB) was shown to destabilize CO in the morning (Valverde et al., 2004). The photoreceptors cryptochrome 1 (cry1), cry2, phytochrome A (phyA) and FKF1 stabilize CO at the end of the day (Song et al., 2012; Valverde et al., 2004). Night degradation of CO is mediated by SUPPRESSOR OF *PHYA105-1* (SPA1), SPA3, SPA4 and the interacting E3 Ubiquitin Ligase CONSTITUTIVE PHOTOMORPHOGENESIS (COP1). SPA1 and COP1 require the CCT domain of CO for interaction (Jang et al., 2008; Liu et al., 2008; Wenkel et al., 2006).

1.4.3. The *CONSTANS-like* gene family shows little redundancy to CO

CO is the founding member of the 17 member *CONSTANS-like* (*COL*) gene family in *A. thaliana*. COL proteins are composed of one or two B-Boxes and a CCT domain. The *COL* gene family is organized in three different groups (I,II and III) according to their domain composition and conservation (Griffiths et al., 2003; Robson et al., 2001). Group I, consisting of *CO*, *COL1*-*COL5*, characteristically has two B-Boxes and a CCT domain which is separated by one intron in the genomic context. Group II genes (*COL6-8, COL16*) in contrast lack one B-Box. Group III genes (*COL9-COL15*) have one B-Box and a second, different zinc finger motif, a CCT domain and characteristically three introns.

Group I can be further subdivided. Subgroup Ia *COL* genes have four conserved regions between its B-Boxes and CCT domain. In *A. thaliana CO*, *COL1* and *COL2* belong to this subfamily and many other plant species have at least one subgroup Ia *COL* homolog (Table 1). *A. thaliana COL4-COL6* were categorized into subgroup Ic but *A. thaliana* has no homologs in subgroups Ib, Id and Ie (Griffiths et al., 2003).

35S::COL5 was shown to partially complement *co* mutant in SD conditions (Hassidim et al., 2009). The two closest homologs of *CO* in *A. thaliana*, *AtCOL1* and *AtCOL2,* failed to accelerate flowering when overexpressed (Ledger et al., 2001) and did not complement *co-9* mutation (Kim et al., 2013). *COL1* and *COL2* were implicated in circadian clock regulation but not flowering and their functions have not been resolved (Kim et al., 2013; Ledger et al., 2001).

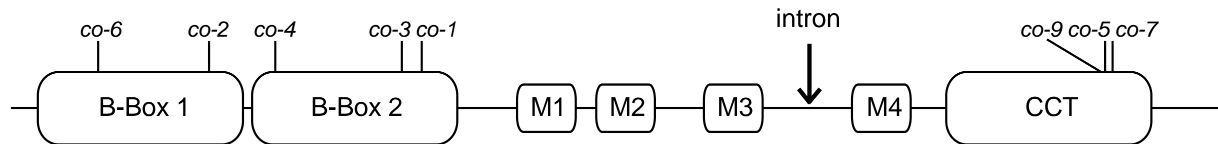


Figure 1: Schematic representation of domains of subgroup 1a *COL* genes using *AtCO* as an example
CO protein consists of two zinc finger domains named B-Boxes in the N-terminal part (B-Box 1 and B-Box 2). Four conserved regions can be identified in the middle part (M1-M4) and the only intron is located between M3 and M4. The *CONSTANS*,*CONSTANS-like*, *TOC1* (CCT) domain is located close to the N-terminus. The causal regions for *co* mutants are indicated. *co-2* and *co-6* locate to B-Box 1 while *co-1*, *co-3* and *co-4* are in B-Box 2. Three mutations were identified in the CCT domain (*co-9*, *co-5* and *co-7*). The Col-0 null mutant *co-10* (*co-sail*) is a T-DNA insertion line and thus its position is not indicated. Figure modified from Griffiths et al. (2003).

1.4.4. Conservation of *CO* function in other species

Conservation of *CO* function in floral induction is the subject of many studies in the plant kingdom. Close homologs of *CO* were identified in many plant species (Table 1) and shown to complement *co* mutants in *A. thaliana* using *35S* overexpression lines. Such studies were carried out with *Pharbitis nil CO* (*PnCO*), *Beta vulgaris COL1* (*BvCOL1*) and *Populus trichocarpa CO1* (*PtCO1*) (Chia et al., 2008; Hsu et al., 2012; Liu et al., 2001). In the model legumes *Medicago truncatula*, *Pisum sativum* and *Lotus japonicus* on the other hand orthologs of *CO* called *COLa* were identified, but no link to flowering was established and the genes were suggested not to be functional *CO* homologs (Hecht et al., 2005, 2007; Yamashino et al., 2013).

Table 1: Selected non-Brassicaceae subgroup Ia *COL* homologs

Gene name	Species	Family	Publication
ThCOL	*Tarenaya hassleriana*	Cleomaceae	(Schranz and Mitchell-Olds, 2006)
LjCOLa	*Lotus japonicus*	Fabaceae	(Yamashino et al., 2013)
MtCOLa	*Medicago truncatula*	Fabaceae	(Hecht et al., 2005, 2007)
PsCOLa	*Pisum sativum*	Fabaceae	(Hecht et al., 2005)
StCO1 *StCO2 (StCO)*	*Solanum tuberosum*	Solanaceae	(González-Schain et al., 2012; Kloosterman et al., 2013; Navarro et al., 2011)
VvCO	*Vitis vinifera*	Vitaceae	(Almada et al., 2009)
PnCO	*Pharbitis nil*	Convolvulaceae	(Liu et al., 2001)
PtCO1 *PtCO2*	*Populus trichocarpa*	Salicaceae	(Bohlenius et al., 2006; Hsu et al., 2012)
OsHD1	*Oryza sativa*	Poaceae	(Yano et al., 2000)
HvCO1 *HvCO2*	*Hordeum vulgare*	Poaceae	(Campoli et al., 2012; Griffiths et al., 2003)

The role of *CO* homologs in flowering time regulation in the *Poaceae* family on the other hand is well established. In the SD plant *Oryza sativa* (Rice) *OsGI* (*AtGI* ortholog) was placed upstream of *HEADING DATE 1* (*Hd1*, homolog of *AtCO*) which in turn regulates *Hd3a* (*AtFT* ortholog). In contrast to *A. thaliana* regulation of *Hd3a* by *Hd1* showed a functional inversion by supressing the *FT* homolog in non-inductive LD conditions but activating it in SD conditions (Hayama et al., 2002, 2003). *Hordeum vulgare CO1* (*HvCO1*), which is a close homolog of *Hd1*, was shown to accelerate flowering in barley when overexpressed but failed to complement *co-2* mutant when overexpressed from the *35S* promoter in *A. thaliana* (Campoli et al., 2012). This suggests that the *GI-CO-FT* regulon is conserved in flowering plants, but genes are not fully interchangeable between species.

The *CO-FT* regulon was shown to be involved in other photoperiodically regulated developmental processes. In *Populus* spp. (Poplar) a genetic link between *PtCO2* and *PtFT1* was found. *PtFT1* is a strong flowering inducer in *Populus* and additionally regulates SD induced bud formation in autumn. In RNA interference (RNAi) plants downregulating either *PtFT1* or *PtCO2* bud formation occurs earlier suggesting both genes work in the same pathway (Bohlenius et al., 2006). Even though *35S::PtCO1* rescued the *co-1* mutant in *A. thaliana, 35S::PtCO1* and *35S::PtCO2* failed to accelerate reproductive onset in transgenic poplar (Hsu et al., 2012). In *Solanum tuberosum L. StCO* (also called *StCO2*) and *StCO1* negatively regulate photoperiodic tuberization by upregulating *StSP5G* (an *AtFT* homolog) which in turn represses *StSP6A* (another *AtFT* homolog) expression in potato (González-Schain et al., 2012; Kloosterman et al., 2013; Navarro et al., 2011). Interestingly, *35S::StCO* only partially rescues *co-2* mutant in *A. thaliana* (González-Schain et al., 2012). These experiments suggest that the *CO-FT* homologs have regulatory function beyond flowering time control and were employed many times in evolution to regulate different photoperiodic responses.

Overexpression of a *CO* homolog from the green alga *Chlamydomonas reinhardtii* (*CrCO*) using heterologous promoters (*35S*, *SUC2*) induced early flowering and/or complemented both *co-10* (Col) and *co-8* (L*er*) mutants (Serrano et al., 2009). These experiments additionally indicate that subgroup Ia *COL* genes might be controlling photoperiodically regulated mechanisms more generally and have an ancient evolutionary origin.

1.5. Regulation of *CO* transcription by external coincidence

CO protein functions as a potent transcriptional activator of *FT* and the regulation of protein abundance is crucial for the function of *CO* (Adrian et al., 2010; Tiwari et al., 2010; Valverde et al., 2004). Nonetheless post transcriptional regulation of CO protein is not sufficient to regulate photoperiodic flowering in *A. thaliana*, which is indicated by early flowering day-length insensitive phenotypes induced by overexpression of *CO* from heterologous promoters (An et al., 2004; Onouchi et al., 2000). Hence regulation of *CO* mRNA levels is additionally necessary to ensure floral induction only under favourable conditions. *CO* mRNA levels are under complex transcriptional regulation in *A. thaliana* involving many transcription factors and its regulators (Fornara et al., 2009; Ito et al.,

2012a). The regulation of *CO* mRNA incorporates signals from the circadian clock and the external signal light (Imaizumi et al., 2003, 2005; Sawa et al., 2007; Suarez-Lopez et al., 2001). The precise output of these mechanisms ensures that *CO* activates *FT* exclusively under inductive conditions. Hence, *CO* mRNA regulation plays a pivotal role in the functionality of the external coincidence model of flowering induction in *A. thaliana* under inductive long day conditions.

1.5.1. *CO* mRNA levels are diurnally regulated

CO mRNA is diurnally expressed through both light and circadian clock control and its temporal expression pattern was studied in detail (Fig. 2 ,Suarez-Lopez et al., 2001; Yanovsky and Kay, 2002). In order to standardize day times for cycling processes the term ZEITGEBER TIME (ZT) was introduced defined as hours after dawn or lights on depending on the experiment. Under SD conditions of 8 hours light and 16 hours darkness *CO* mRNA is only present in the dark hours of the day but very low in the light hours of the day between ZT0 and ZT8. In LD conditions of 16 hours light and 8 hours darkness *CO* mRNA is present at dawn but then quickly degraded. The trough levels of *CO* mRNA are between ZT 3 and ZT9. *CO* is then transcribed again to form an evening peak (ZT12 to ZT16) which is thought to be important for *FT* activation (Suarez-Lopez et al., 2001; Yanovsky and Kay, 2002). *CO* mRNA exhibits an additional night peak of unknown function. *CO* is the major activator of *FT* and integrates signals from the plant circadian clock and day length to ensure that flowering is only activated upon favourable inductive conditions. The tight regulation of mRNA abundance was subject to many studies in the past and activators and repressors of *CO* gene expression were discovered and linked to both the circadian clock and light regulation as described below (Fig. 2, Fornara et al., 2009; Imaizumi et al., 2005; Ito et al., 2012a).

1.5.2. *FBHs* activate *CO* transcription

CO transcription is activated by a subclass of redundant basic helix-loop-helix (bHLH) TFs, called FLOWERING bHLH (FBH, Ito et al., 2012a; Toledo-Ortiz, 2003). Four members, namely *FBH1*, *FBH2*, *FBH3* and *FBH4*, belonging to the same phylogenetic group (bHLH, clade IX) were shown to accelerate flowering and upregulate *CO* mRNA when expressed from the 35S promoter. The quadruple exhibited only a mild late flowering phenotype both in LD and SD and the night peak of *CO* mRNA was impaired. FBH1 and FBH2 proteins were shown to bind specifically to E-Box (CANNTG) elements but not to a G-Box (GAGCTC) element in the proximal *CO* promoter in EMSA studies as well as transient *LUCIFERASE* (*LUC*) expression assays. FBH1 binds to *CO* promoter *in vivo* as shown by ChIP. Strongest binding was identified close to the TS of *CO* but association was also seen at a distal (-2kb) region. Interestingly, *35S::FBH* overexpression lines induce *CO* mRNA up to 20fold higher than wild type levels but the temporal expression pattern in LD is unchanged. This is in agreement with the finding that *FBH* mRNA levels do not cycle as *CO* mRNA does but are rather ubiquitously expressed throughout the day. Additionally, the expression of *β-GLUCORONIDASE* (*GUS*) (Jefferson et al., 1987) driven by promoter *CO* (*pCO*) in *FBH* overexpression lines is restricted

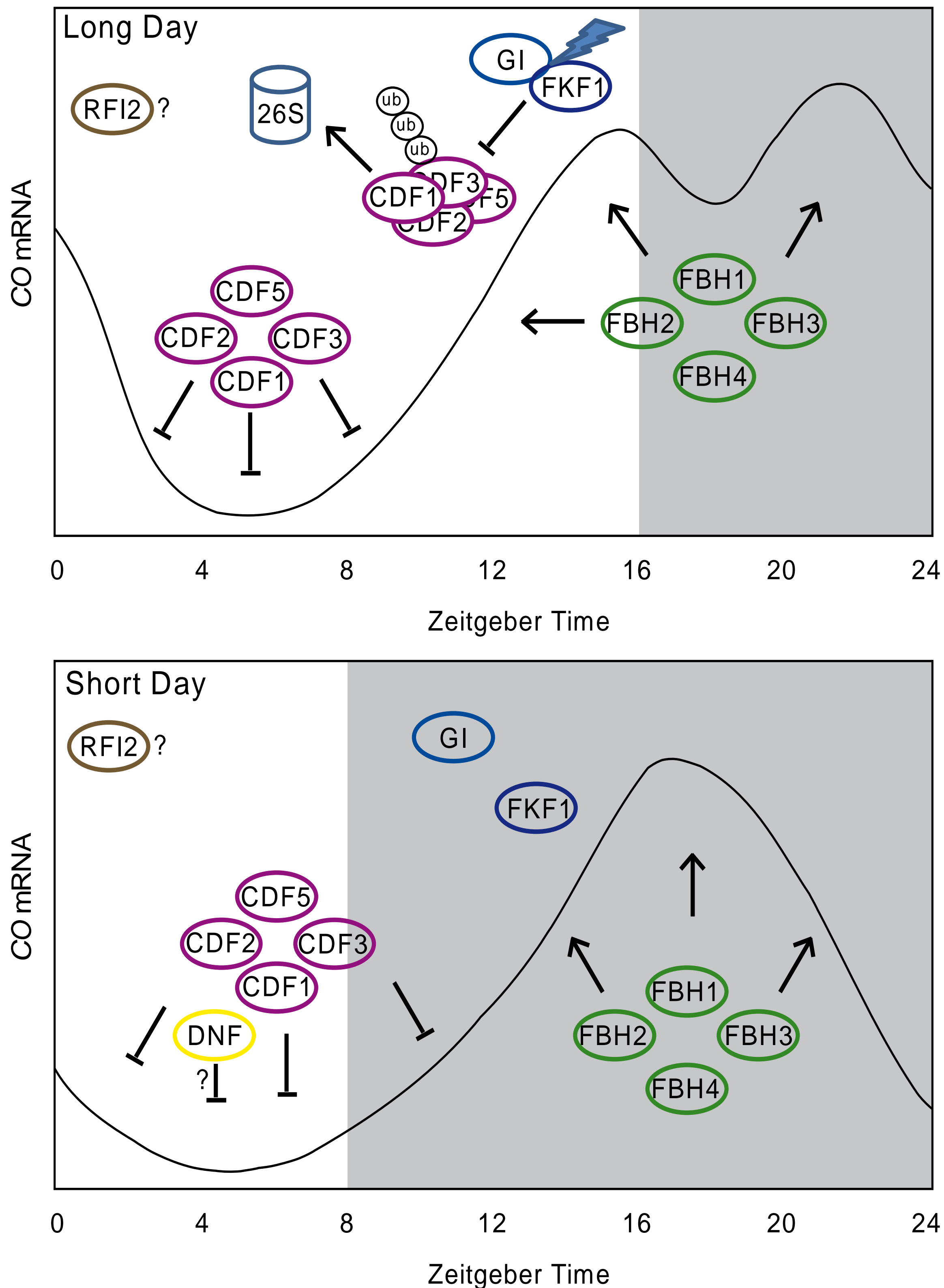


Figure 2: Transcriptional regulation of *CO* in LD and SD

CO is transcriptionally differentially regulated in inductive LD and non-inductive SD conditions. *CO* mRNA is downregulated efficiently in the morning of a LD by a class of redundant DOF TFs namely *CDF1*, *CDF2*, *CDF3* and *CDF5*. In the later part of the day (ZT10-ZT14) the blue light dependent complex of FKF1 and GI protein forms an E3 ubiquitin ligase to mark the CDFs by polyubiquitination (ub) for degradation by the 26S proteasome (26S). *CO* is constitutively activated by a class of redundant BHLH TFs called *FBH1*, *FBH2*, *FBH3* and *FBH4*. They activate CO at the end of the light period in LD. Additionally a night peak forms in LD conditions. In SD conditions light and *CO* expression do not overlap. *CO* mRNA is low before lights on and stays low until the end of the light period. This repression is mediated directly by the DOFs and putatively by DNF. GI and FKF1 complex does not form in darkness, thus DOF protein levels stay high and *CO* levels do not form an evening peak. RFI2 is implicated in repression of *CO* mRNA both in LD and SD conditions by an unknown mechanism.

to only a few cell types including leaf stomata cells, the root tip and importantly vascular tissues. To summarize, FBH TFs can activate *CO* transcription but cannot explain the full diurnal or spatial pattern indicating that *CO* mRNA regulation is more complex both spatially and temporally, and must involve other activators.

1.5.3. *CDFs* are transcriptional repressors of *CO*

Diurnal expression of *CO* mRNA is facilitated by a class of repressive TFs, called CYCLING DOF FACTOR (CDF) which belong to the DNA-binding with one finger (DOF) TF family (Fornara et al., 2009; Imaizumi et al., 2005; Sawa et al., 2007; Umemura et al., 2004). Six members, *CDF1*, *CDF2*, *CDF3*, *CDF4*, *COG1* and *CDF5*, were associated with regulation of photoperiodic flowering via repression of *CO* and *FT* mRNA (Fornara et al., 2009). *CDF* mRNA levels are controlled by the plant circadian clock, showing peaks of expression in the beginning of the day when *CO* mRNA is low (Fornara et al., 2009; Imaizumi et al., 2005). Promoter *GUS* fusion experiments implicated that CDF TFs are expressed in vascular tissue. Overexpression of CDFs from the *35S* or *SUC2* promoter led to a delay in flowering under inductive LD conditions, respectively (Fornara et al., 2009; Imaizumi et al., 2005). The *cdf1-R*, *cdf2-1*, *cdf3-1*, *cdf5-1* quadruple mutant (*cdf1235*) flowered early under LD and SD conditions and had elevated *CO* mRNA levels throughout the day and importantly the downregulation of CO mRNA at dawn is impaired (Fornara et al., 2009). CDF1 was shown to bind to the proximal *CO* promoter (-500bp) in EMSA studies as well as transient *LUC* expression assays (Imaizumi et al., 2005). The DOF consensus binding site is AAAG, while (AAAG)TGT contributes to specificity of CDF1 as shown in EMSA competition assays (Imaizumi et al., 2005; Yanagisawa and Schmidt, 1999). ChIP experiments indicate that CDF1 binds to the *CO* promoter close to the TS as well as in a distal part of the promoter (~-2kb) (Sawa et al., 2007).

1.5.4. FKF1 and GI co-ordinately degrade CDFs and relieve repression of *CO*

According to the current model CDF TFs occupy the *CO* locus to repress its expression. A complex interplay of circadian clock and light regulated genes is necessary to release this repression upon inductive LD conditions and drive transcription of *CO*. The blue light activated proteins FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 (FKF1), and LOV KELCH REPEAT PROTEIN 2 (LKP2) were shown to interact with CDF1, CDF2 and CDF3 both in Yeast two hybrid (Y2H) and pull down assays. FKF1 and LKP2 together with ZEITLUPE (ZTL) form a small protein family (Nelson et al., 2000; Somers et al., 2000). These proteins contain a LIGHT, OXYGEN, VOLTAGE (LOV) domain, which enables them to sense blue light, an F-Box domain and KELCH repeat domain, which allows them to function as E3 ubiquitin ligases (reviewed in Ito et al., 2012b). The *fkf1-2* mutant shows lower *CO* mRNA levels between ZT9 and ZT12 (Imaizumi et al., 2003). In the evening of a long day FKF1 mediates the degradation of CDF1 in a blue light dependent manner (Imaizumi et al., 2005). Blue light is not sufficient for this mechanism: GIGANTEA (GI), a plant specific protein of unknown biochemical function, additionally regulates FKF1 protein abundance and the GI-FKF1 complex is

stabilized specifically by blue light (Fornara et al., 2009; Fowler et al., 1999; Park et al., 1999; Sawa et al., 2007). *GI* is part of the plant circadian clock and its evening specific expression (~ZT10) is highly conserved in flowering plants (Fowler et al., 1999; Hayama et al., 2002; Mizoguchi et al., 2002, 2005). Only when *GI* and *FKF1* mRNA expression and blue light overlap can this protein complex act as an E3 protein ligase to mark CDF TFs for degradation by the 26S proteasome and release repression of *CO* transcription (Fornara et al., 2009; Sawa et al., 2007).

1.5.5. Additional regulators implicated in *CO* transcriptional regulation

Another E3 Ligase, called DAY NEUTRAL FLOWERING (DNF) was identified to regulate *CO* mRNA levels independently of *GI* and *FKF1* but downstream of the circadian clock (Morris and Jackson, 2010; Morris et al., 2010). Mutations in *dnf* led to increased *CO* mRNA levels in the morning in SD conditions which caused early flowering.

Another protein that might regulate *CO* mRNA directly is RED AND FAR-RED INSENSITIVE 2 (RFI2) (Chen and Ni, 2006a). Mutations in this RING-domain zinc finger protein led to higher *CO* mRNA levels under LD, SD and constant light conditions. The *rfi2-1* mutation causes early flowering both in LD and SD conditions and was epistatic to *co-2* mutation without affecting *GI* expression or other parts of the circadian clock. The direct regulatory mechanism for *CO* mRNA regulation was not identified but *RFI2* was suggested to function downstream of the photoreceptors phyA and phyB (Chen and Ni, 2006b).

1.5.6. Involvement of Photoreceptors in *CO* mRNA regulation

The involvement of blue light sensing FKF1 and its close homologs LKP2 and ZTL in *CO* mRNA regulation was studied in detail. But also mutations in other photoreceptors exhibit defects in flowering time regulation that are partially due to changes in *CO* mRNA levels. Mutations in the red light photoreceptor phyA leads to a late flowering phenotype in LD conditions (Cerdán and Chory, 2003) and *CO* was identified as part of the postulated phyA-regulated transcriptional network (Tepperman et al., 2001). Accordingly, *CO* mRNA levels are reduced in incandescent light enriched LD conditions compared to wild type control (Yanovsky and Kay, 2002). The opposite effect was observed in the *phyB-9* mutant which flowered early in LD and showed slightly elevated levels of *CO* mRNA in LD (Cerdán and Chory, 2003). It is not known which regulatory mechanisms or transcription factors mediate the mild effects of phyA and PhyB on *CO* mRNA levels.

1.5.7. Conservation of *CO* mRNA regulation in other species

CO has close homologs in probably all flowering plants investigated (Table 1) and the diurnal regulation of *CO* is similar to rice ortholog *HEADING DATE 1* (*HD1*) (Hayama et al., 2003). Additional similarities in expression of Poplar *PtCO2* led to the assumption that *CO* mRNA regulation is conserved in evolution (Song et al., 2010). In many other species like *Pharbitis nil*, *Vitis vinifera*, *Beta vulgaris* and *Solanum tuberosum* on the other hand expression patterns of *CO* homologs show a

clear morning phase of expression like *LATE ELONGATED HYPOCOTYL* (*LHY*) and *CIRCADIAN CLOCK REGULATED 1* (*CCA1*) (Almada et al., 2009; Chia et al., 2008; Kloosterman et al., 2013; Liu et al., 2001; Mizoguchi et al., 2002). Interestingly the morning expression is also shared by the closest homologs of *CO* in *A. thaliana*, *COL1* and *COL2* (Ledger et al., 2001).

1.6 The sister families Brassicaceae and Cleomaceae as research tools

A. thaliana is the best studied model organism in the plant kingdom and the first plant species of which the genome was sequenced (The *Arabidopsis* Genome Initiative, 2000). In order to define mechanisms in *A. thaliana* and understand the broader significance of discoveries made in *A. thaliana* a comparison with closely related species is useful. *A. thaliana* belongs to the mustard family (Brassicaceae or Cruciferae). Characteristically flowers of Brassicaceaen species have four petals which are cross-shaped (cruciform, Brechner, 2001). Economically important members of this family are the crops rapeseed (*Brassica napus*), Cabbage (*Brassica oleracea*) and black mustard (*Brassica nigra*). The family comprises of 3709 species in 338 genera and probably originated in the Irano-Turanian region (Ansell et al., 2011; Franzke et al., 2011). The family can be divided into two major groups: The genus *Aethionema* falls into a distinct group from the core of the family (Galloway et al., 1998; Koch et al., 2000; Zunk et al., 1999). Estimates about when this split occurred vary between 15mya and 60mya (Fawcett et al., 2009; Franzke et al., 2009; Koch et al., 2000; Schranz and Mitchell-Olds, 2006). While other Brassicaceae species are found all over the world the genus *Aethionema* is limited to a region ranging from Spain and Morocco to Turkmenistan (Franzke et al., 2009). The origin of the Brassicaceae family suggests that long day photoperiodic flowering response as seen in *A. thaliana* might represent the ancestral state.

The Brassicaceae family underwent five independent whole genome duplications (WGD) of which the two most ancient duplications are shared with the seed plants and angiosperms, respectively (Jiao et al., 2011). The *At-γ* represents the third oldest WGD, while the *At-β* occurred ~70mya (Tang et al., 2008). The *At-α* WGD, which happened ~43mya is restricted to the Brassicaceae family and shared with *Aethionema arabicum* (*A. arabicum*, Fawcett et al., 2009; Haudry et al., 2013). Thus the divergence time between the genus *A. arabicum* and the core Brassicaceae is assumed to be less than 43mya.

As the phylogeny of the Brassicaceae family is well developed and the model *A. thaliana* is a member, new model organisms emerge and move into research focus. *Arabis alpina* (*A. alpina*) was established as a model for perennialism (Wang et al., 2009b). *Eutrema parvulum* (also known as *Thellungiella parvula*) is a model for plant adaptation to extreme habitats (Dassanayake et al., 2011). *Arabidopsis lyrata (A. lyrata)* is studied because it is very closely related to *A. thaliana* (10mya divergence time) but shows interesting differences in chromosome number (n=5 in *A. thaliana*; n=8 in *A. lyrata*) and in constrast to *A. thaliana* is an outcrossing species (Hu et al., 2011). Taken together the Brassicaceae family is of special interest for research leading to the availability of nine genome sequences until now

(Haudry et al., 2013). All Brassicaceaen species that were used for comparison in this study are included in Fig. 3 and their phylogenetic relationship is shown relative to *Tarenaya hassleriana* (*T. hassleriana*) that is used as outgroup.

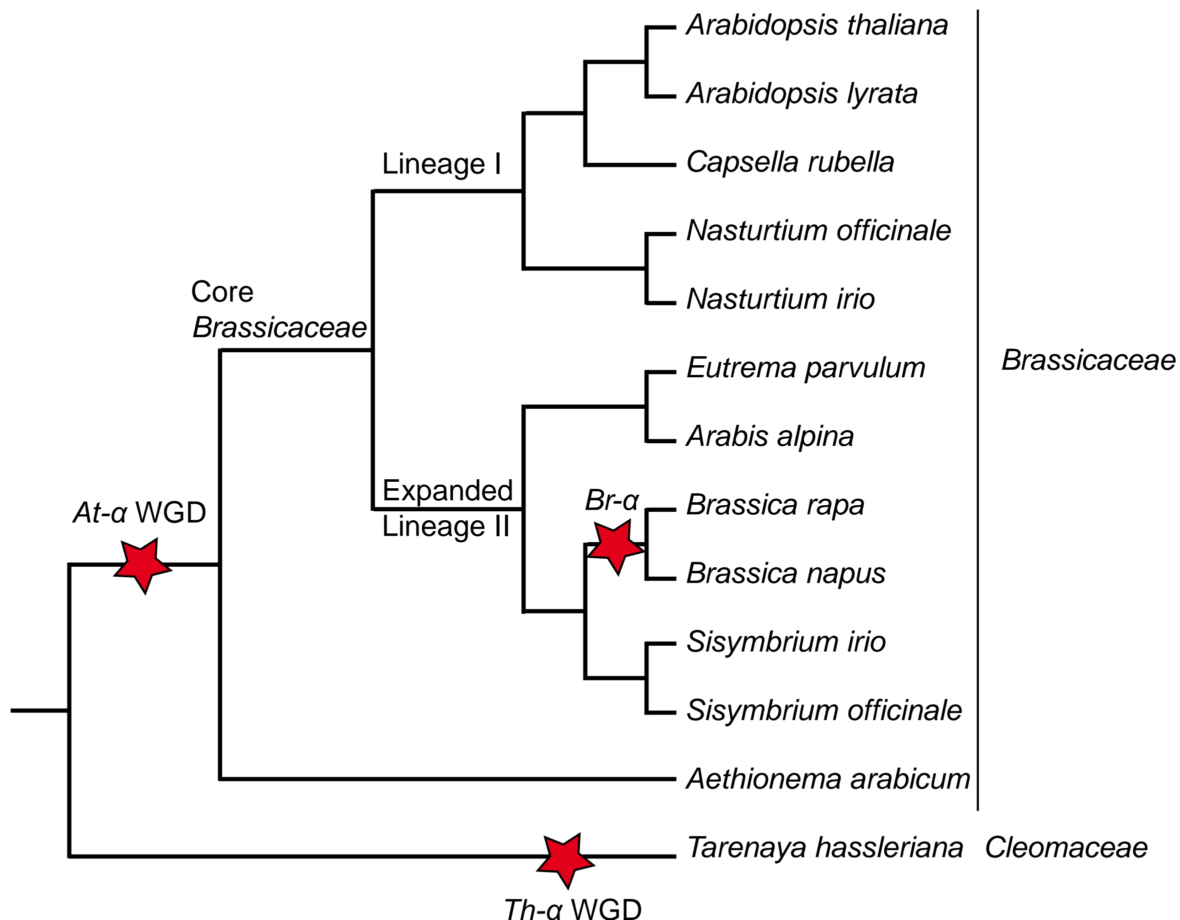


Figure 3: Schematic representation of phylogenetic relationships of species from the order Brassicales used in this study.
The Brassicaceae family can be divided in two major lineages, the core Brassicaceae and the Aethionema lineage, here represented by *Aethionema arabicum*. The core Brassicaceae can be further divided into Lineage I represented by *Arabidopsis*, *Capsella* and *Nasturtium* species, while expanded lineage II consists of *Arabis alpina*, *Eutrema parvulum*, *Brassica* and *Sisymbrium* species. *Tarenaya hassleriana* (formerly *Cleome hassleriana*) is the only member of the Cleomaceae family displayed, which is the sister family of the Brassicaceae in the order Brassicales. Both families underwent independent whole genome polyploidization events. The *At-α* is the latest whole genome duplication (WGD) event identified in *Arabidopsis thaliana* and is shared by all lineages of the Brassicaceae but not the Cleomaceae. The *Th-α* genome triplication is found in *Tarenaya hassleriana*. It is of similar age compared to the Br-α, which was exclusively identified in the genus *Brassica*. *At= Arabidopsis thaliana*, *Th= Tarenaya hassleriana,Br= Brassica* , Red stars indicate whole genome duplications (WGD). References: (Bailey et al., 2006; Cheng et al., 2013; Franzke et al., 2011; Haudry et al., 2013)

T. hassleriana (formerly named *Cleome hassleriana*) is a member of the Cleomaceae family, a sister family of the Brassicaceae in the order *Brassicales* (Hall et al., 2002; Iltis and Cochrane, 2007). Both families diverged between 19mya and 65mya (Beilstein et al., 2010; Franzke et al., 2009). *Cleome* is the largest genus of the Cleomaceae family with 180 to 200 species (Sánchez-Acebo, 2005). The genus *Cleome* is of research interest for studying the evolutionary progression between C3 and C4 photosynthesis (Marshall et al., 2007). The spider flower *T. hassleriana*, which is closely related to *Cleome spinosa*, has its natural habitat in south America (Cheng et al., 2013). Flowering was shown to

be accelerated in LD conditions, indicating that *T. hassleriana* is a facultative LD plant (Koevenig, 1973). *T. hassleriana* was used to decipher the phylogenetic relationship of the sister families. The genome of the *T. hassleriana* ancestor underwent a paleopolyploidization event independent of the *At-α* duplication resulting in a triplication of the genome now referred to as *Th-α* (formerly *Cs-α*) (Fig. 3, (Cheng et al., 2013; Schranz and Mitchell-Olds, 2006). This makes *T. hassleriana* of special interest for this study as the functional divergence of *AtCO*, *AtCOL1* and *AtCOL2* occurred after the *At-α* and the homologies with the one *T. hassleriana* ortholog identified will be discussed (Schranz and Mitchell-Olds, 2006).

2. AIM OF THE STUDY

CO protein is the key activator of transcription of *FLOWERING LOCUS T* in *A. thaliana* for LD induced photoperiodic flowering. Transcription of *CO* itself is highly regulated by light and the circadian clock to ensure that mRNA levels leading to stable CO protein are only present under favourable LD conditions.

Studies of the mechanism of *CO* transcriptional regulation in *A. thaliana* were so far focussed on the regulating Transcription factors (CDFs and FBHs) and their posttranslational regulators (GI and FKF1). In this study evolutionary comparisons of *CO* promoter will be used to identify conserved regions, with a focus on known and new *cis*-elements in the promoter that can then be tested for functionality using reporter lines and performing complementation experiments. One aim is to identify regions of *CO* promoter regulating transcription and analyze their influence on flowering time.

The functional conservation of *CO* and its closest homologs *COL1* and *COL2*, which are implicated in clock function rather than flowering regulation, will be investigated by comparison to the orthologs in the basal Brassicaceae species *A. arabicum*. *T. hassleriana* from the Cleomaceae family has exactly one close *CO* ortholog (*ThCOL*) which shares one common ancestral gene with *CO*, *COL1* and *COL2*. The functional and molecular characterisation of *ThCOL* aims at determining whether the ancestor of group Ia *COL* genes in the Brassicales was a functional flowering time regulator.

In summary, genetic, molecular and phylogenetic studies will be used to define and explain evolutionary changes in both transcriptional regulation and function of the photoperiodic flowering time regulator *CO* in the Brassicaceae family.

3. MATERIAL AND METHODS

Polymerase chain reaction (PCR)

Standard PCR reactions (genotyping) were performed with non-proofreading PCR according to standard lab protocols. Cloning PCRs were performed using LA taq (TAKARA) for long genomic fragments or AccuPrime™ *Taq* DNA Polymerase High Fidelity (Invitrogen) for all other purposes according to suppliers manual. PCR fragments were purified either directly or from agarose gel using NucleoSpin® Extract II (Macherei-Nagel).

Transformation of *Escherichia coli*

Escherichia coli (*E. coli*) was transformed with plasmid DNA using heat shock or electroporation and selected on appropriate antibiotics according to standard lab procedures. For all applications *E. coli* strain DH5α (Invitrogen) was used except for Gateway® (GW, Invitrogen) compatible target vectors containing the *ccdB* gene which were propagated in DB3.1™ strain (Invitrogen). Plasmid DNA was purified from liquid culture using NucleoSpin® Plasmid (Macherey-Nagel).

Plasmid construction

pCDF2::3xHA:GW was constructed based on *35S::3xHA:GW* (*pAlligator2*). Briefly, promoter of *CDF2* was amplified with overhangs for HINDIII and the HA-tag (Table 2). Complementary the HA tag was amplified with overhangs for pCDF2 on the 5´end and MluI restriction site on the 3´end. In a second PCR both fragments were combined. *p35::3xHA* was removed from *pAlligator2* by double digest with HindIII and MluI and *pCDF2::3xHA* was inserted by T4 ligation according to suppliers manual (New England BioLabs) resulting in *pCDF2::3xHA::GW*. The cDNA of *CDF2* was present as a GW construct in *pDONR201* and was recombined using LR reaction according to suppliers manual to obtain *pCDF2::3xHA::CDF2* as used for ChIP.

Table 2: Cloning primer for *pCDF2::3xHA:GW*

Name	Sequence	Purpose
pCDF2-HindIII-Fw	CAAGCTAAGCTTTCCGCGTGGGCTTTGTCC	pCDF2 with HindIII and HA overhangs
pCDF2-HA-RvLong	CGTCGTATGGGTATGCCATGGTAATGATTCCTTTTC	
HA-pCDF2-Fw	ATCATTACCATGGCATACCCATACGACG	HA tag with pCDF2 and MluI overhangs
HA-MluI-Rv	GGATCCACGCGTTTACGCC	

pCO::3xHA:GW was constructed based on *pSUC2::3xHA:GW*. *A. thaliana* (Col-0) *pCO* was excised from *pCO::GW* (*pAlligator2*) using double restriction digest (XbaI and XhoI) and *pSUC2* was subsequently replaced via T4 ligation with *pCO* to obtain *pCO::3xHA:GW*.

Constructs for *co-10* complementation experiments were created by amplification of desired coding sequence (CDS) from cDNA of *A. arabicum* (*AeCO* and *AeCOL2*) and *T. hassleriana* (*ThCOL*) and subcloning into *pDONR201* via BP reaction (Table 4). CDS of *AtCOL1*, *AtCOL2* in *pDONR201* were retrieved from the stocks. After confirmation by Sanger sequencing cDNAs were recombined into respective vectors (*pAlligator2*, Table 3).

Table 3: Plasmids used in this study for cloning

Name	Backbone	Bacterial selection	Plant selection	Purpose
pDONR201	pDONR	Gentamycin	(-)	subcloning of cDNAs
pDONR207	pDONR	Gentamycin	(-)	Subcloning of *pAtCO*
GW::LUC	pGreen0029	Kanamycin	BASTA	*LUCIFERASE* reporter
GW:pnos::LUC	pGreen0029	Kanamycin	BASTA	minimal promoter *LUCIFERASE* reporter
GW::GFP:COcDNA	pGreen0029	Kanamycin	BASTA	*CO* promoter complementation
pSUC2::3xHA:GW	pAlligator2	Spectinomycin	GFP seed coat	Subgroup Ia *COL* complementation
pCO::3xHA:GW,	pAlligator2	Spectinomycin	GFP seed coat	Subgroup Ia *COL* complementation
pCDF2::3xHA:GW	pAlligator2	Spectinomycin	GFP seed coat	ChIP

Table 4: Primer for cloning of subgroup Ia *COL* CDS from *A. arabicum* and *T. hassleriana*

Name	Sequence	Purpose	Final construct
Aeth_COGW_Fw	(attB1) GCATGTCAAAAGAAGAGAGTAACG	*AeCO* cDNA	*pSUC2::3xHA:AeCO* *pCO::3xHA:AeCO*
Aeth_COGW_Rv	(attB2) TCAGAATGAAGGAACAATTCC		
Aeth_COL2GW_Fw	(attB1) GCATGTTGAAAGAAGAGAGTAACG	*AeCOL2* cDNA	*pSUC2::3xHA:AeCOL2* *pCO::3xHA:AeCOL2*
Aeth_COL2GW_Rv	(attB2) TCAAAATGAAGGAACAATTCC		
Cl_COL_GW_Fw	(attB1) GCATGTTGAAGGAAGAGCGAACC	*ThCOL* cDNA	*pSUC2::3xHA:ThCOL* *pCO::3xHA:ThCOL*
CL_CO_GW_Rv	(attB2) TTAAAATGAAGGAACAATTCC		

A.thaliana CO promoter complementation experiments and reporter lines were created using the GW technology. Briefly, the *CO* promoter fragments were amplified from genomic DNA (L*er*) integrating *att*B1 and *att*B2 GW sites (Table 5) and subcloned into *pDONR207* via BP reaction. Desired constructs were recombined by LR reaction into destination vector in the context of *GFP:CO*, *pnos:LUC* and *LUC*, respectively (Table 3).

Table 5: Primer for *CO* promoter constructs

Name	Sequence	Purpose	Final construct
COprom3GW	(attB1) TTAGTATCCACATGAATTTACCTACCAG	2.6kb Fw; DFw	*2.6kb::LUC;2.6kb::GFP:CO* *D:pnos::LUC*
GW_CO_4_RV	(attB2) GCCTTCGGATAACTGTTACGAGTAA	D_Rv	*D:pnos::LUC*
CO_4b_FW_GW	(attB1) TTACTCGTAACAGTTATCCGAAGGC	C_Fw	*C:pnos::LUC*

CO_6b_RV_GW	(attB2) TGATACCCATTATGTGGTCCAAAA	C_Rv	*C:pnos::LUC*
CO_6b_FW_GW	(attB1) TTTTGGACCACATAATGGGTATCA	B_Fw	*B:pnos::LUC*
CO_9a_RV_GW	(attB2) TTTGTGAATGCATCTGCATCAT	B_Rv	*B:pnos::LUC*
GW_CO_9_FW	(attB1) ATGATGCAGATGCATTCACAAA	A/0.7kb_Fw	*A:pnos::LUC;A::LUC* *0.7kb::GFP:CO*
111coGWA	(attB2) CTCAGATGTAGTAAGTTTGATGGTG	Rv (5bp from TS)	All *pCO::GFP:CO* and *pCO::LUC*
COprom5GW	(attB1) ATTTACTCTTCATGAACTCGAACCA	1.8kb Fw	*1.8kb::GFP:CO* *1.8kb::LUC*
GW_CO_7_FW	(attB1) CATTATGAGCATAACACCAACAAGG	1.4kb Fw	*1.4kb::GFP:CO*
M1_FW	(attB1) CCAATAGAATGTTACTATTTGG	CDM Fw	*CDM:pnos::LUC*
M1_RV	(attB2) GGAAAAGTGCAGTGACCTTTGG	CDM Rv	*CDM:pnos::LUC*
M2_FW	(attB1) CTTGTATGTTACGTTCCAATT	CMM Fw	*CMM:pnos::LUC*
M2_RV	(attB2) CTTAAGAGCTGAAAAGTGAGT	CMM Rv	*CMM:pnos::LUC*
M3_FW	(attB1) CATAGGAGGAGATGTTTACAC	CPM Fw	*CPM:pnos::LUC*
M3_RV	(attB2) CGAATCCTGAGTGCCTACACG	CPM Rv	*CPM:pnos::LUC*
M1_Myb_FW	ATTTGGTGTGTAGTCTAGCTAGCACG	CDMlbs Fw	*CPMlbs:pnos::LUC*
M1_Myb_RV	CGTGCTAGCTAGACTACACACCAAAT	CDMlbs Rv	*CPMlbs:pnos::LUC*
M3_Myb_Rv	(attB2) AGACTACTGAGTGCCTACACGTG	CPMlbs Rv	*CPMlbs/gbox:pnos::LUC*
M3_Myb_G_Rv	(attB2) AGACTACTGAGTGCCTATCTAGA	CPMlbs gbox Rv	*CPMlbs/gbox:pnos::LUC*

Plant transformation

Binary vectors were transformed into *Agrobacterium tumefaciens* strain GV3101 which is resistant to rifampicin. Helper plasmids *pSoup* (tetracycline) for *pGreen* based constructs and *pMP90RK* (gentamicin and kanamycin) for *pAlligator2* based constructs were used to establish stable transgenic Bacteria (Hellens et al., 2000a, 2000b). Bacteria were grown to OD_{600} of 0.6 - 1.0 and growth media was replaced by Sucrose solution [5%] supplemented with Silvet L-77 [500µl/l]. Floral dip was used to transform constructs into *A. thaliana* accessions and genotypes indicated (Clough and Bent, 1998). BASTA selection (*pGreen*) or GFP signal in the seed coat (*pAlligator2*) was used to select for transformants, test segregation to confirm single insertions and homozygosity in the following generation. Presence of the transgene was tested via PCR on genomic DNA using standard procedures (Edwards et al., 1991).

Plant material

A. thaliana accessions Col-0 and L*er* were used in this study. All *A. thaliana* mutant lines used in this study are published as indicated in the text. *A. arabicum* and *T. hassleriana* seeds were kindly provided by Eric Schranz (Wageningen, Netherlands).

Generation of *cdf1235phyB-9*

The *phyB-9* mutant was crossed to the *cdf1235* mutant by Fabio Fornara (Universita degli studi di milano, Italy). Approximately 300 F2 plants were screened for early flowering in SD conditions. Homozygosity of all mutations was confirmed by PCR on genomic DNA (Table 6) in the three earliest plants and homozygosity of the CDF1-R transgene was confirmed in the F3 generation. For genotyping the *phyB-9* mutation PCR was followed by digestion of the product with BslI resulting in a 167bp fragment for *phyB-9* but 148bp and 22bp for Col-0.

Table 6: Primer for genotyping

Name	Sequence	Purpose
PhyB-9dCAPup	GTGGAAGAAGCTCGACCAGGCTTTG	phyB-9 Fw
PhyB-9dCAPlw	AACGCGTTTCTCTGACATTTG	phyB-9 Rv
F195-35S-For	CTT CGC AAG ACC CTT CCT	CDF1R Fw
F215-CDF1-1R	GAG CGC TCG TCA TTG TTC	CDF1R Rv
Li18	ATT TTC GAA TCC CCA AAA GC	CDF2 Fw
Li19	CCT GAG CTT GGT GTT GTT GA	CDF2 cdf2 Rv
F279-GKLB	CCC ATT TGG ACG TGA ATG TAG ACA C	cdf2-1 cdf3-1 Fw
FO 37	CGA AAC CTT GGA CAA AGA GG	CDF3 Fw
FO 38	GAT CCG CTT GAG CAA TCA TC	CDF3 cdf3 Rv
L06H09-2R	CCA CCA GCT GTC CAA TAC CT	CDF5 cdf5-1 Fw
F276-CDF5-3R	CCT CTC TCC CTT CTA ATG CAC	CDF5 Rv
FO62 - LBa1	TGG TTC ACG TAG TGG CCA TCG	cdf5-3 Rv

Plant growth conditions and flowering time measurement

Seed were stratified on soil for three days at 4°C in darkness. Plant were grown in climate controlled conditions on soil in growth chambers (Percival or GroBank; CLF plant climatics) at 22°C day and 18-20°C night temperatures under cool white light (approximately 80µE) supplemented with incandescent light. Flowering time of *A. thaliana* was measured as total leaf number (rosette leaf number at bolting + cauline leaf number) in LD (16h light, 8h dark) and SD (8h light, 16h dark). Total leaf number at bolting was determined for *A. arabicum* in growth chambers and for *T. hassleriana* in temperature and light controlled glass house conditions.

LUC expression analysis with TOP COUNT Detection system

Ten day old seedlings (n=24) of homozygote T3 lines grown on ½MS in growth chambers (Percival) under cool white light (approximately 70µE) were transferred to 96 well plates filled with 200µl ½MS + sucrose [1%] and 18µl D-luciferin [2,5mM] was applied. Light emission was measured in Top Count machine (PerkinElmer Inc.; Waltham, Massachusetts, USA) under LD LED conditions (blue + red light, 16 hours light, 8 hours darkness) over five days. Light emission of 24 seedlings at the

indicated ZT was averaged and further normalized to the mean of the overall expression in the experiment of the corresponding line (normalized light emission).

Expression analysis of mRNA levels

Plants were grown on soil in growth chambers (Percival) under the same conditions flowering time was measured. For shift experiments to red and blue light conditions plants were grown on plates (½MS and 1% sucrose) for ten days and shifted to LED blue light, LED red light or darkness, respectively. Ten day old seedlings (15-20 for *A. thaliana*, 5-10 for *A. arabicum* and 2-3 for *T. hassleriana*) grown in the indicated light conditions were collected into liquid nitrogen at the given ZEITGEBER TIME (ZT). RNA was extracted from *A. thaliana*, *A. arabicum* and *T. hassleriana* using RNeasy Plant Mini Kit (Qiagen) according to manual. Samples were treated with Ambion DNase (life technologies) according to manual and 1.5-3µg of RNA was used for cDNA synthesis (SuperScript II Reverse Transcriptase; Invitrogen). Real time PCR was performed using IQ SYBR Green Supermix (Bio-Rad) on LightCycler480 (Roche) using the following conditions for the genes indicated (Table 7) followed by examination of the melting curve:

	95°C	5min
	95°C	20sec
45x	60°C	20sec
	72°C	20sec

Data were normalized against arbitrary dilution series to normalize PCR efficiency. One cDNA sample from the same extraction was diluted 1:1, 1:2, 1:4 and 1:8 and calibration curve was automatically calculated resulting in relative concentrations. Alternatively concentrations were calculated as $2^{(\text{cycle number})}$. Values are presented as ratios to the housekeeping gene *PP2A* for *A. thaliana* and *ACTIN* for *A. arabicum* and *T. hassleriana* to correct for differing cDNA concentrations between samples. For comparison of expression pattern samples were additionally normalized to the mean of each dataset consisting of eight ZT values throughout one day. Mean and standard error of three technical replicates from one real time PCR reaction were calculated coming from the same plate. If not stated otherwise representative experiment of at least two biological replicates is displayed.

Table 7: Primer pairs for determination of mRNA levels

Gene (gene identifier)	Name	Sequence
PP2A (AT1G69960)	F433-PP2a FW	CAG CAA CGA ATT GTG TTT GG
	F434-PP2a RV	AAA TAC GCC CAA CGA ACA AA
Actin	108 ActF	GGTAACATTGTGCTCAGTGGTGG
	109 ActR	AACGACCTTAATCTTCATGCTGC
GFP	C195	TCATGGCCGACAAGCAGAAGAACG
	C196	CGGCGGCGGTCACGAACTC

LUC	LUCRT2fw	AAGCGGTTGCCAAGAGGTTCC
	LUCRT2rev	CGCGCCCGGTTTATCATC
FT (AT1G65480)	F193-FT-qRT-F	CGA GTA ACG AAC GGT GAT GA
	F194-FT-qRT-R	CGC ATC ACA CAC TAT ATA AGT AAA ACA
AtCO (AT5G15840)	F191-CO-qRT-F	TAA GGA TGC CAA GGA GGT TG
	F192-CO-qRT-R	CCC TGA GGA GCC ATA TTT GA
AtCOL1 (AT5G15850)	Col1_RT_FW3	CACCTTACCCTCCAGCTCAG
	Col1_RT_RV	GTGGAGAAAGCTTGGTTTGC
AtCOL2 (AT3G02380)	COL2_CDS_Fw2	CCAGGGAAAAACATCGGTAA
	COL2_CDS_Rv2	TTTTGTTGGCTTTGTTGCAG
ThCO	0101Cl_CO_RT_FW	CGCTAACCCACTCGCTAGAC
	0102Cl_CO_RT_RV	CGTCCTCATCTTCAGCTTCC
AeCO	SA68Ae_CO_RT_FW	AAGGCAGACAACGCATCTCT
	0105Ae_CO_RT_RV	TCTGTTGTCGCTCTTTGCTG
AeCOL2	Aeth_COL2_Fw2	GCCTGTGATGCAGAGATTCA
	Aeth_COL2_Rv2	TCCATTATCTGCATCCGTGA

The expression levels of *A. alpina* homologs of *CO* (Aa_G356690), *COL1* (Aa_G356670) and *COL2* (Aa_G42560) were checked in next generation sequencing of cDNA from a LD timecourse (kindly provided by Eva-Maria Willing, MPIPZ, Cologne). Briefly, data plotted represents sequence reads that were mapped to the reference genome and normalized both to gene length and library size.

Chemical treatment (cordycepin)

10 day old seedlings grown on plates (½MS and 1% sucrose) in growth chambers (Percival) under LD conditions were transferred to ddH2O with 0.6mM final concentration of 3´-deoxyadenosine (cordycepin, Sigma) or mock treatment. Seedlings were vacuum infiltrated for 1min and transferred back to growth chambers and harvested in liquid nitrogen at the indicated timepoints after drying the seedlings (Yakir et al., 2007).

Chromatin Immunoprecipitation

In total three previously described ChIP protocols were performed (Bowler et al., 2004; Gendrel et al., 2002; Searle et al., 2006). Binding of HA:CDF1 and HA:CDF2 to DNA was tested using three different α-HA antibodies (Table 8). Additionally α-CDF2 antibody (Table 8, Fornara et al., 2009) was used for ChIP in Col-0 and *cdf1235* mutants (data not shown). The ChIP results displayed were obtained using the method employed by Sawa et al., (2007), which is based on Bowler et al. (2004) with minor modifications. Briefly, ten day old seedlings grown in LD were harvested at ZT8 and crosslinked with formaldehyde [1% in PBS] (twice 5min vacuum). Reaction was stopped by adding glycine [0.1M final concentration]. Material was dried and frozen in liquid nitrogen. Nuclei were extracted by mixing plant tissue with Extraction Buffer 1 [0.4M Sucrose, 10mM Tris-HCl pH8, 10mM $MgCl_2$, 5mM BME, complete tablets 1/50ml] and filtering through miracloth. Pellet was dissolved in

Extraction Buffer 2 [0.25M Sucrose, 10mM Tris-HCl pH8, 10mM $MgCl_2$, 1% Triton X-100, 5mM BME, 1 complete mini/10ml] and spun full speed. Pellet was again dissolved in Extraction Buffer 3 [1.7M Sucrose, 10mM Tris-HCl pH8, 0.15% Triton X-100, 2mM $MgCl_2$, 5mM BME, 1 complete mini/10ml] and cleaned by gradient centrifugation in the same buffer. Nuclei were lysed in Nuclei Lysis Buffer [50mM Tris-HCl pH8, 10mM EDTA, 1% SDS, 1 complete mini/10ml]. DNA is sheared with Bioruptor (diagenode) twice for 5min (30sec on/ 30sec off position H).

Dynabeads (50µl) were incubated in ChIP dilution buffer [1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris-HCl pH8, 167mM NaCl] with α-HA antibody from rat (Table 8) for at least 2hours at 4°C and subsequently washed three times in ChIP dilution buffer. Sample was diluted 1/10 in ChIP dilution buffer to lower SDS concentration to 0.1%. 10% input was taken and samples were incubated overnight at 4°C with pre-treated beads and washed on the next day with Low Salt Wash Buffer (twice) [150mM NaCl, 0.1% SDS, 2mM EDTA, 1% Triton X-100, 20mMTris-Hcl pH8], High Salt Wash Buffer (once) [500mM NaCl, 0.1% SDS, 2mM EDTA, 1% Triton X-100, 20mMTris-Hcl pH8], LiCl Wash Buffer (once) [0.25M LiCl, 1% NP-40, 1% sodium deoxycholate, 1mM EDTA, 10mM Tris-HCl ph8] and TE Buffer (twice). Tubes were changed before the last wash step. DNA was eluted using elution buffer (twice 250µl) [1%SDS, 0.1M $NaHCO_3$].

Table 8: Antibodies used for ChIP experiments

Antibody	**Supplier**	**Catalogue number / reference**
αHA rabbit	abcam	Ab13834
αHA rat	Roche	3F10 / 11867423001
αHA mouse	Roche	11883816001
αH3 rabbit	Abcam	Ab1791
αCDF2	-	(Fornara et al., 2009)

Protein-DNA crosslinking was reverted by adding NaCl [final 0.2M] and incubation at 65°C overnight (also for input). Protein was removed by incubation with EDTA [final 1mM], Tris pH 6.5 [final 40 mM] and proteinase K [4µg/ml] for 1hour at 42°C. DNA was recovered by phenol/chloroform extraction in two steps using phase separation kit/tubes. Precipitation was done overnight at -20°C is the presence of glycogen [final 25µg/ml], sodium acetate [final 0.3M] and two volumes of EtOH [100%]. Pellet was washed in EtOH [70%] and eluted in 50µl H_2O.

RT PCR was performed with 2µl sample in a 10µl total reaction volume using IQ SYBR Green Supermix (BIO-RAD) on LightCycler 480 (Roche) (Table 9) to detect immunoprecipitated DNA. Precipitated DNA was normalized against input sample after normalization to an arbitrary dilution series as described for mRNA detection.

Table 9: Primer for ChIP RT-PCR

Abbreviation	Name	Sequence
M3	0110 M3_ChIP_Fw	ACGAAGAAGTGCATAGGAGGAG
	0111 M3_ChIP_Rv	AGATCGAATCCTGAGTGCCTAC
20.1	0116 amp20.1_Fw	AAAATGCAATGACAACCACTTG
	0117 amp20.1_Rv	TTTATGCGGATTCTCCACTTTC
FT13	FT amp 13 Fw	AGA GGG TTC ATG CCT ATG ATA C
	FT amp 13 Rv	CTT TGA TCT TGA ACA AAC AGG TG

Western Blot

ChIP nuclear extract was used for detection of 3xHA:CDF2 protein expressed from *pCDF2*. Western blot was performed using α-HA antibody (rat, Table 8) as described before (Simon, 2009).

Isolation of *CO* promoters

CO promoter sequences of *Arabidopsis thaliana, Brassica napus, Brassica rapa, Sisymbrium irio, Cleome spinosa, Carica papaya, Populus trichocarpa, Solanum tuberosum* and *Medicago truncatula* were identified by BLAST search on NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The trace archives section was used to construct the *CO* promoter from *Arabidopsis lyrata* before the genome was assembled. *A. alpina CO* promoter was assembled from short reads identified by BLAST search of the preliminary genome assembly (MPIPZ, Cologne). Sequences were assembled using SeqMan Pro (DNASTAR Lasergene Version 8.0.2).

Table 10: Length of *CO* homologous promoters including start codon of CO and 3´UTR of the upstream gene *COL1*

Species	Promoter length [bp]
Arabidopsis thaliana	3594
Arabidopsis lyrata	3558
Capsella rubella	3756
Nasturtium officinale	2937
Arabis alpina	3477
Brassica napus	3494
Brassica oleracea	4827
Sisymbrium irio	3366
Sisymbrium officinale	3322
Aethionema arabicum	3169

CO Promoters from *Capsella rubella*, *Nasturtium officinale* and *Sisymbrium officinale* were obtained by cloning and sequencing of the intergenic region between *COL1* and *CO* orthologs from genomic DNA (kindly provided by Markus Berns, MPIPZ, Cologne). To design primers in conserved regions of *COL1* and *CO* genes, respectively, sequences of both genes from *A. thaliana, Arabidopsis lyrata,*

Brassica napus, Brassica rapa and *A. alpina* were aligned and a primer set was picked which amplifies about 4.5kb sequence in *A. thaliana* (Table 11). Amplification of ~4.5kb region including the whole intergenic region between *COL1* and *CO* was successful for *Arabidopsis cebennensis, Capsella rubella Nasturtium officinale* and *Sisymbrium officinale* using LA taq (TAKARA) according to suppliers manual with the following program: 94°C 1min; 35*(94°C 15sec; 56°C 30sec; 68°C 7min), 72°C 7min using genomic DNA (200µg per reaction). DNA product of expected size was excised from agarose gel and TA subcloned according to suppliers manual (pGEM-T kit; Promega) into *E.coli DH5α*. Purified vectors were sequenced by Sanger sequencing (MPIPZ Genome Centre, Cologne).

The *CO* promoter of the *A. arabicum* ortholog was identified by BAC screen. A probe for BAC library screen was amplified from *A.thaliana* cDNA (~440bp, Table 11) using standard PCR technique. BAC library was screened by Stefan Wötzel (Wötzel, 2010). BAC vector purification was performed with Plasmid Midi Kit (Qiagen) according to handbook (100 ml saturated Bacteria in LB). Lysis steps including buffer P1, P2 (prepared fresh) and P3 were scaled up 2.5 times. Middle region of *CO* was amplified using standard PCR conditions from the BAC DNA, purified and sequenced. Obtained sequence from BAC 8C and 16N was used for subsequent primer walk.

Table 11: Primers for amplification of Brassicaceae *CO* promoters and *A. arabicum* BAC screen probe

Name	**Sequence**	**Purpose**
COL1_FW to CO	TGACACMGGATATGGAATTG	Conserved COL1 Fw
CO_RV to COL1	TGGCAGAGTGRACTTGAGCA	Conserved CO Rv
CO_MR_Fw	ACTGGTGGTGGATCAAGAGG	BAC probe Fw
CO_MR_Rv	TCTTGGGTGTGAAGCTGTTG	BAC probe Rv

Bioinformatics data analysis tools

Paired end next generation sequencing reads of *A. arabicum* genomic sequence (http://vegi.cs.mcgill.ca/public-data/raw-data/sisymbrium) were *de novo* aligned using CLC Genomics Workbench (CLC Bio) with standard parameters. Desired genes were subsequently searched in the *A. arabicum* genome running BLAST on a local database. Phylogenetic tree for relationship of subgroup Ia COL protein sequences was inferred using the Neighbor-Joining method in MEGA4 with bootstrap values from 10000 replicates (Saitou and Nei, 1987; Tamura et al., 2007). Substitution rates were calculated from a codon alignment of the respective coding sequences followed by calculation of substitution rates using SNAP v1.1.1 (Rodrigo and Learn, 2001, Chapter 4, pages 55-72).

Promoter and locus alignments were performed using LAGAN algorithm on mVISTA (Brudno et al., 2003; Mayor et al., 2000). Output is displayed as conservation in per cent in 100bp sliding windows to the corresponding reference sequence. For motif identification datasets were submitted to MEME motif identification tool (Version 4.6 ,Bailey and Elkan, 1994). Two different setting were used.

Default settings identify motifs with a maximum motif width of 50bp and maximum motif number of three. Alternatively short motifs were identified (6bp to 10bp) that can occur in any number. Alignments are displayed using ClustalX 2.0.11 (Larkin et al., 2007). WebLogo was used to present conserved motifs (Version 2.8.2, Crooks et al., 2004). Phylogenetic tree for promoter comparison was created using Phylogeny.fr (Dereeper et al., 2008). *Cis*-elements were counted using customized Perl script (Geo Velikkakam James, MPIPZ, Cologne). Sequences were shuffled 100 times by the algorithm “shuffle” (Version 1.02, Press et al., 2007 , p.281). One Way ANOVA and tuckey test were performed using SigmaStat 3.5 (Systat Software)

4. Phylogenetic relationship and functional diversification of *CO* Homologs in the Brassicaceae and Cleomaceae

Polyploidization is common throughout the evolution of higher plants (Soltis et al., 2009; Tang et al., 2008). The Brassicaceae and its sister family Cleomaceae in the order *Brassicales* underwent independent paleopolyploidization events, *At-α* and *Th-α*, respectively (Barker et al., 2009). Single gene duplication is another common and important evolutionary mechanism (Wang et al., 2013). For instance *A. thaliana CO*, *COL1* and *COL2* arose through the At-α WGD (*COL2* and *CO/COL1*) followed by a single gene duplication (*CO* and *COL1*) (Bowers et al., 2003). In *T. hassleriana* only one subgroup Ia *COL* gene was identified (Schranz and Mitchell-Olds, 2006), which shares one common ancestral gene with its three *A. thaliana* orthologs. Here the conservation of subgroup Ia *COL* genes in the Brassicaceae is investigated by comparison to their orthologs in the basal Brassicaceaen species *A. arabicum* (Bailey et al., 2006) with the aim of assigning regulatory and functional features to the *T. hassleriana COL* ortholog to define the function and regulation of the last common ancestral gene of *CO*.

4.1. *A. arabicum* and *T. hassleriana* are long day plants

Photoperiodic flowering response depends on *CO* function in *A. thaliana* (Putterill et al., 1995). A flowering time response to long days is thus an indication of the presence of this pathway and *CO* in related species. Hence flowering time of the basal Brassicaceae species *A. arabicum* gives insights whether long day acceleration of flowering is conserved in the Brassicaceae. *T. hassleriana* belongs to the new world group of the Cleomaceae family which is sister to the Brassicaceae family. *T. hassleriana* is present in South America and shows a mild flowering time response to LD conditions (Koevenig, 1973; Sánchez-Acebo, 2005).

The photoperiodic flowering response of *A. arabicum* was tested. It flowered after a mean of 41.7 days with 21.7 leaves at bolting under LD conditions of 16h light and 8h dark at 20°C (Fig. 4a,b). In controlled conditions plants did not flower under SD conditions (8h light / 16h dark). Thus *A. arabicum* represents an obligate long day requiring plant.

T. hassleriana flowering time was examined by leaf number at bolting under different day length in glasshouse conditions. The first experiment was performed in autumn and winter and *T. hassleriana* flowered with around 30 leaves both in LD (16h light) and SD (8h light) (Fig. 4c,d). The second experiment was performed in summer and while SD grown plants still flowered with ~30 leaves the LD grown plants flowered with around 20 leaves. *T. hassleriana* seems to be either dayneutral or a facultative LD plant with a slight acceleration of flowering under inductive conditions, which is in agreement with its putative origin in day neutral conditions. The weak photoperiodic flowering

response also indicates that the LD pathway as known from *A. thaliana* might be impaired in this species.

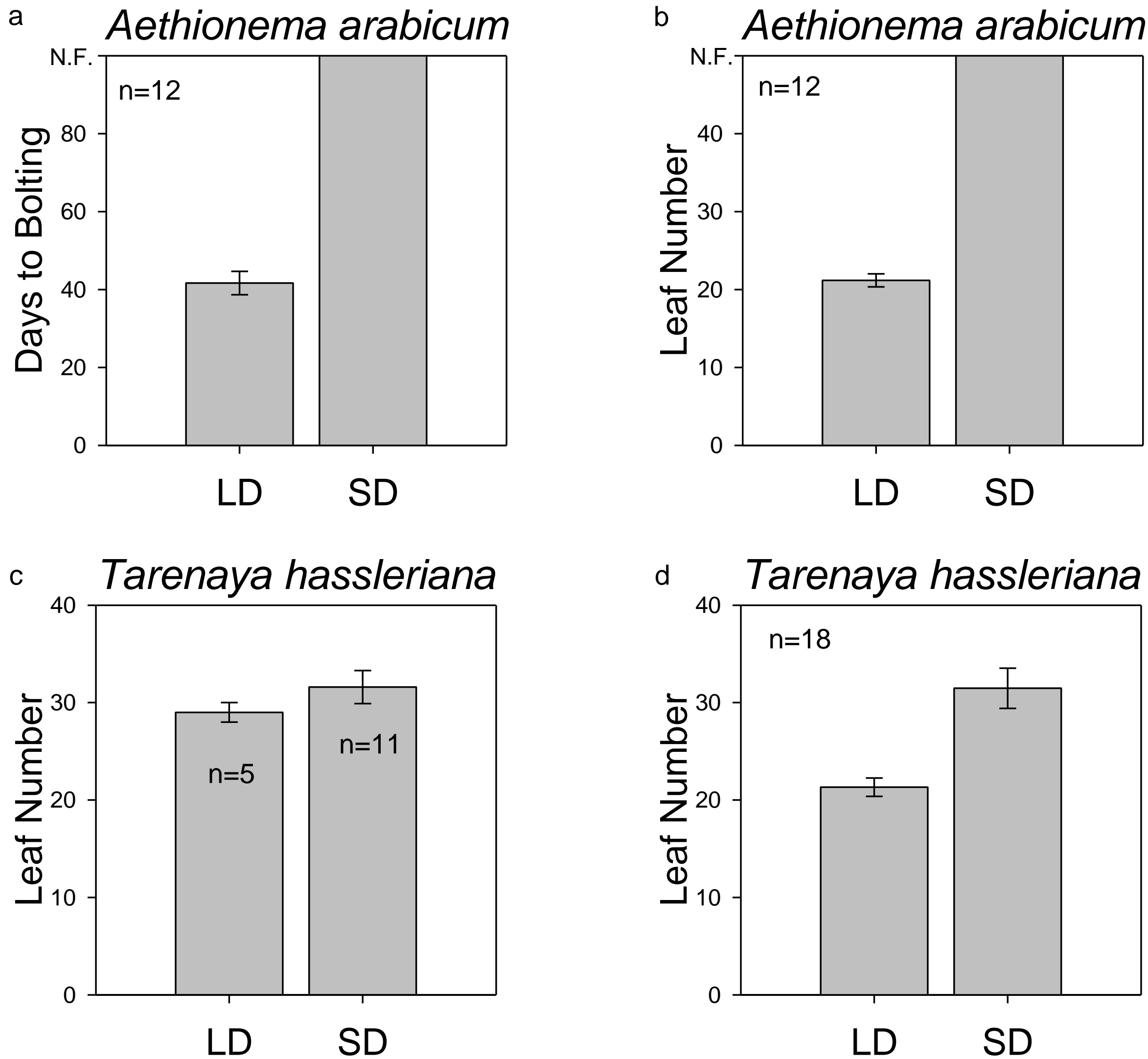


Figure 4: Flowering time of *A. arabicum* and *T. hassleriana*.
Mean flowering time as days to bolting (a) or leaf number at bolting (b) of *A. arabicum* in LD (16h light/8h dark) and SD (8h light / 16h dark) conditions. n=12; N.F. = Not flowering. Mean flowering time of *T. hassleriana* plotted as leaf number at appearance of first flowers measured under controlled greenhouse conditions under artificial lights in winter (c) and summer (d). Plants tested indicated by n. Error bars represent standard error.

4.2. Subgroup Ia *COL* genes are conserved in the Brassicaceae and *T. hassleriana* contains only one such gene that is closely related to *AtCOL2*

A. thaliana CO has two close homologs (*COL1* and *COL2*) while only one subgroup Ia *COL* gene (*ThCOL*) was identified in *T. hassleriana* yet (Schranz and Mitchell-Olds, 2006). In order to examine whether the three subgroup Ia *COL* genes are conserved in the Brassicaceae their presence was checked in the core Brassicaceae species *A. alpina* as well as in the more distally related species *A. arabicum* from the genus Aethionema, which split from the core of the Brassicaceae family early during its radiation. The goal of this study is to define whether the ancestor of subgroup Ia *COL* genes

in *A. thaliana* was a functional CO ortholog and sequence information from *ThCOL* can provide an indication.

For *A. thaliana* the sequence information of the desired three loci was retrieved from public databases (TAIR10 *CO*=At5G15840, *COL1*=At5G15850, *COL2*=At3G02380). For *A. alpina* the preliminary alignment of the genome sequence available at the MPIPZ Cologne was used. All three genes were identified by BLAST search and microsynteny of the hits was checked (data not shown), which confirmed that neighbouring genes in *A. alpina* show homology to the corresponding genes in *A. thaliana*.

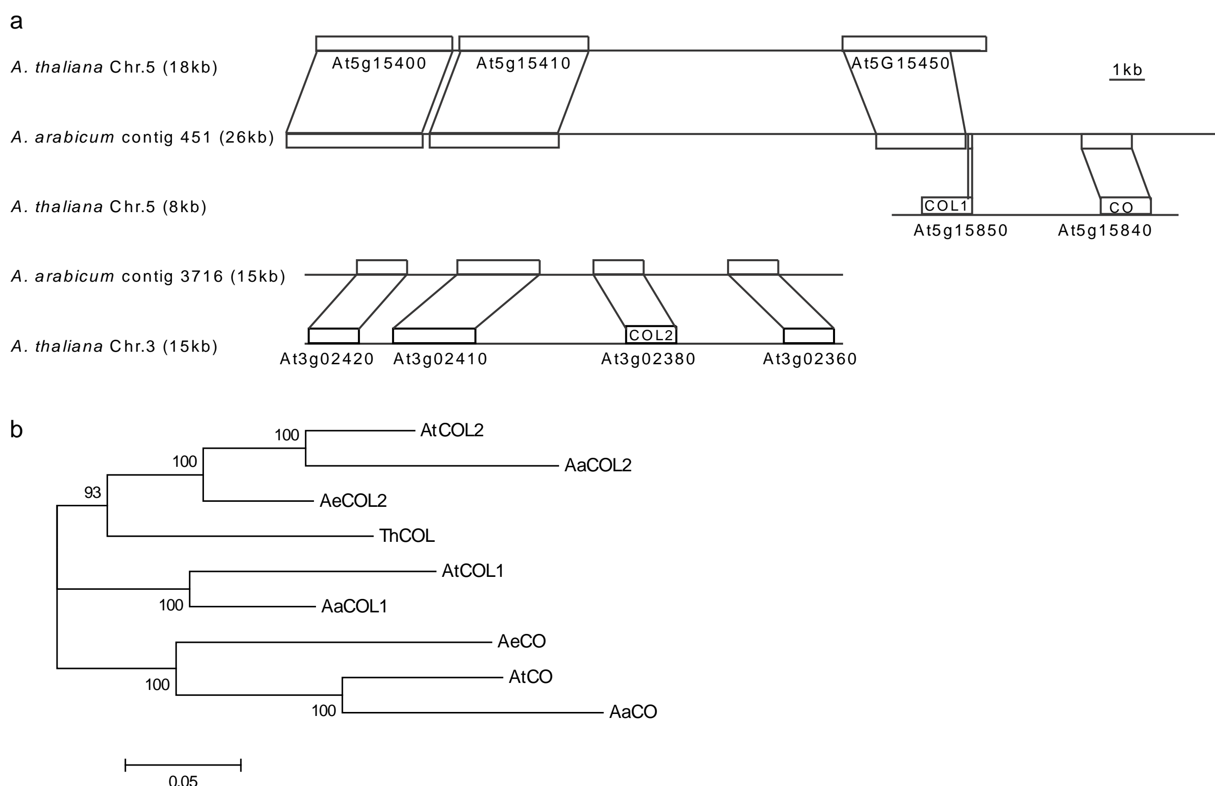


Figure 5: Phylogenetic Relationship of COL group 1a proteins in *A. thaliana*, *A. alpina*, *A. arabicum* and *T. hassleriana*

Syntenic relationship between *A. thaliana* and *A. arabicum CO*, *COL1* and *COL2* homologous genes was checked in their chromosomal context (a). No full ortholog of *AeCOL1* is present due to a rearrangement. Neighbor-joining tree (b) of amino acid sequence from subgroup Ia COL proteins from *A. thaliana* (At), *A. alpina* (Aa), *A. arabicum* (Ae) and *T. hassleriana* (Th). Numbers indicate bootstrap values of 1000 iterations.

The genomic sequence of *A. arabicum* is now publicly available (Haudry et al., 2013). However, the next generation sequencing reads from a public database were available earlier and were used to create a rough assembly of the genome using CLC workbench (CLC Bio). Data from four next generation paired end sequencing runs with an insert size of 280bp were aligned. The assembly reached a N50 value of ~11.9kb with the largest contig being ~105kb. Using this dataset two contigs containing BLAST hits for *CO* and *COL2* were identified. Contig 3716 (~15kb) contained sequence homologous to *COL2* and microsynteny was conserved with *A. thaliana* (Fig. 5a). Contig 451 (18kb) included a homolog of *CO*. The downstream sequence of *CO* was identified to be homologous confirming the

microsynteny. The whole intergenic region between *CO* and *COL1* was identified as well but only the distal 59bp of the coding region of the *COL1* homolog are present in the assembly. The remaining upstream region seems to have undergone chromosomal rearrangement. In parallel *CO* was identified by BAC screen (Chapter 5) and sequenced by SANGER sequencing. Both sequences, spanning the intergenic region from *COL1* to *CO* as well as *CO* coding region were aligned without a mismatch confirming that the result is not an artefact of either the screening of a BAC library of *A. arabicum* genomic DNA or the next generation sequencing approach. The region present at the *COL1* homolog in *A. arabicum* is homologous to a different part of Chr.5 in *A. thaliana* (Fig. 5a). The proximal part of *COL1* in *A. thaliana* was not identified by BLAST search in *A. arabicum*.

For comparison to *T. hassleriana* the sequence from BAC Cs1 *COL* homolog as identified in a previous study was used (Schranz and Mitchell-Olds, 2006). It was also confirmed by BLAST search that no *COL* orthologs could be identified on the two other BAC sequences from orthologous regions to the other *COLs* in *A. thaliana* as investigated in Schranz and Mitchell-Olds (2006).

The protein sequences of all subgroup Ia *COL* genes identified were aligned and a phylogenetic tree was constructed using the neighbour-joining method (Fig. 5b). Remarkably only one paralogous gene is present in *T. hassleriana*, here referred to as *ThCOL* and no orthologous gene to *COL1* is present in *A. arabicum*. CO, COL1 and COL2 orthologs group together with 100% bootstrap support. ThCOL groups with COL2 orthologs with a bootstrap support of 93%, but COL2 orthologs from the Brassicaceae family are closer related to each other than to ThCOL. The tree did not resolve the relationship between the gene subfamilies themselves. Taken together the ThCOL protein has a closer phylogenetic relationship to COL2 compared to CO or COL1 orthologs. As ThCOL has an equal evolutionary distance to all other genes it is unexpected that it is more closely related to one group unless it shares functionality with this gene. The findings thus indicate that *ThCOL* might be a functional homolog of *COL2* and not *CO* or *COL1*.

The nucleotide substitution rates of *AtCO*, *AtCOL1*, *AtCOL2* and *ThCOL* were compared in a codon alignment of the corresponding CDS (Table 12). The comparison of non-synonymous and synonymous substitution rates (dn/ds) gives an indication of the selective pressure that acts on coding sequences. The comparisons of *A. thaliana* subgroup Ia genes with *ThCOL* showed dn/ds values below 1 indicating stabilizing selection. ThCOL and AtCOL2 proteins, which are phylogenetically closer related compared to AtCOL1 and AtCO (Fig. 3b), additionally exhibited lowest total (117.8) and corrected (0.16) number of non-synonymous substitutions in their respective DNA sequences. This resulted in the lowest dn/ds (0.21) for *AtCOL2* compared to *ThCOL* indicating that these two genes are under stabilizing selection and additionally highlighting their similarity.

The number of substitutions can also be used to estimate the relative time when the evolutionary split between sequences occurred and two assumptions can be tested in this dataset. First, the idea that the

three subgroup Ia genes in *A. thaliana* and *ThCOL* have one last common ancestral gene is supported by the finding that the comparison of *AtCO*, *AtCOL1* and *AtCOL2* to *ThCOL* led to the highest total synonymous substitutions observed (>110). Second, lowest number of total (223) and synonymous changes (79.5) was observed for *AtCO* and *AtCOL1* supporting the current hypothesis that the tandem duplication of *CO* and *COL1* occurred after the At-α WGD and thus after the split from *AtCOL2*.

Table 12: Non-synonymous and synonymous substitution rate in *A. thaliana* and *T. hassleriana* subgroup Ia *COL* genes

The number of observed synonymous substitutions (Sd) and the number of observed non-synonymous substitutions (Sn) as well as total substitution number (total) is displayed in a pairwise comparison. The values were corrected using Jukes-Cantor method for synonymous (ds) and non-synonymous (dn) substitutions and ratio (dn/ds) was calculated. Lowest values in each class are marked in red.

Gene name		**Sd**	**Sn**	**total**	**ds**	**dn**	**dn/ds**
AtCO	*ThCOL*	111.50	148.50	260	0.726	0.194	0.267
AtCOL1	*ThCOL*	113.50	142.50	256	0.772	0.196	0.254
AtCOL2	*ThCOL*	111.17	117.83	229	0.774	0.162	0.210
AtCO	*AtCOL1*	79.50	143.50	223	0.457	0.198	0.433
AtCO	*AtCOL2*	100.67	135.33	236	0.679	0.191	0.281
AtCOL1	*AtCOL2*	109.83	130.17	240	0.808	0.193	0.239

4.3. Expression patterns of *A. thaliana*, *A. arabicum* and *T. hassleriana* group Ia *COL* genes are conserved

The expression pattern of *CO* is important for its function as a day length sensor and is distinct from those of *COL1* and *COL2* in *A. thaliana* (Fornara et al., 2009; Ledger et al., 2001). The expression patterns of *A. alpina* and *A. arabicum* were checked to confirm conservation to expression patterns found in *A. thaliana* homologs. Examining expression pattern of *ThCOL* gives indications of how the last common ancestor of all Brassicaceae subgroup Ia *COL* genes was expressed.

Expression pattern of subgroup Ia *COL* genes were examined in seedlings grown under LD and SD conditions. All mRNA levels examined in *A. thaliana* in this study were normalized to the housekeeping gene *PROTEIN PHOSPHATASE 2A* (*PP2A,* Eisenberg and Levanon, 2003; Wassarman et al., 1996). For comparison of expression patterns the values were normalized to the mean of each experiment (normalized expression). *A. thaliana CO* (*AtCO*) mRNA levels changed over the course of a day and were different in LD and SD. Rapid downregulation of mRNA at dawn (between ZT0 and ZT3) and the rise of mRNA levels starting at ZT9 are characteristic for expression in seedlings grown for ten days under LD conditions. Under LD conditions in soil grown seedlings a strong night peak (ZT18 to ZT21) compared to evening peak (ZT9 – ZT15) was observed (Fig. 6a). In SD conditions the peak expression was similar. The evening peak was more pronounced than the night peak in SD compared to LD. Importantly, *CO* mRNA was not present at ZT0 under SD conditions while under LD conditions mRNA was detected (ZT0 was harvested exactly at lights on in this study).

The expression of *A. thaliana COL1* (*AtCOL1*) showed highest expression at the beginning of the day in both LD and SD. The mRNA levels were reduced but clearly detectable at ZT3 and ZT6 in contrast to *AtCO*. The trough expression was reached between ZT9-ZT15. *AtCOL1* expression started rising again afterwards. No major differences between LD and SD were observed.

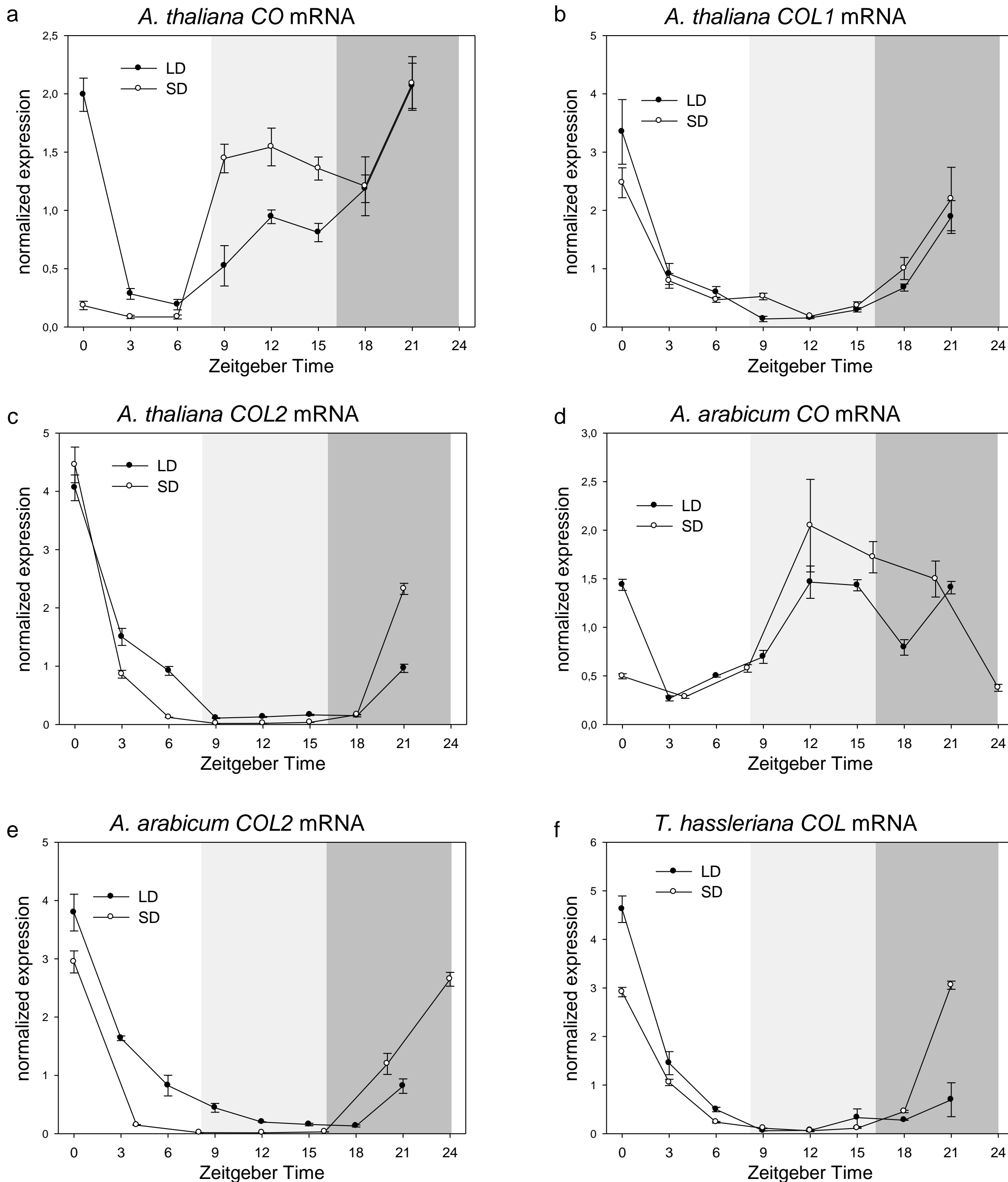


Figure 6: Normalized mRNA levels of subgroup Ia *COL* genes in *A. thaliana*, *A. arabicum* and *T. hassleriana*.
Time course experiment determining mRNA levels of *AtCO* (a), *AtCOL1*(b), *AtCOL2*(c) *AeCO* (d) *AeCOL2* (e) and *ThCOL* (f) in 10 day old *A. thaliana*, *A. arabicum* or *T. hassleriana* seedlings at the indicated Zeitgeber Time in long day (LD) or short day (SD) conditions. cDNA concentrations were normalized to *PP2A* (*A. thaliana*) or *ACTIN* (*A. arabicum* and *T. hassleriana*) followed by normalization to the average of each experiment (normalized expression). Dark grey and light grey areas indicate darkness in LD and SD, respectively. The mean of three technical replicates of the same cDNA is displayed with error bars indicating standard deviation.

The expression patterns of *A. thaliana COL2* (*AtCOL2*) were similar to *AtCOL1* (Fig. 6c). Its mRNA levels stayed higher at ZT3 and ZT6 in LD compared to SD but expression rose later in LD at ZT18 and ZT21 compared to SD. This pattern indicated that the expression might be adjusted to the beginning of the night (dawn tracking). In contrast to *AtCOL1* the mRNA of *AtCOL2* only rises after ZT18.

A. arabicum CO (*AeCO*, Fig. 6d) mRNA was expressed in both LD and SD conditions in ten day old seedlings. Its expression shared key features with *AtCO* but not with *AtCOL1* and *AtCOL2*. In LD conditions its mRNA was efficiently downregulated upon lights on and its trough of expression was reached at ZT3 and ZT6. The mRNA started to increase again afterwards and the evening and night peaks were identified. In SD conditions only one broad peak was identified between ZT12 and ZT20. Importantly the trough of expression started at ZT24 and continues until ZT8 spanning the complete light period, a pattern similar to expression of *AtCO* where it is important for gene regulation. The overall mRNA expression pattern of *A. arabicum COL2* (*AeCOL2*, Fig. 6e) showed high similarity to expression of *AtCOL1* and *AtCOL2* both in SD and LD. As seen for *AtCOL2* the peak of expression was shifted between LD and SD similar to *AtCOL2*. Levels of mRNA rose earlier in the end of the SD and also fell earlier in the beginning of the day, meaning the trough of mRNA was reached at ZT3 in SD but only at ZT9 or ZT12 in LD, which was even more pronounced than the effect observed in *AtCOL2*.

Additionally the expression of *A. alpina* subgroup Ia *COL* genes was checked in a long day timecourse. RNA seq data (kindly provided by Eva-Maria Willing, MPIPZ, Cologne) showed that *AaCO* is downregulated in the morning upon lights on and has both the evening and the night peak (Fig. 7a). *AaCOL1* and *AaCOL2* show morning expression, peaking at ZT0 (ZT24) (Fig 7b). *AaCOL1* has its trough expression between ZT8 and ZT16 while *AaCOL2* mRNA is lowest between ZT4 and ZT20.

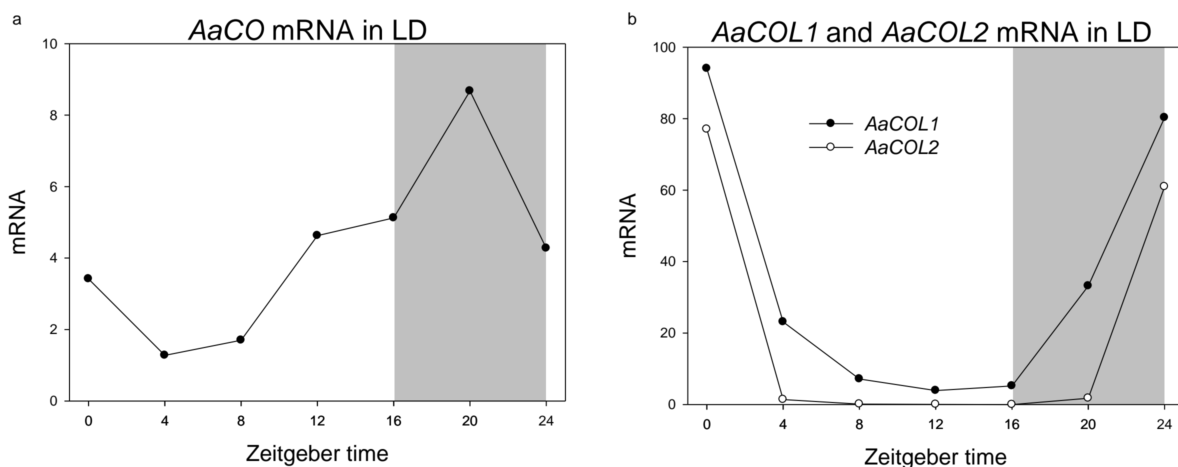


Figure 7: mRNA Levels of subgroup Ia orthologs from *A. alpina*
mRNA levels of *A. alpina* (*Aa*) *AaCO* (a), *AaCOL1* (b) and *AaCOL2* (b) orthologs as determined by RNA-seq in LD conditions. Grey areas specify hours in the dark.

As differences were clear and conserved between *CO*, *COL1* and *COL2* close orthologs in *A. thaliana, A. alpina* and *A. arabicum* it was of interest to test the expression pattern of the only member of subgroup Ia *COLs* in *T. hassleriana*, as this likely represents the expression pattern of the last common ancestral gene. *T. hassleriana COL* (*ThCOL*) mRNA was detected in seedlings (Fig. 6f) and leaves of flowering plants (data not shown). The mRNA pattern clearly shared all the key features described so far for a Brassicaceae *COL1* or *COL2* ortholog and not of *CO*. Expression of *ThCOL* was highest around ZT0 and reached its trough at ZT9. Expression significantly rose at ZT21 but mRNA levels increase earlier under SD compared to LD, a feature described for both *AtCOL2* and *AeCOL2* expression.

In general the expression patterns of subgroup Ia *COL* genes in the Brassicaceae can be assigned to two classes one being similar to *AtCO* while the other is similar to *COL1/COL2*. The night peak is common between both patterns but differences like the rapid morning downregulation of *CO* are not shared. The phylogenetic relationships between orthologs correspond to the conservation of expression patterns for all genes investigated. *ThCOL* expression groups with *COL1/COL2* expression confirming previous findings that ThCOL protein is most closely related to COL2 orthologs compared to COL1 and CO orthologs. In summary this suggests that the *COL1/COL2* expression patterns represent the ancient state and that *CO* mRNA pattern evolved in the Brassicaceae.

4.4. Subgroup Ia *COL* genes from *A. thaliana*, *A. arabicum* and *T. hassleriana* are able to complement the *co-10* mutant

Phylogenetic analysis and expression studies indicate that *ThCOL* is not a functional homolog of *CO*. In order to test this hypothesis as well as confirm conservation of *CO* function in the Brassicaceae complementation experiments using *AeCO* and *ThCOL* CDS were performed. *AtCOL1* and *AtCOL2* as well as *AeCOL2* were included as negative control as overexpression of *AtCOL1* or *AtCOL2* from the CaMV 35S promoter was shown to have no effect on flowering time regulation in *A. thaliana* (Kim et al., 2013; Ledger et al., 2001). Complementation experiments were performed with 3xHA (HA) tagged versions of DCSs expressed from the native *AtCO* promoter (2.6kb) or the phloem specific *SUC2* promoter (*pSUC2*) in *A. thaliana co-10* background (co-sail) (Imlau et al., 1999; Laubinger et al., 2006). Flowering time of T1 plants was scored in LD conditions.

Flowering time experiments of *pSUC2* driven *COL* lines derived from *co-10* mutant were performed in the glasshouse in LD conditions. Col-0 plants flowered with mean of 12.9 leaves while *co-10* control flowered with a mean of 26.7 leaves (Fig. 8a). A homozygous *pSUC2::HA:AtCO* (in *co-10*) line was tested as a positive control and plants flowered earlier compared to wild type with a mean of 5.5 leaves. The flowering time of *pSUC2::HA:AeCO* was as predicted earlier than Col-0 with a mean of 7.9 leaves. Interestingly *AtCOL1* and *AtCOL2* as well as *AeCOL2* and *ThCOL* behaved in a similar way. Flowering time ranged between a mean of 8.6 leaves for *AtCOL2* and 13.3 leaves for *ThCOL*. A one way ANOVA followed by a tuckey test (pairwise comparison) was performed to identify

statistical significance. In this test Col-0, *co-10* and *pSUC2:HA:AtCO* plants fall into statistically distinct groups ($p<0.05$). Interestingly the only line that was statistically not distinguishable from *pSUC2::HA:AtCO* was *pSUC2::HA:AeCO*. The result also confirmed that all other constructs were able to at least rescue the flowering phenotype of *co-10* to wild type levels. Flowering time of *pSUC2::HA:AtCOL2* was statistically earlier than Col-0 but later than *pSUC2::HA:AtCO*. Thus overexpression of all subgroup Ia *COL* genes in the phloem companion cells rescued the late flowering phenotype of *co-10* mutant indicating that all the *COL* genes tested have the ability to promote flowering in *A. thaliana*.

The flowering time of complementation lines driven by the native *A. thaliana CO* promoter (*pAtCO*) was tested in growth chambers (Fig. 8b). In this experiment the control Col-0 flowered with a mean total leaf number of 23.3 while *co-10* flowered with a mean of 44.7. A homozygous control line carrying *pCO::HA:AtCO* in *co-10* background was used as a positive control and flowered with a mean total leaf number of 8.8 leaves. The mean total leaf number of lines carrying *pAtCO::HA:AeCO* was 14.8 leaves. Flowering time of *pAtCO* driven *AtCOL1* and *AtCOL2* was mean of 32.5 and 36.3 leaves, respectively. Lines carrying *AeCOL2* flowered similarly with a mean of 33.1 leaves. *pAtCO::HA:ThCO* lines flowered with a mean of 27.6 leaves. One Way ANOVA and tuckey test was performed for this experiment as well and indicated that *AeCO* lines are significantly ($p<0.05$) earlier flowering than Col-0 control but not statistically different from the *AtCO* control line. Interestingly all other genotypes fall into one group being statistically later than Col-0 but also statistically earlier than *co-10*, except *ThCOL* which is indistinguishable from Col-0.

In summary, complementation with *AeCO* yielded a similar flowering time compared to the *AtCO* lines tested, which confirmed the hypothesis that *CO* function is conserved down to the basal Brassicaceae species *A. arabicum*. The *T. hassleriana* ortholog failed to give the same result. The capacity of substituting *CO* function in *A. thaliana* was comparable to the other subgroup Ia *COL* genes tested. Interestingly the use of different promoters namely *pCO* and *pSUC2* provided different results from previous publications using the CaMV 35S promoter. It is possible to complement the co-10 mutation using all subgroup Ia *COL* genes tested when driven by the *SUC2* promoter. A partial complementation was even observed using *pCO*. This is in agreement with a previous study in which a *Chlamydomonas reinhardtii CO* homolog was used to complement *co* mutation. The authors observed an earlier flowering phenotype when using the *pSUC2* promoter as well as the *p35S* (Serrano et al., 2009). The expression of *CO* (An et al., 2004) as well as its target *FT* (Adrian et al., 2010) and its upstream regulators, *CDFs* (Fornara et al., 2009; Sawa et al., 2007) and *FBHs* (Ito et al., 2012a), is limited to the vasculature where *pSUC2* is active. Using the *SUC2* promoter thus creates a situation in which the transgene is highly expressed in the appropriate cell type. The same is true for the native CO promoter. However when expressed from the native promoter mRNA levels are lower than when expressed under the control of *pSUC2* but significantly higher than for wild type *CO* (Song et al.,

2012 Fig. S3). In summary all *COL* genes tested have the capacity to promote flowering but nonetheless Brassicaceaen *CO* orthologs are functionally more active. Together with the phylogenetic history these data indicate that *CO*, *COL1* and *COL2* have the capacity to promote flowering in *A. thaliana* but *CO* became functionally more potent during the early evolution of the Brassicaceae.

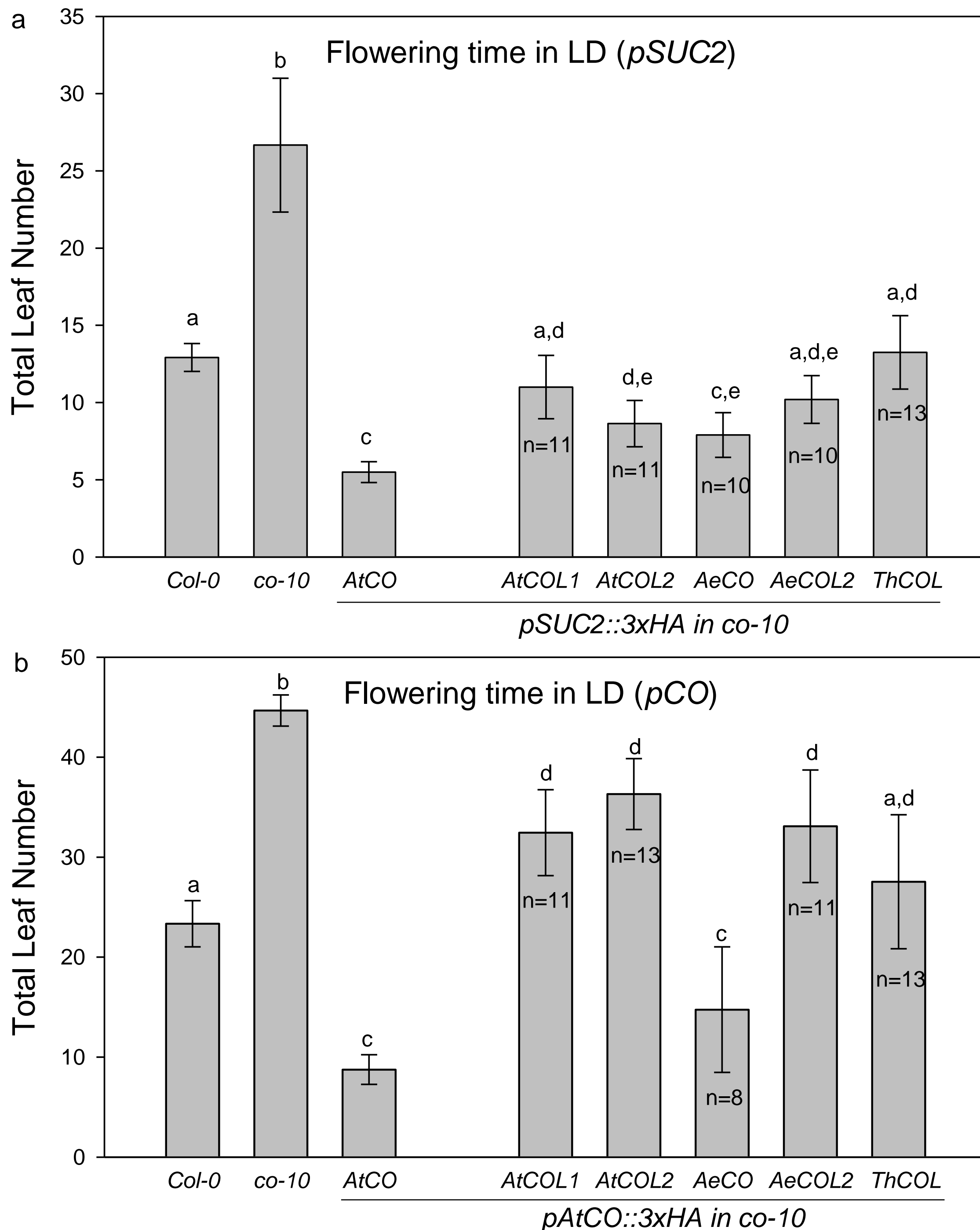


Figure 8: Flowering time of complementation lines using cDNA from subgroup Ia *COL* genes genes from *A. thaliana*, *A. arabicium* and *T. hassleriana*.

A.thaliana co-10 null mutant was complemented with cDNA of *AtCOL1*, *AtCOL2*, *AeCO*, *AeCOL2* and *ThCOL* expressed from *pSUC2* (a) or *pAtCO* (b). Flowering time was tested using T1 plants selected by GFP signal in the seed coat of indicated number (n) under LD conditions in either glasshouse (*pSUC2*) or climate chamber (*pCO*). *wt*, *co-10* and homozygous *pCO::HA:CO* and *pSUC2::HA:CO* are shown as controls (n=12). Error bars represent Standard deviation. Statistical significance was calculated using a One Way ANOVA followed by a Tuckey test (pairwise comparison) $p<0,05$, genotypes showing statistically similar flowering phenotypes are grouped (small letters: a-d).

4.5. Amino acid sequence changes may explain functional differences between CO, COL1, COL2 and ThCOL

The above results suggest that ThCOL is more similar to COL2 than to CO or COL1 in phylogeny based on amino acid sequence, mRNA expression patterns and function. Thus the amino acid alignment of all subgroup Ia COL proteins was examined to identify differences that could explain why CO orthologs are the most active. The sequence alignment between COL1, COL2 and CO orthologs (Fig. 9) reveals high conservation. The two B-Boxes and the CCT domain (Robson et al., 2001) are highly conserved over long stretches (Fig. 9). When comparing the CO homologs to all other proteins at least nine amino acid positions are conserved in all COLs but changed in all CO orthologs (see arrows in Fig. 9). Of these, four are found inside the conserved B-Box and CCT domains. One substitution from serine in CO to alanine in COLs is found in B-Box I (AtCO,S49A). Having a serine at this position seems to be the derived state, as e.g. AtCOL4, *Pharbitis nil* CO (PnCO) and OsHD1 all carry an alanine at this position (Griffiths et al., 2003). This exchange was already tested in a transgenic approach and no effect was observed when this amino acid was changed in AtCO (Kim et al., 2013). The same is true for the first substitution in B-Box II (AtCO, D79A), where not only COL1 and COL2 orthologs have an alanine but also other subgroup 1 COL proteins AtCOL3-5, PnCO and OsHD1 (Griffiths et al., 2003). The second substitution in B-Box II is a change from Isoleucine to valine (AtCO, I90V). In this case AtCOL3-5 carry an Isoleucine, meaning that Isoleucine could represent the ancient state (Griffiths et al., 2003). The amino acid substitution in the CCT domain (AtCO, N342K) where CO orthologs have an Asparagine (N) but all others carry a Lysine (K). Carrying a Lysine at this position is the ancient state as AtCOL3-5, PnCO, OsHD1 have the same amino acid on that position. Conservation goes further in evolution as related CCT domain proteins like PRR3, PRR5, PRR9 and TOC1 also have a conserved Lysine residue at that position (Gendron et al., 2012).

All conserved amino acid changes are observed between CO and the other COLs. No conserved amino acid changes are common between ThCOL and Brassicaceaen CO indicating that ThCOL protein might not function as a CO ortholog but rather as a COL1/2 ortholog. In agreement with both expression and complementation data ThCOL is more closely related to COL2 orthologs in the phylogenetic tree reconstruction but there are also amino acid changes that are shared with either COL1 or COL2 supporting the idea that all subgroup Ia genes in the Brassicaceae and *ThCOL* derived from one last common ancestral gene and might have undergone sub functionalization later in evolution. In conclusion amino acid differences between CO and closely related COL proteins were identified that might be causal for the functional change from the putative ancestral protein.

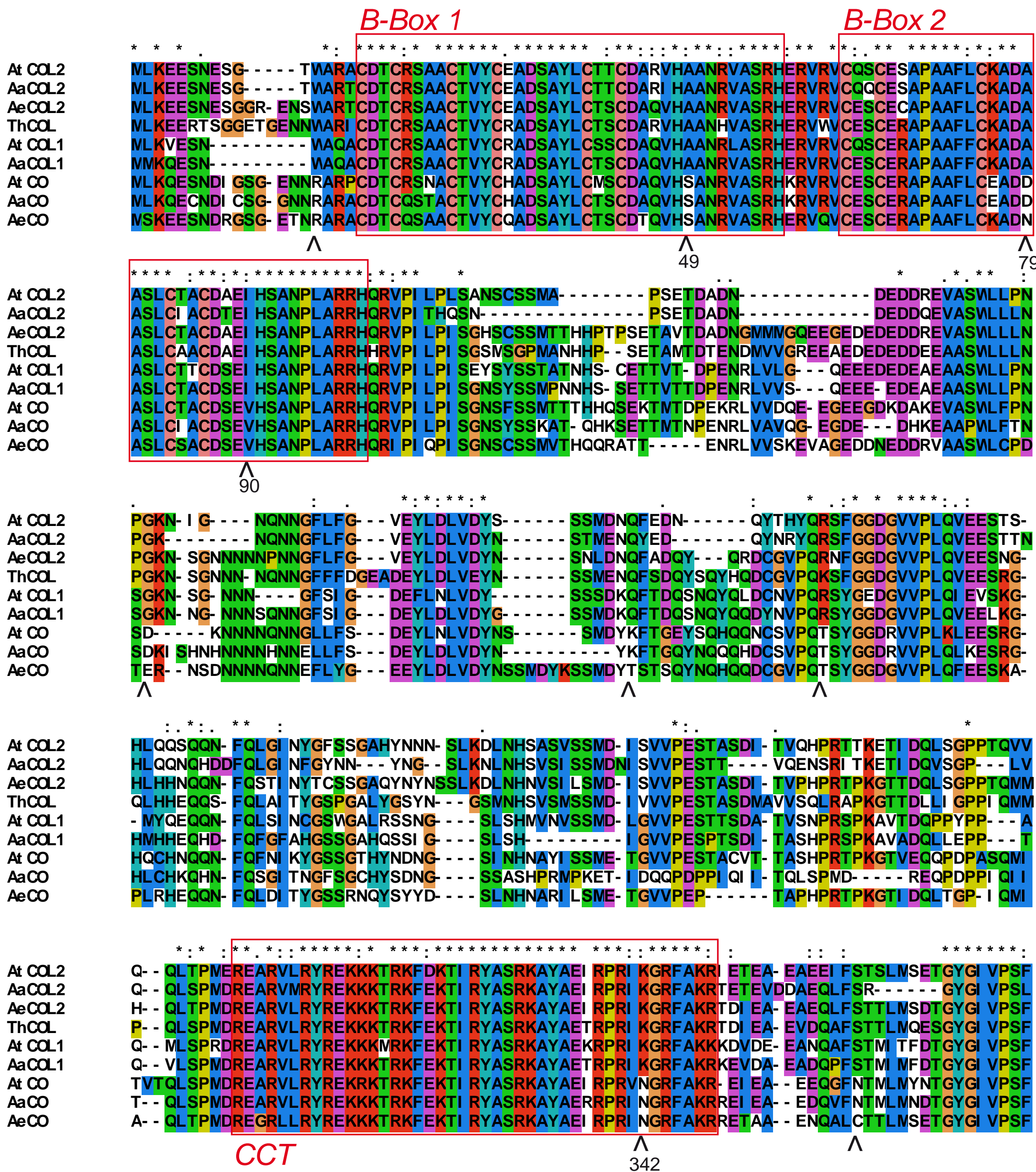


Figure 9: Sequence alignment from subgroup Ia *COL* genes in *A. thaliana*, *A. alpina*, *A. arabicum* and *T. hassleriana*
Sequence alignment (amino acids) from *CO*, *COL1* and *COL2* orthologs *A. thaliana (At)*, *A. alpina (Aa)*, *A. arabicum (Ae)* in comparison with *COL* gene from *T. hassleriana (Th)*. Red frames mark B-Boxes and CCT domain, respectively. < indicates amino acids that are conserved in all *COL* orthologs but different in *CO* orthologs. Asterisks indicate conserved amino acids.

4.6. Discussion Chapter 4

The flowering time of *A. arabicum* was tested in controlled conditions. *A. arabicum* is an obligate LD plant in greenhouse conditions that did not flower in SD. This is in contrast to the facultative LD plant *A. thaliana*. The response to cold exposure (vernalization) was not checked but *A. arabicum* flowered after ~40 days in LD already, and therefore is unlikely to show a strong response to vernalization.

Two different behaviours were observed for *T. hassleriana* in the two biological replicates presented. The first experiment found no significant differences in leaf number at bolting between LD and SD

conditions as seen for day neutral plants. The second experiment performed in summer conditions revealed earlier flowering in LD compared to SD indicating that *T. hassleriana* is a facultative LD plant. This finding is supported by a previous study in which *T. hassleriana* is described as a facultative LD plant (Koevenig, 1973). Similar to this study mean leaf numbers of 18.4 in LD and 30.6 in SD at flowering were observed. *T. hassleriana* which is found in South America (Sánchez-Acebo, 2005) shows a response to long photoperiods.

Identification of subgroup Ia *COL* genes revealed the lack of a *COL1* ortholog in *A. arabicum*. It might be found in another position in the genome but will likely be non-functional as the last 59bp remain adjacent to *CO*. Hence, the *CO/COL1* duplication likely existed and *COL1* was later lost in *A. arabicum*. Thus *A. arabicum* is probably able to survive and develop without a functional *COL1* ortholog, perhaps because it is largely redundant to *COL2*. This is also supported by the finding that *T. hassleriana* only carries one subgroup Ia *COL* ortholog even though it has undergone an independent genome triplication (Cheng et al., 2013; Schranz and Mitchell-Olds, 2006). The other two copies must have been lost after the triplication event. It is possible that the triplication of the ancestral gene in the Brassicaceae led to subfunctionalization. Hence the original function could be maintained by one copy probably *COL2*, while the two other copies, in this case *COL1* and *CO* changed function.

The phylogenetic reconstruction revealed that ThCOL protein is more closely related to COL2 orthologs compared to CO and COL1 orthologs, respectively. A limitation of this analysis is that it is only based on a small number of genes. Increasing the number of genes included in the analysis, especially orthologs from other Cleomaceae species would increase the resolution of the tree. Also previous analyses were performed only for the B-Boxes or the CCT domain, excluding the conserved middle part (M1-M4, Fig. 1). This could be tested as well, even though there might be too few differences in amino acid sequence to obtain high resolution in the phylogenetic tree.

The analysis of nucleotide substitution rates (dn/ds) indicated that stabilizing selection is acting on subgroup Ia genes in all comparisons tested. Interestingly dn/ds values of comparisons to AtCO were the highest observed (Table 12). This indicates that a selection process might be acting on *AtCO* gene which is masked by the overall evolutionary force to keep the gene a functional transcriptional regulator. Specific parts of *CO* gene from different Brassicaceae species could be tested to identify regions that might be under positive selection.

The expression of *A. thaliana*, *A. alpina*, *A. arabicum* and *T. hassleriana* subgroup Ia *COL* orthologs gave clear results, which support the phylogenetic relationships discussed above. *CO* mRNA regulation evolved in the Brassicaceae after the At-α whole genome duplication and is conserved down to the basal species *A. arabicum*. This study aimed at finding the similarities between related genes but it is also of interest to identify differences between species. For example the expression of

AeCO could be analysed in *A. arabicum* in intermediate day length such as 12h light and 12h dark to find differences.

The expression pattern of *ThCOL* suggests that the gene is not involved in discrimination between photoperiods as mRNA levels are high in the morning both in LD and SD which is shown to abolish day length discrimination by AtCO (Fornara et al., 2009). This is supported by the finding that acceleration of flowering in *T. hassleriana* under LD conditions is mild.

It would be of interest to compare the spatial expression pattern of the genes. *COL1* and *COL2* were generally expressed more highly than *CO* in *A. thaliana* (>100fold, data not shown), which was confirmed by the *A. alpina* RNA seq data (Fig. 7). If *AtCOL1* and *AtCOL2* were expressed highly in the vasculature this would lead to early flowering in both LD and SD, which is not observed in *A. thaliana*. Thus the differences in expression will most likely involve changes in the spatial expression. To examine the spatial expression fusions of different promoters like *pAtCOL1*, *pAtCOL2*, *pAeCOL2* and *pThCOL* to *GUS* gene will be analysed in transgenic *A. thaliana*.

The functional investigation of these closely related orthologs in complementation of *co-10* mutant revealed unexpected results. All genes investigated complement the *co-10* mutation to a certain degree. The stronger *pSUC2* enables more complementation compared to *pCO*. Of course these flowering time experiments need to be repeated with homozygous T3 plants of representative lines. In previous studies use of the CaMV *35S* promoter led to different conclusions (Kim et al., 2013; Ledger et al., 2001). The acceleration of flowering as observed using the *pSUC2* promoter is an artificial situation that does not reflect the wild type condition. Similarly *CDF2* causes late flowering when expressed from the *pSUC2* promoter (Fornara et al., 2009) but not necessarily when driven by the CaMV *35S* promoter (Imaizumi et al., 2005). Nevertheless using *pSUC2* strengthens the current model as *CDF2* is assumed to be a direct repressor of *CO* and loss of function *cdf2* alleles increased *CO* expression (Fornara et al., 2009). Analysis of the *pSUC2* driven lines might thus enable more accurate predictions about the function of vascular expressed genes. The experiments using *pCO* support these findings. It was unexpected to observe a partial complementation of the *co-10* mutation when using *pCO* driven *COL* because CaMV 35S driven *COLs* failed to accelerate flowering or complement *co* mutants (Kim et al., 2013; Ledger et al., 2001). It can also not be excluded that *ThCOL* is an activator of an *FT* homolog in *T. hassleriana* and does this as efficiently as CO activates *FT* in *A. thaliana*. Evolutionary changes most likely did not only occur in CO and its orthologs but also its many interactors, which might be different in *T. hassleriana* and *A. thaliana* compared to the last common ancestral genes. Furthermore, these experiments aimed at comparing *COL* homologs in their ability to replace AtCO. Unfortunately it cannot be tested whether ThCOL is able to replace the AtCOL1 and AtCOL2 because their function is not described in detail and double mutants have not been isolated (Kim et al., 2013; Ledger et al., 2001).

Functionally CO orthologs and COL1/COL2 orthologs including ThCOL fall into distinct groups as shown in the *pCO* complementation experiments. This led to the detailed analysis of differences between CO and other COL orthologs. The focus was on differences in four residues in the conserved B-Boxes and CCT domains of the proteins. The residues were conserved in either groups but different between groups. Three of these changes (AtCO, S49A, D79A and N342K) suggest that CO is the derived state. CO seems to represent the ancient state only in one change in B-Box II (AtCO, I90V), where amino acids with the same properties (nonpolar) are exchanged. Deviating amino acid changes are additionally conserved in AtCO, AaCO and AeCO, meaning they could be of importance, especially when taking into account that the *A. arabicum* and core Brassicaceae split occurred after the *At-α* duplication meaning up to 43~ mya ago (Fawcett et al., 2009). These changes could influence function, which allow sub functionalization of CO. Even though one position (S49A) was already excluded by mutational analysis (Kim et al. 2013), the other changes or even an interaction of different amino acids could be responsible for the differences between AtCOL1, AtCOL2 and AtCO.

5. Phylogenetic Shadowing of *CONSTANS* promoter and functional test

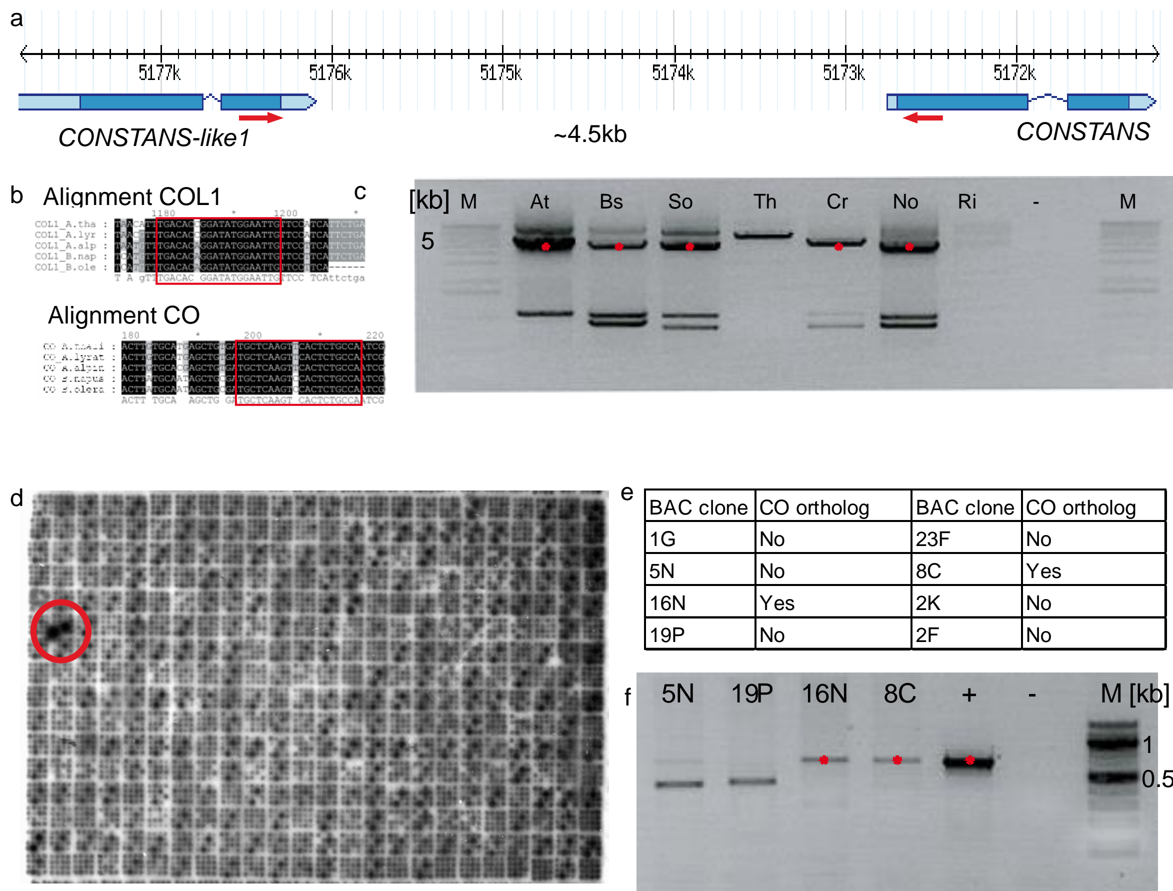
Transcription is regulated by interaction of TFs with DNA sequence motifs at the locus of the gene of interest. *CO* mRNA is differentially regulated in changing day length and cycles during the day suggesting complex transcriptional regulation. Earlier studies revealed that in *A. thaliana* 2.5kb upstream of *CO* start codon is sufficient to complement the *co-2* mutation when driving transcription of *CO* cDNA (An et al., 2004). Spatial expression studies using the *GUS* reporter gene are also either based on 2.5kb promoter studies or genomic constructs of *CO* locus (An et al., 2004; Takada and Goto, 2003). Even though binding of *CO* regulators was shown to occur at more distal parts of *CO* promoter (Ito et al., 2012a; Sawa et al., 2007) most studies focus on *cis*-elements in the proximal part of the promoter (up to -1kb). These include DOF binding sites (Imaizumi et al., 2005) and E-Boxes (Ito et al., 2012a). Additionally the finding that the *CO* gene is only present in the Brassicaceae and that its transcriptional regulation evolved in this family (Chapter 4) can be used to identify conserved motifs. Thus here *CO* promoter sequences from different Brassicaceae species are used to identify conserved regulatory sequence. The aim of this study is to define a minimal promoter for *CO*, describe its regulatory regions and compare *cis*-element changes between *CO* and its closely related paralogs.

5.1. *CO* promoters are present in all Brassicaceaen species investigated

Phylogenetic footprinting and shadowing are powerful techniques to identify conserved sequences in intergenic regions that can subsequently be tested for their functional impact in biological assays. In order to identify important regulatory sequences by phylogenetic shadowing the intergenic space between *CO* and its upstream paralog *COL1* was collected from different Brassicaceae species. In *A. thaliana* the intergenic region spans ~3.6kb. By searching databases using the BLAST algorithm the sequences of *Brassica napus*, *Brassica rapa* and *Sisymbrium irio* were retrieved. Additionally the sequence from *Arabidopsis lyrata* was assembled using short reads from the NCBI trace archives. *A. alpina* sequence was identified by searching the preliminary genome alignment by BLAST (MPIPZ, Cologne, unpublished).

To identify *CO* promoter sequences from species where sequence information was not available a conserved primer set was anchored in *CO* and its upstream paralog *COL1*. To obtain a conserved region suitable for anchoring primers the coding sequences of *CO* and *COL1* orthologs from *A. thaliana, Arabidopsis lyrata, Brassica napus, Brassica oleracea* and *A. alpina* were aligned. Primers were chosen close to the C-terminus of *COL1* and in the first exon of *CO* to minimize PCR fragment size. Fig. 10b shows part of the alignment used to select the primer set. The resulting PCR product has a length of ~4.5kb in *A. thaliana* (Fig. 10a). Amplification of the desired fragment was successful for five of seven species tested (genomic DNA provided by Markus Berns, MPIPZ, Cologne), including *A. thaliana* as a positive control (Fig. 10c). PCR fragments of the desired length were subcloned and sequenced from *Boechera stricta*, *Sisymbrium officinale*, *Capsella rubella* and

Nasturtium officinale. No sequence could be amplified from *T. hassleriana*, *Reseda luteola* (Fig. 10c) and *A. arabicum* (data not shown).



BAC clone	CO ortholog	BAC clone	CO ortholog
1G	No	23F	No
5N	No	8C	Yes
16N	Yes	2K	No
19P	No	2F	No

Figure 10: Isolation of *CO* orthologous promoter sequences from different Brassicales species.
A. thaliana CO and *COL1* are located in tandem on Chromosome 5 (a). Primers were placed in conserved sequences of *COL1* and *CO* to amplify intergenic regions from other species in the order Brassicales. Alignment of *COL1* and *CO* orthologous DNA sequences (b) from five Brassicacean species was used for the design of primers indicated in a. PCR products (c) using conserved primers (a,b). Intergenic regions were confirmed by sequencing (marked with asterisks). M Marker; *At A. thaliana*; *Bs Boechera stricta*; *So Sisymbrium officinale*; *Th T. hassleriana*; *Cr Capsella rubella*; *No Nasturtium officinale*; *Ri Reseda luteola* (–) water control. *A. arabicum* BAC screen (example in d) using an *A. thaliana CO* cDNA probe (439bp) spanning the conserved middle region resulting in 8 BACs tested (e). Two adjacent spots indicate same bacterial clone in a square (d). Identical *A. arabicum CO* ortholog was identified in 16N and 8C as confirmed by PCR (f) using *A. thaliana* genomic DNA as positive control (+ amplifies 672bp), while no product was amplified in water control (-).

To obtain the *CO* promoter of *A. arabicum* a BAC screen was performed using the central region of *AtCO* as a probe. This region was chosen as it is conserved in *CO* orthologs and other subgroup Ia *COL* genes (M1-M4) but varies in other members of the *COL* family (Griffiths et al., 2003). The primers used amplified 439bp of *CO* gene from cDNA but 672bp from genomic DNA as the primers span the intron. As intron sequence is in general more variable compared to coding region the probe was generated from cDNA, which excludes the intron. The BAC screen (Fig. 10d) yielded eight candidate BACs (Fig. 10f). Of these two were confirmed to contain a *CO* ortholog by amplification, subcloning and sequencing. The intergenic region between *CO* and *COL1* was sequenced using primer walk (Material and Methods) and both BACs (16N and 8C) carried identical sequence. The intergenic region spans ~3.4kb in *A. arabicum* but only 59bp of *AtCOL1* ortholog could be sequenced and the

upstream region was homologous to another region on Chr.5 of *A. thaliana* (see Chapter 4). This result was later confirmed by analysis of whole genome sequencing data. This rearrangement explained why amplification using a *COL1* anchored primer was not successful for *A. arabicum*. The other six BACs were not investigated further but they might contain other members of the subgroup Ia *COL* genes. A *COL2* ortholog was identified in the *A. arabicum* genome (Chapter 4). *CO* homolog of *T. hassleriana* was identified by Schranz and Mitchell-Olds (2006). The finding that no *COL1*/*CO* orthologous fragment could be amplified from *T. hassleriana* supports earlier discoveries, that *ThCOL* is the only member of subgroup Ia *COL* genes in this species. Additionally, no *CO* promoter could be amplified from *Reseda luteola*, a member of the Resedaceae, a family in the order Brassicales. The *CO* promoter might thus not be present in other families of the order Brassicales. In summary orthologous *CO* promoters were identified in all Brassicaceaen species investigated and ten *CO* promoters were isolated for further studies.

5.2. The promoter of *CO* is conserved in ten Brassicaceaen Species and features a unique *cis*-element composition.

To identify conserved sequence regions in the promoters of *CO* orthologs, the promoters isolated from the Brassicaceae species were used for a multiple sequence alignment (Appendix). The sequences from ten Brassicaceaen *CO* promoters were aligned using the LAGAN algorithm in the mVISTA suite (material and methods). For the alignment the intergenic region between *CO* and *COL1* orthologs was used including both *CO* 5´UTR as well as *COL1* 3´UTR. The VISTA plot representation is shown using *A. thaliana* (Fig. 11a) as well as *A. arabicum* (Fig. 11b) as a base genome. The conservation to the respective base genome is shown in percent of identical base pairs in a sliding window of 100bp. Values above 75% conservation were marked in light red. From the comparison of all sequences with *A. thaliana* it can be observed that the overall conservation is higher when sequences from closely related species are compared to *A. thaliana*. When comparing it to more distantly related sequences from for example *Brassica*, *A. alpina* or *Sisymbrium*, single blocks of conservation between ~100bp and 1kb emerge from the analysis. Comparison of the *A. thaliana CO* promoter to that of *A. arabicum*, which is considered to be the most basal Brassicaceae species used in this study (Bailey et al. 2006), still reveals conserved blocks. A phylogenetic tree was constructed from the dataset (Fig. 11e). In this tree *A. thaliana*, *Capsella* and *Nasturtium* species group together while *Brassica*, *Sisymbrium* and *A. alpina* species form a second subgroup. *A. arabicum* is located as an outgroup. This tree reflects the phylogeny of the respective species in the Brassicaceae family and distinguishes *A. arabicum* as a member of a distinct lineage outside the core Brassicaceae (compare Fig. 3, Bailey et al., 2006).

A. arabicum CO promoter shows similar conservation to all other orthologous sequences investigated (Fig. 11b), because the radiation of all other species (core Brassicaceae) occurred after the split from the *A. arabicum* lineage and it shares one last common ancestor with all of them (Bailey et al., 2006; Franzke et al., 2011). On the other hand *A. thaliana* and *Capsella* lineages for example split later than

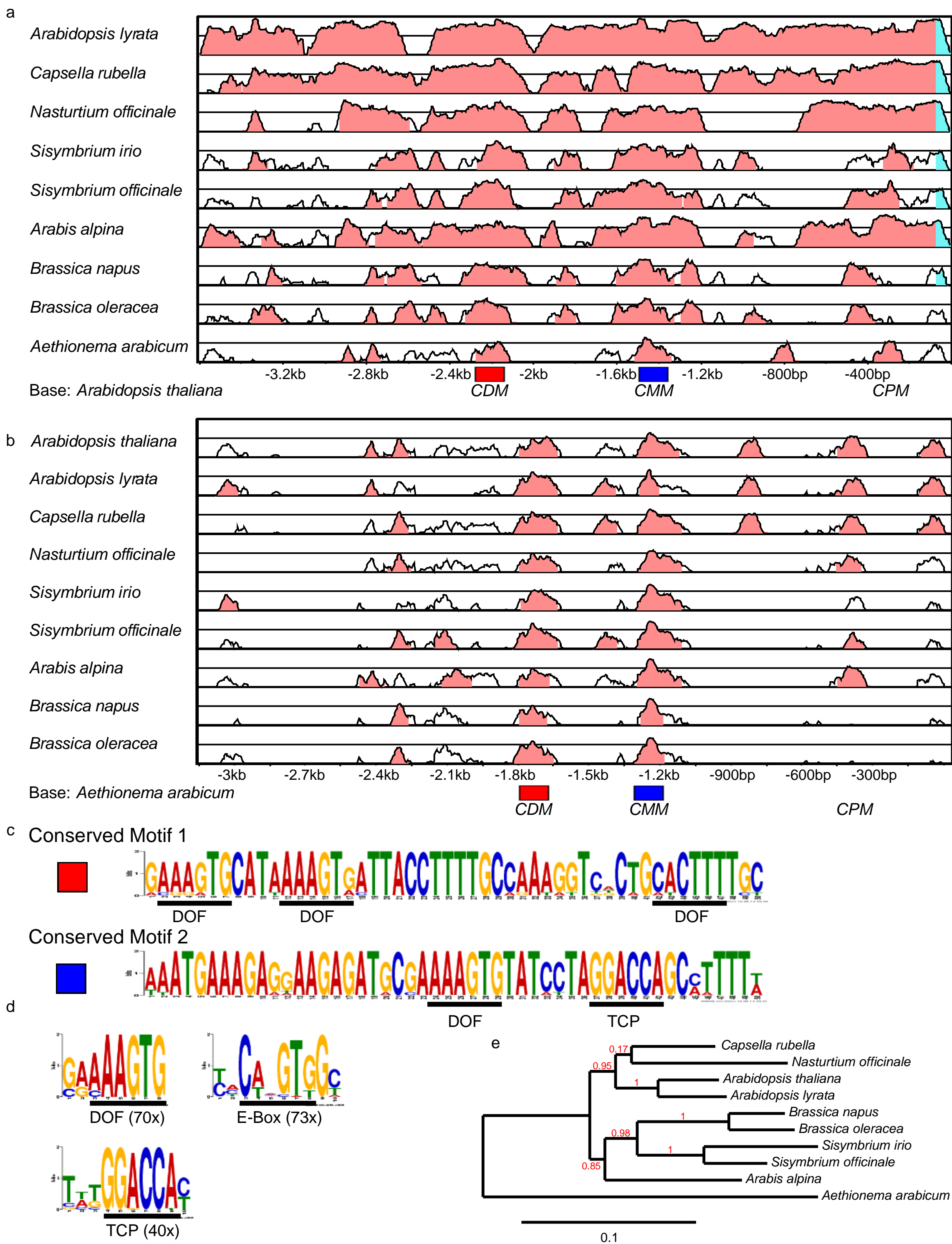


Figure 11: Conservation of *CO* promoter in the Brassicaceae
CO homologous promoters (intergenic region between *COL1* and *CO* including 5´UTR) from ten Brassicaceae species were used for comparison. Alignment (LAGAN) is displayed as mVISTA plot with the base genome *A. thaliana* (a) or *A. arabicum* (b). Sliding window displays conservation between 50-100% in 100bp sliding window. Conservation above 75% is highlighted in light red color. 5´UTR of *CO* in turquoise as annotated in TAIR10. The approximate positions of Conserved Proximal Motif (*CPM*), Conserved Middle Motif (*CMM*) and Conserved Distal Motif (*CDM*) as described in Fig. 20 are indicated. Two highest conserved promoter parts (50bp) as identified by MEME motif identification (c), Red and Blue areas represent position in b. DOF (AAAGTG) and TCP (GGACCA) binding sites are underlined. Three highest conserved short motifs (6bp to 10bp) are displayed as identified by MEME (d) motif identification tool, which occur with a high number of repetitions (40 to 73 times) in the ten promoters. Consensus TF binding sites are underlined. Neighbour joining tree (e) reflects the phylogenetic relationship between *CO* promoters from ten Brassicaceaen species, red numbers indicate bootstrap value.

A. thaliana and *A. alpina*, thus *A. thaliana* shares more sequence similarity with *Capsella* than with *A. alpina* (Fig. 11a).

Additionally variation among conserved regions is observed. For example a small conserved block at around -800bp and -900bp (Fig. 11b) was only present in the *Arabidopsis*/*Capsella* lineages but absent in all other sequences while a block between -300bp and -400bp was absent only in the *Brassica* lineage but present in all other sequences. These blocks might be interesting for further analysis of putative differences in the regulation of *CO* orthologs in these specific species.

Conservation of the *CO* promoter was successfully detected between Brassicaceaen species. To identify functional regulatory sequences the promoter set was further analysed with a focus on CRMs and especially their *cis*-element composition. For this purpose the whole set of orthologous *CO* promoters was analysed using the MEME motif identification tool (Bailey et al., 2009). Fig. 11c shows the strongest two hits when searching for a conserved motif with a maximum length of 50bp that is found in all sequences. In this representation the height of the Base represents its frequency, while a full letter represents 100% conservation. The Conserved Motifs 1 and 2 matches with the core of the broadest peaks identified (Fig. 11b). Conserved Motif 1 contains three DOF binding sites (AAAGTG), while Conserved Motif 2 contains one DOF binding site and one TCP binding site (GGACCA, see below). Interestingly the highest conserved CRMs were not identified in the proximal part of the promoter but ~1.4kb and ~2.1kb upstream of the TS, indicating that these regions might be of importance for regulation of *CO* mRNA levels in *A. thaliana* and related species.

Reanalysis of the dataset with parameter for short motifs (6-10bp), which could occur at any number, identified three highly overrepresented short sequence stretches in the whole promoter (Fig. 11d). Even though the algorithm was not looking for known *cis*-element the best three hits contained known *cis*-elements. The first element, being present 70 times in all ten promoters, contained the DOF binding site (AAAGTG). The AAAG motif was described as the core DOF binding site (Yanagisawa and Schmidt 1999), while the additional TG residues, completing the hexamer, were found to give additional specificity for CDF1 (Imaizumi et al., 2005) and putatively for CDF2, CDF3 and CDF5 as they belong to the same subgroup of the DOF family (group II, Fornara et al., 2009) . This additional specificity might be important for target discrimination as the DOF transcription factor family contains 26 members (Yanagisawa, 2002; Yanagisawa and Schmidt, 1999). The second element (CANGTG) was identified 73 times and contains the E-Box consensus (CANNTG) which is a bHLH binding site (Toledo-Ortiz, 2003). A group of bHLH transcription factors called *FBHs* were previously identified as activators of *CO* transcription (Ito et al., 2012a). FBHs were also shown to interact *in vitro* and *in vivo* with *CO* promoter sequence (Ito et al., 2012a), thus identifying this motif in the promoter set supports the current model for *CO* mRNA regulation. The third motif discovered has a GGACCA hexamer as its core sequence, which was found to be present 40 times (3-5 times per sequence) in the dataset. This motif was shown to be a TEOSINTE BRANCHED/CYCLOIDEA/PCF 4 (TCP4) TF

binding site by *in vitro* studies (Schommer et al., 2008). The GGACCA core motif was independently identified by searching for overrepresented motifs in genes that are misregulated in *tcp* mutant or *TCP* overexpressor (Schommer et al., 2008). Thus, the MEME motif identification approach was successful in identifying the so far unlinked class of TCP transcription factors (Cubas et al., 1999) as putative *CO* transcriptional regulators. Hence, this analysis supports the direct interaction of *CO* transcriptional activators (FBHs) and repressors (CDFs) with *CO* locus not only in *A. thaliana* but also in the whole Brassicaceae family.

Whether the *cis*-element composition of the promoters is related to expression pattern differences between *CO* orthologs and *COL1/COL2* orthologs was then tested. To examine differences the abundance of *cis*-elements in the different promoters was determined. The aim was to identify *cis*-elements that were statistically overrepresented in *CO* orthologous promoters compared to the other subgroup Ia *COL* genes, which do not share the mRNA expression pattern. For this purpose only species were included where sequence information about all three subgroup Ia *COL* genes was available in the Brassicaceae. The analysis was made with *A. thaliana*, *A. alpina* (genome sequence unpublished: MPIPZ, Cologne), *Brassica rapa* (Wang et al., 2011), *Eutrema parvulum* (*Thellungiella parvula*, Dassanayake et al., 2011) and *A. arabicum* (Haudry et al., 2013). Please note that no *COL1* ortholog is present in *A. arabicum*. For *Brassica rapa* two *COL1* and *COL2* orthologs but only one *CO* ortholog were identified by BLAST search and sequence comparison. This results in five *CO*, five *COL1* and six *COL2* orthologous promoters available for analysis. For comparison outside the Brassicaceae *COL* homologs belonging to subgroup Ia were used (Promoter sequences and accession numbers in the appendix). In several cases, the gene expression was tested and shown to be circadian regulated with a strong morning peak like *AtCOL1*/*AtCOL2* (Fig. 6). *T. hassleriana COL* as a member of the Cleomaceae family was included (Schranz and Mitchell-Olds, 2006) because it is the only member of group Ia *COL* genes identified and morning expression is described in this study in detail (Fig. 6). The *Medicago truncatula* homolog *MtCOLa* was investigated as it phylogenetically groups with subgroup Ia *COL* genes from *A. thaliana* (Hecht et al., 2005). *Vitis vinifera CO* homolog (*VvCO*), two homologs from *Populus trichocarpa*, *PtCO1* and *PtCO2*, and another two homologs from *Solanum tuberosum*, namely *StCO1* and *StCO2,* were classified as subgroup Ia *COL* genes and additionally share a morning mRNA expression (Almada et al., 2009; Hsu et al., 2012). *StCO1* mRNA was shown to be regulated by StCDF1, a potato CDF ortholog (Kloosterman et al., 2013). In order to minimize bias the -4kb sequence from the start codon of all these genes was selected for analysis.

The aim was to examine the frequency of *cis*-elements identified by MEME motif identification in the promoter of *CO* and compare this to the other *COL* subgroup Ia promoter in order to relate the change in transcriptional pattern to changes in *cis*-elements. Three more *cis*-elements were included in the analysis. The G-Box (CACGTG) was included as it is present in the *A. thaliana* promoter at two positions one close to the TS (Ito et al., 2012a) and the other around ~2kb upstream (Fig. 13). MYB

binding sites (GATWCG) are present in *CO* promoter with two being in close proximity to the G-Boxes described above (Fig. 13, 20). This specific MYB binding site is involved in the association of the evening complex (EC, Nusinow et al., 2011) to DNA by providing a binding site for the LUX ARRHYTHMO (LUX) TF (Helfer et al., 2011). The finding that *AtCO* promoter has three LUX binding sites (LBS) in the first 500bp of its promoter (Chow et al., 2012) and that GI interacts with a member of the EC (ELF3, Yu et al., 2008) makes it an interesting putative regulatory element for *CO* transcriptional regulation. The CCAAT-Box (CCAAT) motif was also investigated as it might be involved in restricting transcription to vascular cells (Siefers et al., 2008; Wenkel et al., 2006). How overexpression lines like *35S::FBH1* and *35S::FBH2* cause elevated *CO* expression (*pCO::GUS*) restricted to the vasculature (as well as stomata in leaves and the root tip) remains unclear (Ito et al., 2012a). A similar phenomenon is observed with *35S::CO* enhanced expression of *pFT::GUS* (Adrian et al. 2010). In total six elements were investigated which might regulate *CO* mRNA expression in an orchestrated way.

Cis-elements were counted for all promoters and the mean and standard deviation was calculated (Fig. 12). The random occurrence of such short motifs changes with the GC content of both the motif itself and the sequence investigated. For example in *A. thaliana* the GC content deviates from 50%, being ~44% in coding sequence and dropping in the intergenic space to 32-33% (The *Arabidopsis* Genome Initiative, 2000). Additionally GC content also varies between different plant species (Šmarda et al., 2012). In order to correct for this bias in predicting the random occurrence of each *cis*-element the promoter sequences were shuffled 100 times and reanalysed. To confirm significant differences a One Way ANOVA followed by a tuckey test was performed and significantly different groups with $p<0.05$ were identified.

In this analysis the mean occurrence of DOF binding sites in *CO* promoters was 13.4 (Fig. 12a). The mean occurrence in the other promoters was statistically lower at 2 (*COL1*), 2.7 (*COL2*) and 4.3 (non-Brassicaceaen *COL*). The values for the randomized dataset ranged between 2.6 and 3 which is not statistically different from the *COL1*, *COL2* and *COL* promoters. This confirms that *CO* orthologous promoters have statistically more DOF binding sites than their *COL* homologs, and these could contribute to the changes in *CO* transcription that occurred during evolution (Chapter 4). Additionally the overall occurrence of DOF binding sites in *A. thaliana* promoters (-3kb sequence from TS) was checked and the *CO* promoter along with one other gene (AT3G51870) shows the highest number of DOF binding sites (14x AAAGTG, Karl Nordström, MPIPZ, Cologne) in the entire genome.

TCP binding sites (GGACCA) were identified by MEME motif identification approach. Their mean abundance per *CO* promoter was higher (4.8 binding sites) than found in *COL1* (1.2), *COL2* (1.8) and non-Brassicaceaen *COL* (1.6) (Fig. 12b). Even lower numbers were observed for the shuffle control (0.7-1). The statistical test revealed that TCP binding sites are significantly enriched in *CO* orthologous promoters compared to the other subgroup Ia *COL* promoters and the shuffle control. No

statistical difference between the number of TCP binding sites in *COL* promoters and the shuffle control was observed.

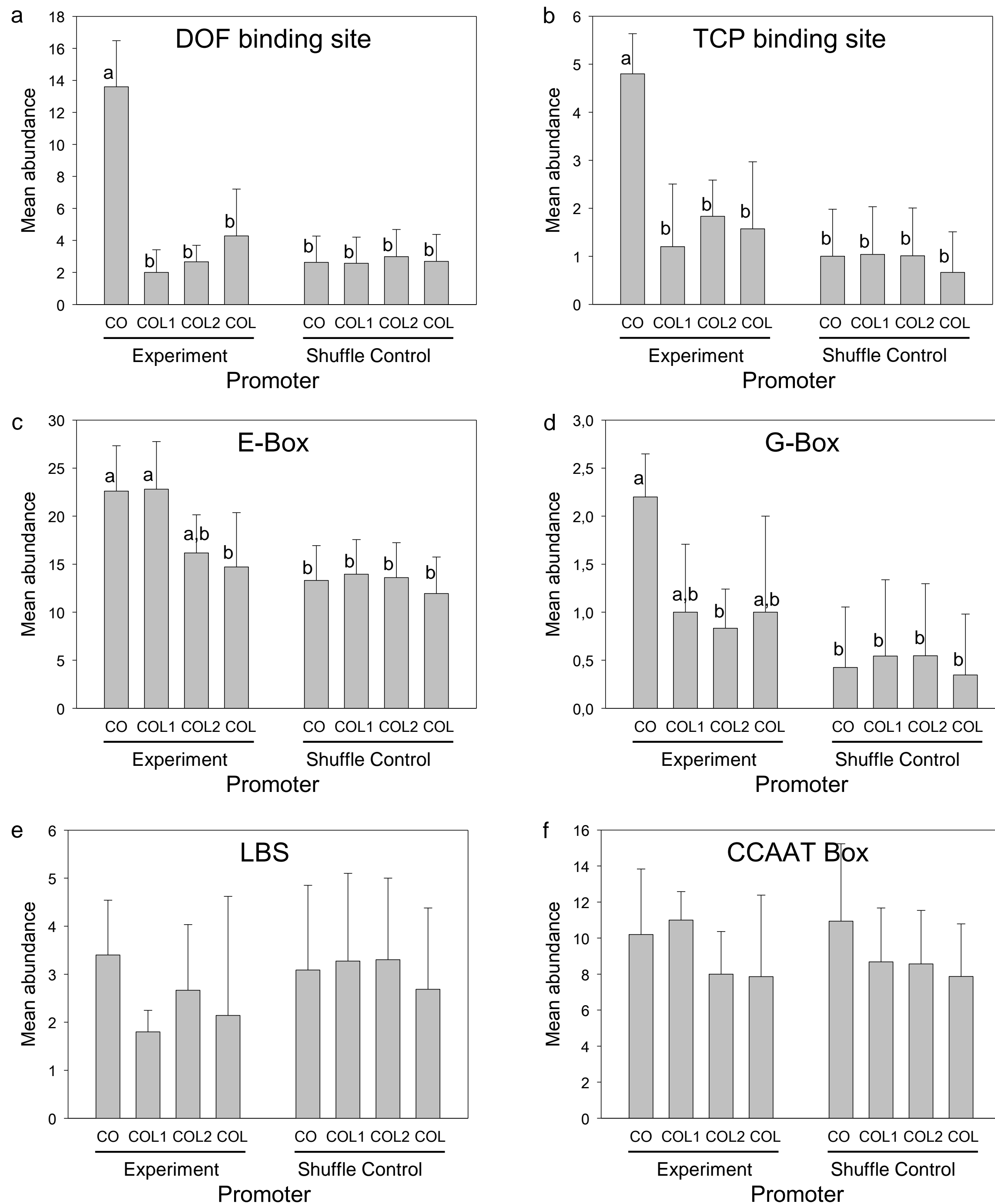


Figure 12: Statistical analysis of TF binding sites in subgroup Ia *COL* genes

Mean Abundance of *cis*-elements (a-f) in the -4kb upstream region of Brassicaceaen *CO*, *COL1* and *COL2* as well as non-Brassicaceaen *COL* orthologs (*COL*) as indicated (Experiment). Same analysis was performed 100 times with shuffled datasets as negative control (Shuffle Control). One-Way-Anova followed by Tuckey test was performed to test for significance. Letters a,b indicate that data sets fall into significantly different groups (p<0,05), while error bars represent standard deviation. Brassicacaean sequences include homologs from *A. thaliana*, *A. alpina*, *Eutrema parvulum*, *Brassica rapa* (1x*CO*, 2x*COL1* and 2x*COL2*) and *A. arabicum* (no *COL1*). Upstream sequences from non-Brassicacean species include *T. hassleriana (ThCOL)*, *Medicago truncatula* (*MtCOLa*), *Populus Trichocarpa* (*PtCO1; PtCO2*), *Solanum tuberosum* (*StCO1; StCO2*) and *Vitis vinifera* (*VvCO*). a) DOF=AAAGTG, b) TCP=GGACCA, c) E-Box= CANNTG, d) G-Box=CACGTG, e) LBS=GATWCG and f) CCAAT-Box=CCAAT.

Next the occurrence of E-Boxes was investigated (Fig. 12c). As this element has only four fixed positions and two non-specific bases (CANNTG, Toledo-Ortiz, 2003) a higher number of E-Boxes was expected per promoter compared to hexamers. The mean number in *CO* promoters was found to be 22.6. *COL1* promoters showed virtually the same number (22.8), while *COL2* with 16.2 and COL promoters with 14.7 sites per promoter showed lower mean abundance. The shuffled data set showed lower mean numbers between 11.9 and 14. Statistically, the number of E-Boxes in *CO* and *COL1* promoters is higher compared to non-Brassicaceaen *COL* and the shuffle controls. *COL2* promoters did not show significant differences with *CO/COL1* or all other datasets tested.

The mean number of G-Boxes (CACGTG) in *CO* promoters is 2.2, while the mean is 1 for *COL1*, 0.8 for *COL2* and 1 for *COL* promoters, respectively (Fig. 12d). The random occurrence of this motif ranged between 0.6 and 0.8 in this dataset. *CO* promoters have significantly more G-Boxes than *COL2* promoters but not as many as *COL1* and non-Brassicaceaen *COL* promoters. Interestingly *CO* promoters are the only group that shows significantly more G-Boxes compared to its corresponding shuffle control. The influence of *cis*-element position in a promoter was not addressed, but for an element that only occurs once or twice position may be important. G-Boxes are found close to the TS and are involved in transcriptional regulation (reviewed in Terzaghi and Cashmore, 1995). Most G-Boxes identified in this study are also positioned close to the TS. Interestingly the second G-Box in *CO* promoters can be found mostly in close proximity with Conserved Motif 1 (Fig. 11, 13, 20) and its role will be discussed later in this chapter.

Analysis of LBSs (GATWCG) revealed no significant difference in abundance in the *CO* promoter and the other promoters or the shuffle control with a mean number of 1.6 and 3.8 binding sites per promoter (Fig. 12e). This does not exclude a role for this *cis*-element as it is present in all *CO* orthologous promoters investigated, but a function would need to be established by experimental analysis.

A mean of 10.2 CCAAT-Boxes (CCAAT) was found for *CO* promoters (Fig. 12f). This number was not significantly different in all other promoters and shuffle controls investigated. As for the LBSs a biological role cannot be ruled out.

In summary this experiment revealed that a dramatic increase in DOF binding sites and a smaller increase in TCP binding sites are the major evolutionary changes in *cis*-element number between promoters of *CO* and its close relatives, while differences for LBS and CCAAT-Boxes were not found. The number of E-Boxes is higher than statistically expected in *CO* promoters but this is also true for *COL1* promoters. They might play a regulatory role that is shared with *COL1* and maybe even *COL2*. Similarities were observed for G-Boxes, which are overrepresented in *CO* compared to shuffle control but not overrepresented compared to *COL1* and *COL* , suggesting that they could play a role in all three promoters. In this case other parameters like positional conservation would need to be

considered in evaluating their significance. In general whether a change in abundance of *cis*-elements in a promoter has influence on the transcriptional regulation of the associated gene needs to be examined by experimentation.

5.3. Strategy for functional analysis of the *A. thaliana CO* promoter

Three different approaches were taken to define regulatory regions in promoter of *A. thaliana CO*. Promoter deletion constructs were tested in complementation experiments using *AtCO* cDNA and the firefly *LUCIFERASE* (*LUC*) reporter (Millar et al., 1992). In a second gain of function approach *CO* promoter was divided into four parts and fragments were fused to the *NOPALINE SYNTHASE (NOS)* minimal promoter (*pnos*, Puente et al., 1996; Shaw et al., 1984) and *LUC*. The *pnos* minimal promoter provides a transcriptional start site that enables testing the capacity of upstream promoter regions to enhance transcriptional initiation. The third approach combines phylogenetic shadowing with the gain of function approach. So *AtCO* promoter regions that are conserved between Brassicaceae species were fused to *pnos::LUC* to test their functionality.

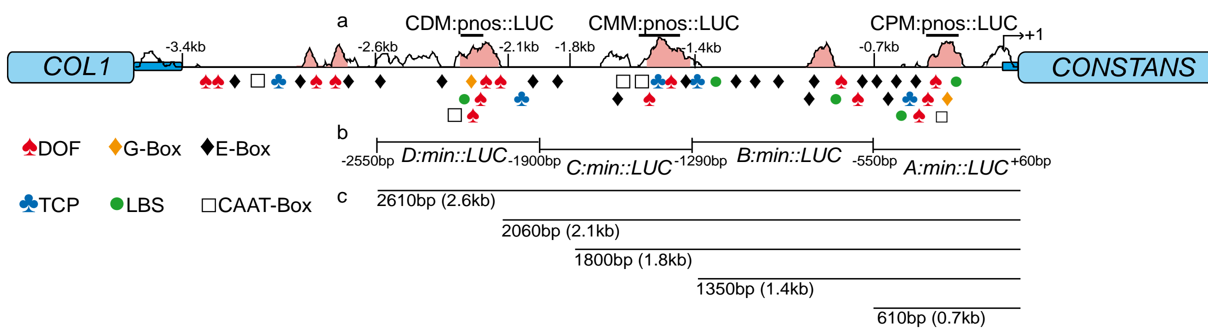


Figure 13: Model of *CO* locus in *A. thaliana* indicating position of *cis*-elements in the context of reporter lines.
Model of *CO* locus is presented from the second exon of *COL1* to first exon of *CO*. In orange the conservation to *A. arabicum CO* promoter is displayed as represented in Fig 11a. Each *cis*-element is symbolized indicating its approximate position in the locus. DOF=AAAGTG, TCP=GGACCA, E-Box=CANNTG, G-Box=CACGTG, LBS=GATWCG and CCAAT-Box=CCAAT. Approximate positions of conserved sequence stretches used for promoter analysis (*pnos::LUC*) are indicated (a). CDM= Conserved Distal Motif, CMM= Conserved Middle Motif, CPM= Conserved Proximal Motif. Approximate position of ~700bp fragments (A-D) analysed (*pnos::LUC*) are displayed (b). Position relative to TS, D includes 60bp 5´UTR of *CO*. Relative position of promoter deletion constructs used for complementation experiments and *LUC* reporter assays are shown (c). 5´UTR is included in all constructs thus length (bp or kb) refers to upstream region from the ATG. The position of the constructs is additionally indicated in the model.

5.4. Analysis of *A. thaliana CO* regulation by promoter deletion constructs identifies a minimal promoter.

In order to define the minimal promoter of *CO*, different lengths of *CO* promoters were fused to *GFP:CO* cDNA and transformed into Landsberg *erecta* (L*er*) *co-2* mutant (Mark Rühl, Umeå universitet, Sweden). The full length promoter was described to be ~2.5kb from the TS in *A. thaliana* (An et al., 2004). Thus, a 2.6kb construct was tested as a positive control and shorter fragments of 2.1kb, 1.8kb, 1.4kb and 0.7kb including the 5´UTR of *CO* were designed using L*er* genomic DNA as a template. Additionally the 2.6kb and the 1.8kb *AtCO* constructs were fused to *LUC* and introduced into Col-0 wild type plants. Fig. 13 illustrates the source of these fragments. Notably, one *cis*-element

rich region that is further upstream of -2.1kb, later referred to a Conserved Distal Motif, is absent in the 2.1kb and 1.8kb constructs.

The flowering time of homozygous T_3 complementation lines was examined in LD and SD conditions (Fig. 14). L*er* control flowered with 12.9 and *co-2* mutant control with 23.8 leaves in LD. Plants carrying the full length promoter driven transgene (2.6kb) flowered with ~6.5 leaves (three independent lines, Fig. 14a). *GFP:CO* driven by 2.1kb promoter accelerates flowering in LD to 7.3, 9.4, 9.5 and 13 leaves, respectively. The lines carrying the 1.8kb fragment flowered with 12.5, 12.9 and 20.2 leaves in LD, respectively. Therefore in LD conditions the 2.6kb *pCO* driven complementation lines flowered significantly earlier compared to 1.8kb *pCO* driven constructs. The result for the 2.1kb fragment was more heterogeneous. While one early line (#2.1) was not significantly different from the 2.6kb lines the flowering time of another line (#5.1) resembled 1.8kb *pCO* complementation lines. Two other lines showed an intermediate (#6.2 and #8.1) flowering time phenotype compared to 2.6kb and 1.8kb. This result suggested that longer promoter fragments resulted in a gradual increase in promoter activity. Flowering time was also tested in SD conditions. All except three tested lines flowered similar to L*er* and *co-2* control plants with around 27.5 and 29 leaves, respectively. Three lines though (2.6kb #6.9 and 7.3, 2.1kb #2.1) showed early flowering in SD conditions with leaf number between 11.6 and 14.6. These lines were among the earliest in LD conditions as well. In summary, 1.8kb *pCO* is sufficient to complement *co-2* mutation but upstream sequences up to 2.6kb additionally accelerate flowering in LD conditions and thus include important regulatory sequence.

Independently, the flowering time of transgenic lines expressing *GFP:CO* from 1.4kb and 0.7kb promoter CO constructs were tested (Fig. 14b). In LD conditions only a mild acceleration of flowering compared to *co-2* mutant (28.9 leaves) was observed for 1.4kb pCO (24.6 leaves). Two independent lines employing a 0.7kb *pCO* fragment flowered with 26.8 and 28.1 leaves, respectively. L*er* and 2.6kb *pCO* controls showed similar flowering times compared to the other experiments (Fig. 14a). Thus *CO* promoter fragments of 1.4kb or shorter are not sufficient to complement the *co-2* mutation, when used to express *GFP:CO*.

GI binds to *CO* promoter at at least two positions in *A. thaliana* and acceleration of flowering by GI is mainly through activation of *CO* transcription (Mizoguchi et al., 2005; Sawa et al., 2007; Suarez-Lopez et al., 2001). Interestingly one binding site spans the conserved region identified in this study upstream of -2.1kb, later referred to as Conserved Distal Motif. This region is excluded in the 1.8kb construct but included in the 2.6kb. In order to test the interaction of GI with the distal part of *CO* promoter the 2.6kb and 1.8kb lines were crossed to *35S::GI* in *co-2* mutant and flowering time was examined in both LD and SD conditions (Fig. 14c). L*er* control flowered with 14.1 leaves in LD and 35.3 leaves in SD. While *35S::GI* flowered earlier in LD (9) and SD (15.8), *35S::GI* in *co-2* mutant flowered later in LD (22.3) but earlier in SD (29.1) compared to L*er*. The flowering time of transgenic

complementation lines was very consistent for the 2.6kb constructs. All four lines flowered with 5.7 to 6.7 leaves both in LD and SD conditions. 1.8kb constructs showed later flowering in LD with 9.4, 9.7 and 18.3 in LD and slightly later flowering in SD with 10.4, 11.6 and 23.2 leaves, respectively. This experiment indicates that *35S::GI* accelerates flowering via these *CO* promoter constructs and that both the 2.6kb *CO* promoter as well as the 1.8kb *CO* promoter are under the control of *35S::GI.*

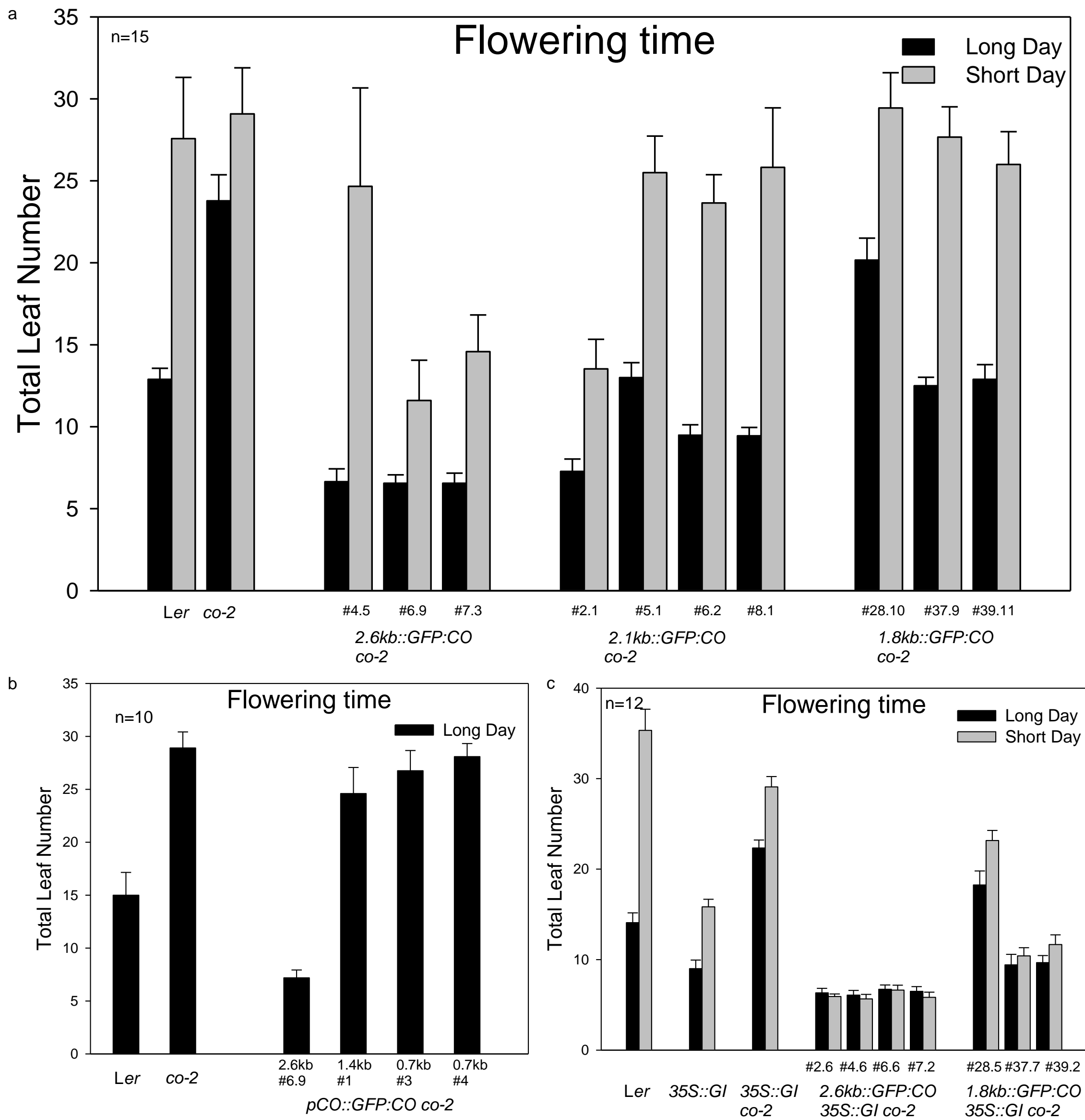


Figure 14: Flowering Time of *CO* promoter complementation lines.
Mean flowering time (Total leaf number) of *pCO::GFP:CO* lines transformed into L*er co-2* mutant. Flowering time was determined in long day (black) or short day (grey) conditions. L*er*, *co-2* and *35S::GI* in *co-2* are shown as controls, respectively. Flowering time of 2.6kb, 2.1kb and 1.8kb *pCO::GFP:CO* lines (a) as indicated (n=15). Flowering time of 1.4kb and 0.7kb *pCO::GFP:CO* transgenic lines in LD (b, n=10) . Flowering time of 2.6kb and 1.8kb *pCO::GFP:CO* lines crossed to 35S::GI in *co-2* in LD and SD (c, n=12). All values are displayed as mean and standard error. # indicates independent transgenic homozygous T_3 line.

The mRNA levels of *GFP:CO* and *FT* were examined in the 2.6kb, 2.1kb and 1.8kb *pCO* complementation lines to identify differences in temporal mRNA regulation. Seedlings were grown for 10 days and mRNA from one representative line of each construct was sampled every three hours in LD and SD conditions (Fig. 15).

FT mRNA levels in LD (Fig. 15c) and SD (Fig. 15f) were highest in the 2.6kb *pCO* complementation line, followed by the 2.1kb and lowest in the 1.8kb line. In LD conditions *FT* mRNA showed two peaks, one smaller morning peak at ZT3 and the major evening peak at ZT18 (Fig. 15c). Differences in mRNA levels of *GFP:CO* transgene were not as clear, thus two biological replicates are shown (Fig. 15a,b). The pattern of expression of *CO* resembled expected pattern of wild type *CO* (compare Fig. 6). In the first biological replicate the evening peak (ZT12) was more pronounced in the 2.6kb line compared to 2.1kb and 1.8kb (Fig. 15a) and also mRNA levels were higher in the morning (ZT0). In the second replicate the expression of *GFP:CO* was higher in the 2.6kb line compared to 2.1kb and 1.8kb at ZT18 (Fig. 15b). Additionally the expression of *pCO::LUC* (2.6kb and 1.8kb) introduced into Col-0 and mRNA levels were examined in a time course experiment for one representative line for each construct (Fig. 15d). Only small differences in expression at ZT0 and ZT21 were observed between transgenic lines tested consistent with the findings in complementation lines.

The *FT* mRNA in SD conditions showed only one peak that occurs between ZT6 and ZT9 (Fig. 15f). Differences in total FT mRNA levels clearly reflect the flowering time of these lines with the 2.6kb *pCO* line having highest levels followed by the 2.1kb and 1.8kb lines. The *GFP:CO* mRNA levels show lowest levels in the first three time points of the SD as expected (Fig. 15e). The mRNA levels rise between ZT6 and ZT9 in the 2.6kb and 2.1kb while the expression stays low in the 1.8kb line. This is consistent with the rise in *FT* levels (Fig. 15f) around this time of the day.

Differences in FT mRNA and flowering time were often more pronounced in this material then differences in *GFP:CO* mRNA. Possibly, *GFP:CO* mRNA levels that give rise to differences in *FT* mRNA could have occurred earlier in development than *FT* mRNA levels and therefore were not observed in the experiment (Corbesier et al., 2007). To test this mRNA levels of *GFP:CO* transgene and *FT* in 2.6kb and 2.1kb transgenic lines were examined in a developmental time course series at ZT18 after 5, 6, 8 and 9 long days (Fig. 15g,h). Differences in *FT* mRNA between the lines were already obvious after 5 LDs and got more pronounced with each long day (Fig. 15h, Corbesier et al., 2007).

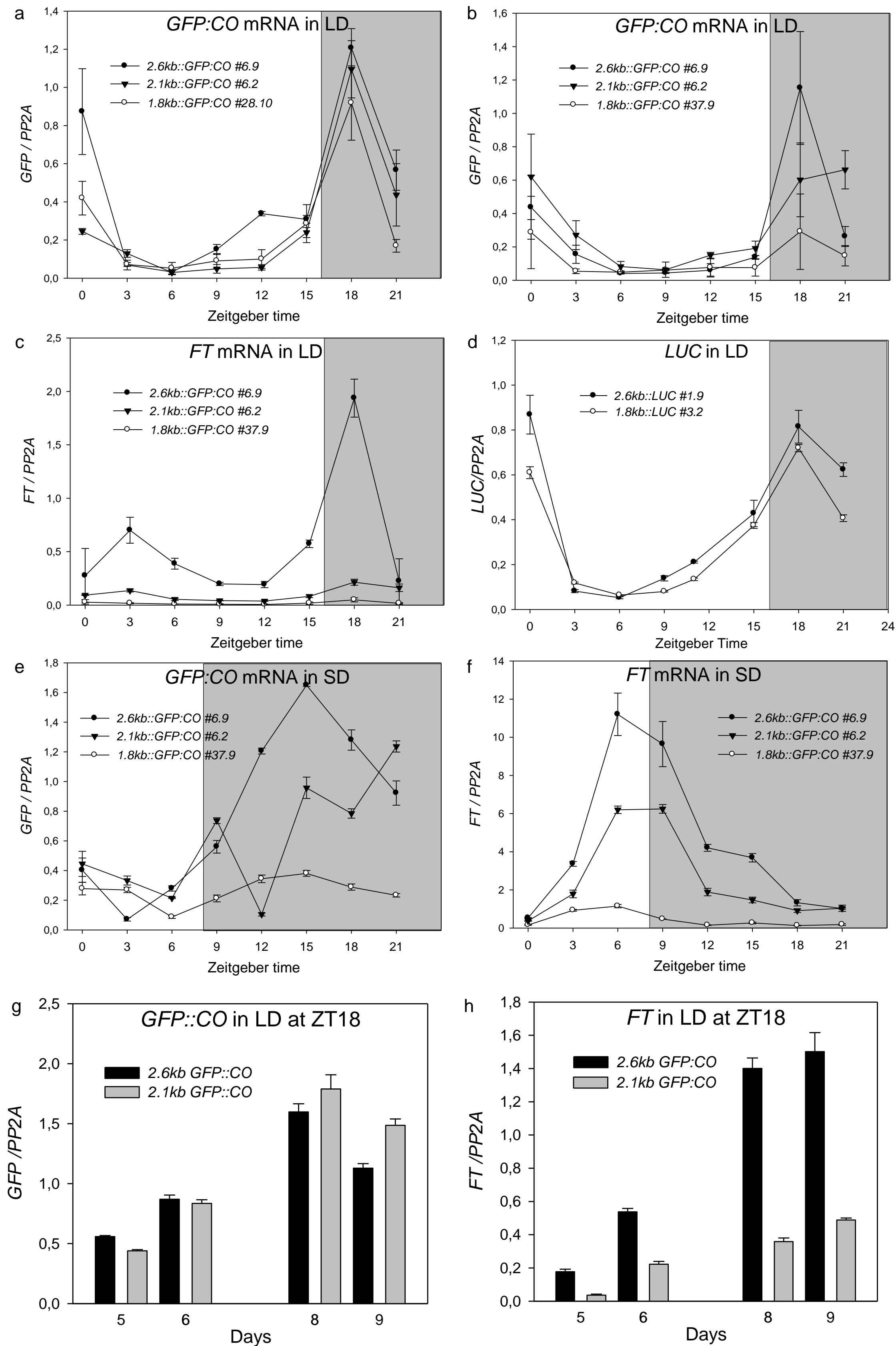


Figure 15: Analysis of mRNA levels in *CO* promoter complementation and reporter lines

mRNA levels of *GFP:CO* (a,b,e), *FT* (e,g) and *LUC* (d) normalized to *PP2A* in representative 2.6kb, 2.1kb and 1.8kb *pCO::GFP:CO* or *pCO::LUC* transgenic lines was examined in LD and SD in timecourse experiments every three hours as indicated. Seedlings were grown for ten days on soil. Additionally, mRNA levels of *GFP::CO* (h) and *FT* (i) normalized to *PP2A* in 5, 6, 8 and 9 day old seedlings at ZT18 grown on plates was examined. The mean of three technical replicates of the same cDNA is displayed with error bars indicating standard deviation. Grey areas specify dark hours of the day.

The *GFP:CO* mRNA levels increased over time as well but no significant difference was observed between the lines. Indeed mRNA levels of the *GFP:CO* transgene appeared to be even lower in the 2.6kb line compared to the 2.1kb line. Possibly *FT* is activated only in a subset of cells expressing *CO*, and in these cells the 2.1kb promoter is less active than the 2.6kb.

In summary, *CO* promoter does not represent a straightforward switch but was shown to have many functional parts leading to a progressive increase in activity with increasing length as represented by acceleration of flowering as well as corresponding *FT* mRNA levels. The examination of *CO:GFP* mRNA levels led to ambiguous results and no clear difference in mRNA pattern was detected. The most likely scenario is that amplitude is lower in shorter promoter fragments. An *A. thaliana CO* promoter longer than 2.1kb is necessary to fully complement the *co-2* mutation.

5.5. Dissection of *CO* promoter by the gain of function approach reveals separable functions for subelements in shaping mRNA expression patterns

To describe the contribution of different promoter parts to *CO* mRNA regulation the ~2.6kb sequence was divided into four parts of ~700bp in length (Fig. 13b). The short elements A (-550bp to +60bp from the TS), B (-1290 to -550), C (-1900 to -1290bp) and D (-2550bp to -1900bp) were fused to the *pnos* minimal promoter and used to drive expression of the *LUC* gene. The full length promoter as well as element A fused to *LUC* without the *pnos* minimal promoter were used as control lines. Element A contains the 5´UTR of *AtCO* and thus does not necessarily require the additional TS of the *pnos* promoter. All constructs were introduced into Col-0 plants.

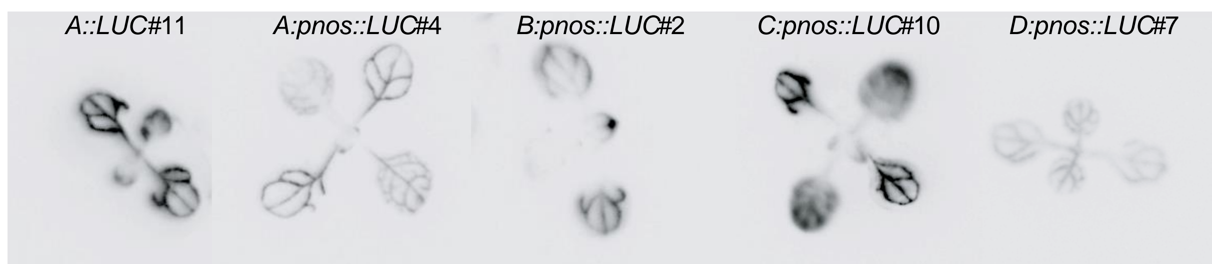


Figure 16: Spatial expression of *A. thaliana pCO* gain of function constructs.
Light emission (*LUC* signal) in seedlings containing ~700bp fragments of *CO* promoter (A-D, Fig. 2.4b) fused to *pnos::LUC* as detected by CCD camera. For each construct one representative line is shown.

The spatial expression conferred by *A, B, C* and *D* fused to *pnos::LUC* was examined by imaging light emission in LD grown seedlings. Light signal was observed in the vasculature of cotyledons and first two true leaves for all constructs tested (Fig. 16). No spatial difference was observed between *A::LUC* and *A::pnos:LUC*. Differences in intensity of the signal were observed between transgenic lines (data not shown) and between constructs. Constructs A and C gave strong light signals while B and D showed weaker signals, indicating that these different parts of the promoter lead to different levels of *mRNA* production *in planta*. All constructs were also fused to *pnos::GUS* and stable transgenic lines were established but no signal was detected for all constructs tested (data not shown). To conclude,

although expression amplitudes were different all four constructs conferred information for the expected spatial expression pattern, comparable to that of endogenous *CO* mRNA.

In order to investigate their temporal expression patterns, activity of the *pnos::LUC* constructs was followed over time to identify similarities and differences to endogenous *CO* mRNA levels. Light emission was measured over a timecourse experiment in LD conditions (Fig. 17a,c,e,g). Both *A:pnos::LUC* and *D:pnos::LUC* showed a diurnal light emission pattern in LD conditions. Light emission in *A:pnos::LUC* was high at the beginning of the day and reached a trough signal around ZT10 and then rose until ZT24 (Fig. 17a). *D::pnos:LUC* light emission decreased upon lights on and started to rise from around ZT10 to reach a plateau at ZT16, similar to expression of the endogenous gene (Fig. 17d). The normalized light signal of *B:pnos::LUC* did not vary over the course of a day, while the signal of *C:pnos::LUC* slowly decreased with time (Fig. 17c,e). Because the light emission in TopCount experiments was close to the detection limit for some lines and constructs (e.g. *B:pnos::LUC*, data not shown), *LUC* mRNA levels were also examined through a Time Course experiment for all four constructs. Two homozygous lines were tested for each construct and one representative biological replicate is shown for each (Fig. 17b,d,f,h). The expression of *A:pnos::LUC* mRNA differed from the light signal observed for the same construct (Fig. 17a,b). Cycling of expression was confirmed by qRT-PCR but mRNA levels fall immediately upon lights on in the morning and stay low until ZT15 and then form an evening peak at ZT18. The expression pattern of *B:pnos::LUC* also cycled throughout the day, but *B:pnos::LUC* mRNA levels reached only 1% (*LUC/PP2A*=0.06) of maximum levels observed for *A::pnos:LUC* (*LUC/PP2A*=0.6), indicating that fragment B on its own confers very low or no expression (Fig. 17b,d). The mRNA levels of *C:pnos::LUC* cycled more strongly than suggested by the luminescence data (Fig. 17e,f). *LUC* mRNA decreases from ZT0 to reach a trough at ZT6 and rises slowly until ZT15. Upon darkness *C:pnos::LUC* mRNA levels rose steeply to reach an evening peak at ZT21 (Fig. 17f). Overall *LUC* expression had a maximum level of 0.2 (*LUC/PP2A*), which is lower but still comparable to *A:pnos::LUC* levels. The downregulation in the morning and upregulation at night conferred by fragment *C* are also features of endogenous *CO* mRNA levels. The mRNA levels of *D:pnos::LUC* were in agreement with the findings in the TopCount luminescence experiments (Fig. 17g,h). Downregulation in the morning with a trough at ZT3 and ZT6, then upregulation reaching a plateau at ZT9 or ZT12 depending on the line examined. Interestingly in this construct the night peak did not dominate the diurnal pattern, which discriminated it from the other lines investigated.

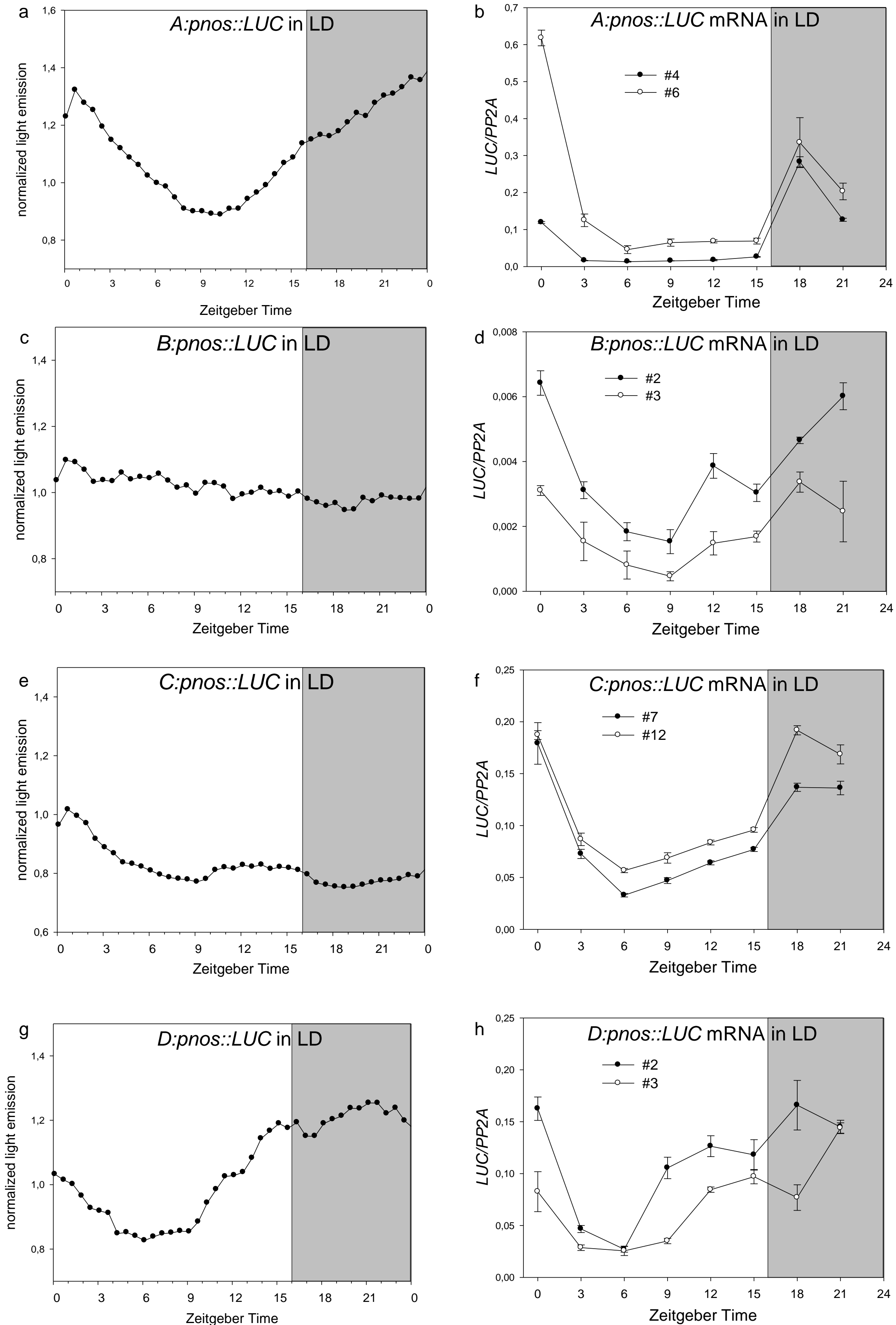


Figure 17: Temporal expression of *pCO* fragments fused to *pnos::LUC* in LD conditions.
Top Count experiment showing normalized A (a), B (c), C (e), and D (g) *pnos::LUC* expression in LD conditions (LED) of ten day old seedling grown on plates. mRNA levels of *LUC* normalized to *PP2A* of A,(b), B (d), C (f), and D (h) *pnos::LUC* (two independent lines) in LD conditions at the indicated ZT in ten day old seedlings grown on soil are presented. The mean of three technical replicates of the same cDNA is displayed with error bars indicating standard deviation. Grey areas specify hours in the dark.

Normalized *LUC* mRNA levels driven by the ~700bp fragments fused to *pnos::LUC* were compared to endogenous *CO* mRNA levels (Fig. 18). In this experiment *CO* mRNA was downregulated between ZT0 and ZT3 so that the trough in expression extended until ZT6. Also *CO* mRNA showed an evening plateau and a peak at ZT18. These comparisons demonstrated that the different parts of *CO* promoter confer different aspects of *CO* mRNA pattern. *A:pnos::LUC* (Fig. 18a) shows efficient downregulation in the morning and the night peak at ZT18. However in this construct evening activation between ZT9 and ZT15 is lacking. *B:pnos::LUC* and *C:pnos::LUC* expression patterns are similar (Fig. 18b,c). The comparison of these constructs with *CO* mRNA showed that the downregulation of *LUC* in the morning is not efficient and trough expression is only reached at ZT6 (*C:pnos::LUC*) or ZT9 (*B:pnos::LUC*), respectively. The expression of *LUC* and *CO* overlaps after ZT9. *D:pnos::LUC* had its trough of expression from ZT3 to ZT9 (Fig. 18d) and mRNA levels then rose. It was the only construct where expression in the evening (ZT12 and ZT15) was higher than observed in normalized *CO* expression.

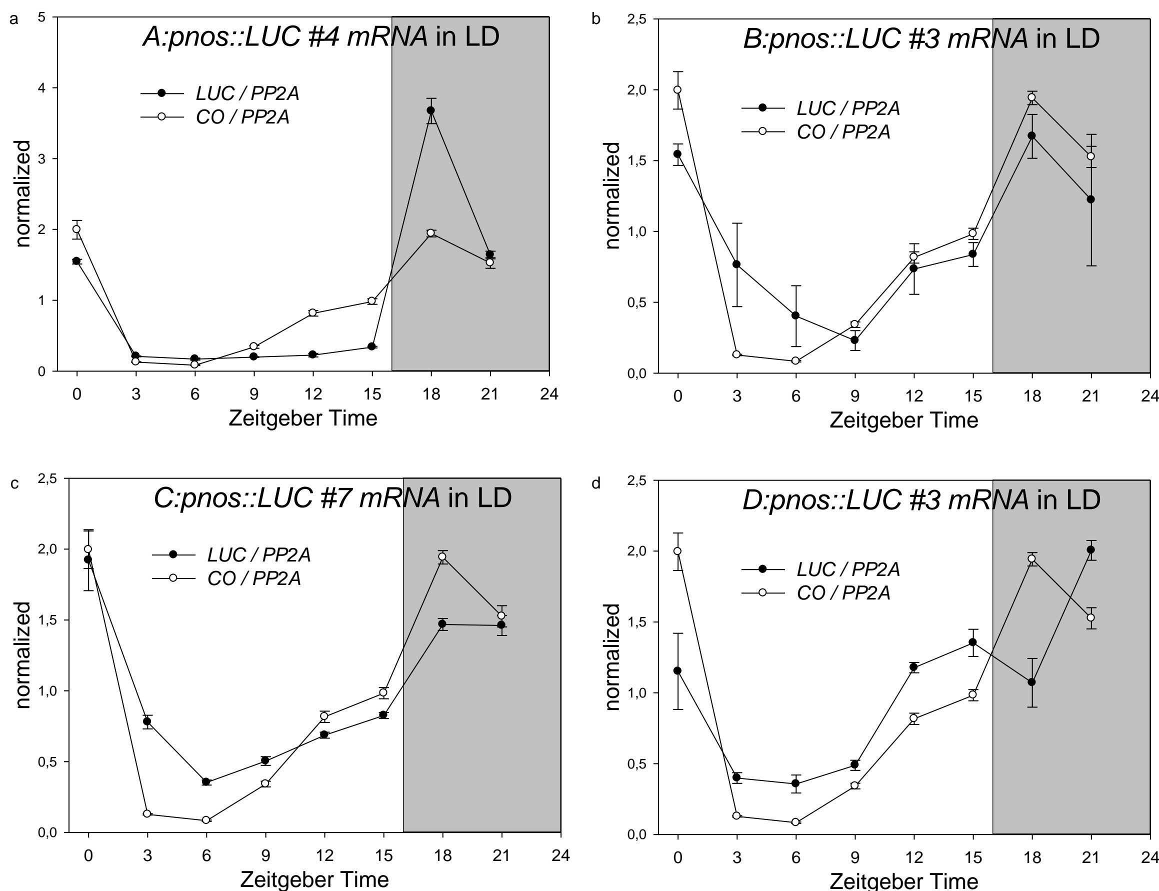


Figure 18: Comparison of *LUC* mRNA driven by 700bp *CO* promoter fragments fused to *pnos::LUC* to endogenous *CO* mRNA levels.
Expression levels of A (a), B (b), C (c) and D (d) fused to *pnos::LUC* were compared to endogenous *CO* mRNA levels in ten day old seedlings grown on soil in timecourse experiments in LD conditions. mRNA levels of *LUC* and *CO* are displayed as ratios to *PP2A* followed by normalization to the average of each experiment (normalized expression). One representative line is displayed. The mean of three technical replicates of the same cDNA is displayed with error bars indicating standard deviation. Grey areas specify hours in the dark.

Expression levels conferred by the promoter fragments were also compared to a *pCO 2.6kb::LUC* control line (Fig. 19a). The mRNA levels of the *pnos::LUC* constructs were significantly lower compared to the full length construct. This result was supported by the finding that 0.7kb (A) *pCO::GFP:CO* cannot complement the *co-2* mutant (Fig. 14b). All promoter fragment constructs produced very low mRNA levels in the conditions tested. The 0.7kb fragment (A) was also fused to *LUC* without the *pnos* minimal promoter to examine whether the artificial promoter introduced a bias. Two independent lines were tested for *A::LUC* (Fig. 19b) and *A::pnos:LUC* (Fig. 17b), respectively. The different constructs produced similar mRNA patterns and in overall levels were similar as well. This confirms earlier studies showing that the *pnos* minimal promoter provides a transcriptional start site but does not influence transcription levels by autonomously activating transcription (Harmer and Kay, 2005; Puente et al., 1996).

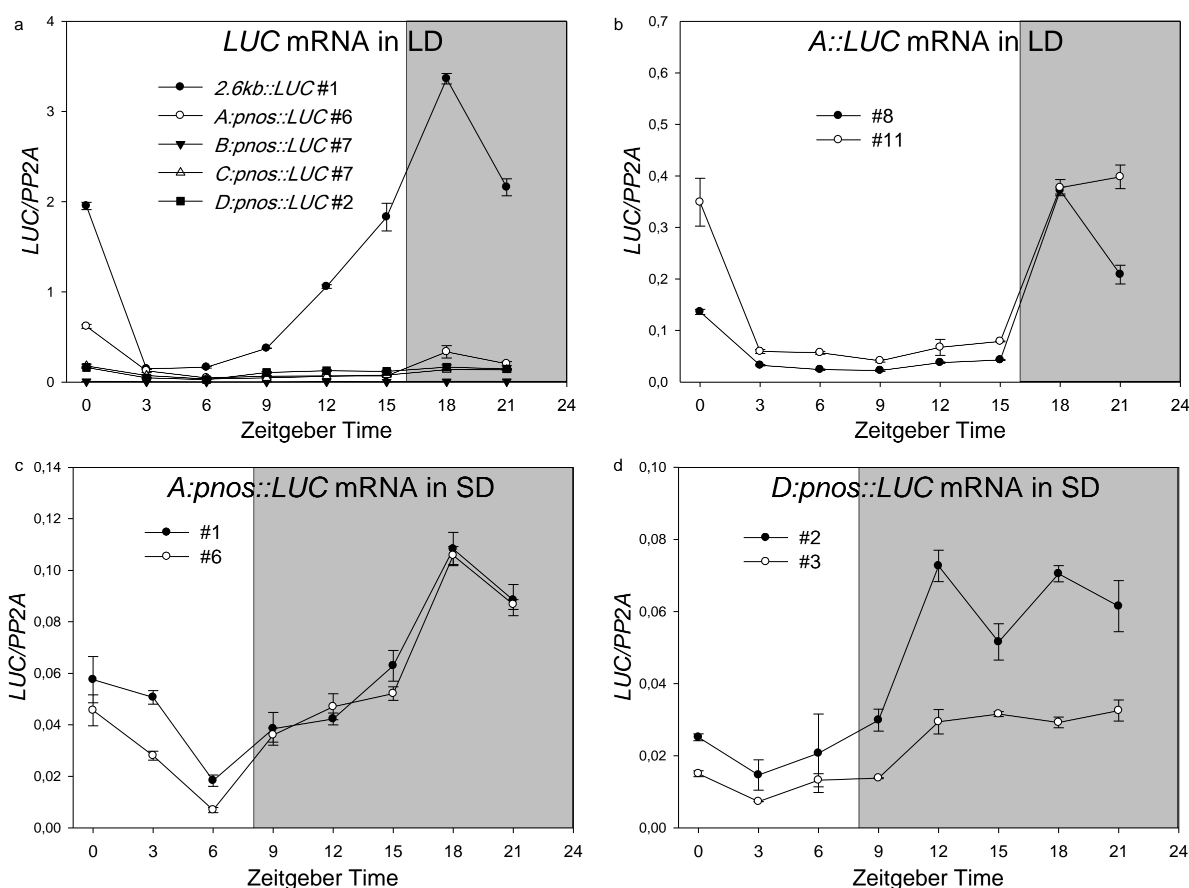


Figure 19: Comparison of *LUC* mRNA levels driven by 700bp *CO* promoter fragments fused to *pnos::LUC* to *2.6kb::LUC* and *0.7kb::LUC* and expression analysis in SD.

LUC mRNA levels of A, B, C and D fused to *pnos::LUC* were compared to *pCO::LUC* (*2.6kb::LUC*) in LDs (a). Additionally, expression of proximal 0.7kb *CO* promoter fused to *LUC* (*A:LUC*) in LD of two representative lines is displayed (b). mRNA levels of *A:pnos::LUC* (c) and *D:pnos::LUC* (d) were examined in SD conditions. All seedlings were grown on soil in the indicated day length for ten days. The mean of three technical replicates of the same cDNA is displayed with error bars indicating standard deviation. Grey areas specify hours in the dark.

As endogenous *CO* is expressed differentially in different day length conditions mRNA of *A:pnos::LUC* and *D:pnos::LUC* was tested in SD (Fig. 19c,d). *A:pnos::LUC* showed a trough at ZT6 and a peak at ZT18 but not as pronounced as in LD conditions. Similar to endogenous *CO* mRNA

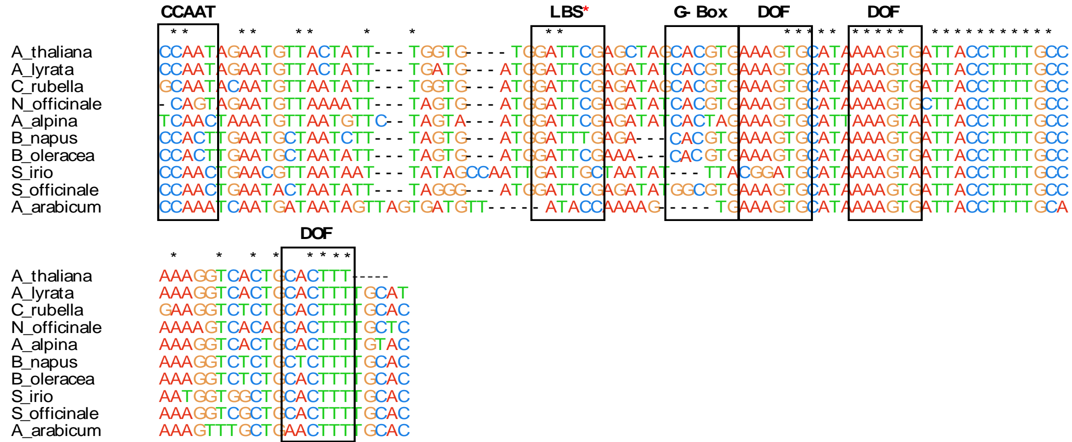
(Fig. 6) the *LUC* mRNA was low at ZT0. The broad plateau during the dark period and very low levels in the light period observed for *CO* mRNA were not reproduced, but this could be due to low overall levels in SD (data not shown). The mRNA of *D:pnos::LUC* varied in the two lines tested (Fig. 19d) but the pattern was similar. The expression was low already upon lights on (ZT0) and stayed low until ZT9 and then rose to reach a plateau between ZT12 and ZT21. This resembles endogenous *CO* in SD conditions and would probably allow repression of *CO* activity under SDs as *LUC* mRNA was only present in the dark hours of the day (ZT8-ZT24) when CO protein is destabilized.

The dissection of the promoter performed by fusing ~700bp fragments to *pnos::LUC* and analysing their function in Col-0 confirmed the redundant role of different promoter parts. The proximal 0.7kb (A) seems to provide the background pattern, which lacks the evening peak, but shows a strong night peak. Element B provides rhythmic but low expression levels. This is consistent with the finding that there are no significant differences in flowering time of *co-2* mutant carrying the 1.4kb or 0.7kb promoter fusions. Region C provides expression when the evening and night peaks occur but the morning downregulation is not efficient. Region D on the other hand is efficiently downregulated in the morning and confers an evening and night peak. Additionally all fragments provide the expected spatial information. Summarizing the findings it is possible that additive effects of the independent *CO* promoter parts on transcription generate the endogenous *CO* expression pattern at wild type levels, especially when a synergistic effect of the elements is assumed. Each of the *CO* promoter fragments provides two of the three key regulatory features of endogenous *CO* mRNA.

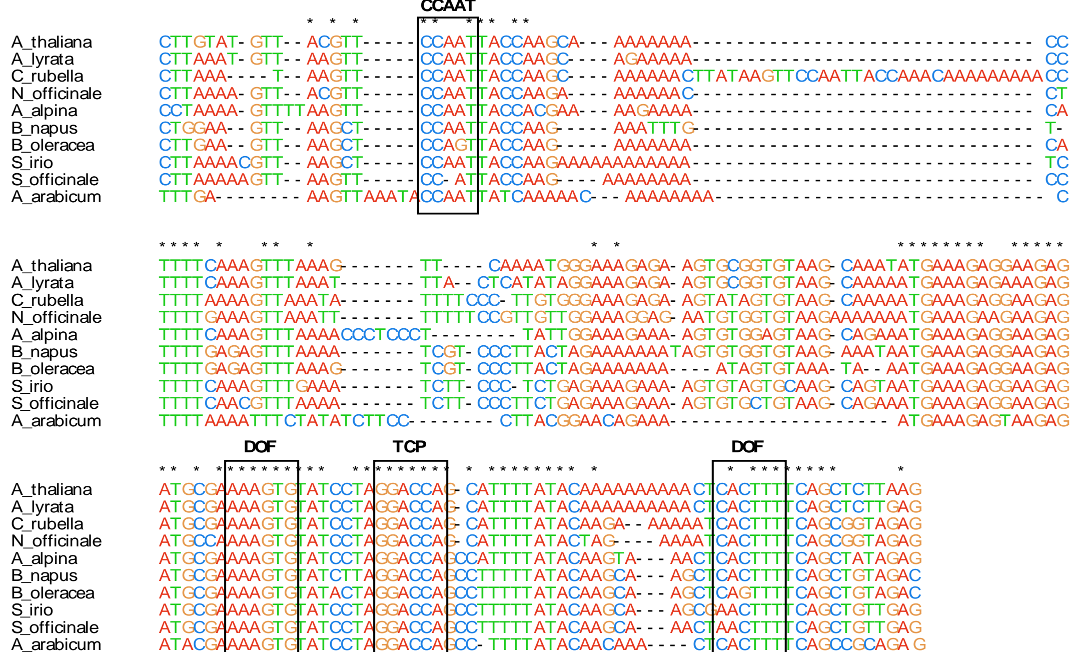
5.6. Analysis of conserved regulatory motifs by gain of function approach supported findings from phylogenetic shadowing and *cis*-element analysis

In order to refine the promoter regions that confer expression patterns mimicking endogenous *CO* mRNA, three short conserved promoter regions were tested (Fig. 13, 20). The three short motifs named Conserved Distal Motif (CDM, 93bp), Conserved Middle Motif (CMM, 171bp) and Conserved Proximal Motif (CPM, 135bp) were fused to *pnos::LUC* and introduced into Col-0 (Fig. 20). In this approach not only conservation but also the *cis*-element composition was taken into account for design of the constructs. CDM has one CCAAT-Box, one LBS, one G-Box and three conserved DOF binding sites in 93bp (Fig. 20). It also overlaps with the binding region of GI, FKF1 and CDF1 to *CO* promoter as defined by ChIP (Sawa et al. 2006). The CMM motif has, of all cis-elements investigated in this study, a CCAAT-Box, two DOF binding sites and one TCP binding site. The CPM consists of three DOF binding sites, one CCAAT-Box, one G-Box and one LBS, but shows less conservation among species (Fig. 20). It shares its *cis*-element content with CDM but the order and distance between the elements varies between the two motifs. The LBS and G-Box are found both in CDM and CPM, either in an important regulatory region of the promoter together with repressive DOF binding sites or in close proximity to the TS, respectively. Both LBSs were changed by site directed mutagenesis in the

Conserved Distal Motif 93bp (-2308 to -2214): *CDM:pnos::LUC*



Conserved Middle Motif 171bp (-1663 to -1491): *CMM:pnos::LUC*



Conserved Proximal Motif 135bp (-237 to -102): *CPM:pnos::LUC*

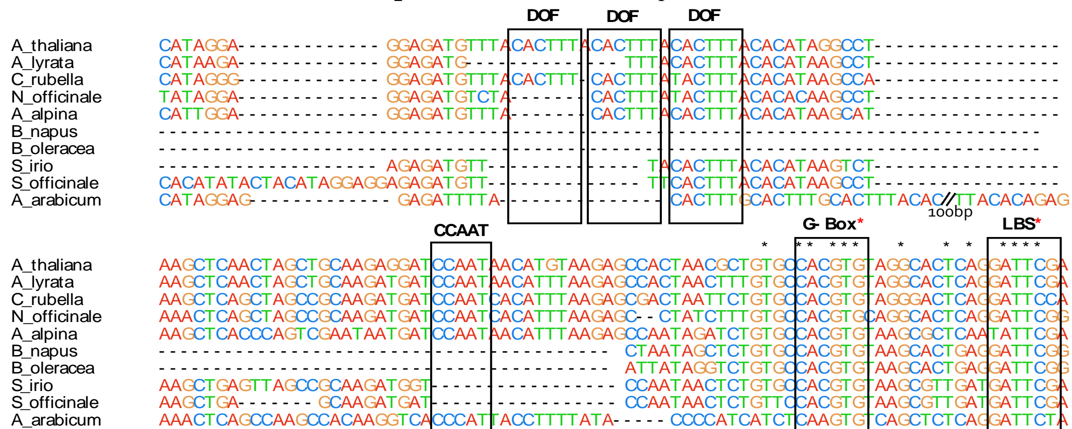


Figure 20: Sequence alignment representing three conserved regions in *CO* promoter.
Alignment of homologous regions of *CO* promoter as displayed in Fig 11 in the indicated species. Conserved Proximal Motif (*CPM*), Conserved Middle Motif (*CMM*) and Conserved Distal Motif (*CDM*) were deduced from this alignment (Fig 13). Length and position of the Motifs is indicated. Identified *cis*-elements in *A. thaliana* are represented (CCAAT-Box, LBS, G-Box, DOF, TCP). Red asterisks indicate *cis*-elements that were changed by site directed mutagenesis in the corresponding constructs. Black asterisks indicate 100% base pair conservation between all ten Brassicaceaen *CO* promoters.

respective constructs. The mutation of the G-Box in the context of the whole promoter did not have an obvious effect on gene expression (data not shown), thus it was mutated in the context of CPMlbs.

The 700bp fusions to *pnos::LUC* constructs described previously resulted in low light emission in TopCount experiments (Fig. 17). The CDM, CMM and CPM fusions to *pnos::LUC* were therefore directly analysed by extracting mRNA of ten day old seedling every three hours for 24h. The comparison with *pCO 2.6kb::LUC* showed that these constructs do confer expression but lower than the full length *CO* promoter (Fig. 21a). Interestingly the mRNA levels produced by CPM and CDM were lower but comparable to those caused by *A:pnos::LUC* while element CMM expression appeared undetectable when plotted on the same scale (Fig. 21b)

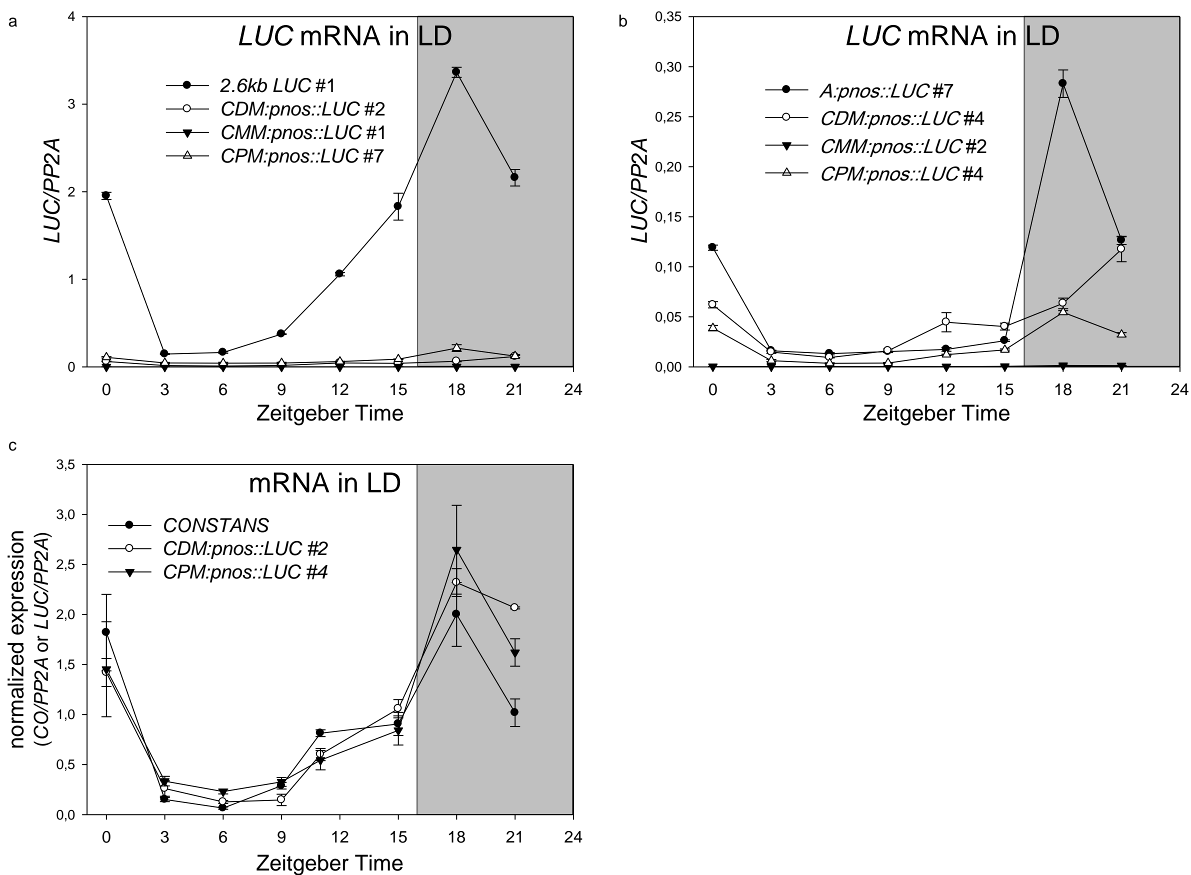


Figure 21: The *LUC* mRNA levels conferred by short conserved *CO* promoter motifs fused to *pnos::LUC* compared to *2.6kb::LUC*, *A:pnos::LUC* and endogenous *CO*.
LUC mRNA levels of CPM, CMM and CDM fused to *pnos::LUC* were compared to mRNA levels of *pCO::LUC* (*2.6kb::LUC*) (a) and *A:pnos::LUC* (b) in LDs, respectively. Additionally endogenous *CO* and *LUC* mRNA levels were compared in *CPM:pnos::LUC* and *CDM:pnos::LUC*. Values are displayed as ratios to *PP2A* followed by normalization to the average of each experiment (normalized expression). Seedlings were grown for ten days on soil in LD conditions. The mean of three technical replicates of the same cDNA is displayed with error bars indicating standard deviation. Grey areas specify hours in the dark.

Time Course experiments of two independent lines were compared, to examine the expression pattern conferred by these short promoter sequences. The mRNA levels of *LUC* were normalized against

PP2A to compare the relative expression levels and additionally normalized to the dataset to be able to compare gene expression patterns.

mRNA levels of *CDM:pnos::LUC* and *CPM:pnos::LUC* were very similar to endogenous CO expression pattern (Fig. 21c). The pattern is characterised by a strong downregulation at dawn followed by trough expression. The mRNA forms an evening peak between ZT12 and ZT15 and rises again to exhibit a night peak around ZT18 and ZT21 which is usually higher compared to the evening peak (Fig. 21c, 22a,b,g,h). The expression was lower in the night compared to *A:pnos::LUC* (Fig 21b). The pattern of expression of *CMM:pnos::LUC* was characterised by low levels in the light period (ZT0 to ZT15) and a rise by about three fold after lights off at ZT18 and ZT21 (Fig. 22c,d).

The mutation of the LBS in the context of CDM (CDMlbs) and CPM (CPMlbs) did not cause significant differences in expression levels or pattern (Fig. 22 a,b,e,f,). Expression levels vary between lines and experiments but a clear trend was not observed. While one wild type CPM line exhibited higher levels compared to two CDMlbs lines similar levels were detected for another wild type line (Fig. 22a). Additional indication that there is no significant difference comes from the CPMlbs lines. In these experiments one wild type line showed reduced amplitude indicating opposing effects in both lines with the same *cis*-element composition (Fig. 22e). The pattern of expression was also not changed for CPMlbs and CDMlbs compared to their respective controls as indicated by comparison of normalized values (Fig. 22b,h). Mutating the G-Box in CPMlbs (CPMlbsg/box) on the other hand resulted in significantly lower mRNA levels (Fig 2.12g) while the expression pattern was unchanged (Fig 22h). Even though variation in expression levels was observed for these lines all experiments suggested that the G-Box contributes to an elevation in overall mRNA production in this context.

Notably expression patterns that recreate endogenous *CO* are maintained even without the G-Box and the LBS. These experiments show that the G-Box might play an activating function in regulation of *CO* mRNA. But as suggested before (Ito et al. 2012), the pattern might be shaped by the repressors of *CO*, the CDFs, and thus activation through E-Boxes as shown before or the G-Box, as suggested in this study, might only change amplitude of expression. In conclusion short conserved sequence stretches in *CO* promoter are sufficient to mediate an mRNA pattern that mimics endogenous *CO*. The most prominent cis-element in this analysis remains the DOF binding sites while no effect was assigned to the LBS. The G-Box close to the TS in *CO* promoter might contribute amplitude of *CO* mRNA in *A. thaliana*.

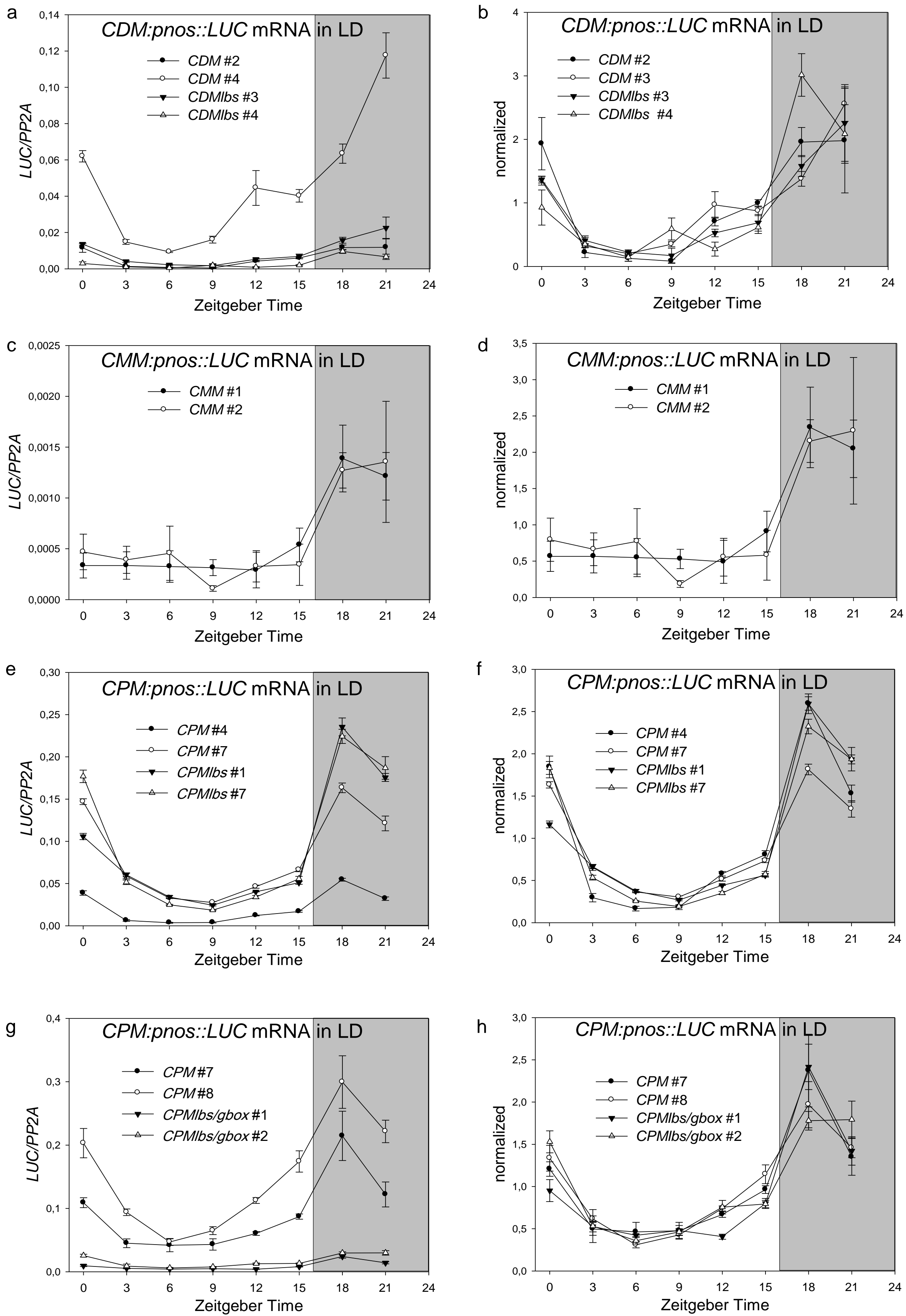
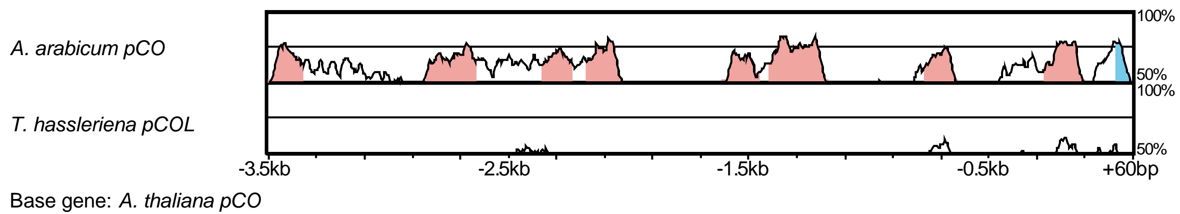


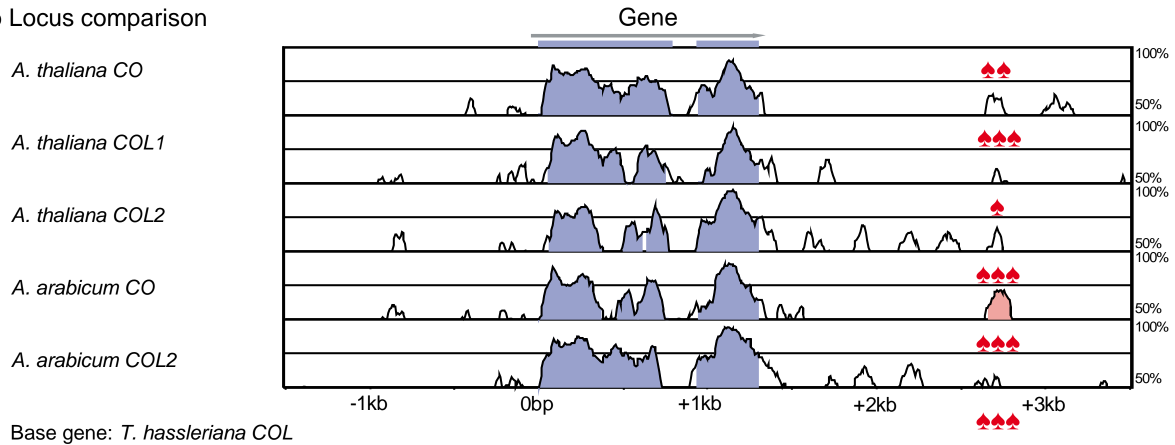
Figure 22: *LUC* mRNA levels in CPM, CMM and CDM *pnos::LUC* reporter lines.
LUC mRNA levels in ten day old seedling grown in LD of *CDM::pnos:LUC* including *lbs* mutation (a,b), *CMM::pnos:LUC* (c,d), *CPM::pnos:LUC* including *lbs* mutation (e,f) and *CPM::pnos:LUC* including *lbs/gbox* mutation (g,h). Expression was either normalized to *PP2A* (a,c,e,g) or additionally followed by normalization to the average of each experiment (normalized expression) (b,d,f,h), respectively. The mean of three technical replicates of the same cDNA is displayed with error bars indicating standard deviation. Grey areas specify hours in the dark.

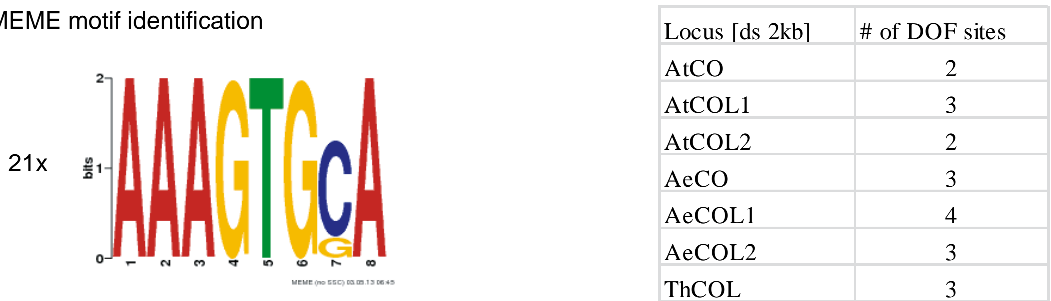
5.7. Tandem duplication leading to emergence of *CO* caused appearance of a regulatory DOF binding site cluster in the distal *CO* promoter

This study shows that the promoter of *CO* is conserved in the Brassicaceae. The conserved regions of *CO* promoter were shown to contain functionally important regulatory sequence in *A. thaliana*. The promoter of *CO* is only present in the Brassicaceae and most likely evolved after the tandem duplication of *COL1* and *CO* ancestor. While the *COL1* expression pattern remained similar to *COL2* after duplication the *CO* transcriptional pattern changed. The conservation between promoter *CO* and *pCOL* from *T. hassleriana* was low (Fig. 23a), which confirmed observed changes in *cis*-element composition (Fig. 12). Similarly conservation between the promoter of *ThCOL* and those of other subgroup Ia *COL* genes in *A. thaliana* and *A. arabicum* was low (Fig 23b). A downstream conserved region was identified in all subgroup Ia *COL* genes from *A. thaliana*, *A. arabicum* and *T. hassleriana*. This region overlaps with the Conserved Distal Motif (CDM) of the *CO* promoter. Detailed analysis revealed that three DOF binding sites are present in the downstream region of *ThCOL* and between one and three DOF sites in all other *CO* and *COL* loci analysed (Fig. 23b,c). This region is evolutionarily conserved and present in the ancestor of *T. hassleriana* and Brassicaceaen subgroup Ia *COL* genes. Data from this study implies that this conserved sequence stretch might be important for proper *CO* regulation in *A. thaliana* (Fig. 14), being involved in contribution to *CO* mRNA pattern but not *COL1*/*COL2* pattern (Fig. 15). Additionally this part on its own is sufficient to confer an expression pattern highly similar to endogenous *CO* in a *LUC* reporter assay (Fig. 21c). In conclusion this conserved intergenic region between *CO* and *COL1* most likely regulates *CO* but not *COL1* transcription. The role of this sequence stretch in the downstream region of the other *COL* gene remains uninvestigated. During the tandem duplication that led to *CO* and *COL1* the downstream region became duplicated as well and one copy was subsequently located upstream of what became *CO* (Fig 23d). This mechanism introduced DOF binding sites into the upstream regulatory region of *CO*, which additionally increased in other positions of the promoter. Thus this duplication not only allowed the functional diversification of the two genes but also provided the first step in evolution of *CO* transcriptional pattern by introducing certain *cis*-elements. The change in expression pattern presumably evolved to reduce overall expression of *CO* and abolish expression in the first part of the day, which could both be mediated by CDF transcription factors. Such repression was essential for the photoperiodic function of *CO*. The relief of repression in the evening by the FKF1-GI module might have evolved later to enhance activation under LDs.

a Promoter comparison







Locus [ds 2kb]	# of DOF sites
AtCO	2
AtCOL1	3
AtCOL2	2
AeCO	3
AeCOL1	4
AeCOL2	3
ThCOL	3

d Model for *COL* tandem duplication

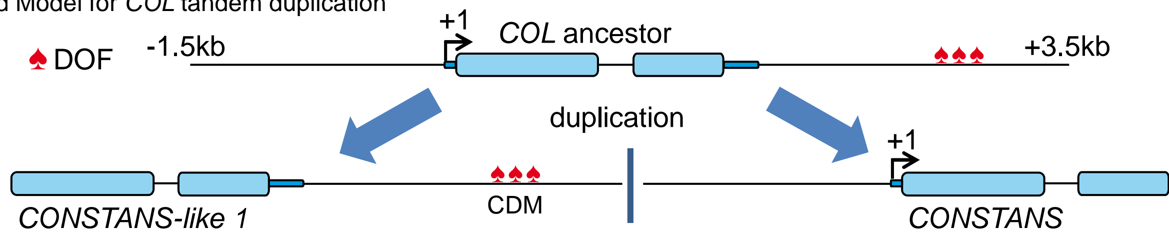


Figure 23: Evolution of *CO* promoter in the Brassicaceae
mVISTA representation of alignment of *AtCO* to *AeCO* and *ThCOL* promoter is displayed (a). Light red color represents regions of more than 75% bp conservation in a sliding window of 100bp. Turquoise indicates 5´UTR of *AtCO* in as annotated in TAIR10. mVISTA representation of alignment of subgroup 1a *COL* genes (-1.5kb to +3.5kb) is represented (b). *AtCO*, *AtCOL1* and *AtCOL2* as well as *AeCO* and *AeCOL2* are compared to the base sequence *ThCOL*. ♠ indicates presence of conserved DOF binding sites (AAAGTG) in each sequence. Consensus short motif identified by MEME motif identification tool (c) in the downstream region (ds, 2kb) of genes indicated in b and their abundance in the respective ds. Blue color indicates protein coding region of *AtCO* in as annotated in TAIR10. A Model for the evolution of *COL* tandem duplication is displayed (d). *CO/COL1* gene became duplicated including approximately 1.5kb upstream and 2kb downstream sequence (up to +3.5kb). The duplication led to the appearance of DOF binding sites (♠) in the distal promoter (CDM) of the ancestor gene of all Brassicaceae *CO* orthologs.

5.8. Discussion Chapter 5

5.8.1. *CO* promoter conservation in the Brassicaceae

This study highlights the conservation of the *CO* promoter between distantly related Brassicaceaen species. This approach assumes that important sequences are conserved among species, and thus only identifies similarities. Even though the overall expression pattern is conserved, most likely other differences in the regulation of *CO* evolved. Every *CO* ortholog is probably regulated in a slightly different way, thus the phylogenetic shadowing approach can only give a starting point for further, more detailed analysis. Additionally only identifying sequences that are conserved between all species results in regulatory regions that are specific to some species and new evolutionary inventions being overlooked. One region mentioned to be only conserved between *Arabidopis*, *Capsella* and *A. arabicum* was not included in the analysis but could be used to investigate putative differences in regulation in these species compared to others missing this conserved promoter part. Another limitation of this approach is the resolution between species. In the Brassicaceae this approach was successful but it failed to detect any conserved regions when promoters were compared to *T. hassleriana* (Fig. 23). *T. hassleriana* is still a closely related species as is belongs to a sister group of the Brassicaceae in the order Brassicales (Franzke et al., 2011). It was expected that *AtCO* and *ThCOL* do not share common promoter regions as their regulation is different. But as the expression of *ThCOL* resembles *AtCOL1* and *AtCOL2* expression promoter conservation would be expected in these species. The range of the phylogenetic shadowing approach is thus limited and varies between genes. The Phylogenetic shadowing in contrast to the related approach of phylogenetic footprinting is intended to find *cis*-regulatory motfis between closely related species (Boffelli et al., 2003), and its success depends on different factors. One reason is that coding regions and *cis*-regulatory regions evolve in different ways. The exact position of a *cis*-element might not be as crucial for function as it is for an amino acid. In the *A. thaliana CO* promoter for example a G-Box is located close to the TS and is generally conserved. *A. arabicum* also contains this G-Box but at an upstream location. One of the E-Boxes nearby evolved into a G-Box, while the G-Box in the proximal part became an E-Box. These changes maintained the *cis*-element composition but changed their positions, which might have little effect on *CO* regulation, but makes identification by sequence comparison harder. Another factor influencing the efficiency of phylogenetic shadowing is the functional importance of the sequences. If a regulatory module is redundant changes can be compensated for and obvious functional conservation does not correlate with direct sequence conservation. The *ThCOL* promoter for example is not more similar to *AtCOL1* and *AtCOL2* promoters compared to *pAtCO* (Fig. 23) but the transcriptional regulation is comparable (Fig. 6). These limitations led to the second approach taken, where *cis*-elements were identified disregarding their direct position but putting weight on the overall occurrence of a single element.

Promoter conservation was previously studied for other *A. thaliana* genes and differing degrees of conservation were observed. The promoter of *FT* is longer than the *CO* promoter but comparison to *A. alpina* already failed to detect highly conserved regions and *FT* promoter can thus be regarded as less conserved (Adrian et al., 2010). Comparison of *GI* promoter from *A. thaliana* and *A. alpina* highlighted that these two species have a suitable evolutionary distance to identify conserved block in this gene (Berns, 2012). The *CO* promoter is even more highly conserved across its whole length between *A. thaliana* and *A. alpina* compared to *GI*. The promoter of *PRR9* on the other hand showed similar conservation between *A. thaliana* and *A. alpina* compared to *CO* promoter (Herrero et al., 2012). The number of genes investigated is probably too low to draw any conclusions about functional conservation of these regulatory mechanisms across the whole Brassicaceae family.

5.8.2. The composition of *cis*-elements in the *CO* promoter

Previously only two *cis*-elements in the *CO* promoter, the E-Box and the DOF binding site were shown to regulate *CO* transcription based on analysis of their cognate TFs (Imaizumi et al., 2005; Ito et al., 2012a). This study revealed by MEME motif identification that another family of transcription factors, the TCP family might also regulate *CO* transcription. This *cis*-element recognized by TCP TFs was highly overrepresented in the promoters of *CO* orthologs but not in the promoters of the other subgroup Ia *COL* genes investigated. Increase in *cis*-element number can only imply a regulatory function, which needs to be confirmed in biological experiments. It can only be speculated which of the 24 predicted TCP proteins in *A. thaliana* (Kosugi and Ohashi, 2002) might be involved in the regulation of *CO*, but two different studies found a connection between TCP TFs and flowering time. The *tcp4* single mutant, the *tcp2/tcp4/tcp10* triple mutant as well as the jaw-D line, a mutant in which mir319a is overexpressed leading to downregulation of *TCP* mRNA levels, was shown to flower significantly later in LD conditions (Schommer et al., 2008). I repeated this experiment and found that under our conditions the *tcp2/tcp4/tcp10* triple mutant flowered with 14.7 leaves in LD (n=10) while the Col-0 control flowered with 13.4 leaves (n=12). The triple *tcp* mutant was not genotyped but showed the obvious leaf morphology phenotype as described (Schommer et al., 2008). Another report observed a significant late flowering phenotype for the *tcp3/4/5/20/13* quintuple mutant (Koyama et al., 2010), even though the difference of ~2-3 leaves is more mild then observed by Schommer et al. (2008) consistent with my observations. However, the difference in flowering time measured in days was more pronounced meaning the plants develop slower, which might suggest an indirect effect on leaf number at bolting as a result of another developmental effect. Additionally, it was found that *CO* is downregulated in a microarray experiment using a chimeric TCP3 protein that carries a transcriptional repressor domain and was expressed from an inducible promoter to identify direct targets (Koyama et al. 2010). In this experiment *CO* was repressed when the chimeric TCP3 protein was induced. Whether TCP transcription factors regulate flowering time via *CO* transcriptional regulation could be further tested by for example disrupting TCP binding sites in *CO* promoter by site directed mutagenesis and testing the effect on transcription in a transgenic approach or by analysing

the diurnal *CO* expression pattern in the various single or higher order mutants available (e.g. *tcp3* or *tcp4*).

The analysis of *cis*-elements that putatively regulate *CO* transcription was broadened to six elements. DOF and TCP binding sites show a clear overrepresentation, while LBSs and CCAAT-Boxes did not. G-Boxes and E-Boxes are significantly more frequent in *CO* compared to their random occurrence but other subgroup Ia *COL* genes show similar numbers. A limitation in this approach is lack of a bioinformatics pipeline, as well as the knowledge of the identity of *cis*-elements to test. The DOF consensus for example is described as AAAG (Umemura et al., 2004), but for CDF1 there appears to be extra specificity for AAAGTGT (Imaizumi et al., 2005). Such effects probably hold true for most TFs and thus the core binding site is often easily found but TFs present higher specificity for additional Bases around the core region. Similarly it is difficult to test all possible combinations, especially if different members of the same family regulate the same gene but show a slightly different *cis*-element preference.

5.8.3. Functional analysis of the *CO* promoter in *A. thaliana*

Functional analysis of the *CO* promoter revealed important regions for the regulation of flowering time and that the full promoter of 2.6kb is necessary and sufficient to drive *CO* expression. Constructs containing 1.4kb or less failed to complement the *co-2* mutation. This transgenic approach fusing a promoter fragment of *CO* to a tagged version of *CO* cDNA via GATEWAY cloning causes unforeseen difficulties. The flowering time of transgenic *co-2* lines carrying *CO* fused to promoters longer than 1.8kb was earlier than wild type, so that the constructs led to overcomplementation of the *co-2* mutant. It proved difficult to pinpoint exactly which differences in *CO* mRNA levels led to differences in *FT* mRNA and flowering time as observed. In this transgenic approach the genomic locus of *CO* was not only isolated from its context in the genome of *A. thaliana* but also inserted in a binary vector, which might additionally influence the regulation of the gene. The GFP tag, which was used to differentiate *GFP:CO* from endogenous *CO* mRNA, made it possible that posttranslational modification of the protein *in planta* might also be altered. It cannot be excluded, that GFP:CO fusion protein is more stable compared to endogenous CO protein and that this might explain the early flowering. Previously a *pCO::HA:CO* line with 2.5kb promoter was shown to express *CO* mRNA higher compared to wild type and a *cdf1235* line (Song et al., 2012). Possibly regulatory sequences required to reduce *CO* expression are missing in this study. The influences of the intron and the downstream 3´ sequences of *CO* on its mRNA regulation were also not investigated. In summary, the distal promoter fragment, which has a high density of *cis*-elements and is conserved (Fig. 11, 13), has an effect on flowering time and *FT* levels even though the causal changes in *CO* expression could not be identified.

The *CO* promoter was additionally investigated by fusion of fragments to *LUC* adjacent to the *pnos* minimal promoter. In general these experiments helped to define regulatory parts of *A. thaliana CO* promoter and support the results from phylogenetic comparisons. The *LUC* reporter system has

technical limitations. In general the *LUC* reporter system was developed in *A. thaliana* to investigate real time transcriptional changes of clock regulated genes that give robust rhythms exhibiting huge differences over the course of a day (Millar et al., 1992). However, the system is suboptimal for low expression levels, as found for *CO*. Pushing the limits of the system led to uncertainty. Therefore I used *LUC* mRNA measurements to support the luciferase activity data. It was not investigated so far how expression behaves on the mRNA level in empty vector controls (*LUC* and *pnos::LUC*). Thus an important negative control is still missing. All constructs investigated in this study exhibit rhythmic expression (Fig. 17, 19), which could therefore still be an artefact of the system. On the other hand all constructs created come from the diurnally regulated gene *CO* and it cannot be excluded that all parts of the promoter contribute to the diurnal pattern in one or the other way, even if important *cis*-elements (e.g. G-Box) are mutated. Transgenic plants carrying the empty vector controls were recently created and can be assessed in the future. The expression patterns of *A::LUC* and *A:pnos::LUC* were very similar suggesting that the *pnos* minimal promoter does not create a bias but confers only the transcriptional start site. Nonetheless two transcriptional starts are present in *A:pnos::LUC*. It was not investigated which TS is used *in planta*. If the *CO* TS is used in both *A::LUC* and *A:pnos::LUC* the control fails to exclude a bias by the minimal promoter. This needs to be examined in the future to validate this control.

The subdivision of the *CO* promoter into four parts was useful to assess functions to each promoter part. On the other hand it fails to detect cross talk between elements as well as synergistic effects of promoter parts. In the future different parts of the promoter such as A and D which yield interesting mRNA patterns individually could be fused together. Element C on the other hand seems to only have a weak effect on expression levels and could be excluded to help define a minimal promoter for *CO*.

Investigating the expression conferred by the conserved short motifs (CDM, CMM and CPM) led to additional conclusions. A short motif stretch is sufficient to give expression that resembles endogenous *CO*. It was also shown that the LBSs did not contribute to expression but that a G-Box and LBS mutation combined were important for amplitude in this reporter system. This set of experiments suggests that short stretches, which are enriched in DOF binding sites, are sufficient to confer an expression pattern like *CO*. When the 2.6kb, 2.1kb and 1.8kb constructs were designed it was unclear whether these regions would contribute to activation or repression of transcription. The experiments performed suggest that DOF binding sites might also act as sites of activation by multiple use of the same binding site. Further experiments, like mutating all DOF binding sites in *CO* promoter, need to be performed. This would also underline the importance of these elements in the evolution of *CO* promoter, but is complicated by the high frequency with which this element occurs.

5.8.4. The model for *CO* promoter evolution

A mechanism by which the duplication of the *COL1/CO* ancestor brought *CO* mRNA levels under the control of DOF binding sites was suggested. This hypothesis is currently supported by data from

A. thaliana CO promoter studies. Additionally the influence of downstream regions from other subgroup Ia *COL* genes on their expression patterns can be tested in reporter assays. Alternatively reporter constructs available can be expanded by fusions to the 3´end of the cDNA. By such approaches the influence of downstream regions of *AtCOL1* or *AtCO* could be tested in the context of different promoters. This would also help to define the role of this conserved stretch in the downstream region of subgroup Ia *COL* genes.

This study provides the first data on a detailed analysis of *CO* promoter and more work is necessary to fully understand the transcriptional regulation of this important flowering time regulator. An evolutionary approach was taken which proved powerful as *CO* was the target of recent duplication and sub functionalization events. More functional analysis is necessary to complete our understanding of the *CO* promoter. For example all *cis*-elements could be mutated in the context of the promoter to understand their effect on *CO* mRNA regulation. Additionally new interactors could be identified by Y1H to isolate for example specific TCPs that regulate *CO*. A biological role for newly identified TFs can later be examined by analysis of mutant or overexpressor lines. EMSA and ChIP experiments can be performed to confirm binding. The *CO* promoter was also found to be regulated by a distal conserved site. In order to investigate how transcriptional regulation is facilitated by this part of the promoter a physical interaction with the TS could in the future be identified by tested by using methods such as chromosome confirmation capture (Dekker et al., 2002).

6. Regulation of *CO* transcription by CDFs is a novelty that arose in the Brassicaceae

CO transcription is regulated by a class of redundant Transcription factors called *CYCLING DNA BINDING WITH ONE FINGER* (*CDF*) (Imaizumi et al., 2005). Four members, namely *CDF1*, *CDF2*, *CDF3* and *CDF5*, of this subfamily of DOF TFs were identified as major regulators of *CO* by repressing its transcription (Fornara et al., 2009). Mutations in CDFs lead to upregulation of *CO* as well as *FT* mRNA causing the higher order *cdf1-R*, *cdf2-1*, *cdf3-1* and *cdf5-1* (*cdf1235*) quadruple mutant to flower early in both LD and SD conditions (Fornara et al., 2009). CDF1 binds to *CO* promoter at at least two positions as well as to *FT* promoter (Sawa et al., 2007; Song et al., 2012). Additionally, the increase in DOF binding sites in *CO* promoter is suggested to be an important evolutionary change that discriminates it from other subgroup Ia *COL* genes (Chapter 4). The aim of this study is to examine subgroup Ia *COL* mRNA regulation in the absence of CDFs in *A. thaliana* and make comparisons to other subgroup Ia *COL* genes in *A. arabicum* and *T. hassleriana*.

6.1. CDF1 and CDF2 binding to the *A. thaliana CO* promoter could not be detected *in planta*

Even though four members of the CDF family were shown to regulate *CO* transcription only CDF1 was confirmed to physically interact with the *CO* promoter *in planta* (Sawa et al., 2007). Chromatin Immunoprecipitation (ChIP) experiments were performed to confirm binding of CDF1 to *CO* promoter using an epitope tagged version of CDF1 driven by its own promoter (*pCDF1::3xHA:CDF1*, Sawa et al., 2007). Additionally the binding of CDF2 to *CO* locus was tested using a *pCDF2::3xHA:CDF2* construct introduced into the *cdf1235* mutant. An established line of epitope tagged and overexpressed TERMINAL FLOWER2 (TFL2) protein (*35S::TFL2:HA*), which was shown to associate with *FT* locus was used as a positive control (Turck et al., 2007).

For all ChIP experiments three binding regions were tested by qPCR. The primers *20.1* and primer *M3* cover conserved proximal (CPM) and distal (CDM) binding sites in the *CO* promoter and overlap with primers that gave positive results in ChIP experiments (Fig. 24a, Sawa et al., 2007). The third primer pair was located in the proximal promoter part of the *FT* promoter (*FT13*), shown to be bound by CDF1 (Song et al., 2012). Enrichment was plotted as the ratio of DNA in the sample versus the input for all experiments shown. Binding of TFL2:HA to *FT* promoter was tested to confirm that the ChIP experiments worked technically and was successfully detected (Fig. 24c,d). No or only mild enrichment compared to primer set *FT13* was observed for TFL2 on *CO* promoter (Fig. 24c,d), which is in agreement with the finding that *CO* locus is neither a TFL2 (Turck et al., 2007, MPIPZ genome browser) nor H3K27 trimethylation target (Zhang et al., 2007).

No enrichment was detected for 3xHA:CDF1 for all three primers tested compared to negative control (Col-0) and no antibody control (Fig. 24c), indicating no binding of CDF1 to the *CO* or the *FT* promoter could be detected.

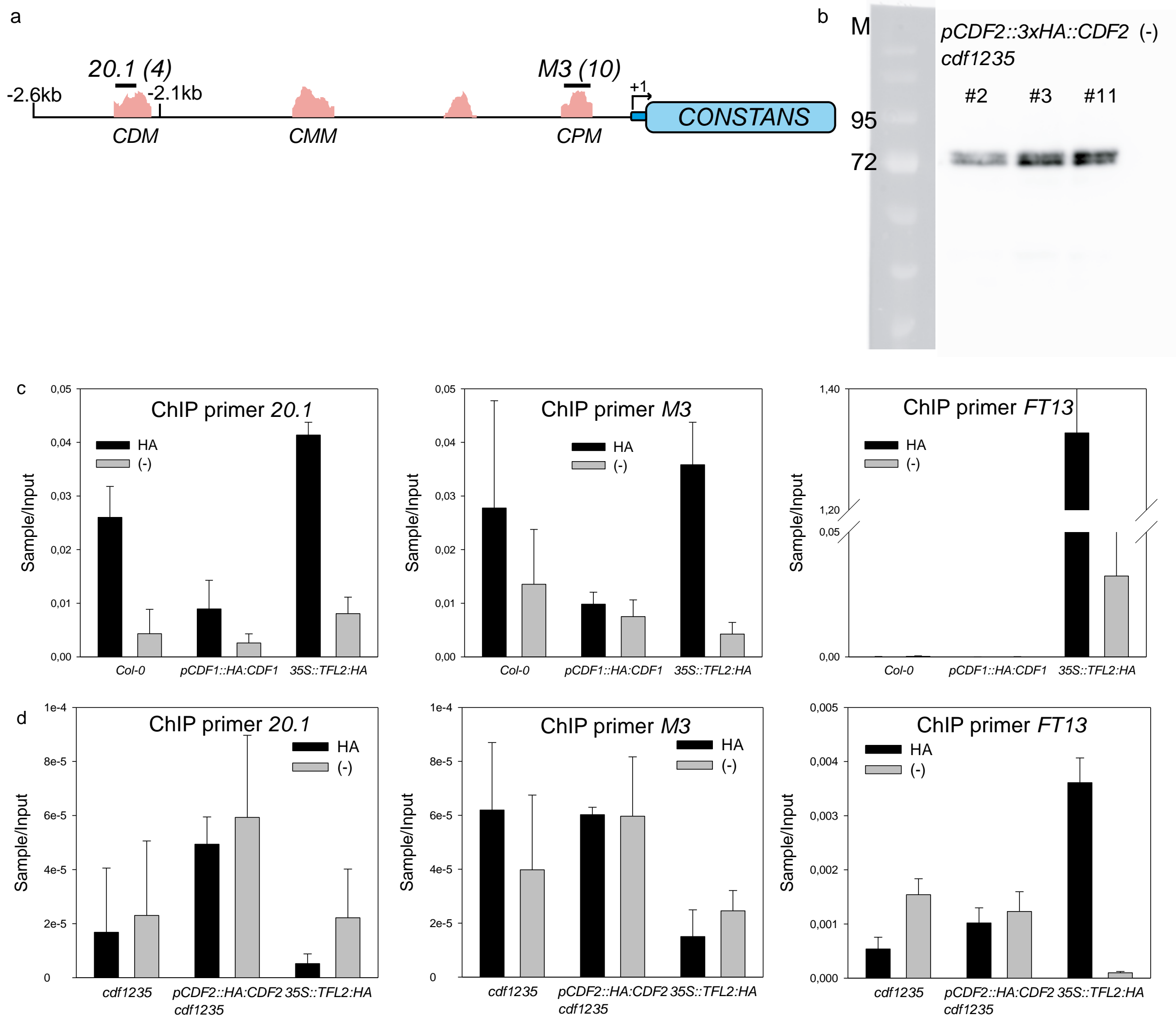


Figure 24: Chromatin Immunoprecipitation experiments demonstrate no binding of HA:CDF1 and HA:CDF2 to the *CO* and the *FT* promoter

Schematic representation of primer location in *CO* locus (a) with primer location, (4) and (10) indicate amplicon number (Sawa et al., 2007). Light red color visualizes DNA conservation compared to the *A. arabicum CO* promoter. The approximate positions of Conserved Proximal Motif (*CPM*), Conserved Middle Motif (*CMM*) and Conserved Distal Motif (*CDM*) as described in Fig. 20 are indicated. Western blot (b) of nuclear extracts from seedlings expressing *pCDF2::3xHA:CDF2* used for ChIP experiments. Three independent transgenic lines (#2, #3 and #11) in the *cdf1235* mutant were tested using α-HA antibody, *cdf1235* (-) mutant served as negative control. M=marker, size indication in kDa. ChIP experiment (c) with *pCDF1::3xHA:CDF1*. Antibody treated sample (α-HA) is compared to no antibody control (-) for transgenic line (*pCDF1::HA:CDF1*), negative control (Col-0) and positive control for *FT13* (*35S::TFL2:HA*). Binding is checked to two primers in the *CO* promoter (*20.1* and *M3*, a) as well as one in the *FT* promoter (*FT13,* Song et al., 2012). Representative experiment of two biological replicates is displayed. Ratio of Sample to input is shown with error bars representing the standard deviation of three technical replicates from the same qPCR. ChIP experiment performed with *pCDF2::3xHA:CDF2* (#2) (d) in *cdf1235*as in c, but and *cdf1235* as negative control.

Additionally *pCDF2::3xHA:CDF2* lines were created to test for binding of CDF2 to the *CO* and *FT* promoter. The construct was transformed into *cdf1235* quadruple mutant and rescued the early flowering phenotype of the mutant (data not shown). All three independent transgenic lines tested

showed similar levels of CDF2 protein in nuclear extracts used for ChIP (Fig. 24b). Two bands were detected in Western blot analysis around 72 kDa while no band was detected for the *cdf1235* negative control. The size of the detected proteins is in agreement with previous experiments and two bands can be explained by a previous finding that CDF2 protein is probably a target of phosphorylation (Simon, 2009). Reproducible results were observed for ChIP in all three *pCDF2::HA:CDF2* lines tested and only one representative line and experiment is shown (Fig. 24d): No enrichment was observed for ChIP of 3xHA::CDF2 to the *CO* or the *FT* promoter compared to no antibody and *cdf1235* negative control, respectively. In summary binding of CDF1 and CDF2 to *CO* and *FT* loci could not be detected in the conditions tested.

6.2. CDF transcription factors regulate *CO* but not *COL1* and *COL2* in *A. thaliana*

CDFs are important transcriptional regulators of *CO*, but *CO* mRNA regulation evolved in the Brassicaceae and increase of DOF binding sites was suggested as a major change discriminating the *CO* promoter from those of its close homologs, *AtCOL1* and *AtCOL2* (Chapter 5). Hence the analysis was expanded to test the influence of DOFs on the other subgroup Ia *COL* genes in *A. thaliana*. The expression of *COL1* and *COL2* in the *cdf1235* mutant was tested in LD and SD conditions (Fig. 25). LD data is only available from ZT0-ZT12. In LD conditions *CO* expression is higher in the *cdf1235* mutant compared to Col-0 control (Fig. 25a) for all timepoints, as shown previously (Fornara et al., 2009). By contrast the expression of *AtCOL1* (Fig. 25c) and *AtCOL2* (Fig 25e) did not change in the quadruple mutant. In SD conditions *CO* expression was higher in the *cdf1235* mutant compared to Col-0 (Fig 25b, Fornara et al., 2009). Additionally changes in mRNA pattern were observed in the *cdf1235* mutant compared to Col-0. *CO* mRNA in Col-0 in SD is very low between ZT0 and ZT6 and only expressed in the dark hours of the day (ZT9-ZT21). The mRNA levels in the *cdf1235* on the other hand displayed highest expression at ZT0 and ZT21 and low expression at all other timepoints (ZT3 - ZT15) comparable to morning peaks observed for *AtCOL1* and *AtCOL2*. As in LD no differences between Col-0 and *cdf1235* were observed for *AtCOL1* and *AtCOL2* expression in SD. At ZT9 in SD (Fig. 25d) expression of *AtCOL1* in Col-0 is higher compared to ZT6. This was not observed for other experiments in SD and might represent a technical problem with this replicate. The expression analysis presented strongly suggests that *AtCOL1* and *AtCOL2* genes are not under transcriptional control of *CDF1, CDF2, CDF3* or *CDF5*.

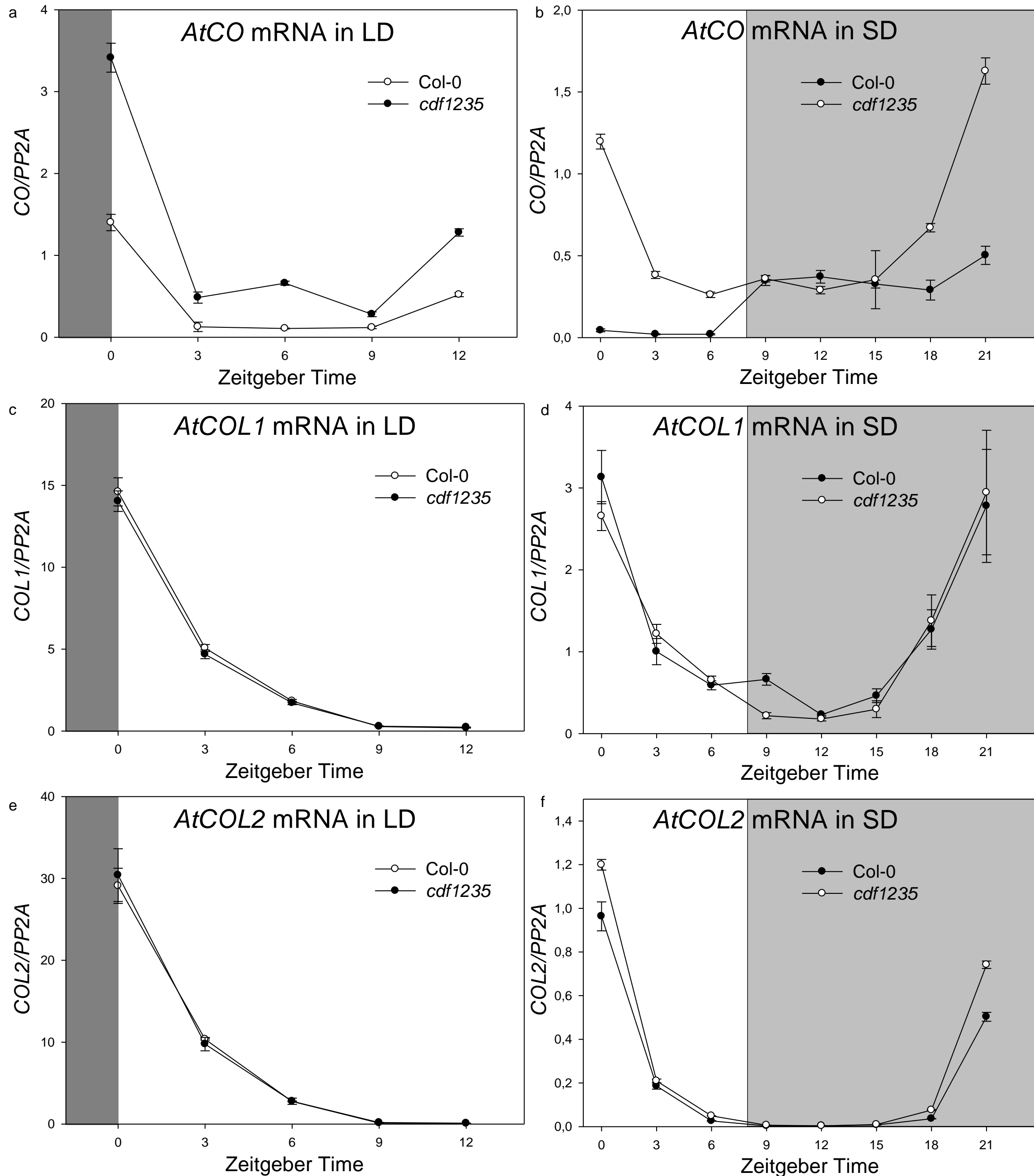


Figure 25: Expression of *AtCO* but not *AtCOL1* and *AtCOL2* is altered in the *cdf1235* quadruple mutant in LD and SD.

Expression of *AtCO* (a,b), *AtCOL1* (c,d) and *AtCOL2* (e,f) normalized to *PP2A* in ten day old seedling entrained in LD (a,c,e) or SD (b,d,f) conditions. The mean of three technical replicates of the same cDNA is displayed with error bars indicating standard deviation. LD data is displayed for ZT0-ZT12, SD data for ZT0-ZT21. White areas indicate light periods while Grey areas indicate dark periods.

6.3. Light downregulates *CO* mRNA in the morning in the absence of CDF Transcription factors

CO mRNA levels are elevated in the *cdf1235* mutant compared to wild type (Fig. 25, Fornara et al., 2009). Efficient downregulation of *CO* mRNA upon lights on is impaired, which contributes strongly to the early flowering phenotype of the *cdf1235* mutant. Thus the regulation of *CO* mRNA by light at

the beginning of the day was examined. *CO* mRNA was tested in both Col-0 and *cdf1235* in LD and dark conditions. On the day of harvest plants were transferred to dark before lights were switched on and mRNA levels were compared to light control. The *CO* mRNA levels were higher in the *cdf1235* mutant compared to wild type at all timepoints (Fig 26a). The shift to darkness did not change *CO* mRNA in Col-0 samples compared to light control. *CO* mRNA was downregulated efficiently and stayed low until ZT12 in Col-0. A slight upregulation in LD compared to dark was observed at ZT12. In the *cdf1235* mutant *CO* mRNA levels were different in light and dark treated samples. *CO* expression was found to be higher in the dark in all timepoints observed compared to LD control (Fig 26a). This experiment suggests that exposure to light downregulates *CO* mRNA in the morning in the *cdf1235* mutant.

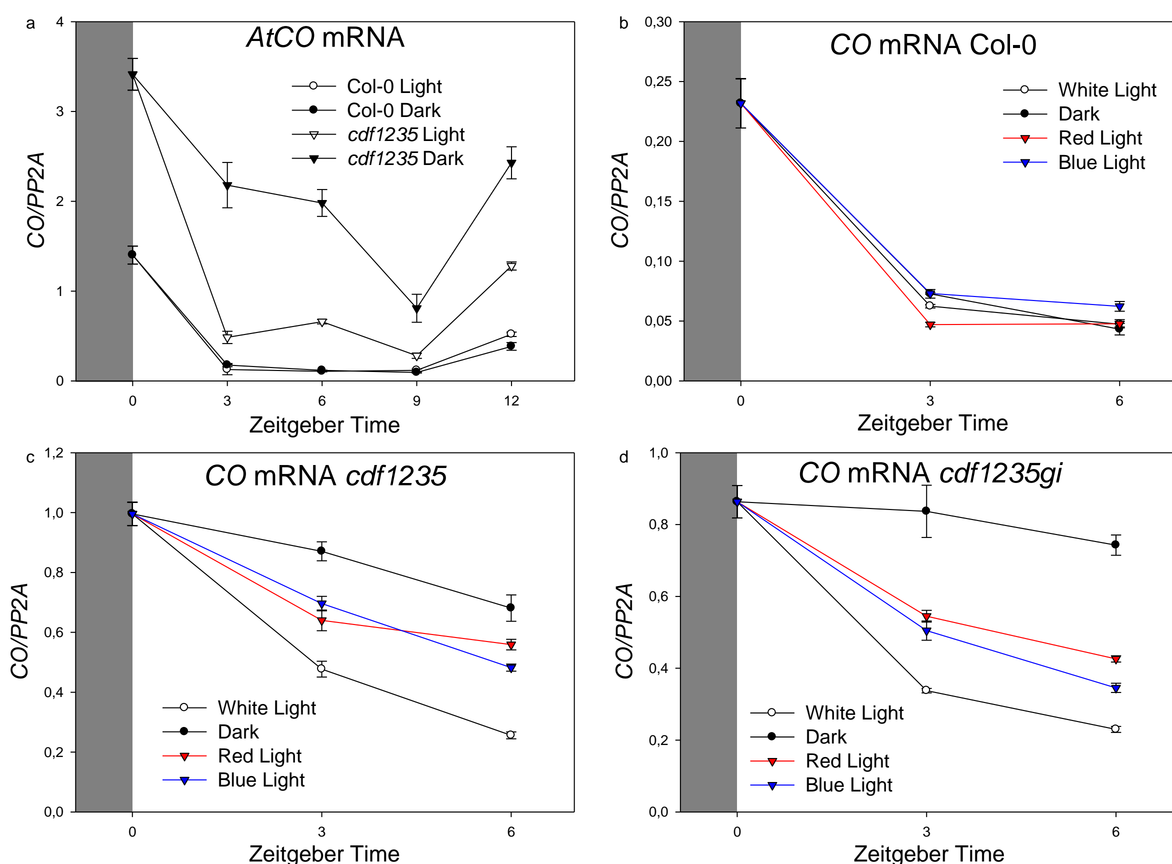


Figure 26: White, red and blue light supress *CO* mRNA in *cdf1235* and *cdf1235gi-100* mutants in the morning.
a) *CO* mRNA of ten day old seedling in LD conditions in Col-0 and *cdf1235*. Plants were either kept in Light (LD) or transferred to darkness before ZT0. Effect of white light, darkness, red light and blue light (LED) on *CO* mRNA in Col-0 (b), *cdf1235* (c) and *cdf1235gi-100* (*cdf1235gi*) (d) was tested. The mean of three technical replicates of the same cDNA is displayed with error bars indicating standard deviation. Grey areas specify dark hours of the day.

6.4. Light shapes the expression patterns of *AtCOL1* and *AtCOL2* as well as *ThCOL* in the morning

The effect of light on *CO* mRNA is detectable in the *cdf1235* mutant but not in Col-0. Even though *AtCOL1* and *AtCOL2* are not regulated by CDFs both might still share the dark phenotype observed. To test this hypothesis the effect of light versus darkness on mRNA levels of subgroup Ia *COL* genes

in the morning was examined. If the downregulation of mRNA levels by light in the morning is conserved between subgroup Ia *COL* genes it should be possible to detect an effect in wild type seedlings as *AtCOL1* and *AtCOL2* are not regulated by the CDFs, which mask this effect in *AtCO*. Therefore, *AtCOL1* and *AtCOL2* mRNA levels were determined in a dark shift experiment (Fig. 27a,b). Both genes were expressed differentially in the dark compared to light at the beginning of the day. In both LD and SD conditions the expression of *AtCOL1* and *AtCOL2* was higher at ZT0 compared to ZT3 (Fig. 25). In the dark shifted samples in both Col-0 and *cdf1235* the mRNA expression of *AtCOL1* and *AtCOL2* reached a peak at ZT3 instead of ZT0 (Fig 27a,b). This indicates that light mediated downregulation of mRNA levels in the morning occurs for all three subgroup Ia *COL* genes in *A. thaliana*.

Expression patterns of subgroup Ia *COL* genes are conserved in the Brassicaceae. It was therefore of interest whether the downregulation of mRNA levels in the morning is a conserved feature of their transcriptional regulation. Hence, the expression of *A. arabicum CO* and *A. arabicum COL2* was tested in dark shift experiments both in LD and SD conditions. In this experiment the effect of light on *AeCO* mRNA was reversed compared to *A. thaliana* genes tested (Fig 27c,d): the expression levels of *AeCO* mRNA in LD and SD conditions were lower in the dark samples compared to light samples. Only one biological replicate of each experiment was performed, so it remains to be tested whether this finding can be confirmed. *AeCOL2* expression exhibited a response to dark shift. While *AeCOL2* expression was decreasing from ZT0 to ZT6 in the light an increase in expression was observed in the dark treated sample (Fig 27e). The effect was even stronger in SD conditions (Fig. 27f). At ZT3 the difference between light and dark treated sample appeared to be ~10fold. The effect of light on mRNA levels of the only subgroup Ia *COL* gene in *T. hassleriana*, *ThCOL* was also tested. A clear elevation of *ThCOL* mRNA levels in dark versus light treated samples was observed in seedlings grown in LD conditions (Fig. 27g). In contrast to findings in *A. thaliana* and *A. arabicum COL* gene expression was higher at ZT6 compared to ZT3 in the dark treated samples in LD. In SD on the other hand expression of *ThCOL* (Fig. 27h) was very similar to *AtCOL1* and *AtCOL2* as well as *AeCOL2* with higher mRNA levels observed between dark and light at both ZT3 and ZT6. These experiments suggest that light at dawn and early morning represses subgroup Ia *COL* genes to shape the morning peak expression observed. Additionally this regulation is conserved in the subgroup Ia *COL* genes in *A. thaliana, A. arabicum* and *T. hassleriana*.

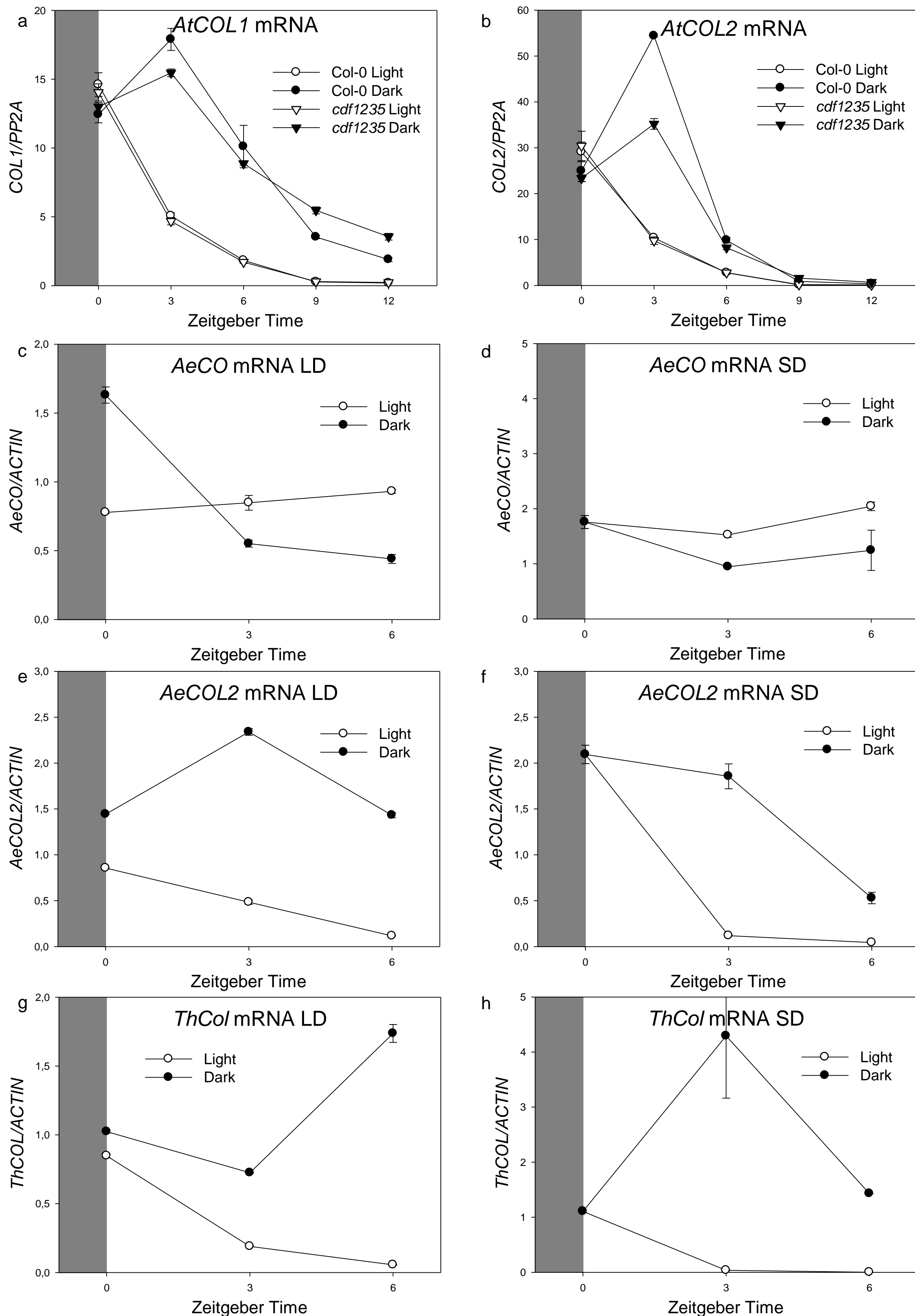


Figure 27: Light repression of mRNA is conserved in subgroup 1a *COL* genes from *A. thaliana*, *A. arabicum* and *T. hassleriana*

Expression *AtCOL1* (a) and *AtCOL2* (b) expression in LD grown seedlings either kept in white light or shifted to dark in Col-0 and *cdf1235* mutant background. The mean of three technical replicates of the same cDNA normalized to *PP2A* is displayed with error bars indicating standard deviation. *AeCO* (c,d), *AeCOL2* (e,f) and *ThCOL* (g,h) expression in LD (c,e,g) and SD (d,f,h) as well as the respective shifts to dark. *A. arabicum* and *T. hassleriana* experiments were performed once; expression is normalized to *ACTIN* gene. Error bars show standard deviation from at least three technical replicates of the same cDNA. All experiments were performed on soil. Grey areas specify dark hours of the day.

6.5. The mRNA of *AtCO, AtCOL1* and *AtCOL2* is highly unstable

All subgroup Ia *COL* transcript levels tested are regulated by light in the morning (except *AeCO*). It is possible that this effect is not associated with transcriptional regulation but a result of mRNA stability. To examine a possible effect of mRNA stability the influence of the chemical compound Cordycepin on *A. thaliana CO*, *COL1* and *COL2* mRNA was checked. Cordycepin is an adenosine derivate. The incorporation of this 3´-Deoxyadenosine by RNA Pol II into an actively synthesized mRNA leads to the termination of transcription. By this mechanism Cordycepin blocks mRNA synthesis reversibly in all cells treated with the compound (Siev et al., 1969) and it was previously shown to work in *A. thaliana* (Gutiérrez et al., 2002, Yakir et al., 2007).

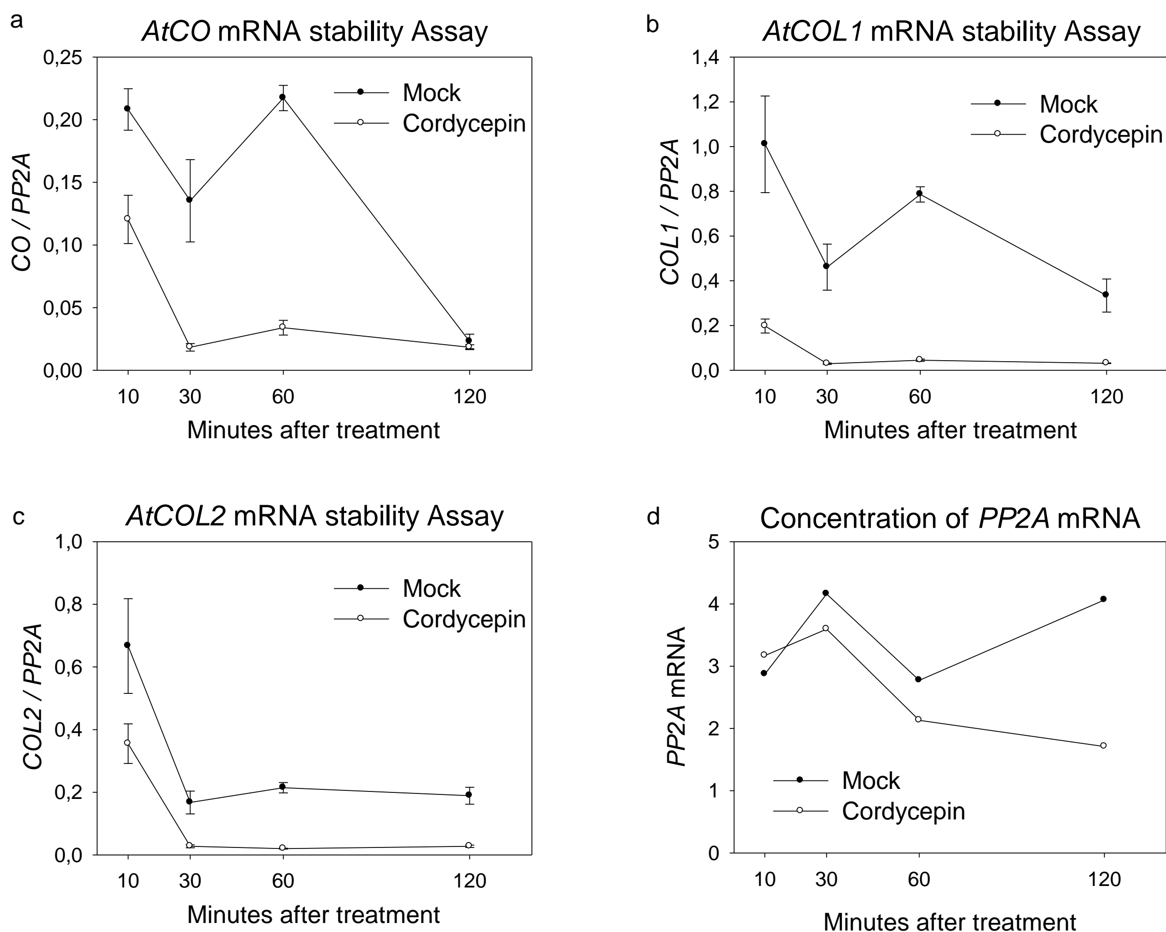


Figure 28: mRNA stability assay of *A. thaliana* subgroup Ia *COL* genes

Transcription was blocked by treatment with cordycepin in 10 day old *A. thaliana* Col-0 seedlings at ZT24 before lights on. mRNA levels of *AtCO* (a), *AtCOL1* (b) and *AtCOL2* (c) were determined 10, 30, 60 and 120min after treatment. cDNA concentrations were normalized to *PP2A*. Concentrations of PP2A (d) normalized to an arbitrary standard curve is shown as negative control. The mean of three technical replicates of the same cDNA is displayed with error bars indicating standard deviation.

In this experiment the housekeeping gene *PP2A* mRNA levels was used as a negative control (Fig. 28d). *PP2A* levels were normalized against an arbitrary standard curve (dilution series of a highly concentrated cDNA sample generated from the same material) and plotted to exclude major effects on

the stability of this mRNA species. In all samples (Cordycepin and mock treated) the concentration of *PP2A* varied in the range of two fold (between two and four arbitrary units; Fig. 28d). The biggest difference (two-fold) was exhibited after 120min of treatment. Thus *PP2A* might have a relatively stable mRNA (Gutiérrez et al., 2002) and is a good control for this experiment.

AtCO, *AtCOL1* and *AtCOL2* mRNA levels were determined (Fig. 28a-c) and normalized against *PP2A* mRNA. Analysis of the subgroup Ia *COL* mRNA levels revealed that all three mRNA species are unstable. The normalization showed between 1.5 and 10 fold reductions in mRNA already after 10min. After 30min most of the mRNA was degraded. This experiment was performed on day 10 starting at ZT24 in the dark. Transcripts were previously assumed to be "unstable" when presenting a half-life of less than 60min (Gutiérrez et al., 2002). At this timepoint for *AtCO* a block of transcription by CDF1, CDF2, CDF3 and CDF5 was predicted by Fornara et al. (2009) meaning that mRNA levels of *AtCO* would already be reduced over time. The effect of the Cordycepin treatment was much more dramatic, leaving the possibility that *AtCO* mRNA is degraded quickly without continuous *de novo* synthesis. In general the same is true for *AtCOL1* and *AtCOL2*. As described in Chapter 4.3 the amount of *AtCOL1* and *AtCOL2* mRNA had its highest levels around ZT0 and decreased in the first hours of a day. The instability of the mRNA species suggested that the downregulation of *AtCO* mRNA in the morning could be established by blocking transcription of *AtCO* while the slow decline of *AtCOL1* and *AtCOL2* mRNA likely involves active transcription, otherwise mRNA levels would decrease at a faster rate than observed in Fig. 28.

6.6. Red and blue light are partially responsible for downregulating *CO* mRNA in the morning in the absence of CDF Transcription factors

Downregulation of *CO* mRNA in *cdf1235* was mediated by white light. It was thus examined which light qualities mediate the transcriptional repression of *CO* at the beginning of the day. Plants were grown in LD and either shifted to dark, blue light, red light or kept in white light. The mRNA levels of *CO* did not change in Col-0 in all light regimes tested, confirming earlier observations in white light (Fig. 26b). In the *cdf1235* mutant *CO* mRNA levels were more than two fold higher in the dark shifted seedlings compared to light shifted seedlings at ZT3 and ZT6 (Fig. 26a). Both red light and blue light treatment resulted in intermediate phenotypes: in both conditions *CO* mRNA levels were higher compared to white light but lower compared to dark treated samples. No difference between blue and red light was observed.

GI is involved in the blue light mediated degradation of CDFs in the evening in long day conditions. It is hence possible that GI regulates light mediated *CO* mRNA regulation at other times of the day. It was thus tested whether the *gi-100* mutation influences *CO* repression in the morning. In the *cdf1235 gi-100* (*cdf1235gi*) quintuple mutant *CO* expression is higher compared to Col-0 but lower compared to *cdf1235* alone (Fornara et al., 2009). Higher levels of *CO* mRNA in dark treated seedlings compared to light treated seedlings was observed in the *cdf1235 gi* quintuple mutant, comparable to

the *cdf1235* mutant (Fig 26d). Also exposure to red and blue light lead to intermediate phenotypes in the quintuple mutant. The *gi* mutant was not tested on its own as it is known to exhibit very low *CO* mRNA levels already (Fornara et al., 2009; Suarez-Lopez et al., 2001). Thus light downregulation of *CO* expression in the *cdf1235* mutant was mediated by both blue and red light and was independent of GI activity.

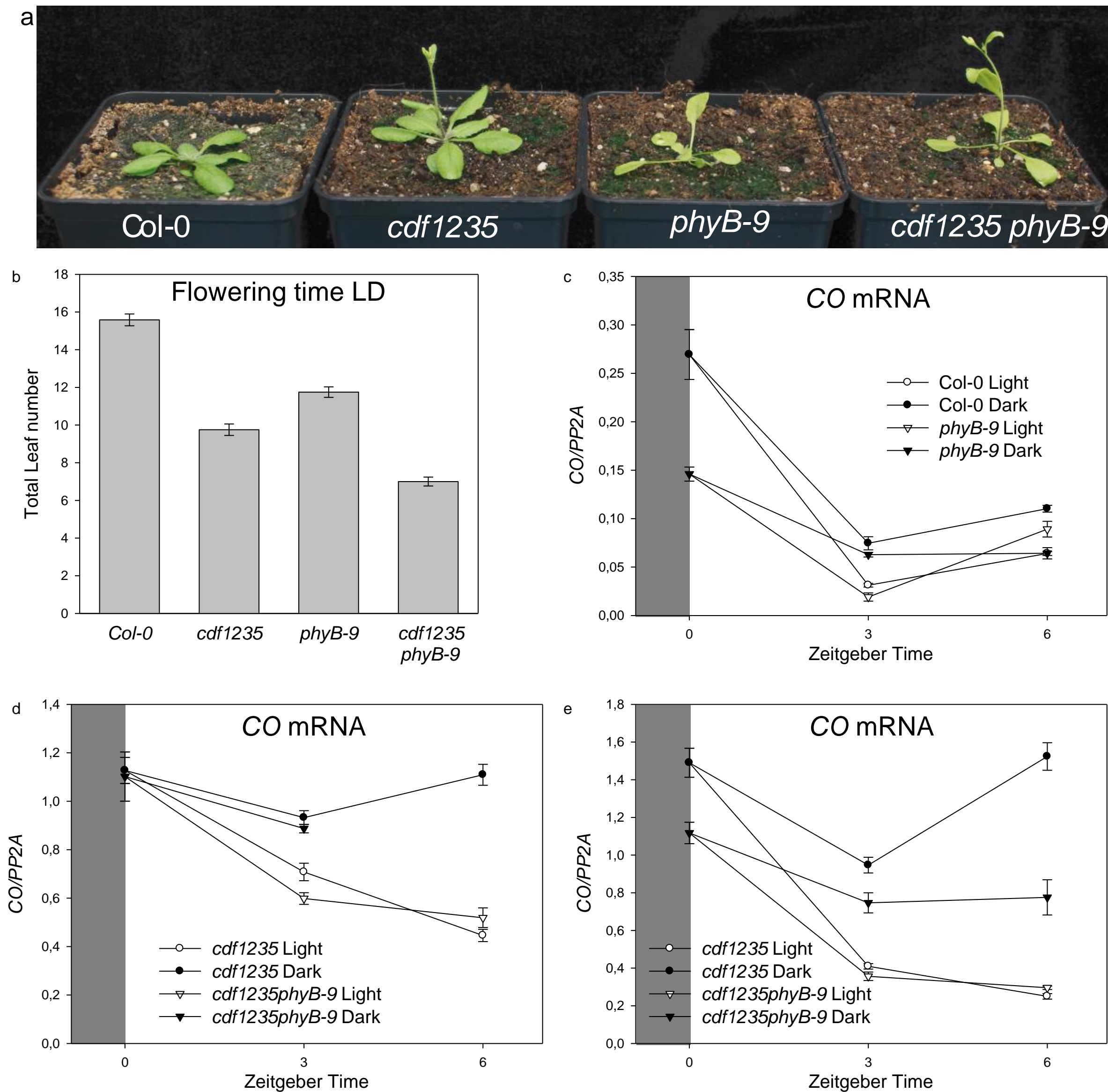


Figure 29: The red light photoreceptor phyB is not responsible for light dependent *CO* mRNA repression in the morning in the *cdf1235* mutant
Picture of Col-0, *cdf1235*, *phyB-9* and *cdf1235phyB-9* plants grown in LD conditions for ~14days (a). Mean flowering time (b) in LD conditions as total leaf number (n=12, except *cdf1235phyB-9* n=9), error bars display standard error. mRNA levels of *CO* normalized to *PP2A* in ten day old seedling grown in LD of Col-0 and *phyB-9* (c) as well as *cdf1235* and *cdf1235phyB-9* (d,e) in light and dark. The mean of three technical replicates of the same cDNA is displayed with error bars indicating standard deviation. Error bars show standard deviation from at least three technical replicates of the same cDNA. Grey areas specify dark hours of the day.

phyB is the major photoreceptor in red light and was already suggested to be involved in *CO* mRNA regulation (Aukerman et al., 1997; Cerdán and Chory, 2003; Robson et al., 1993). To identify a role for red light in *CO* mRNA regulation in the morning and upstream signalling the *phyB-9* null mutation

was introduced into the *cdf1235* (Reed et al., 1993). Both *phyB-9* and *cdf1235* mutant plants exhibit early flowering phenotypes when compared to Col-0 (Fig. 29a,b). The flowering time of the *cdf1235phyB-9* mutant was tested in LD conditions. The effects on flowering time were found to be additive as the *cdf1235phyB-9* mutant flowered ~6 leaves earlier compared to the *cdf1235* quadruple mutant (~10 leaves) and *phyB-9* mutant (~12 leaves), respectively (Fig. 29b). *CO* expression was tested in the *phyB-9* mutant compared to Col-0. This experiment showed lower *CO* levels at ZT0 in *phyB-9* mutant but no obvious differences between dark and white light (Fig. 29c). Testing the *cdf1235* mutant versus *cdf1235phyB* led to conflicting results. In the first experiment no difference between genotypes was observed in dark or white light (Fig. 29d). The second experiment showed slightly lower *CO* mRNA levels in the *cdf1235phyB* mutant compared to *cdf1235* control after shift to darkness (Fig. 29e). *CO* mRNA was downregulated in the morning in the *cdf1235phyB* mutant compared to *cdf1235* in both experiments. The red light photoreceptor PhyB is most likely therefore not solely responsible for the downregulation of *CO* mRNA upon exposure to white light in the morning independent of the CDFs.

6.7. Discussion Chapter 6

6.7.1. ChIP of CDF1 and CDF2 to *CO* and *FT* locus did not detect binding

Unexpectedly the ChIP experiments which show binding of CDF1 to the *CO* and *FT* loci could not be repeated (Sawa et al., 2007; Song et al., 2012). The experiments shown follow closely the method of the published experiment after communication with the Imaizumi lab (Franziska Turck, MPIPZ, Cologne, personal communication). In total three different protocols were tested (material and methods), which all worked for the positive control (*35S::TFL2:HA* or Histone3, not shown) but failed to show enrichment of CDF1 and CDF2 to *CO* locus. Different HA antibodies were used to perform the immunoprecipitation step (material and methods). The native CDF2 antibody (Fornara et al., 2009) was also tested and no enrichment was found (data not shown). The following general technical difficulties were encountered in the ChIP experiments presented: the quality of antibodies (e.g. native CDF2 antibody) as well as of commercial batches varies which can influence the immunoprecipitation step leading to high background levels. The purity and low amounts of DNA determined by qPCR can lead to technical errors that lead to variability and prevent detection of enrichment. These problems can be solved by using a tag that can be more reliably immunoprecipitated. A GFP fusion of CDF2 driven by its native promoter (*pCDF2::GFP:CDF2*) was constructed and transformed in the Col-0 background. This construct is at the T_1 stage and can be used for ChIP experiments using a GFP antibody, which has proven more effective for ChIP with other TFs (Gregis et al., 2013; Kaufmann et al., 2010).

The CDFs are mainly expressed in the leaf vasculature where they repress *CO* expression (Fornara et al., 2009). Thus the expression domains of CDF1 and CDF2 proteins under their native promoters are mainly this tissue. This might dilute enrichment by ChIP as DNA from other cell types is

overrepresented but does not give a ChIP signal, this was already suggested to cause problems in performing ChIP of CO protein on *FT* locus (Franziska Turck and Lyangu Liu, MPIPZ, Cologne, personal communication). Enrichment of target cells by Fluorescence-activated cell sorting (FACS, Iyer-Pascuzzi and Benfey, 2010) could lead to detection of CDF binding to *CO* locus. Additionally future technical advances might lead to easier identification of *in vivo* binding of proteins to DNA.

The experiments presented suggest that CDF1 and CDF2 do not bind to the *CO* or the *FT* promoter *in planta*. By contrast direct binding of CDF1 to the *CO* promoter was detected by ChIP experiments and also in EMSA studies and transient *LUC* activation assays (Imaizumi et al., 2005; Sawa et al., 2007). The finding of an increase in DOF binding sites during the evolution of the *CO* promoter and its conservation in the Brassicaceae family (Chapter 5) additionally supports direct binding of CDFs to *CO* locus. The alternative possibility that CDF regulation of *CO* and *FT* is indirect could be tested by screening direct targets of CDFs by ChIP-seq, and using these as positive controls in the analysis of *CO*.

6.7.2. Regulation of mRNA levels by CDFs is not conserved between subgroup Ia *COL* genes

AtCOL1 and *AtCOL2* in contrast to *AtCO* were not misregulated in the *cdf1235* mutant compared to Col-0 wild type. This was unexpected as conserved DOF binding sites were found downstream in close proximity of the coding regions of both genes (Fig. 23). Only two conditions (LD and SD) were tested and it is possible that the CDFs regulate *AtCOL1* and *AtCOL2* in other conditions such as high or low temperature. Additionally no CDF overexpression line was tested. Moreover the influence of other DOF transcription factors on *AtCOL1* and *AtCOL2* mRNA levels could be tested as the family consists of 37 members (Fornara et al., 2009; Yanagisawa, 2002).

To test conservation of *CO* regulation by CDFs in *A. arabicum* new materials need to be created. *A. arabicum CDF1*, *CDF2*, *CDF3* and *CDF5* orthologs were identified by BLAST search (data not shown) and could be target using an RNAi approach, after a plant transformation assay is developed for this species. In *T. hassleriana* on the other hand a plant transformation protocol is available (Tsai et al., 2012). Thus an RNAi approach knocking down CDFs could be established in *T. hassleriana* after identification of CDF homologs in the recently published genome sequence (Cheng et al., 2013).

Taken together all experiments suggest that regulation of *CO* transcription by CDFs in the Brassicaceae is a mechanism that has evolved within the family and is specific to *CO* among *COL* Ia genes. Opposing findings come from Potato: *StCO1* was shown to be downregulated as a result of overexpression of a *CDF* homolog (StCDF1.2), suggesting that the regulation of subgroup Ia *COL* by *CDF* transcription factors might be conserved in many plant families (Kloosterman et al., 2013). It is also possible that other members of the DOF transcription factor family regulate genes such as *AtCOL1* and *AtCOL2* and that regulatory mechanism changed during evolution. Alternatively, the regulation of *AtCO* and *StCO1* by CDFs could be the result of convergent evolution.

6.7.3. Light downregulation of subgroup Ia *COL* genes is conserved and depends on light quality

The downregulation of *CO* transcription in the first hours of the day is crucial for its function as a day length sensor (Fornara et al., 2009). This downregulation is impaired in the *cdf1235* mutant. Transfer of *cdf1235* plants to darkness at dawn identified a CDF-independent mechanism that represses *CO* expression. This phenotype was not observed in the wt control, most likely because it is partly redundant with the activity of CDF TFs (Fornara et al., 2009).

The repression of *CO* mRNA was identified in the higher order *cdf1235* mutant but was clearly detected in wild type plants for the other subgroup Ia *COL* genes. This regulation was not only found in *A. thaliana* but also in *A. arabicum* and *T. hassleriana*. Light repression of *AeCO* mRNA in the morning was not detected in wt *A. arabicum* seedlings, probably because as conservation of *CO* temporal transcription pattern suggests, CDFs repress *AeCO* mRNA independently of light (Chapter 4 and 5). Loss of function mutants of CDFs in *A. arabicum* could confirm this hypothesis (see above).

AtCO, *AtCOL1* and *AtCOL2* mRNA was observed to be unstable when compared to *PP2A*. Hence without transcription, mRNA levels fall very rapidly. This experiment was only performed at the beginning of the day and stability kinetics could very well change over the course of a day. Further experiments could be performed to check for this possibility. Nonetheless these experiments suggest that subgroup Ia *COL* genes are actively transcribed in the morning so that mRNA levels do not fall rapidly and that light represses transcription resulting in reduction of mRNA levels as observed.

The dark shift phenotype was additionally tested in two light conditions (red and blue) and intermediate phenotypes were observed compared to white light in the *cdf1235* mutant as well as the *cdf1235gi* quintuple mutant. Thus both red and blue light and possibly light of other wavelengths contributes to this regulation in a so far uninvestigated way. Red and Blue light experiments were performed with LEDs. These LEDs provide less light when compared to white light provided by fluorescent tubes and bulbs. Hence, the effect of light intensity should be tested in different white light intensities in the future. The effect of the red light receptor phyB was tested in the context of the *phyB-9* mutant. Only mild or no differences were observed between *cdf1235* and *cdf1235phyB* quintuple mutant in white light. The levels of *CO* in the *phyB-9* mutant were slightly changed but without effect on the light phenotype. The same was true for the comparison of *cdf1235* and *cdf1235phyB*. A 24h timecourse in different day length regimes is necessary to examine the effect of *cdf1235phyB-9* mutation on *CO* mRNA in the conditions tested. Nonetheless the light downregulation is not mediated by *phyB* alone. It is still possible that the red light photoreceptors act redundantly and that other red light sensing phytochromes (PhyC, PhyD, PhyE) play an important role in the regulation of this pathway (Balasubramanian et al., 2006; Franklin et al., 2003). The effect of other light conditions on the phenotype can be tested in different photoreceptor mutants such as *cry1* and *cry2* for blue light.

The function of the dark shift phenotype in nature is unknown but it is of physiological importance for plants to keep track of the time of day by the circadian clock (Dodd et al., 2005). Plants are for example able to track dusk and dawn (Edwards et al., 2010). Thus clock genes such as *GI* exhibit a shift in peak expression between LD and SD (Fowler et al., 1999). A similar phenotype was observed for *AtCOL2*, *AeCOL2* as well as *ThCOL* (Fig. 6) as their expression rose earlier in SD compared to LD. In agreement with this finding mRNA was also reduced earlier in the morning in SD leading to a shift in peak time. The results presented suggest that light is an important signal for regulating all *COL* genes investigated. In *COL1* and *COL2* orthologs light might be used to track dawn in the morning and downregulate the mRNA upon this signal. In agreement with this overexpression of *AtCOL1* accelerates the circadian clock in a light fluence rate dependent manner (Ledger et al., 2001). This suggests that *AtCOL1* is implicated in or regulated by the circadian clock. Maybe this gene has a role in the light input of the circadian clock. Expression of *AtCOL1* was tested in DD and LL and the gene is misregulated in both conditions (Ledger et al., 2001). The peak of expression is shifted in DD by four hours to ZT4 and expression stays higher in constant darkness throughout the day, which is in agreement with the results presented here (Fig. 27). *AtCOL2* expression is higher in DD compared to LD around the first dawn but then attenuates in DD and could not be detected in the following dusk (ZT24, Ledger et al., 2001). This suggested differential regulation of *AtCOL1* and *AtCOL2* in DD, so that *AtCOL1* expression is present under these conditions while *AtCOL2* expression is absent. In summary, describing the so far unclear functions of *AtCOL1* and *AtCOL2* will help to assign a role for this dawn tracking mechanism (Kim et al., 2013; Ledger et al., 2001).

In *CO* orthologs the meaning of light regulated downregulation is not obvious as the phenotype is masked by CDF transcription factors in wild type. It can be speculated that this phenotype represents an evolutionarily ancient mechanism which regulated the ancestor of subgroup Ia *COL* genes in the Brassicaceae and Cleomaceae and was not lost in evolution of *CO*. Repression of *CO* mRNA in the morning by light is too leaky for *CO* to perform as a photoperiodic sensing module. Hence the regulation of *CO* mRNA by CDFs evolved as an additional layer to ensure *CO* mRNA does not overlap with light in the morning.

7. CONCLUSIONS AND PERSPECTIVES

CO is the major activator of *FT* in the photoperiodic flowering pathway in *A. thaliana*. In this study a systematic evolutionary approach was taken to investigate the evolution of both the function and the transcriptional regulation of *CO* in comparison to its closest homologs in *A. thaliana COL1* and *COL2*. The evolutionary history of these three genes is linked to their genomic location. *CO* and *COL1* are located in tandem on chromosome 5 while *COL2* is a single gene on chromosome 3. These two chromosomal regions arose from the latest whole genome duplication detected in *A. thaliana* (At-α). Phylogenetic analysis strongly suggests that the tandem duplication leading to occurrence of *CO* and *COL1* occurred after the At-α.

Conservation of *CO* function was checked by complementation of *co-10* mutant by *A. arabicum CO* ortholog. *A. arabicum*, an obligate long day plant is a member of the ancient *Aethionema* lineage in the Brassicaceae family, which split from the core Brassicaceae early in the evolution of the family. These experiments indicate that *CO* function and its role in photoperiodic flowering are conserved in the whole Brassicaceae family.

7.1. Insights into transcriptional regulation of *CO*

The transcriptional pattern of *CO* is unique and important for its function as a flowering time regulator. Key features of *CO* mRNA regulation in LD like rapid downregulation in the morning and peak levels of expression in evening and night were identified in expression analysis of *AeCO*. The close homologs of *CO* in *A. thaliana* namely *COL1* and *COL2* but also *AeCOL2* from *A. arabicum* are differentially expressed. Their mRNA levels are higher compared to *CO* and form a clear morning peak, which contradicts a putative function as a day length sensor.

The conservation of *CO* expression pattern is reflected by the conservation of regulatory sequences in the promoter of *CO* in all Brassicaceaen species investigated including the basal lineage *Aethionema*. The promoter of *CO* is partially redundant and consists of regulatory regions up to 2.6kb upstream from the TS important for mediating full flowering response. The highest conserved *cis*-elements identified in Brassicaceaen *CO* promoters are DOF binding sites. These DOF binding sites were found to be significantly enriched in the promoter of *CO* compared to the promoters of its close paralogs or to random sequences with identical base composition. Additionally they are functional in conveying *CO* mRNA expression pattern as parts of conserved regulatory modules, putatively shaping the temporal pattern of CO independently of other regulatory regions. In summary the increase in DOF binding sites at the *CO* locus shapes the expression of *CO* to function as a long day sensor. In the future a possible dual function of DOF binding sites in repression and activation of *CO* transcription could be examined.

If the function of CDFs is impaired *CO* expression shares additional features with *COL1* and *COL2*. In the higher order *cdf1235* mutant *CO* expression levels are higher in all timepoints but especially morning downregulation is impaired, consistent with the high expression levels of the *CDF* genes in the morning. Additionally light downregulates *CO* mRNA exclusively in the *cdf1235* mutant but not in wt. *COL1* and *COL2*, which are not regulated by CDFs, share this aspect, indicating that it might be ancient and shared between paralogs. The regulation of *CO* by CDFs thus appears to be an additional layer of regulation that evolved on top of existing and conserved regulatory mechanisms, which suggests a pattern of regulation of the last common ancestor of all three genes.

Two classes of *cis*-elements were previously shown to be involved in regulation of *CO* mRNA in *A. thaliana*, the DOF binding site and the E-Box (Imaizumi et al., 2005; Ito et al., 2012a; Sawa et al., 2007). This study additionally suggests further classes of *cis*-elements to be involved in *CO* mRNA regulation namely TCP binding sites and G-Boxes. A future goal is to identify the TFs associated with these putative regulatory regions.

Furthermore transcriptional regulation is not mediated by single TFs but by complexes associated to chromatin. It was shown already that CDF1 and its interactors FKF1 and GI co-occupy the same positions on the *CO* locus (Sawa et al., 2007). More experiments are needed not only to identify interaction of *cis*-elements in *CO* mRNA regulation but also how these interactions are mediated on the protein level. The interplay between regulatory factors is presumably much more complex and thus goes beyond the scope of this study.

7.2. A model for *CO* promoter evolution

A model for the evolution of the *CO* promoter was proposed with a focus of how DOF binding sites appeared as novel regulators of *CO* transcription. The tandem duplication that led to *CO* and *COL1* resulted in conserved CDF binding sites that were present in the downstream sequence of the ancestor to appear distally upstream of *CO* forming a conserved cluster referred to as Conserved Distal Motif (CDM). It can be imagined that this CDM served as a blueprint for the *cis*-element changes that occurred close to the TS leading to similar TF binding site composition. The ancestor of *CO* thus became sandwiched between DOF clusters after the tandem duplication. The detailed analysis of upstream regulatory regions also revealed that the promoter of *CO* is leaky in transgenic plants, leading to overcomplementation phenotypes observed previously and also in this study (Song et al., 2012). In the future an interaction of upstream and downstream regulatory regions could be examined, which might indicate a role for both in the regulation of *CO*. In general the physical interactions of chromatin at the *CO* locus between distant regulatory motifs might reveal new regulatory regions.

7.3. *CO* as a model for promoter evolution

CO presents a suitable model for studying the evolution of a promoter as its evolutionary split from *COL1* can be dated after the At-α WGD but before the split of the Aethionema lineage from the core-

Brassicaceae. Additionally *COL1* can be assumed to represent the ancient conserved state of expression while *CO* is the derived novel state. In future experiments the differences in regulation which were correlated with changes in *cis*-element composition in this study could be tested in promoter reporter studies to trace or even reconstruct the evolutionary steps leading to establishment of the promoter of *CO*.

7.4. Interplay of *CO* Function and Transcription

In contrast to previous reports *COL1* and *COL2* can replace the function of *CO* as discovered by overexpression from the *SUC2* promoter. However their function is weaker as both fail to fully complement the *co-10* mutant when expressed from the *CO* promoter. The functions of *COL1* and *COL2* are not clarified in *A. thaliana* yet, but all three genes are at least partially functionally redundant. Probably all three members regulate transcription and their distinction is caused by few differences in amino acid probably leading to changes in posttranslational modification as well as interaction partners. The finding that *COL1* and *COL2* can fully complement *CO* function when being highly expressed in the vasculature implies that high levels of COL1 and COL2 can overcome protein differences when compared to CO. Additionally it can be concluded that regulatory differences do not only affect temporal expression but inevitably must affect spatial expression, because high expression of *COL1* and *COL2* as observed in wt would unavoidably lead to early flowering if present in the expression domain of *CO*. In summary the higher expression levels of *COL1* and *COL2* probably result from a greater spatial expression domain compared to *CO*. Thus after determining mechanisms for temporal expression the mechanism mediating spatial expression of all three genes has to be investigated.

7.5. The last common ancestral gene of *CO*

All three subgroup Ia *COL* genes share one common ancestral gene with the only homolog present in *T. hassleriana*. Thus features that are shared between *ThCOL* and *AtCO* or *AtCOL1*/*AtCOL2* can be assigned as ancient while differences should be assigned as new or derived. *ThCOL* is morning expressed like *AtCOL1* and *AtCOL2* indicating that expression of *CO* mRNA evolved in the Brassicaceae and is evolutionarily young. This is supported by the finding that novelties like *cis*-element changes that occur in the Brassicaceae for *CO* are not shared with *ThCOL*. Functionally *ThCOL* behaved similarly to *AtCOL1* and *AtCOL2* in transgenic complementation assays. Thus it has to be stated that the strong flowering initiation mediated by *CO* homologs in the Brassicaceae is a novelty. The ancestor of *CO* only presented a moderate ability to promote flowering via the photoperiodic flowering pathway. This might suggest that the different functions of *AtCO*, *AtCOL1* and *AtCOL2* are all mediated by *ThCOL* and subfunctionalization only occurred in the Brassicaceae.

7.6. Function of *CO* orthologs beyond the Brassicales

CO has homologs in many plant species. There is additionally evidence that the *CO-FT* regulon is conserved in flowering plants and was employed to sense day length and regulate downstream mechanisms both in flowering time control and other mechanisms like tuberization. *A. thaliana CO* and Rice *HD1* expression is very similar but evolved through convergent evolution. This is not only the result of this study but furthermore supported by the finding that *CO* orthologs from many species that are evolutionarily placed between the Brassicaceae and the Poaceae show a classical morning expression pattern like *AtCOL1/AtCOL2*. *CO* is a strong transcriptional regulator in *A. thaliana*. The complementation experiments presented in this study could in future be used to identify the causal differences for the increase in function of true *CO* homologs in the Brassicaceae family compared to its paralogs as well as *CO* orthologs from other plant families.

In summary *CO* is a potent activator of photoperiodic flowering that evolved in the Brassicaceae probably during adaptation to favourable long day photoperiods. Importantly the transcriptional regulation of *CO* adapted to allow increased *CO* mRNA specifically in LDs and therefore activation of flowering under favourable conditions. These mechanisms might be linked to the spread of the Brassicaceae family into temperate climates allowing adaptation to strong differences between seasons.

LITERATURE

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, J., Yamaguchi, A., Ikeda, I., Ichinoki, H., Notaguchi, M., Goto, K., and Araki, T. (2005). FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex. Science *309*, 1052–1056.

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 Online *22*, 1425–1440.

Alberts, B., Johnson, A., Walter, P., Lewis, J., Raff, M., and Keith, R. (2008). Molecular biology of the cell (New York: Garland Science).

Almada, R., Cabrera, N., Casaretto, J.A., Ruiz-Lara, S., and González Villanueva, E. (2009). VvCO and VvCOL1, two CONSTANS homologous genes, are regulated during flower induction and dormancy in grapevine buds. Plant Cell Rep. *28*, 1193–1203.

An, H., Roussot, C., Suarez-Lopez, P., Corbesier, L., Vincent, C., Pinero, M., Hepworth, S.R., Mouradov, A., Justin, S., Turnbull, C., et al. (2004). CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis. Development *131*, 3615–3626.

Andrés, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. Nat. Rev. Genet. *13*, 627–639.

Ansell, S.W., Stenoien, H.K., Grundmann, M., Russell, S.J., Koch, M.A., Schneider, H., and Vogel, J.C. (2011). The importance of Anatolian mountains as the cradle of global diversity in Arabis alpina, a key arctic-alpine species. Ann. Bot. *108*, 241–252.

Aukerman, M.J., Hirschfeld, M., Wester, L., Weaver, M., Clack, T., Amasino, R.M., and Sharrock, R.A. (1997). A deletion in the PHYD gene of the Arabidopsis Wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing. Plant Cell Online *9*, 1317–1326.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. *2*, 28–36.

Bailey, C.D., Koch, M.A., Mayer, M., Mummenhoff, K., O'Kane Jr, S.L., 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.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208.

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

Barker, M.S., Vogel, H., and Schranz, M.E. (2009). Paleopolyploidy in the Brassicales: Analyses of the Cleome Transcriptome Elucidate the History of Genome Duplications in Arabidopsis and Other Brassicales. Genome Biol. Evol. *1*, 391–399.

Beckerman, R., and Prives, C. (2010). Transcriptional Regulation by P53. Cold Spring Harb. Perspect. Biol. *2*, a000935.

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. *107*, 18724–18728.

Berns, M.C. (2012). Transcriptional regulation of the Arabidopsis thaliana flowering-time gene GIGANTEA.

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.

Bohlenius, H., Huang, T., Charbonnel-Campaa, L., Brunner, A.M., Jansson, S., Strauss, S.H., and Nilsson, O. (2006). CO/FT Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. Science *312*, 1040–1043.

Borner, R., Kampmann, G., Chandler, J., Gleissner, R., Wisman, E., Apel, K., and Melzer, S. (2000). A MADS domain gene involved in the transition to flowering in Arabidopsis. Plant J. Cell Mol. Biol. *24*, 591–599.

Bowers, J.E., Chapman, B.A., Rong, J., and Paterson, A.H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature *422*, 428–433.

Bowler, C., Benvenuto, G., Laflamme, P., Molino, D., Probst, A.V., Tariq, M., and Paszkowski, J. (2004). Chromatin techniques for plant cells. Plant J. *39*, 776–789.

Brechner, E. (2001). Kompaktlexikon der Biologie (Heidelberg; Berlin: Spektrum, Akad. Verl.).

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

Campoli, C., Drosse, B., Searle, I., Coupland, G., and von Korff, M. (2012). Functional characterisation of HvCO1, the barley (Hordeum vulgare) flowering time ortholog of CONSTANS: Functional characterisation of HvCO1 in barley. Plant J. *69*, 868–880.

Cerdán, P.D., and Chory, J. (2003). Regulation of flowering time by light quality. Nature *423*, 881–885.

Chen, M., and Ni, M. (2006a). RFI2, a RING-domain zinc finger protein, negatively regulates *CONSTANS* expression and photoperiodic flowering. Plant J. *46*, 823–833.

Chen, M., and Ni, M. (2006b). RED AND FAR-RED INSENSITIVE 2, a RING-Domain Zinc Finger Protein, Mediates Phytochrome-Controlled Seedling Deetiolation Responses. PLANT Physiol. *140*, 457–465.

Cheng, S., van den Bergh, E., Zeng, P., Zhong, X., Xu, J., Liu, X., Hofberger, J., de Bruijn, S., Bhide, A.S., Kuelahoglu, C., et al. (2013). The Tarenaya hassleriana Genome Provides Insight into Reproductive Trait and Genome Evolution of Crucifers. Plant Cell *25*, 2813–2830.

Chia, T.Y.P., Muller, A., Jung, C., and Mutasa-Gottgens, E.S. (2008). Sugar beet contains a large CONSTANS-LIKE gene family including a CO homologue that is independent of the early-bolting (B) gene locus. J. Exp. Bot. *59*, 2735–2748.

Chow, B.Y., Helfer, A., Nusinow, D.A., and Kay, S.A. (2012). ELF3 recruitment to the PRR9 promoter requires other Evening Complex members in the Arabidopsis circadian clock. Plant Signal. Behav. *7*, 170–173.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. Plant J. *16*, 735–743.

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., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., et al. (2007). FT Protein Movement Contributes to Long-Distance Signaling in Floral Induction of Arabidopsis. Science *316*, 1030–1033.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561–563.

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

Cubas, P., Lauter, N., Doebley, J., and Coen, E. (1999). The TCP domain: a motif found in proteins regulating plant growth and development. Plant J. *18*, 215–222.

Dassanayake, M., Oh, D.-H., Haas, J.S., Hernandez, A., Hong, H., Ali, S., Yun, D.-J., Bressan, R.A., Zhu, J.-K., Bohnert, H.J., et al. (2011). The genome of the extremophile crucifer Thellungiella parvula. Nat. Genet. *43*, 913–918.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing Chromosome Conformation. Science *295*, 1306–1311.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. *36*, W465–W469.

Dodd, A.N., Salathia, N., Hall, A., Kevei, E., Toth, R., Nagy, F., Hibberd, J.M., Millar, A.J., and Webb, A. (2005). Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage. Science *309*, 630–633.

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, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. *19*, 1349.

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., et al. (2010). Quantitative analysis of regulatory flexibility under changing environmental conditions. Mol. Syst. Biol. *6*.

Eisenberg, E., and Levanon, E.Y. (2003). Human housekeeping genes are compact. Trends Genet. *19*, 362–365.

Fawcett, J.A., Maere, S., and Van de Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. Proc. Natl. Acad. Sci. *106*, 5737–5742.

Fornara, F., Panigrahi, K.C.S., Gissot, L., Sauerbrunn, N., Rühl, 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., Lee, K., Onouchi, H., Samach, A., Richardson, K., Morris, B., Coupland, G., and Putterill, J. (1999). GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in

Arabidopsis and encodes a protein with several possible membrane-spanning domains. EMBO J. *18*, 4679–4688.

Franklin, K.A., Praekelt,, U., Stoddart,, W.M., Billingham,, O.E., Halliday,, K.J., and Whitelam, G.C. (2003). Phytochromes B, D, and E Act Redundantly to Control Multiple Physiological Responses in Arabidopsis. PLANT Physiol. *131*, 1340–1346.

Franzke, A., German, D., Al-Shehbaz, I.A., and Mummenhoff, K. (2009). Arabidopsis family ties: molecular phylogeny and age estimates in Brassicaceae. Taxon 425–437.

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 Plant Sci. *16*, 108–116.

Galloway, G.L., Malmberg, R.L., and Price, R.A. (1998). Phylogenetic utility of the nuclear gene arginine decarboxylase: an example from Brassicaceae. Mol. Biol. Evol. *15*, 1312–1320.

Garner, W.W., and Allard, H.A. (1920). Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. J. Agric. Res. *18*, 553–606.

Gendrel, A.-V., Lippman, Z., Yordan, C., Colot, V., and Martienssen, R.A. (2002). Dependence of Heterochromatic Histone H3 Methylation Patterns on the Arabidopsis Gene DDM1. Science *297*, 1871–1873.

Gendron, J.M., Pruneda-Paz, J.L., Doherty, C.J., Gross, A.M., Kang, S.E., and Kay, S.A. (2012). Arabidopsis circadian clock protein, TOC1, is a DNA-binding transcription factor. Proc. Natl. Acad. Sci. *109*, 3167–3172.

González-Schain, N.D., Díaz-Mendoza, M., Żurczak, M., and Suárez-López, P. (2012). Potato CONSTANS is involved in photoperiodic tuberization in a graft-transmissible manner: Regulation of tuberization by potato CONSTANS. Plant J. *70*, 678–690.

Gregis, V., Andrés, F., Sessa, A., Guerra, R.F., Simonini, S., Mateos, J.L., Torti, S., Zambelli, F., Prazzoli, G.M., Bjerkan, K.N., et al. (2013). Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis. Genome Biol. *14*, R56.

Griffiths, S., Dunford,, R.P., Coupland, G., and Laurie,, D.A. (2003). The Evolution of CONSTANS-Like Gene Families in Barley, Rice, and Arabidopsis. PLANT Physiol. *131*, 1855–1867.

Gumucio, D.L., Heilstedt-Williamson, H., Gray, T.A., Tarle, S.A., Shelton, D.A., Tagle, D.A., Slightom, J.L., Goodman, M., and Collins, F.S. (1992). Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human gamma and epsilon globin genes. Mol. Cell. Biol. *12*, 4919–4929.

Gutiérrez, R.A., Ewing, R.M., Cherry, J.M., and Green, P.J. (2002). Identification of unstable transcripts in Arabidopsis by cDNA microarray analysis: rapid decay is associated with a group of touch-and specific clock-controlled genes. Proc. Natl. Acad. Sci. *99*, 11513–11518.

Hall, J.C., Sytsma, K.J., and Iltis, H.H. (2002). Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. Am. J. Bot. *89*, 1826–1842.

Harmer, S.L., and Kay, S.A. (2005). Positive and Negative Factors Confer Phase-Specific Circadian Regulation of Transcription in Arabidopsis. Plant Cell Online *17*, 1926–1940.

Hassidim, M., Harir, Y., Yakir, E., Kron, I., and Green, R.M. (2009). Over-expression of CONSTANS-LIKE 5 can induce flowering in short-day grown Arabidopsis. Planta *230*, 481–491.

Haudry, A., Platts, A.E., Vello, E., Hoen, D.R., Leclercq, M., Williamson, R.J., Forczek, E., Joly-Lopez, Z., Steffen, J.G., Hazzouri, K.M., et al. (2013). An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat. Genet. *45*, 891–898.

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 Cell Physiol. *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.

Hecht, V., Foucher, F., Ferrandiz, C., Macknight, R., Navarro, C., Morin, J., Vardy, M.E., Ellis, N., Beltran, J.P., Rameau, C., et al. (2005). Conservation of Arabidopsis Flowering Genes in Model Legumes. PLANT Physiol. *137*, 1420–1434.

Hecht, V., Knowles, C.L., Vander Schoor, J.K., 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 Physiol. *144*, 648–661.

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

Hellens, R., Mullineaux, P., and Klee, H. (2000a). Technical Focus: A guide to Agrobacterium binary Ti vectors. Trends Plant Sci. *5*, 446–451.

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

Herrero, E., Kolmos, E., Bujdoso, N., Yuan, Y., Wang, M., Berns, M.C., Uhlworm, H., Coupland, G., Saini, R., Jaskolski, M., et al. (2012). EARLY FLOWERING4 Recruitment of EARLY FLOWERING3 in the Nucleus Sustains the Arabidopsis Circadian Clock. Plant Cell Online *24*, 428–443.

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.

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 Online *15*, 1296–1309.

Hsu, C.-Y., Adams, J.P., No, K., Liang, H., Meilan, R., Pechanova, O., Barakat, A., Carlson, J.E., Page, G.P., and Yuceer, C. (2012). Overexpression of Constans Homologs CO1 and CO2 Fails to Alter Normal Reproductive Onset and Fall Bud Set in Woody Perennial Poplar. PLoS ONE *7*, e45448.

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., et al. (2011). The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat. Genet. *43*, 476–481.

Huang, W., Perez-Garcia, P., Pokhilko, A., Millar, A.J., Antoshechkin, I., Riechmann, J.L., and Mas, P. (2012). Mapping the Core of the Arabidopsis Circadian Clock Defines the Network Structure of the Oscillator. Science *336*, 75–79.

Iltis, H.H., and Cochrane, T.S. (2007). Studies in the Cleomaceae V: A New Genus and Ten New Combinations for the *Flora of North America*. Novon J. Bot. Nomencl. *17*, 447–451.

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.

Imaizumi, T., Schultz, T.F., Harmon, F.G., Ho, L.A., and Kay, S.A. (2005). FKF1 F-Box Protein Mediates Cyclic Degradation of a Repressor of CONSTANS in Arabidopsis. Science *309*, 293–297.

Imlau, A., Truernit, E., and Sauer, N. (1999). Cell-to-cell and long-distance trafficking of the green fluorescent protein in the phloem and symplastic unloading of the protein into sink tissues. Plant Cell Online *11*, 309–322.

Ito, S., Song, Y.H., Josephson-Day, A.R., Miller, R.J., Breton, G., Olmstead, R.G., and Imaizumi, T. (2012a). FLOWERING BHLH transcriptional activators control expression of the photoperiodic flowering regulator CONSTANS in Arabidopsis. Proc. Natl. Acad. Sci. *109*, 3582–3587.

Ito, S., Song, Y.H., and Imaizumi, T. (2012b). LOV Domain-Containing F-Box Proteins: Light-Dependent Protein Degradation Modules in Arabidopsis. Mol. Plant *5*, 573–582.

Iyer-Pascuzzi, A.S., and Benfey, P.N. (2010). Fluorescence-activated cell sorting in plant developmental biology. Methods Mol. Biol. Clifton NJ *655*, 313–319.

Jaeger, K.E., and Wigge, P.A. (2007). FT Protein Acts as a Long-Range Signal in Arabidopsis. Curr. Biol. *17*, 1050–1054.

Jang, S., Marchal, V., Panigrahi, K.C., Wenkel, S., Soppe, W., Deng, X.-W., Valverde, F., and Coupland, G. (2008). Arabidopsis COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. EMBO J. *27*, 1277–1288.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. *6*, 3901.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature *473*, 97–100.

Jung, J.-H., Ju, Y., Seo, P.J., Lee, J.-H., and Park, C.-M. (2012). The SOC1-SPL module integrates photoperiod and gibberellic acid signals to control flowering time in Arabidopsis. Plant J. Cell Mol. Biol. *69*, 577–588.

Kaufmann, K., Wellmer, F., Muino, J.M., Ferrier, T., Wuest, S.E., Kumar, V., Serrano-Mislata, A., Madueno, F., Krajewski, P., Meyerowitz, E.M., et al. (2010). Orchestration of Floral Initiation by APETALA1. Science *328*, 85–89.

Kim, S.-K., Park, H.-Y., Jang, Y.H., Lee, J.H., and Kim, J.-K. (2013). The sequence variation responsible for the functional difference between the CONSTANS protein, and the CONSTANS-like (COL) 1 and COL2 proteins, resides mostly in the region encoded by their first exons. Plant Sci. *199-200*, 71–78.

Kloosterman, B., Abelenda, J.A., Gomez, M. del M.C., Oortwijn, M., de Boer, J.M., Kowitwanich, K., Horvath, B.M., van Eck, H.J., Smaczniak, C., Prat, S., et al. (2013). Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature *495*, 246–250.

Koch, M.A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). Mol. Biol. Evol. *17*, 1483–1498.

Koevenig, J.L. (1973). Floral Development and Stamen Filament Elongation in Cleome hassleriana. Am. J. Bot. *60*, 122.

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. MGG *229*, 57–66.

Kosugi, S., and Ohashi, Y. (2002). DNA binding and dimerization specificity and potential targets for the TCP protein family. Plant J. *30*, 337–348.

Koyama, T., Mitsuda, N., Seki, M., Shinozaki, K., and Ohme-Takagi, M. (2010). TCP Transcription Factors Regulate the Activities of ASYMMETRIC LEAVES1 and miR164, as Well as the Auxin Response, during Differentiation of Leaves in Arabidopsis. Plant Cell Online *22*, 3574–3588.

Kumimoto, R.W., Zhang, Y., Siefers, N., and Holt, B.F. (2010). NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in Arabidopsis thaliana: NF-YC function in flowering time. Plant J. *63*, 379–391.

Laloum, T., De Mita, S., Gamas, P., Baudin, M., and Niebel, A. (2013). CCAAT-box binding transcription factors in plants: Y so many? Trends Plant Sci. *18*, 157–166.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Laubinger, S., Marchal, M., Le Gourrierec, J., Wenkel, S., Adrian, J., Jang, S., Kulajta, C., Braun, H., Coupland, G., and Hoecker, U. (2006). Arabidopsis SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability. Development *133*, 4608–4608.

Ledger, S., Strayer, C., Ashton, F., Kay, S.A., and Putterill, J. (2001). Analysis of the function of two circadian-regulated CONSTANS-LIKE genes. Plant J. *26*, 15–22.

Lee, H. (2000). The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. Genes Dev. *14*, 2366–2376.

Lee, J., Oh, M., Park, H., and Lee, I. (2008). SOC1 translocated to the nucleus by interaction with AGL24 directly regulates *LEAFY*. Plant J. *55*, 832–843.

Levine, M., and Davidson, E.H. (2005). Gene regulatory networks for development. Proc. Natl. Acad. Sci. U. S. A. *102*, 4936–4942.

Li, L., Wang, X., Stolc, V., Li, X., Zhang, D., Su, N., Tongprasit, W., Li, S., Cheng, Z., Wang, J., et al. (2006). Genome-wide transcription analyses in rice using tiling microarrays. Nat. Genet. *38*, 124–129.

Lifschitz, E., and Eshed, Y. (2006). Universal florigenic signals triggered by FT homologues regulate growth and flowering cycles in perennial day-neutral tomato. J. Exp. Bot. *57*, 3405–3414.

Liu, J., Yu, J., McIntosh, L., Kende, H., and Zeevaart, J.A. (2001). Isolation of a CONSTANS ortholog from Pharbitis nil and its role in flowering. Plant Physiol. *125*, 1821–1830.

Liu, L.-J., Zhang, Y.-C., Li, Q.-H., Sang, Y., Mao, J., Lian, H.-L., Wang, L., and Yang, H.-Q. (2008). COP1-Mediated Ubiquitination of CONSTANS Is Implicated in Cryptochrome Regulation of Flowering in Arabidopsis. Plant Cell Online *20*, 292–306.

Marshall, D.M., Muhaidat, R., Brown, N.J., Liu, Z., Stanley, S., Griffiths, H., Sage, R.F., and Hibberd, J.M. (2007). Cleome, a genus closely related to Arabidopsis, contains species spanning a

developmental progression from C3 to C4 photosynthesis: C4 photosynthesis in Cleome. Plant J. *51*, 886–896.

Mathieu, J., Warthmann, N., Küttner, F., and Schmid, M. (2007). Export of FT Protein from Phloem Companion Cells Is Sufficient for Floral Induction in Arabidopsis. Curr. Biol. *17*, 1055–1060.

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. Bioinforma. Oxf. Engl. *16*, 1046–1047.

Michael, T.P., Mockler, T.C., Breton, G., McEntee, C., Byer, A., Trout, J.D., Hazen, S.P., Shen, R., Priest, H.D., Sullivan, C.M., et al. (2008). Network Discovery Pipeline Elucidates Conserved Time-of-Day–Specific cis-Regulatory Modules. PLoS Genet. *4*, e14.

Michaels, S.D. (2005). Integration of Flowering Signals in Winter-Annual Arabidopsis. PLANT Physiol. *137*, 149–156.

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 Online *4*, 1075–1087.

Mizoguchi, T., Wheatley, K., Hanzawa, Y., Wright, L., Mizoguchi, M., Song, H.-R., Carré, I.A., and Coupland, G. (2002). 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., et al. (2005). Distinct Roles of GIGANTEA in Promoting Flowering and Regulating Circadian Rhythms in Arabidopsis. Plant Cell Online *17*, 2255–2270.

Morris, K., and Jackson, S.P. (2010). *DAY NEUTRAL FLOWERING* does not act through *GIGANTEA* and *FKF1* to regulate *CONSTANS* expression and flowering time. Plant Signal. Behav. *5*, 1105–1107.

Morris, K., Thornber, S., Codrai, L., Richardson, C., Craig, A., Sadanandom, A., Thomas, B., and Jackson, S. (2010). DAY NEUTRAL FLOWERING Represses CONSTANS to Prevent Arabidopsis Flowering Early in Short Days. Plant Cell *22*, 1118–1128.

Ben-Naim, O., Eshed, R., Parnis, A., Teper-Bamnolker, P., Shalit, A., Coupland, G., Samach, A., and Lifschitz, E. (2006). The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. Plant J. *46*, 462–476.

Nakamichi, N., Kiba, T., Henriques, R., Mizuno, T., Chua, N.H., and Sakakibara, H. (2010). PSEUDO-RESPONSE REGULATORS 9, 7, and 5 Are Transcriptional Repressors in the Arabidopsis Circadian Clock. Plant Cell *22*, 594–605.

Navarro, C., Abelenda, J.A., Cruz-Oró, E., Cuéllar, C.A., Tamaki, S., Silva, J., Shimamoto, K., and Prat, S. (2011). Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature *478*, 119–122.

Nelson, D.C., Lasswell, J., Rogg, L.E., Cohen, M.A., and Bartel, B. (2000). < i> FKF1</i>, a Clock-Controlled Gene that Regulates the Transition to Flowering in< i> Arabidopsis</i>. 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.

Nusinow, D.A., Helfer, A., Hamilton, E.E., King, J.J., Imaizumi, T., Schultz, T.F., Farré, E.M., and Kay, S.A. (2011). The ELF4–ELF3–LUX complex links the circadian clock to diurnal control of hypocotyl growth. Nature *475*, 398–402.

Onouchi, H., Igeño, M.I., Périlleux, C., Graves, K., and Coupland, G. (2000). Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among Arabidopsis flowering-time genes. Plant Cell Online *12*, 885–900.

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

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.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (2007). Numerical recipes: the art of scientific computing (Cambridge, UK ; New York: Cambridge University Press).

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.

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.

Rédei, G.P. (1962). Supervital mutants of Arabidopsis. Genetics *47*, 443.

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

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., et al. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science *290*, 2105–2110.

Robson, F., Costa, M.M.R., Hepworth, S.R., Vizir, I., Reeves, P.H., Putterill, J., and Coupland, G. (2001). Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. Plant J. *28*, 619–631.

Robson, P.R., Whitelam, G.C., and Smith, H. (1993). Selected components of the shade-avoidance syndrome are displayed in a normal manner in mutants of Arabidopsis thaliana and Brassica rapa deficient in phytochrome B. Plant Physiol. *102*, 1179–1184.

Rodrigo, A.G., and Learn, G.H. (2001). Computational and evolutionary analysis of HIV molecular sequences (Boston, MA: Kluwer Academic Publishers).

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425.

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., and Coupland, G. (2000). Distinct Roles of CONSTANS Target Genes in Reproductive Development of Arabidopsis. Science *288*, 1613–1616.

Sánchez-Acebo, L. (2005). A phylogenetic study of the new world Cleome (Brassicaceae, Cleomoideae). Ann. Mo. Bot. Gard. 179–201.

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.

Schmid, M., Uhlenhaut, N.H., Godard, F., Demar, M., Bressan, R.A., Weigel, D., and Lohmann, J.U. (2003). Dissection of floral induction pathways using global expression analysis. Development *130*, 6001–6012.

Schommer, C., Palatnik, J.F., Aggarwal, P., Chételat, A., Cubas, P., Farmer, E.E., Nath, U., and Weigel, D. (2008). Control of Jasmonate Biosynthesis and Senescence by miR319 Targets. PLoS Biol. *6*, e230.

Schranz, M.E., and Mitchell-Olds, T. (2006). Independent Ancient Polyploidy Events in the Sister Families Brassicaceae and Cleomaceae. Plant Cell Online *18*, 1152–1165.

Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Kröber, S., Amasino, R.A., and Coupland, G. (2006). The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. Genes Dev. *20*, 898–912.

Serrano, G., Herrera-Palau, R., Romero, J.M., Serrano, A., Coupland, G., and Valverde, F. (2009). Chlamydomonas CONSTANS and the Evolution of Plant Photoperiodic Signaling. Curr. Biol. *19*, 359–368.

Shaw, C.H., Carter, G.H., and Watson, M.D. (1984). A functional map of the nopaline synthase promoter. Nucleic Acids Res. *12*, 7831–7846.

Siefers, N., Dang, K.K., Kumimoto, R.W., Bynum, W.E., Tayrose, G., and Holt, B.F. (2008). Tissue-Specific Expression Patterns of Arabidopsis NF-Y Transcription Factors Suggest Potential for Extensive Combinatorial Complexity. PLANT Physiol. *149*, 625–641.

Siev, M., Weinberg, R., and Penman, S. (1969). The selective interruption of nucleolar RNA synthesis in HeLa cells by cordycepin. J. Cell Biol. *41*, 510–520.

Simon, S. (2009). Transcriptional regulation of the flowering-time gene CONSTANS in Arabidopsis thaliana.

Šmarda, P., Bureš, P., Šmerda, J., and Horová, L. (2012). Measurements of genomic GC content in plant genomes with flow cytometry: a test for reliability. New Phytol. *193*, 513–521.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., dePamphilis, C.W., Wall, P.K., and Soltis, P.S. (2009). Polyploidy and angiosperm diversification. Am. J. Bot. *96*, 336–348.

Somers, D.E., Schultz, T.F., Milnamow, M., and Kay, S.A. (2000). < i> ZEITLUPE</i> Encodes a Novel Clock-Associated PAS Protein from< i> Arabidopsis</i>. Cell *101*, 319–329.

Song, Y.H., Ito, S., and Imaizumi, T. (2010). Similarities in the circadian clock and photoperiodism in plants. Curr. Opin. Plant Biol. *13*, 594–603.

Song, Y.H., Smith, R.W., To, B.J., Millar, A.J., and Imaizumi, T. (2012). FKF1 Conveys Timing Information for CONSTANS Stabilization in Photoperiodic Flowering. Science *336*, 1045–1049.

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 Online *21*, 2606–2623.

Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Mas, P., Satchidananda, P., Kreps, J.A., and Kay, S.A. (2000). Cloning of the Arabidopsis Clock Gene TOC1, an Autoregulatory Response Regulator Homolog. Science *289*, 768–771.

Suarez-Lopez, P., Wheatley, Robson, 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.

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

Takada, S., and Goto, K. (2003). TERMINAL FLOWER2, an Arabidopsis Homolog of HETEROCHROMATIN PROTEIN1, Counteracts the Activation of FLOWERING LOCUS T by CONSTANS in the Vascular Tissues of Leaves to Regulate Flowering Time. Plant Cell Online *15*, 2856–2865.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol. *24*, 1596–1599.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., and Paterson, A.H. (2008). Synteny and Collinearity in Plant Genomes. Science *320*, 486–488.

Taoka, K., Ohki, I., Tsuji, H., Furuita, K., Hayashi, K., Yanase, T., Yamaguchi, M., Nakashima, C., Purwestri, Y.A., Tamaki, S., et al. (2011). 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. Nature *476*, 332–335.

Teper-Bamnolker, P., and Samach, A. (2005). The Flowering Integrator FT Regulates SEPALLATA3 and FRUITFULL Accumulation in Arabidopsis Leaves. Plant Cell Online *17*, 2661–2675.

Tepperman, J.M., Zhu, T., Chang, H.-S., Wang, X., and Quail, P.H. (2001). Multiple transcription-factor genes are early targets of phytochrome A signaling. Proc. Natl. Acad. Sci. *98*, 9437–9442.

Terzaghi, W.B., and Cashmore, A.R. (1995). Light-regulated transcription. Annu. Rev. Plant Biol. *46*, 445–474.

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 797–815.

Tiwari, S.B., Shen, Y., Chang, H.-C., Hou, Y., Harris, A., Ma, S.F., McPartland, M., Hymus, G.J., Adam, L., Marion, C., et al. (2010). The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. New Phytol. *187*, 57–66.

Toledo-Ortiz, G. (2003). The Arabidopsis Basic/Helix-Loop-Helix Transcription Factor Family. Plant Cell Online *15*, 1749–1770.

Torok, M., and Etkin, L.D. (2001). Two B or not two B? Overview of the rapidly expanding B-box family of proteins. Differ. Res. Biol. Divers. *67*, 63–71.

Tsai, Y.-T., Chen, P.-Y., and To, K.-Y. (2012). Plant regeneration and stable transformation in the floricultural plant Cleome spinosa, a C_3 plant closely related to the C_4 plant C. gynandra. Plant Cell Rep. *31*, 1189–1198.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.-L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007). Arabidopsis TFL2/LHP1 Specifically Associates with Genes Marked by Trimethylation of Histone H3 Lysine 27. PLoS Genet. *3*, e86.

Umemura, Y., Ishiduka, T., Yamamoto, R., and Esaka, M. (2004). The Dof domain, a zinc finger DNA-binding domain conserved only in higher plants, truly functions as a Cys2/Cys2 Zn finger domain. Plant J. *37*, 741–749.

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.

Wang, J.-W., Czech, B., and Weigel, D. (2009a). miR156-Regulated SPL Transcription Factors Define an Endogenous Flowering Pathway in Arabidopsis thaliana. Cell *138*, 738–749.

Wang, R., Farrona, S., Vincent, C., Joecker, A., Schoof, H., Turck, F., Alonso-Blanco, C., Coupland, G., and Albani, M.C. (2009b). PEP1 regulates perennial flowering in Arabis alpina. Nature *459*, 423–427.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.-H., Bancroft, I., Cheng, F., et al. (2011). The genome of the mesopolyploid crop species Brassica rapa. Nat. Genet. *43*, 1035–1039.

Wang, Y., Tan, X., and Paterson, A.H. (2013). Different patterns of gene structure divergence following gene duplication in Arabidopsis. BMC Genomics *14*, 652.

Wassarman, D.A., Solomon, N.M., Chang, H.C., Karim, F.D., Therrien, M., and Rubin, G.M. (1996). Protein phosphatase 2A positively and negatively regulates Ras1-mediated photoreceptor development in Drosophila. Genes Dev. *10*, 272–278.

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., and Coupland, G. (2006). CONSTANS and the CCAAT Box Binding Complex Share a Functionally Important Domain and Interact to Regulate Flowering of Arabidopsis. Plant Cell Online *18*, 2971–2984.

Wigge, P.A., Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U., and Weigel, D. (2005). Integration of Spatial and Temporal Information During Floral Induction in Arabidopsis. Science *309*, 1056–1059.

Wötzel, S. (2010). Analysis of flowering-time genes in Brassicaceae species - Characterization of a SHORT VEGETATIVE PHASE homolog in Arabis alpina.

Yakir, E., Hilman, D., Hassidim, M., and Green, R.M. (2007). CIRCADIAN CLOCK ASSOCIATED1 Transcript Stability and the Entrainment of the Circadian Clock in Arabidopsis. PLANT Physiol. *145*, 925–932.

Yamaguchi, A. (2005). TWIN SISTER OF FT (TSF) Acts as a Floral Pathway Integrator Redundantly with FT. Plant Cell Physiol. *46*, 1175–1189.

Yamaguchi, A., Wu, M.-F., Yang, L., Wu, G., Poethig, R.S., and Wagner, D. (2009). The MicroRNA-Regulated SBP-Box Transcription Factor SPL3 Is a Direct Upstream Activator of LEAFY, FRUITFULL, and APETALA1. Dev. Cell *17*, 268–278.

Yamashino, T., Yamawaki, S., Hagui, E., Ueoka-Nakanishi, H., Nakamichi, N., Ito, S., and Mizuno, T. (2013). Clock-Controlled and FLOWERING LOCUS T Dependent Photoperiodic Pathway in Lotus japonicus I Verification of the Flowering-Associated Function of an FT Homolog yamashino 2013.pdf.

Yanagisawa, S. (2002). The Dof family of plant transcription factors. Trends Plant Sci. *7*, 555–560.

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

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, M., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura,, Y., et al. (2000). Hd1, a Major Photoperiod Sensitivity Quantitative Trait Locus in Rice, Is Closely Related to the Arabidopsis Flowering Time Gene CONSTANS. Plant Cell Online *12*, 2473–2484.

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

Yu, J.-W., Rubio, V., Lee, N.-Y., Bai, S., Lee, S.-Y., Kim, S.-S., Liu, L., Zhang, Y., Irigoyen, M.L., Sullivan, J.A., et al. (2008). COP1 and ELF3 Control Circadian Function and Photoperiodic Flowering by Regulating GI Stability. Mol. Cell *32*, 617–630.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007). Whole-Genome Analysis of Histone H3 Lysine 27 Trimethylation in Arabidopsis. PLoS Biol. *5*, e129.

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

Zunk, K., Mummenhoff, K., and Hurka, H. (1999). Phylogenetic relationships in tribe Lepidieae (Brassicaceae) based on chloroplast DNA restriction site variation. Can. J. Bot. *77*, 1504–1512.

APPENDIX

Promoter sequences from Brassicaceae CO used for phylogenetic shadowing

>Aethionema_arabicum
aagcattccctacaatgctaatcttcgacaccggatatagaattgttcctacattttgatgatcctgtaataaaactagactgtatgcttgtaaattaattatgatttgaattcctacaagtttttaagacagttatgcatata
ggtctgtagttttgtctaacaaattgaaattaatgagatatgttaatattttgttgcagtttttaaacaaactatgtttaactggtagaaatctacctatttgtgaaatatttcctattaaaactccgtcctaatgtctaattgcaa
attgatgctacaggaaagagatatgcattactctgttctgtcctaatgtctaatggctgtaacaagaaaactaaaaatgtacttcaccttgtcaatgttcaatgtcgttagctattagtgagattccacatcacttgaaca
ctttagaagatagaaaatttccaagaacattttcatgcaaccctgaggaaaacaagattagacaacaaagtggattacataaaccaatagtaaataagtatactttagaaatctaataaatctctttaaaccatccact
aaatgatcaattgttgtcacaaactcagaggttattatgaatgattacatcgaacatattctcacgataagaagattctaacaccttttttcccttatcactatattactcacaaccaagtaagtagcacaatacttcaatgt
ctatggaccatttgtatacataacgatacttcatgaaacctatttttttgctatttcttttagaataaagtttatgctttgctttaccaaggtatatgctaagattaaagaatgacgaggcaagaaaaaaaaacaatgacactg
tacctaagacttcactttttgccacctggataaggccctagagttcaacttagcaaaacatttccatggtttaagcagctgaattcctgaaccattcaataaacattggaatcaaaagattagtttctgctagagcttaat
atccacatgaattttaccactagattgtcttaaaatctgttctttttggaccactttaactgacctcatccatccactgatcccagaagtgcctactgcagcaagctctcagaccacaagaaaattagtcagcaatttgg
tttgttcatggaatccagctgaatacgtcaagtaaataaatcgcagatttccatggcctgtacatattactagaatatcatcaacaaggaaatttgaagtagtagagttcccactggaattttcttcaggaaacttaaaa
tgcaataacaagcacttaaactaaaatacttaaggatctaggagccaaatcaatgataatagttagtgatgttataccaaaagtgaaagtgcataaaagtgattaccttttgcaaaagtttgctgaacttttgcactttt
ccacgtgtagttctcacagccctgcaaagtggagaacccatataaaacctacaatacaagctttacgtggtccctataccccttattctctttttagtcttcggcaggaaaaaaaacaaatttctccttaattgtagatga
tatttgaataaaatgatcatctaaagcaccaaagccttgctcaaattatccaattactcctatggattagacatcaaacaatggctagctaaaacatttgtaaagaaacctatgatgtaaggaacctattatctttttcttt
agaaattaaatgataacggaacagttccataacattttacattacattcctcaaaaaaccagtaaattcgacaaagaggaatttgaaagttaaataccaattatcaaaaacaaaaaaaacttttaaaatttctatatcttc
ccttacggaacagaaaatgaaagagtaagagatacgaaaagtgtatcctaggaccagccttttatacaacaaactcacttttcagccgcagaggcatagagttgaagtaaaccacatgaatttggcgactacagt
gtcgaagcatcctgcgctttttggaccacaaaataggtatcatacatccactgaacttaaaagtgactggagcaagtactcaaacctaaagagtatatacactatagcttccaaataataaataattgtactccatcc
acacaataaaatttggcatgtcacttctaagtagtatgcattagatttctatgaaacatatacggctaccttgctacatactttgctactatgatggtgatcatgagaaaaacaatatttctacgaaacatatacggctac
ttgatataactaaccttgttctcctttggactccaagatgctgttcagcttgtggctcaggtgcatcctaatctctaggctcaagactagagggcggatccaaattgtgtcctagagaatccatcatgtctacgtttgcc
atgtcagcttcctaatacgaaaacactttccaacaaacaaaccatttcaaaatgtcaattacctaatgccctgtttcatgagtttctctatcaatgtaaatccagtgctttggtgatgaccctggtagctgaaaccttaca
cgggtaaacctgaagtaaaattacaaccacacaatgttaaaaactaatattcaacagcataacctattgtggagccataaagattcaaaagtatctaactctacttcactccatccgaaaatagtaaaatgtaaaac
gaaaatagtaatttttatacgaagagactaaatatgtgtctgcaacacaatgacgttaaggggaacacttcctcttcagaatctccttgtgccacgtggcaaaacctaagcaacctgtggataagtcccaccgaaa
aatacataggaggagattttacactttgcactttgcactttacactttacacagagagaggcgtccgtaagtcttggtatggcttttttttctctattgccaagatatatgaaattgaagtttttcctaataattaataatttcc
aatagccactgtgggcaactgtggtccctatacttgagcttttctgaaaaactcagccaagccacaaggtcacccattaccttttataccccatcatctcaagtgtcagctctcaggattctatctttcctctgtttattct
cttccagcagatatagccttaattattagccccttccatcagatacaatttcctacaccattagactacttcactgagttataatg

>Arabidopsis_thaliana
atgctgctatggcaaaactagactgcatgcttgtaaattacttctgttttgaaattcctgcaagttattaccacggttttatgcatatagggtctgtagtttactagttttgcctaactagtacgagttaattagatatacact
catacttagctgctgtaagttaatacaaattttgcccacttggtagaagtcctcttatctatgaaatactgccctacagtctaaattcaaactgatgctacaaagagatatgcattagtatagctagctgtaccaagaac
agacaacacgtaactcaccttgcctaagttcaaagttaatgattcctttcttggtgacactttccacatcactttatcacgtagaagataaatgattccacttatgccgccccaaagacaatttcgcacatccctgaga
aaagaaagttgtactcaatatgattttataccaaaagtagattacagagacaagtgtaaacatgtaaaactccaataaaatcacacactgaatagtcaggtaacaaataaccaagaatcctgagtgaacaatgcta
ataggtacaatcacatcaacgcatattcttgtactgattagtgattagcattggtgctaaatccattcccacgaggttgtctcggataacaagattctatcactcacaaccactactagcagaatatgtcaatgcctatg
gaccatccaaatgaacaaagatattcagtgtgagaaatctatttcttggtctatcttttagaataatagctcactttttccaaggtatatactggcactaaagaataacgacgcaaccaaaagaaaaagggtgacact
gtaaagccttcactttttgccacctggaaagggccagagatcaacttatttaccactaggatagcttcccacttcaacggttaagctgatgaattcatgtaccaaatcaatactttttagccataaatgagtcagtttta
gtatccacatgaatttacctaccagagtgttgtaaattatgttcttttttgggcccacttacatggatctcattcattcactgcagcgagttctcagaccaccagaaaatttattcagtgatctgttttgatcatgcaacataaa
cttataagccacacaagcaaaacaaagatatcccatgttgcatataatacgagctagcatatcataaagaaggaaacttgaagtagcaaagtttctactaaatttcttgtcaggaatttttaaaatgcaatgacaacc
acttggagcactatgagtttcagagccaatagaatgttactatttggtgtggattcgagctagcacgtgaaagtgcataaaagtgattaccttttgccaaaggtcactgcactttcctcagatagtttctcacagccat
ggaaagtggagaatccgcataaacgtacaattacaagctttatatggtccctcgactcttattctcttctcagtctttgcaactaaatagggttttcgttaatctgaaagaagcaaagtattcgcaaccacggaaacct
gataaagaatgaaaacaaataagcaatagtgttttcttgaaaatctcgatgcaactttcaggatattgttacatatgatctattactcgtaacagttatccgaaggcctacacatgtgagagaagttccaaaccgctac
aacaataaacttaattagagactgtcaacgagcaataataagcaaaactactttttcttgagctacaagtgaaaaggccaatacacaatttactcttcatgaactcgaaccacgttacaatctccaaaaaaatttcatc
accaaagcactaaaagccaaagatgcctcaacttatccaacttggcaggataaagatctccaaaaatgcttactaaagaacctagaatcttttcttttagaattcaatgatcatataaccatttcataacaattctaaatg
ccattacattcatcgtaaaaccagtaaataacaagaacttgtatgttacgttccaattaccaagcaaaaaaaacccttttcaaagtttaaagttcaaaatgggaaagagaagtgcggtgtaagcaaatatgaaagagg
aagagatgcgaaaagtgtatcctaggaccagcattttatacaaaaaaaaaaactcacttttcagctcttaaggcatagagtgaaggtagccatatgaatttggccactagagcgtccgtcaaatctcattctttttgga
ccacataatgggtatcatacattcactggacccaaaagcgtaactggagctagtcctcaaacctagagagtatcgtatcctgtagcttccacatagtaaacattatgagcataacaccaacaaggcaactccaagt
actaatggttattagtacaggaaaacccacgatgctaaacacatgaatgggtcaccaaatagagtgaagatggttaaattgcatctatggatcatgtggactagtaaatgagtgtagcagaaaacttcacaattac
ctctgtgatcttagaaacatgtcctgaaaattccatacaagtgtcgtttgtattagattacttccacaggttgagatctaataaagctacaataaatagtatagagtatcatcataaacccaaattacagagatgtgaaa
acactcatgagtcatgttttgtaactacttactataatggttaccaagtgcaaatttctacatactatatatgataaatctaattattgctcatgtggactccaaaatgccttttaagttttaacttgtgcgtcaggtaaatcct
aatttgtagtctcaagactacttggcggattcgagtttgatcctagaaaatccaccgtctctatgtttttcatgtcacttttccgatatgattctcattaccatgactttatgaaccagattaaacattataacacttttcatca
gaaaatccttcgaaagtttcaattgcaaatctttctaaatgatgcagatgcattcacaaataatggaacaacaactataccatattcacgagtttgtctaacctttgtataggtagtcaacccataacagttggtgatgg
ctctgacactcgaagccttactcggagagatacctgaacagtaatcacaaggttcaggatgaatattcaaccacttaaactttgtataaagccaaagagataaaacgaatctagctttacttaaataaaatgcatat
gaaaatagtaaaaggtgatacgaaaaaatagtaacaatttgcctgcaacaccatggcattatccggaccacttcctcttgagaatctcagtatggcaagtggcaaaacctaagcaacttgtgaacgggtcccaac
gaagaagtgcataggaggagatgtttacactttacactttacactttacacataggccttcccaaaagctcaactagctgcaagaggatccaataacatgtaagagccactaacgctgtgccacgtgtaggcact
caggattcgatctcccctctacttattctctcacaccagatataagctttattagccccttctttcagatacaagctcccacaccatcaaacttactacatctgagttattatg

>Arabidopsis_lyrata
tgctgctatggcaagactagactgcatgcttgtaaattacttctggtttgaaattcctgcaagttgttaccacggttatgcatataggcctgtagttttgtctaactagtacgagttaattagatatatattcatatttagcta
ctaagttaaaacaaaattttcccacttggtagaaatcctcttcatttatgaaatattaccaccacatagaatattgccctacagtctaaattcaaactgatgctacaaagagatatgcattagtttagctagctgtaccaa
gaacagtcaacacgtaactcaccttgcctaagttcaaagcaaataattccactgttcattcctttcttggtgacactttccacatcactgtctttatcgtagaagataaatgattccacttatgccgccccatgacattttc
atacctccctgagaaagaaggttttaatcaatagtatgattttataaccaaactagattacagagactagtgtgaacatgtaaaacttaagagatccaacaaaatcaacataaaaccacacactgaatagtctttttc
ggtaatgtactcagataagtgggactcagataacacgcaaccaagaaccctgagtgaacaatgctaatgggtataattacatcgacgcatattcttgttcctgattagcattggtgccaaatccattcccacgtggt
tgtctcggatgacaagattctatcactcacaaccaagtgctaacagaatatgtcaatgcctatggaccatcctaatgaacaaagatattcagtgtgagaaatctttttcttgctctatcttttagaataatagctctttttc
caattatatgctggcactaacaaataacaaggcaagaaaagaaaaagggtgacactgtaaagccttcactttctgccacctggaaagggccagatctcaacttatctaccactaggatagcttcccacttccatg
gttaagctgatgaattcctgtaccaaatcaatatttttttagccataactgattcagttttaatacccaagtgtgttgcaaaatcagtgatctgttttcatcatgcagcatgaatttataagccacacaagcaaaccaaaga

tgagccctgttgcatatagtactagctagcatatcaggaaacttgaagtagtaaagtttctactaaatttcttgtcaagaaattttaagatgcaatgacaaccacttggagcactaacagagtttaagcgccaatagaa
tgttactatttgatgatggattcgagatatcacgtgaaagtgcataaaagtgattaccttttgccaaaggtcactgcactttgcatttttcctcagatagtttctcacagccatggaaagtggagaatccgcacaaaac
gtacaattacagctttatatggtccctcgactctcttattctcttctcagtctttgcagctaaatacgcttttcgtaaaatctgaaagaaggaaagtatacgcaacacggaaaactgataactaatcaaaacaaataagca
atattgttttctttcctcaatctctagaattgagtcacagtggaagaaaaaaatgcattttcctcccacataatctcaatgtcatgttatctgaaggcctacacatgtgagagaagttccaaaccgctacaacaataaactt
acagattgtcaaggagcaataataagccaaatacctttttcttgaacaacaagtgaaaaaggccaatacacaattcactcttcatgaactcaaacgcgttatttctccaaaaaaatttcataacaaaagcactaaaag
ccaaagatgcctcaacttatccaacttggcaggataaagatctccaaatatgcatagtaatgtaaggaacctagaatctttttcttcagaattcaatgatcatgaaaccatttcataacaattctacatgccattacattc
ctcataaaaccagaaagtaacaagaacttaaatgttaagttccaattaccaagcagaaaaaccttttcaaagtttaaatttactcatataggaaagagaagtgcggtgtaagcaaaaatgaaagagaaagagatg
cgaaaagtgtatcctaggaccagcattttatacaaaaaaaaaactcacttttcagctcttgaggcatagagtgaaggaagccactagagcgtccgtcaaatctcattcttttggaccacaaaatgggtatctacatc
cactggacccaaaagcgcaactggagctagtcctaaaacctagagagtatcatatcctgtagcttccacatagtaaacattatgagcataacaacaacaaggcaactcaaagtactaatggttactagtacaagc
aaacccaagatgctgaacacatgaatgggtcaccaaatagcttgcatttgctaaccaagaaactagcaatggtgcttaaagatactgtggactagtaaacgagtgtagcacaaaactttacaattacctctgtgat
cctataaacaagttcctgaaattttccaaacaggtgtagtatgaattagattacttccacaggttgagatctaaaaagctacaataaataatatagagtatcatcataaactcaaaatgttacagagatgtgacaacac
tcatgagtcatgttttcttactacttactatgatggggatcaagtgcaaatttctacataccatctacggacacctttgctcatgtggactccaacatgcctttcaacttgtgcatcaggtaaatcccaatctttagactca
agactacttggcggattcgagttggatcctagaacgcccacagtctctatgtttttatgtcttattttccgatatgcttctcattgccattactttatgaaccagactaaacattaaaacacttttgatcagaaaatcattag
gaagttccaattgcaaatctttctaaatgatgcagatgcaatcacaaataattgaacaactttaccatattcacgagtttgtctaaccttctataggtagtacacccgtaacagttggtgatggctctgagactcgaag
cctaactcagagatgtacctggacagtaatcacaaggttcaaaatgaatattcaactacttaaacttcgtatgaagccaaaaaaaggtaaacgaatctagctttacttttaaaattcatatgaaaatagtaaaagatga
cacgtaaaatagtaacaatgtgtctgcacagaattttcagtttccggaccacttcctcttgagaatctatgtatggcaagtggcaaaacctaagcaacttgtgaacaggtaccaatgaagaagtgcataagaggag
atgtttacactttacacataagccttcccaaaagctcaactagctgcaagatgatccaataacatttaagagccactaactttgtgccacgtgtaggcactcaggattcgatctaccctctgcttattctctcacacca
gatataagctttattagccccttctttcagatacaatctcccacaccatcaaacttactacacctgagttattatg

>Arabis_alpina
tatgttcctatagcaaaaccagattgcttgcttgtaaattacttatggtttgaattcctgcaagttattaccatggttatgcaatataggtctgtaatttttgtctaaactaattagatatattcatattttgctgctctagttaaa
acaaaatttgccaacttggtagaaattctattgaaatatcaccaccacatagaataaacctaatttcaaattgatgctataatgagatgaatcactttataacaagaacagttaacacaccttgcctaggttccaagca
actactgattcctttcttagtgaccattgctacatcgcttgaacaagttgaagataaatgattccacttatgtcatccctgagaaaagaaggttatacacttatgattttctaatcaaaaatagattgcgtagtccaattgt
aaacatgtaaatagagaaatttaaaccaaaagcagacataatatcacacattaaatagtcattttcagtaacatactcagataagtggcagtcatatagcaaacaaccgagaaccatgagtaaacaatgctacttg
gtataattacatcaaagcatattcttgttcctgattagcattggagcatggttgcgttaccaaaccatttcccacgtggctctctgataacaaagattctaacacctttgtcgctcacattatatcactaacaaccaagttc
tagcagaatatgtcgatgcctatggaccatccaagtaagcaaagatattcagtttgagaaatctttttctggctatatcttttagaataaagaaaaataaaagctcccttttccagggtatatgctagcaataaagaata
atgaggcactgtaactaatccttcactttttgccacctggaaagggccatagagctcaacttatctaccacttggatggcttcccattaagctcatgaattcctataccaatcaaaagatacgtttagacataactcag
cgagtgttagtattcatgtaaatatagccaccaaagtgtcatgatatgtcctttttggaccacataaatggatctcatacatccactgaagccagaagtgtccactgcagcaaagttctcaagaccccgagaaaatc
aagtcagtgatctggtttgatcacgcagcacaaaattataaaccacgcaagaaaatcgaagataccccctggttggatataatactagcatatcatcaacaaggaaacttgaagtaccagaaagtttgttactagatt
tcttttcaggaaatttaagatgcaatgacaactagtttgagtaccaacgaatgtattggagtcaactaaatgttaatgttctagtaatggattcgagatatcactagaaagtgcattaaagtaattaccttttgccaaagg
tcactgcacttttgtacttttccttagatagttacagccatggaaagtggagaatccgcacaatacttacaattacacgcttcatatggtccatggactcttgttctcttctctgcctttttaagaaaagaagcttttcacctt
aaatacgaaagaaggaaagtattcacaaccacggtaaactgataaataatgaaaacagataaaccatatagttttttgtcaatatctaaaattgagtcacaacagatcaaaatacattttcttccatttaattccaatg
caactttcatgtaagatggaggtttatacagaggtctacacatgtgagagaagttgcaaaccgctacaacaataaactcacagattggcaaagagcaacaataaacaaaatacttctttcttgaactagaaccactt
taaaacctccataattttgaccattaatgcccaaacgccaagatgcctcaaatatatccaattacttctatggtttagacaccaaactttgatgacataaagatctccaaagatgcttactaatgtaataaacctagattt
tttttcttagagttcagtgatgatgaaaccatttcataacatttctgcatactattacattcctcatacaaacaggtaaacaataaaaacctaaaagttttaagttccaattaccacgaaaagaaaacattttcaaagtttaa
aaccctcccttattggaaagaaaagtgtggagtaagcagaaatgaaagaggaagagatgcgaaaagtgtatcctaggaccagccattttatacaagtaaactcacttttcagctatagaggcatagagttgaag
gaagccacatgaatttggccactaacatctcgttctttttggaccacaaaatgggtatcatacatccactgaacacaaaagcgtaattggagatagtcctcaaaacctacagagtatcatatcctgtagcttccgcat
agtaaacattgttagcatatcacgaccaagacaactctaagtactggtggctactagtacaagcaaacccaagaggctgaataaaagaacgaatcaccaaatggaaatagacatgggcgcgagacaactctc
gttaatgacatgccaaaacttagtatgtataccacatattgaaatctaataaatttacaataaataataaagagggtgatcatattctcaaaatgtaaacgagatgtgacaatactcatcactggagccatgttttgttac
tatgatggtggtcaagtgcaaatgtctacttaacatacacagacatgtagtataactaaccttgctcctttggtcttcaaaatgctttcaacttgtaactcaggtaaaatacaccatgtatacgttatccgtgtcagcatc
ccgatatgcttctcattaccattactttattagccacactaaatactaatgcacatttcatcagaaaatcacttcaaagtttcaattgcaaatcattccaactgatgcagatgcaatcatcgataactgaataacaacttta
ccatttcacgagtacgtctaaacttcgttgggtagaaaccccataacagttggtgatggttctctgacttgaagccttactcggagatatacctggacagttagcacaaggttcaagaagaatattcaactacatga
acatttactagtatctagcttaaatttaaataaacccatatgaaaatagtaaaaggcgatacgaaaaatagtaacaatgtgtctgcaacaccatggcgttgtccagaccacttcctcttgagaatctctgtatggcaag
tggcaaaaccaaagcaacttgtgaacggggcccatccaagaaatacattggaggagatgtttacactttacactttacacataagcattcccaaaagctcacccagtcgaataatgatccaataacatttaagagc
caatagatctgtgccacgtgtaagcgctcaatattcgatctctcctctgcttattctctcacaacagatatagccccttctttcagatacaatctcccacaccatcaaacttactacatttgagtaattatg

>Brassica_napus
taatactcgtatggtttaactagattccatgcttgtagttatcaccactgtaatgcaaataggtctctatttttgtctaactagtacgaattaattagatatatttttattttgctgttgtagttactatccttaagtctaaattcaa
accgatgctatagagatatgtgcgttactttagctagatgtaccaagaacagttaacaatgaagcctgttctttaatggtcgctgatcgcagcggttgcgtcaacataaaatcactatgacagaaacaaatgaatgg
attttgaagatgaaatcagtcgccaacaagtttccggagacaaaaatatcggatgccgctgtcgctgcttttatgggcgtcaaatttttcttttattttcaccgcagcgactgagaaagcaaattgatattgctccgcc
agcagtgtatttgccgtggtgagagatgtagttttcggtcgctcccgctaatcgcggcgtctatgaaacgaacacgcctgaactcaccttgcaaaaaggttcagacaagtagtgattgctttcttggtgaaacttcc
acatcagcttgataaacaagtagaagataaatgattccacttgtgcctccctgagaaaataaggttttacacaaatatgattttccaaacaaaagtagactacatagtctaactgtacacttataaaacttgagcaaac
ccaataaaaaaacattgtaaccatcacacattgacttgagtgaacaaatgctactatgtaaaattacatcaaaaaacatattcttgtttctggttagcattggagagcatggctgtgtaaccaaacccatgcattccca
cgtggctctctctgataaacaagaatctaaacacctttgtcccatccagtatattaacactcataagtaacaagcaaccactagcagaatatgtcaatgcctatgtttagttacttgctctaaatctattagaataaagg
gctatagtgtgtcacctttccgaggaatataatagcattaaagaataaggaagcaaagaaaaatagaagtgacactgtaactaagacttcactttttttgccacctgtaaaaggccatagagcctcagcttatctacc
actttccatgggttaagctgatgaattcctgcaccaaccaaaagattcgttttagccatcactgagtgagtgttgatatccatacgaattaagccactagattgagtgtaaaatatgttattttggaccacataaatgaa
tctcattcatccactgaaaccagaagtgtgacaagcgccaccactgcattcagaacccatcagaaaatctaatctgtgatctgtttaatttgatcatgtagcacaaacttataaaccacgcaagaaaatggatatgat
attagcatatagagagtttgatctagatatctttccaggaaatttaaagatatgcaataacaattttagtaaatcctaacgagtgtaaagagcaaaatgttaataaaatgaatgttaacatttagtttgctctagatttctttt
caggaaaatgaagatacagtaacaattagttgaagtatactagcaaatgtaagagccacttgaatgctaatctttagtgatggatttgagacacgtgaaagtgcataaaagtgattaccttttgccaaaggtctctgc
tcttttgcacttttccttagatagtttttttacagccatataaagtggagaatcctcacatagcttacaattgcacgctttgtatggtccctagactctaagattctaggttctcttctcagtcctttcaagaaaagaaggtttt
ctatttaaatctgaaagaaggaaagtattcgcaacacagaagctgaaaaataatgaaacaaataagtgacatagttttctttccataatatctctaaaactgagtttaattggcttaaaattgaattttccttcacataatt
ccgatgcaacttccatgtaaggtggaggtctacacttgtgaatgaagttccgaaccacagatacacaaagagaaataataagcaaaactatatttcttgaactacaagtgataaaagggtcaatacacaactcttc
atgagctagaaccactataaaatccccactttttcatatcattaatgtaatgaacctagaatccttctcttctataagatttatccattacattcatcataaaccagtaaataacaacaactggaagttaagctccaattac
caagaaatttgtttttgagagtttaaaatcgtcccttactagaaaaaaatagtgtggtgtaagaaataatgaaagaggaagagatgcgaaaagtgtatcttaggaccagcctttttatacaagcaagctcacttttcag
ctgtagactcatagagttgaaggaagccacatgaatttggccactagggtcttctaacctttggaccacataaatgggtatcattagtacatccactgaacccaaaagtgtcctcaaactctagagagtatcatattc
atacactgtagctttccacatagtaaacattattaacatatcaccaacaaggcaactcaaagcactggtggttattagtacaaaagcaaacccaagatgatggatataccgtacacgaacggtcatcagatagttttt
ttttttttgctaaatggtcatcagatagtttgtagctgctaaccaagaaatagtaaacgaaggtgcttttaaatctgcatctactgaccatatggactagtaaacgagagtagctaaaaacgctacaatttcctccgtga

tcctataaagataagtcatgtaacttccatctctggcctgagacatagctgaagtgtataaaagctaatatggaaacatatataaagctgcaataaacaatatattatcatcaaattgtcaaattgttacagaaaaaatt
gtcaaactgttacagagacatttgacaacacccatgagtcattctttattaccatgatggtgatcaagtgcaaatttctacataaaatgtagacatgtgatacagactaacctagctccttttggactactaatttgctta
ccacttgtggttcgggtaacagttggagttggtgatggctctggaactctgaagccttactcggagatatacctggacagttatcacaaagttcaagaaggaatattcaactatgtaaacttttcctggagccacca
cagagataaagtatctagctttcctataaataaatccaagtgaaaatagtaaaaggtggtacgaaaatagtaacaccatatggtgttttatcctcttgagaatctctcagtggcaagtaagcaacgtgtgaatagttc
ccaactaatagctctgtgccacgtgtaagcactgaggattcggattctctcacaacaagataaagctttattagcccctgctcttcagacacaatttcccacacaccatcacacttaattactacatcagtgagctaa
catg

>Brassica_oleracea
ttctgataatactcctatggttcaactagattccatgcttgtaaatgcattactggaagttatgaccactgtaatgcatataggtctctattttgtctaactagtaccaattaattagatatatttatgtttttctgttgtatttaa
aacagaattttcacacagttactatccttcaatctaagttcacaccgatatgtgcgttactttagctagatgtaccaagaacagttaacaatgaactcaccttgcaaaaaggttcagacaagtagtgattgctttcttgg
tgacacttccacaccactttgataaacaagtagaagataaatgattccacttgtgtctccctgagaaaataaggttttacacaaatatgattttccaaacaattacatagtttaactttagacatatacaacttggaaaatt
caatttaaaaaaaaaaagacatcataacatcacacattgaatagtcatttcctgtaacatacttagaaaagtggcagtcagatagcaaacaaccaagaacccggagtgaacaaatgctactctgtaaaattacatca
aaaacatattcttgtttctggttagcattggagagcatggttgtgtaaccaaacccatgcattcccacctggctctctctgataaacaagaatctaaacacctttgtcccatccagtatattaacactcataagtaacaa
gcaaccactggcagaatatgtcaatgcctatgtttagttacttgctctaaatctattagaataaagggctatagtgtgtcacctttccaaggaatataatagcattaaagaataaggaagcaaagaaaaatagaagtg
acactgtaactaagccttcactttttttgccccactgtaactaagccctcaccactttccatgggttaagctgatgagcaccaaccaaaagattcgttttggccatcactgagtgagtgttgatatccatatgaatttag
ccactagattgagtgtaaaatatgttctctttggaccacataaatgaatctcattcatccactgaaaccagaagtgtgacaacccgccaccactgcattcaaaacccaccagaaaatctagtctgtgatctggtttga
tcatgtaggctgtagcacaaacttataaactactcaagaaaatggacgatatctctggttacatatgatattagcatatcattatcaacaaggaagctagtagagagttgatctagatatctttccaggaaatttaaag
atatgcaataacaattttaggctaaatcctaacgagtgtaaagagccaattgagtgttaacatttagtttgctctagatttcttttcaggaaaatgaagatgcaataacaattagtttaggtatactagcaaatgtaagac
ccacttgaatgctaatatttagtgatggattcgaaacacgtgaaagtgcataaaagtgattaccttttgccaaaggtctctgcactttttgcacttttcctcagatagttttctcacagccatataaagtggagaatcctac
atagcttacaattgcaagctttgtatggtccatggactctaagattctaggttctcttatcagtcttttcaagaaaagaaggttttctatttaaatctgaaagctgaaaaaaataataatgaaacaaataagtgacatgtta
taaaagtaatagaaaagtctatctttattcataacatagaggttctttatataggagattacatcgtcatagataaatagaaagattacaaatcataatctcttggttatgagccatccacaatctggttcataatatcaca
tagtttttttcccttaatatctctaaaactgagtctaattggcatagaattgaattttccttcacataattccaatgcaacttccatgtaaggtggaggtctacacttgtgaatgaagttccaaaccacagataggcaaaga
gaaataataagcaaaactatatttcttgaactacaagtgataaaaggggccaatacacaactctcttcatgagctagaaccactataaaatccccactttttcatatcattaatgtaatgaacctagaattcttctccttc
atgagatttctccgttacattcatcaaaaaaaccagtaaataacaagaacttgaagttaagctccagttaccaagaaaaaaacattttgagagtttaaagtcgtcccttactagaaaaaaaaatagtgtaaataaatgaa
agaggaagagatgcgaaaagtgtatactaggaccagcctttttatacaagcaagctcagttttcagctgtagactcatagagttgaaggaagccacatgaatttggccactagggtcttctaacctttggaccaca
taaatgggtatcattagtacatccaccgaacccaaaagtgtcctcaaactctaaagagtatcatattcatatactgtagctttccacttagtaaacattattaacatatcaccaacaaggcaactcaaagcactggtgg
ttactagtacaaaagcaaacccaagatgatggatataccgtacacgaactgtcatcagatagtttgcagctgctaacccaagaaatagtaaacaaaggtgcttttaaatatgcatctactgatcctatggagtaaacg
agagcatagctcaaaacgctacttttctctgtgatcctataaagacaagtcatgaaacttcccatatctggcctgagacataacagtataaaattaagctaatatggaaacatatatacagctgcagaaaacaatata
ttcatcatattgtcaaattgttacagagacatttgacacacccatgagtcattctttatttaccatgatggtgatttaagttcaaatttctacattaaatataaacatgtgatatactaacactaaccttgctcctttttggactc
ctaatttcctgcaacttgtggcccaggtaacagttggtgatggctatgagtagggatgttacccatgggtttacccaaacccactaataatgggttgggtttttacccattgggtcaaccatgggtattaattttaaaa
ctcatatttatgttgggtaaatacaaaagggtaatggtgtacccatgggttttcaattatatatatagattttcaacattttaattaaattatataaaatttgcaaaaacatttttttttgtcagaaccaagaaaataaaaattt
ccgccaaaaccgtaaaataaattttctcacagaaaccaaaaacgagtttttccaccaaaaccgaaaaatcgagtttttctgccgaaaccacaaaaccgagtttttccgccaaaaccgcaaaatagaatttttccgc
aaataccgtataattgggtttttttcgctaaaatcgcaaaatcgagtttttcgactgaaaaccgtaaaatcgagtattcccgccaaaaccgcaaaatgtgtttttccgccaataccgtaaaacccaattttcccgccaaa
acgatttttcccgccgaaaccgaaaaaccgagttttcccgccaaaaccgaaaaatcgagttttcccgccaaaaccgaaaaaccgagttttcccgtcaaaataaaaacgagttttcccgccaaaaccgaaacacc
gagtttttccgccaaaaccgcaaaaccgagttttcccgtcaaaatcaaaaacgagttttcccgccaaaaccacgaaaccgagttttcctgccaaaaccgctaaaccgaattttctcgcaaaaaccgtaaaaccga
gtttttccgccaaaaaaaatcgcaaaaagattgttcccgccaaaatcgaaaacaggtttcccgccaaaatctaacaatttttttccgtcaaaaccgtaaatacaagtttttccgccaaaaccgcaaaaacgaattttcac
gccaaaaccagaaagattattttttccgcaaaccgcaaaaatcgagttttccaaccaaaaccacaaaaaaaaaaaatttaccgaccaaaaaccgcaaaaagaaaattttccgctaaaaccacaaaaacaacttaac
aagttttctcgttaaaaccataaatgagttttcctgtcaaaacgatgtttactgctaaaaccgcaaaatgagttttccggttaaaattgcaacataattttttttccgaaaaaaacgagtttttccgccaaaatcgtaaacg
agtttttgcgctaaaacaagttttctataaaattgcaaaatcttgtttatatattaaagctcacattaatgatattaatattttacatgatcttcagagtaaataagataatattttgggatattgggttattttttttggtttggattt
caaccttgatgtttctgggtttttgctggttgggtataaacagggttgggttatttgtgggttgggctgggctgggttattttaccacgttaacatccctagctatgagacatgaaaccttactcggagatatacctgga
cagttatcacacggttcaagaagaatactcaactatgtaaacttttcctggagccacatagatagagtacctacctagctttactataaataaatccaagtgaaaatagtaaaaggtggcacgaatatagtaacacc
acatggtgttttatcctcttgagaatctctcagtggcaagtagctgaaaatctataataagcaatgtgtgaaccgttcccaaattataggtctgtgccacgtgtaagcactgaggattcggattccctcacaacaagat
aaagctttattagccccttctctcagacgcaatctcacacaccaccaaacaaactacatcagtgagcaaaatatg

>Capsella_rubella
gtgctcctatggcaaaactagattgcatgattagttattaccactgtaatgcatataggtctatagttttgtctaactagtaccgtagtacggattaatgagacatattcataatttgctgctgtagctgaaacaaaattttc
ccacctggtagaaatcgtcttatttgtaaaatatttaccaccacatgtaatactgctcttcagtctaaattcaattttttgatgctatagagagatatcgtttactttagctagctgtaccaagaacagttaacacggaactc
accttgccaaaggttcatagaaaataactactgattcgtttcttggtgacaccttccacatcacttgaacacctcaaagctaaatgattccacttatgccgccgttggacattttcatacatcccctaagaaaataaggt
tttactcaatacgattttataatcataaatagattacagagacgagtaaacatgtaaaactttggaaatccaataaaatcacacactgaatagttattttcggtaatgtactcagatagcaaataaccaagaacccaga
gtgaacattgttactaggtatatttacatcaatgcatattctctttcctgattaacattgatgccaaatccattcccacgtggctctctctgataaacagattctattactcacaaccaagtgctagcagaatatgtcaatg
cctatggaccatccaaatgaacaaacatattcagtttgagaaattcatatctttctctatcttttggaaggatagcttcacttctccaaggtatatgctggcattaaagaataacgagccaagaaaagaaaaagggtg
aaactgtaacgaagccttcactttttgccgcctgaaaagggccatagatctcaactcagtacctctactatagcttcccacttccatagttaattaagctgatgaatttctgtaccaatcaagatattcgtttcagccata
agtgagtgagtgttaatatccatatgaatttagccactagagtgtcagaagtatttttttggacctcataaatggatctcattcatccactgaagccagaagtgtccactgcagtcaagttatcagaccaccagaaattc
tagtcagtgatctggtttgatcatgctgtacaaacttataaaccacggaagcaaaacaaagatattcccggttgcatataatagtagcatatcatcaatattgaaacttgaagcagtagagtttctactaaatttcttgtc
aggaaattttaagatgcaatgataatcacttggagcactaactaatttaagaggcaatacaatgttaatatttggtgatggattcgagatagcacgtgaaagtgcataaaagtgattaccttttgccgaaggtctctgc
acttttgcacttttcctcaaatatagtttctcacagccatggaaagtggagaatccgcacaaacgtacaattacaagctttatatggtccctcgactcttgttctcttctcagtctttgcaaggaaataggcttttccttaa
atctgaatgcgggaaagtattcgcaaccacaaagaaatgataaacaatgaaaataagcgatatagttttctttcctcaatctctagaactgagtcacaatagatcaaaaatgcagctttcgccttataatctcatgca
actttcatgttatgtggaggactacacatgcgagagaactttcaaacagcaacaacaataaattcacagattcaaagagcaataataagcgaaaaactaatttcttgaactacaagtgaaaaaaaggccaataca
cagaacacagtttactcttcatgaacttaaaccaccttaaaaaaaaaatttcatcattaaatcactaaaagccaagatgcctcaaattatccaattacttcgtggcatagacatcaaactttagcgagatataaagatctc
caaatatgcttactaatgtaaggaacctagtatcatttttctttagagttcaatgatcatgaaaccatttcatcacatttctccatgccattacattcctcataaaaccagtaaataacaagaagagcttaaataagttcca
attaccaagcaaaaaacttataagttccaattaccaaacaaaaaaaaaaccttttaaaagttaaatattttcccttgtgggaaagagaagtatagtgtaagcaaaaatgaaagaggaagagatgcgaaaagtgtatc
ctaggaccagcattttatacaagaaaaaatcacttttcagcggtagaggcatagagcgaaggaagccacatgaatttggccactagagtgtccattaagtctcgttctttttggaccacaaagtgggtataatacat
ccactgaacccaaaagcgtcactggagctagtcgtcaacttagagagtatcatatcatgtagcttccacgtagttaattattattagcatatcactaccaaggcaactcgaagtactagtggttaattactagtacaa
gcaaaccaaagatgccgaacacatgaataggtcaccaaatagcttgcattttctaaccaagaaactagcaagcgaaggtgcttttaaaatttgcatctactgatcctgtggactagtaaacgagtggagctcaaa
attgtacaattacctctgttatcctaaagacaagtcctgaaattttccacacaagtgtagtacgtattacatatcttccacaggttgagatctaataaagctacaatatatagtatagtatagaataacaacataaactca
aaatgttacagagatgtgacaacactcatgagtcatgctatgttactacgatggtgatcaagtgcaactttcttctacataccatatacagacatgtgatataattaactttgctcatttgaactccaaaatgcctttcaa
cttgcgggtcaggtaaatcctaatctttagactcaagactacttggcggattccagtagtcctagaaaatctaccatttctactttttcgatgtcagcgttccgatagctttgacattgccattaatttattaaccagatac

atactaaaacacttttcatcagaaaatcatttggaagttctattgcaaacgttctaagtgatgcagaagcaatcataaaaattcttgtacttacccattttcacgagttcgtctagccttgataggtagtgaacccgtaa
cagttggtgctggctctgatgagagactgatactcgaagccttactcggagatatacctgaacaattatcacaaggttaaggatgactattcagctacgtaaactttgtgtgaagccaaagagatgaaagaatatct
agctttacttttaataaatccatatgaaaatagtaaaaggtgatacgaaaaatagtaacgatgtctgcaacaccatggcttcatccagaccactttctcttgagaatctctgtatggcaagtggcaaaacttaagcaac
ttgtgaacgggtcccagtgaagaagtgcatagggggagatgtttacactttcactttatactttacacataagccatcccaaaagctcagctagccgcaagatgatccaatcacatttaagagcgactaattctgtg
ccacgtgtagggactcaggattccatctcccctctgcttattctctcagaacagatataagctttattagccccttctttcagatacaatttcccacaccaccaaacttactatatctgagttagtatg

>Sisymbrium_irio
tataccgctatggcagaactagattgcatgctagtaaatcacttactggttcgaattcctgcaagttattaccgctgtaatgcatatataggtcagtattttgtctatatactacggattaagtcgatatatattcatagtta
ggtaaaacactaacaccaaatctatatactctgatatgtgaaatgataccaccacatagaatactacccttcattcttagttcaaattgatactataaatatatatacaattactttaataccaagaacagttaacacgga
actcaccttgtgaaaggttcaaagcaaacaacctattgattccacatcacttggaaacgtagaagataaatgattccacttgttccgccctgagaaaataaggttttacacaatacgattttccaaacaaagtagact
acataatcgaattgtaagcatttaaaacttgagaagttcaataaaaaacaacataacaccacacagtgaatagtcattttcagtaacatactcagataaatggcagtcagatagcaagcaaccgagaaccttgagt
gaacaatgctactaggtctaattacattaagagcatattgctgttcctgattagcattggagcatggtttcggtactaagccaatccattcccacgtggttctctctgagaagcaagattctaacacctttgtcgcatcc
agtatattatccactcacaagtcacaaccaactactagcagaaaatgtcaatgcctatgtttacttgctttaatattttaattagctcaatttccaaggaatatgctgccatttaagaataagaaaacaaaagagaaaa
aaaagcaaaggaaacaaagaaaaattaaggtgacactgtaactaagccttcacttctcgccacctggaaaaggccgtagagctcagcttatatatctaccactttccacggtgaagctgatgaattcctgtacca
atcaaaagattcgctttggccatcactgagtgagtgttaatgtccatataaatttagccactagattgagtgtaaaataagttatttttggaccatataaatggatctcattcatccactgaaacccacaactgtgacaagt
gtcaccactgcagaaatttagtctgtgatctgattcgataatatagaacaaacttataaaccacgcaagaaaaatcgaagatatctctggtttcatatgatattagcgtatcatcatcaacaagaaagcttgaagtagt
agagactttgttctagatttcttttgaggaaattttaaagatgcaataacaatttaattttaagtaggaacgagtgtaagagccaactgaacgttaataattatagccaattgattgctaatatttacggatgcataaaagt
aattaccttttgccaatggtggctgcacttttgcactttttctcagatagtttctcacagccatggaaagtggagaatccgcacaaaacttacaattccaagctttatatggtccctggactcttgttctcgactcaacctt
ttctaaaaaagaagcttttctccttaaatctgaaagaaggaaagtattcgcaaccacggaaactgagaataatgaaatcaaatatgtgacatagttttcttccttaatctctaaaattggattgaaaatgtattttcctac
cacaacaaaccgatgcaactttcatttaaggtggaggactacacatgtgacagaagttccaaaccacagataggcaaacagaaataataagcaaacctattcttcttgaactacaattgaaaaaggcctataca
cattactcttcatgaactagaaccacttcaaaatctccactattttcatatcattaaagcgctaaaaaagcagagacacctcagagttatccaattacttctaaacatcaaactttggcgggataaagatctccaaatat
gtttactattaatgtcaatgacattcctcataaaaccagtatatatataacagcaacttaaaacgttaagctccaattaccaagaaaaaaaaaaaatcttttcaaagtttgaaatcttccctctgagaaagaaaagtgta
gtgcaagcagtaatgaaagaggaagagatgcgaaaagtgtatcctaggaccagcctttttatacaagcaagcgaacttttcagctgttgaggcatagagttgaaaaaagccacatgaatttttcgccactagggt
ctcctaaaatctcgttccttttggaccacaaaaatgggtatcagacatccactgaacccaaaagtgccctccaaacctagagagtatcatattcatatactgcagcttccacatagtaaacattattagcatatcacca
acaaggcaactcaacggcctggtggttactagtacaaataaacccaagatgatcgatacacttatggtctccagtctgctaaccaagaaacaagtaaacaaaggtgctttgaaatttgcatctactgattagtaaa
cgaacgtagctcaaaactcttacaatttcctctgtgatcttgaagacaagtcatgaaacgttttatctgaccttaagcacagaatataaacgctaatggcagaaataacattccgaatttgatttagaatgtgttggatt
actttcacagattaaaatctaataaagctgcattaaacaatagagtatcatcatatcctcaaaatgttacagagacatgtgacaatataacactcatgaaccatgttttattactcttatgtgattcaagtgcaatattctg
cataaatatacggagatatatgataataactaacctttctcctttggactccaaatttctttacaacttgttgctcaagtaacagttagtgatggctctgggactcgaaggccgaagcgttaccctgacctgagatata
cctggccagttattacgggttgaagaaaaacattcacctacgtaaaaaccttcctcgaaccgaagccaaagagataaactatgtagctttactttaaataattacaaatgaaaatagcaagtggtgatacgtaaaat
ggtaacatcgggaccacttcctcttgggaatctccgtatggcaagtggcaaatcctaagcaactagtgaacgggtcccaacgaaggagtagagatgtttacactttacacataagtcttcctaaaagctgagttag
ccgcaagatggtccaataactctgtgccacgtgtaagcgttgatgattcgaaatctcttcctctgcttattctctcacaacagatataatttattagcccctcactttcagagccaatctctccacaccatcaaactagtt
actacatcagagtgtaagcattatg

>Sisymbrium_officinale
tatatccctacggcagaactatattgcttacgtgtaaatcactaatttcgaatacctgcaagttataaccgctgtaatgcatatatacgccattatttttgtctgatgtactaccggataattactatacatattcatattttgc
tgcggtaacttaggtaaaagcaaatctaacatcctccgatatgtgaaatgatcatcaccatatattaatactacccttccttcttatttccaattgatactatcaatatacatgcaattacattaatgttaaaaacagttaac
acggagctcaccttgtgaaaggttcaaagcaaatatcactatcgattccacttgtgccgccctgaaaaacaaggttttacacaatatgattttccaaacaaagaacattacataatcgaattgtaaacattgtaaaact
tgagaaattcaatacaaaacaacataacatcacacattgaattgtcattttcagtaacatactcagataaatggcagtcagatagcaatcaaccgasaacattgagtgaacaatgctactaggtatagttacatcaaa
agcatattcttgttcttgattakaattggagcatggtttcggtactaaaccaatgcattcccacrtgaatctctctgakaagcaagattttaacaccttagtatattatcactcacaagtcacaaccaactactagcagaa
tatgtcaatgaatgcctatgtttacttgctctagtattttagaataaagagctatagagtagctcacttttccaaggcatatatgctggcattaaagaataaggaaaaaagaaaaactaaggtgacacggtaactaag
ccttcacttttcgccacctggaaaaggccactcagcttatatatctaccactttccatggttaagctgatgaattcctgtaccaatcaaaagattcgctttggccatcactgagtgagtgttaatatccatatgaatttag
ccactagattgagtgtaaaatatgttctttttggaccacataaatggatatcattcatccactgaaaccagaagagtgacaagtgtcaccactgcagaaaatccagtctgtgatctgagtcgatcatatagcacaaac
ttaaaaaccacgcaagaaagtcgaagatatctctggtttcatatgatattagcatatcatcatcaacaagaaagcttgaaatagtagagagtttgttctagatttattttcaggaaattgaagatgcaataacaattagt
atacgtacgctaacgaatgtaagagccaactgaatactaatatttagggatggattcgagatatggcgtgaaagtgcataaaagtgattaccttttgccaaaggtcgctgcacttttgcacttttcctcagatagtttct
cacagccatggaaagtggagaatccgcacaaaacttccattacaagctttatatggtccctggagtcttgttctcgtctcagtctttttattaaaaaagaagcttttctccttaaatctgaaagaaggaaagtattcgca
aactgataagtgacatatttttctttccttaatctctaaaattgagtcgcattggattgataatgcattttcctctacaaaattcccatgtaactttcatgtaagagaagttctaaaccacagataggcaaacagaaataat
aagcaaaccaattctttcttgaacgacaagtgaaaaaggcctatacacattgctcttcatgaactagaaccacttcaaaatctccactattttcatatcactaaagcgcaaaaatgcagagatgcctcgaatttatcca
attacttctagacatcaaactttgacgggataaataaagatctccagatatgtttactagtgtaaggaacctagaatatttttattaagttcaatgatcatgaacaatttcataagatttctccatgacaatgctcataaaa
ccagtaaatatataacagcaacttaaaaagttaagttccattaccaagaaaaaaaaccttttcaacgtttaaaatcttcccttctgagaaagaaaagtgtgctgtaagcagaaatgaaagaggaagagatgcgaaa
agtgtatcctaggaccagcctttttatacaagcaaactaacttttcagctgttgaggcatagagttgaagaaagccacatgaatttttggccactagggtctcctaaaatctcgttccttttggaccacaaaacgggta
ccatacatccactgaactcaaaagtctcctcaaaacctagagagtatcgtattcacatagtaaacattattagcatatcaccaacaaggcaactcaaagtatagtattactaaacccaagatgattgatacacgaat
ggtcaccagatagtttgcatctgctaaccaagaaacaagtaaacaaaggtgctttaaaatttgcatctactgatcatatgggctagtaaacgaacgtagctcaaaactcttacaatttcctctgtgatcctgaagaca
aatcatgaaacgttttatatgacccgaggcacagagtataaacgctaatggcagaaataacatgccaaattttagttatgtattggatgactgtcacgggttaaaatctaataaagctgcaacaaacaatagagtatt
atcatacagagacttgagacaacagtcatgggccatgttttattactctaatggtgatcaagtgcaaatttctgcataaaatttacagagatgtgatataactaaccttcctcctttttgactccaaatttccttacaacttg
tggcttagtaacagttagtgatggctctaagggactcgaaggccgaagcgttaccctgagatatacctggccagttatcacaaggttgaagaaaaatattcacctacgtaaaaaaattttccaaagccagagaga
taaagtatctagctttactttaaataattccaaatgaaaatagtaaaaagtgctacgaaaaatagtaacaccatggtgttatcgggaccactctctcttgggaatctccgtatggcaagtgacaaatcctaagcaacta
gtgaacgggtcccaatgaaggagtacacatatactacataggaggagagatgttttcactttacacataagccttcctaaaagctgagcaagatgatccaataactctgttccacgtgtaagcgttgatgattcgaa
atcttttcctccgcttattctctcacaacagatatagatttattagcccctcctttcagagccaatctctccacaccatcaaactagttactacatcagagtgttagtattatg

>Nasturtium_officinale
tgctcctatgtctaaactagactgcatgtttagttattattatcactgtaattctgctcttccgtccagaactcagatttatgctatgaagagaaatggtttacttgagccagctataccaagaacagctaacacggaact
cacctttccaaaagttcaaaagaaaataaataacatagactaactgtaaaactctagaaatcaataaagctcaaaataaaatcacatactgaatagtcattttcacgaatgtactcagataagtgggagtcagtcag
atagtaaataatcaagaactctgagtgaacaattctactaggtatagttacattggtgcatattcttgttcctgataacaagattctatcactcacagccaagtactagcagaatataacaatccctatggaccatccaa
atgaacaaagatattcagtttgagaaatctatatctagctctaacttttagaatataataactcacttttccaaggtatatgctggcattaacgaataacgaggcaaggaaagaaaaagggtgatactgtaccttcactt
tttgccacctagaaaggccatagagctttaacttatctacctctacgaagctttccacttccatggttaagctgatgatttcctgtaccagtcaaaagattcgtttcagccataactgagaaagtgataattaacaatat
gaatttagccactaaagttcaaaacacatgttatttttggacacataattgaacctcgttcatctactgaagccagaagtgtccactagaacacaataacaatgttctcagatcaccagaaaatctagtcagtgatccc
gttggatcacataacacaaacttataaaccacacaagcaaatcaaagataccccaggttccatatatcatcatcaagaaaacttgaagtagttgagttcctactaaatttctttttcccaggaaatttaagatgcaatg

aaaacaagttgaagcactaacgaatttaagacagtagaatgttaaaatttagtgatggattcgagatatcacgtgaaagtgcataaaagtgcttaccttttgccaaaagtcacagcacttttgctcttttcttcagataa
tttctctcacagccatggaaagtggagaatccgcataaaactattacaagctttatatggtccctggactcttgttctcttctcactctttacaagaaaatgggcttttctccttaaacctgacagaaggaaagtatttgc
aatcacggaaaactggaaaataatgaaaataaattagcgatataggtttgtctcctcaatctctagaattgagtatcaaaaatgcatttaactccacataatctcgatgcaactttcatgttatatggaggcctacatat
gcgaaagaagttacaaaccgctaacaaaaaataaactcacagattgtcaaagagcaataataagcaaaatagttatttcttgaactactagtgaaaaaggccaacacacttactcttcatgaacttgaaccacctta
aaaccttcgtttttcatcattacagcactaaaagccaagatgtctxxxttagagttcaaagatcatgaaaccatttcgacacatttctccatgccgttacattcntaataaaaccagtaaacgacaagacgaacttaaa
agttacgttccaattaccaagaaaaaaacctttttgaaagttaaattttttttccgttgttggaaaggagaatgtggtgtaagaaaaaaatgaaagaagaagagatgccaaaagtgtatcctaggaccagcattttata
ctagaaaatcacttttcagcggtagagtcatagagagaaggaagccacatgaaattggccactagagtctccattaaatcttcttctttttggaccacaaaacgggtatcatacatccgccgaacccaaaagtgtc
cttcgaacaaagtcctcaacctagagagtatcacatcttgtaggttccacacagtaaagtattattagcatatcactaacacggcaactcaaagtactggagcttactagtacaagcaaactcaagatgctcaacac
atgattggctcaccaaatagctttcttttaccaaccaagaaactatcaagcaaaggtgcttcaacaatttgcatctactgatcagtaaattacctctaagaactgtgatactcaagacaagtctaagttttccatgtcag
cttcccttctcattgccattactttatcaaccagactaaacactagaacaattctcaacaggaaatcatttggaagtttccattgcaaatctttctcaatgatgcagatgcattcataaataatggaacaacaactctacc
attttgacgagtttgtctaacctttgatagggtagtaaacccgtaacagttggtgaaggctctcaaactcgaagccttactcggagatatacctggacagttatcacaaggtttcaagatgaatattcaactatataaa
ctttgtgagaagccaaagagataaaagaatctagctttacgttaaacaactccatatgaaaatagtaaaaggttatacgaaaaatagtaaataagtgtatgcaacaccatggctttatccggaccacttaccctaga
gaatctctgtataacatgtggcaaaactaaagaaacgtgtggtaaagggtcccaatgaagaagtgtataggaggagatgtctacactttatactttacacacaagccttcccaaaaactcagctagccgcaagat
gatccaatcacatttaagagcctatctttgtgccacgtgcaggcactcaggattcggtctcccctctgcttattctctcacagcagatatagctttattagccccttcttccaaacacaatctcccacacaatcaaactc
actacatctaagttattatg

Sequences of subgroup Ia genes as used for locus comparison

>Arabdidopsis thaliana CO
acactcacttttcagctcttaaggcatagagtgaaggtagccatatgaatttggccactagagcgtccgtcaaatctcattctttttggaccacataatgggtatcatacattcactggacccaaaagcgtaactgga
gctagtcctcaaacctagagagtatcgtatcctgtagcttccacatagtaaacattatgagcataacaccaacaaggcaactccaagtactaatggttattagtacaggaaaacccacgatgctaaacacatgaat
gggtcaccaaatagagtgaagatggttaaattgcatctatggatcatgtggactagtaaatgagtgtagcagaaaacttcacaattacctctgtgatcttagaaacatgtcctgaaaattccatacaagtgtcgtttgt
attagattacttccacaggttgagatctaataaagctacaataaatagtatagagtatcatcataaacccaaattacagagatgtgacaacactcatgagtcatgttttgtaactacttactataatggttaccaagtgc
aaatttctacatactatatatgataaatctaattattgctcatgtggactccaaaatgccttttaagttttaacttgtgcgtcaggtaaattctaatttgtagtctcaagactacttggcggattcgagtttgatcctagaaaat
ccaccgtctctatgtttttcatgtcacttttccgatatgattctcattaccatgactttatgaaccagattaaacattataacacttttcatcagaaaatccttcgaaagtttcaattgcaaatctttctaaatgatgcagatgc
attcacaaataatggaacaacaactataccatattcacgagtttgtctaacctttgtataggtagtcaacccataacagttggtgatggctctgacactcgaagccttactcggagagatacctgaacagtaatcaca
aggttcaggatgaatattcaaccacttaaactttgtataaagccaaagagataaaacgaatctagctttactttaaataaaatgcatatgaaaatagtaaaaggtgatacgaaaaaatagtaacaatttgcctgcaac
accatggcattatccggaccacttcctcttgagaatctcagtatggcaagtggcaaaacctaagcaacttgtgaacgggtcccaacgaagaagtgcataggaggagatgtttacactttacactttacactttaca
ctttacacataggccttcccaaaagctcaactagctgcaagaggatccaataacatgtaagagccactaacgctgtgccacgtgtaggcactcaggattcgatctcccctctacttattctctcacaccagatataa
gctttattagccccttctttcagataccagctcccacaccatcaaacttactacatctgagttattatgttgaaacaagagagtaacgacataggtagtggagagaacaacagggcacgaccctgtgacacatgcc
ggtcaaacgcctgcaccgtgtattgccatgcagattctgcctacttgtgcatgagctgtgatgctcaagttcactctgccaatcgcgttgcttcccgccataaacgtgtccgggtctgcgagtcatgtgagcgtgct
ccggctgctttttgtgtgaggcagatgatgcctctctatgcacagcctgtgattcagaggttcattctgcaaacccacttgctagacgccatcagcgagttccaattctaccaatttctggaaactctttcagctccat
gaccactactcaccaccaaagcgagaaaacaatgaccgatccagagaagagactggtggtggatcaagaggaaggtgaagaaggtgataaggatgccaaggaggttgcttcgtggctgttccctaattcag
acaaaaataacaataaccaaaacaatgggttattgtttagtgatgagtatctaaaccttgtggattacaactcgagtatggactacaaattcacaggtgaatacagtcaacaccaacaaaactgcagcgtaccaca
gacgagctacgggggagatagagttgttccgcttaaacttgaagaatcaaggggccaccagtgccataaccaacagaattttcagttcaatatcaaatatggctcctcagggactcactacaacgacaatggttc
cattaaccataacgtaaggcttttgtatatttgttacccccttcaatttagcatcttcccataacgcagcagggtgaattctttcatcatacacacaaatccactgatccactgccaacagttgatctatagcacatagaaa
tttcaccagaagtctataataaaaacaatatatgcttccttttgcatcgactctctttagtcctcttaccagggggattgagaatgtctttgtttctgtcattaggcatacatttcatccatggaaactggtgttgtgccgga
gtcaacagcatgtgtcacaacagcttcacacccaagaacgcccaaagggacagtagagcaacaacctgaccctgcaagccagatgataacagtaacacaactcagtccaatggacagagaagccagggtc
ctgagatacagagagaagaggaagacaaggaaatttgagaagacaataaggtatgcttcgaggaaggcatatgcagagataagaccgcgggtcaatggccggttcgcaaagagagaaatcgaagccgag
gagcaagggttcaacacgatgctaatgtacaacacaggatatgggattgttccttcattctgatactcctgtggcaaaaagaaaaactagattgcaagctgtaaattacttttagtttgagattatgttaggtttggtga
aattcttagcttcaagaagtattactactgttgtgcaaatgggtttgtagttttggctaattaaaactatagtattcttctttctctgcattagttaaacaaaaccagtctaattggtacagatccccatttgtgaaatatggat
catcaagaagtgaaaacgcagatttcaccttgctaactactaagagggaagccaaggttcaaagccaataacttcccgataacctgaacatgtgacactttcacgcctccctgacaacaaaaagattctacaatt
caagttgataacacagactcctcagctgtaaatcacaacttcttactaaaatgagtcatatcacgttcctaattgtagaaaccatcagccatatatgttctcggcttcaagaactgtaagaagaaataagtaataacaa
ccaagtgtatagcttatactatgtgttaatgttaagtgatggattcgagacctcacgtgaaaagttcataaagtgatataaggatgctgcactttgcactttttttgcacttttccaaaaatctaaggatagttcagatcgat
tctcaaagccatcaaattttaaaaccattcccaaactttacatggtccctgggccttcatagattctcactcttttaacttcttgcttacaaaaaaaaaaaccattctccttaatctgaagtaggaaaaagtctgcaactaca
gcatgagtttgacagaaaaattataacactctgctttctaaggttcttatcttgcttcaatgaccatatactaaaatagcaatttcctccacataatccacaaaaaactttggtgtcatgtgaagaaggcctgagagtag
ttcctaactaaccaataaccattacaaaggaaatccacagactgactaagaagaagtaaagacagaacaaatgaaaaatgccaaaacagaacaacacaagcacagctatatctgaacacgccttaaaacacc
atttaatatataatcatctctatacttaccagagagtcagaaaccagtataaaccctaacatttatacaaatcggaatcagtcaccatcaaaatataaaccctaacattcattcaaactgattttcgataaagaaacttac
tttcaacccgtcacaatagatttgggatgattttggataaatttgaggaaaaatatgcgaaattcaatagattgaatatgcaagggtttaaggatgtgaaattcaatagattgaaaaaacacaatttttttgctcaattaga
tccaaattcatcccagatcgattttgtctgacgagctgagattaaaggagtttaccagaaaaccccaaatcgggattgtgaagtgaaaagccctaaattttgagcgacgatcgatcatcaaacttcgaaatattcaa
caatgacgcacgcaactggtcttgtctaatactaacgggctgggccaaaggatctgatataaggcccaattattcttgcgcgataaacttctgtttcaggggacggcccaagtagtcttccgaactaagcttatctt
ctttcaggaagtttcctttttgttgttgcatgatcattttctatctttatcttatcaaaagataataatcttatcattcacgatgagttattctaagcatttcgacaacgaacttcataattgcacaacggatagttttaagtaaa
tgaatgatatttttaatcatctcattgatttcagaaagatcctcctcgaactatcctcactgagaatgcagaataattaacaagatccgcggccttttgaataatttacttttcttcaagtatatcaaacttggcttctctattg
ggtgaagagggtttggtggtagttgtttgtcttctttttatttctctacgtataattattataaaattatactcttctttt

>Arabidopsis thaliana COL1
tactaaaaccacctcgtttgcagcaaaacttttacagacgctctgcaagctgttggagcaaaagccgagcttgttttatacagcggaaagacacatactgacttatttcttcaggtaagccacataaatcagtcaact
taccttacattgtgtcacacacacaaacacttagctcagtgtgagtacaataaaactcatgagagtattgtttataggatcccttacgaggaggtaaagacgaactctttgatgacatagtctctgtgatacatgcgg
aggacaatgatgggctgaccaaagactcgttggctcctcccagaaagcgtcttgtaccagagttgttactgaaactggctcgtgagattagccctttctgactgccaagaaaggaaaacaagctttgtttgttgat
cctcagtttttacctttgttagttgctgtgtttttaattgcatcagtagaacatcatgtaaataaaaacaaactttgttgctccaccaaatttatttgacccatgtagatttgaacacatgaatatgattacttccacaagtttg
aaatctaataaaagctagtaaagataatatcgagttttatcatattctcaaaaaggtgattgagatgtgacaaacacttgatagacttatacatgttttgtaactatgaaatatgatggtgatcataagcaaattggactt
gtataaacagaccacaggtgatgttgatataactatcgttattcaccttaatttggacccaccatttgcttcaacttgtggctcctaatctctagagactcgtagcgacttggcagattccaattggctccttggatctcc
accgttggccaaatcagcttgaccgacttgatgtggaaggaaacaaagttaggacatgccatatctattatctccaacatatcaccaaaaaagacacctagcatcattcgaagataacaaatggttaagagtcata
atctttttaacaatgctaaacagtgatcatttaatttttataattgaaatcatgagacgatgagagaaacctgaacagttagaacatggcaaaagttaagaatcaagataaacattaggctaatggtcaagccaaaga
gctactaaagtttccagctttaccaaaatagtagtagatgatacaaaaatagttagaagttgcacgtttcagaacctaggtaacatgtggatgggtcctatagaagattacataggaagaggtgtttacacttattga
ggaagaaagagagaaaaagtagtaactattggtgccattgcaaagatataagaagacaattagttttttctttctttcacatttcgtgctggttgatgattccaaagctgtaagcatctctgcactggacccccttctca
aaaactcagccagccacaagatgatcactcaacatcaaacccacaaacttctcgccacgtgtccagccctcagaattctatccttcctttccatttccattcaacaactaatatagccttcttagtcccctgctgccttt

acatacgatctcactcaccatcaaacaactacaacatctgcggttagtatgttgaaagtagagagtaactgggcacaagcctgtgatacatgccgatcagccgcctgcaccgtgtactgccgggctgattctgcc
tacttgtgctccagttgtgatgctcaagttcatgctgccaatcgtcttgcttcccgccatgaacgtgttcgagtctgtcaatcatgtgagcgagccccggctgcctttttctgcaaggcagatgctgcatctctatgca
caacctgtgattcagagattcattccgcaaacccacttgctagacgccatcaacgagttccaattctgcccatttctgagtactcttacagttccacggccactaaccattcatgtgagacaacagtgacagatcca
gagaacagacttgtgcttggtcaagaagaagaggatgaagatgaagcagaggcggcttcatggttgttgcctaattcagggaaaaacagtggtaacaacaatggcttctcgattggggatgagtttctgaacct
tgttgattatagttcgagtgataagcaattcacagatcaatccaatcagtatcaactagactgcaacgtacctcagaggagctatggggaagatggagttgttccacttcaaattgaagtatcaaagggcatgtacc
aagagcaacagaactttcagctgagtatcaactgtggctcctggggagctcttcgaagctccaatggttccctcagtcatatggtaagctttcatatatttaattcagatccgttactatgaccagcatgctgcttcttt
cacacaaatccatgaagtaaatttctaagaatgtctttgttactgtcattaggtgaatgtttcatctatggacctgggagttgtgccggagtcaacaacgagtgacgcaacagtatcaaacccaagatcgcccaaag
cggtaacagaccaaccaccttaccctccagctcagatgctcagtccaagggacagagaagctagagtcctgagatacagagagaagaagaagatgaggaaatttgagaagacgataagatatgcttcaagg
aaagcgtatgcagagaaaagaccacggatcaagggccggtttgcaaagaagaaagatgtcgatgaagaggcaaaccaagctttctccacaatgataacatttgacaccggatatggaattgttccatcattctg
atgctgctatggcaaaactagactgcatgcttgtaaattacttctggtttgaaattcctgcaagttattaccacggttttatgcatatagggtctgtagtttactagttttgcctaactagtacgagttaattagatatacact
catacttagctgctgtaagttaatacaaattttgcccacttggtagaagtcctcttatctatgaaatactgccctacagtctaaattcaaactgatgctacaaagagatatgcattagtatagctagctgtaccaagaac
agtcaacacgtaactcaccttgcctaagttcaaagttaatgattcctttcttggtgacactttccacatcactttatcacgtagaagataaatgattccacttatgccgccccaaagacaatttcgcacatccctgagaa
aagaaagttgtactcaatatgattttataccaaaagtagattacagagacaagtgtaaacatgtaaaactccaataaaatcacacactgaatagtcaggtaacaaataaccaagaatcctgagtgaacaatgctaat
aggtacaattacatcaacgcatattcttgtactgattagtgattagcattggtgctaaatccattcccacgaggttgtctcggataacaagattctatcactcacaaccactactagcagaatatgtcaatgcctatgga
ccatccaaatgaacaaagatattcagtgtgagaaatctatttcttggtctatcttttagaataatagctcactttttccaaggtatatactggcactaaagaataacgacgcaaccaaaagaaaaagggtgacactgta
aagccttcactttttgccacctggaaagggccagagatcaacttatttaccactaggatagcttcccacttcaacggttaagctgatgaattcatgtaccaaatcaatactttttagccataaatgagtcagttttagtat
ccacatgaatttacctaccagagtgttgtaaattatgttcttttggggccacttacatggatctcattcattcactgcagcgagttctcagaccaccagaaaatttattcagtgatctgttttgatcatgcaacataaactt
ataagccacacaagcaaaacaaagatatcccatgttgcatataatacgagctagcatatcataaagaaggaaacttgaagtagcaaagtttctactaaatttcttgtcaggaatttttaaaatgcaatgacaaccact
tggagcactatgagtttcagagccaatagaatgttactatttggtgtggattcgagctagcacgtgaaagtgcataaaagtgattaccttttgccaaaggtcactgcacttttcctcagatagtttctcacagccatgg
aaagtggagaatccgcataaacgtacaattacaagctttatatggtccctcgactcttattctcttctcagtctttgcaactaaatagggttttcgttaatctgaaagaagcaaagtattcgaaaccacggaaacctgat
aaagaatgaaaacaaataagcaatagtgttttcttgaaaatctcgatgcaactttgaggatattgttacatatgatctattactcgtaacagttatccgaaggcctacacatgtgagagaagttccaaaccgctacaa
caataaacttaattagagactgtcaacgagcaataataagcaaaactactttttttcttgagctacaagtgaaaaggccaatacacaatttactcttcatgaactcgaaccacgttacaatctccaaaaaaatttcatcac
caaagcactaaaagccaaagatgcctcaacttatccaacttggcaggataaagatctccaaaaatgcttactaaagaacctagaatcttttctttagaattcaatgatcatataaccatttcataacaattctaaatgcc
attacattcatcgtaaaaccagtaaataacaagaacttgtatgttaagttccaattaccaagcaaaaaaaaacttttcaaagtttaaagttcaaaatgggaaagagaagtgcggtgtaagcaaatatgaaagaggaa
gagatgcgaaaagtgtatcctaggaccagcattttatacaaaaaaaaa

>Arabidopsis thaliana COL2
ctcccaaagttcaccaaaatggaagtgtactcaccaagcttaatttcgtctccatggctaagattgacgggcgtatctgagtcgatagtgtcagagttcaatatggtgccgttggaagaaccaagatcgtggattat
ccagttctcggaatcggaaacgatccggaggtgcttggtggagattccggcgtctttgatggcgatttcattaccacggacgattcgaccaattcggatagtggatccgggtttaaaccctagagaatcgccagc
tcttggaccctgggtgaagtcgagcttcagcaatggcggaaccatcgtgcgtctcggggaagctttgctttgattcaaaataatggaattttgaaatttcggatttatgtcaaactcaaaatcgaacggaacagtcct
gaaacaaagaaaccctttcaagcccaatgggcttttgcattgtacaaactatgaatatacagtccagtccataagttaccaaaatatcttcttgttactacagagtacagcccatccaagtttttttttgttgttgttgttg
ttgttgtaaaatgatgatttatgcatgattgactatcaaattgaaacaagaaatgagttatagttgatttagatatattatatatataatatttgttgatataagtaaaaaaccaaaacaaagggagtaaatattcatttcatg
atctaataatttcccaaaaaattaaattacaattttggcaatattgtttcagtgaacaactgaaaaactcaacttatcaacattactccacaattcacaaaaaaaaaaagatttatttcttggtaactggaagtgatgttgat
ataagctaatatataacattataacttgtggctcatgagcatcctaatctcttcactaaaggcttatttgtcacattctcattggtcactctagattcaccactttgccatgtcagctttttcttctcatccacctgcaaaaca
acatttaaaaccaaaatgagtttcacttttatgtattactttgtataatctttggtcactacctgattacttaccatgtatgttccaataagaaaagtttccatctttgcttaaaagcaaaaaactaaaataccaaaagattca
aaaaaaaaccaaaacagtaaaagtaaatagcaatctctgagaatcttatccacgtcagctcacgggtcctacagagaaagctacaataggaagagatgttttcacttacagaagatttacttggtctagatttgttct
cctttgtgaaggttatgagaaaaaaacttttttgatgattcatgaccttgttcaaaactcagacagccacaagatgatccaactacttttaacaaccactaactctgcgccacgtgtatatctccaaagattctatctttc
ccctctcaccacctagatatagtctcattagccttccctcctttctaattacaatctctccttttgccaacaatcattgcttaaaaaacacataccttcttctttagttataaaaatgttgaaagaagagagtaatgagagtg
gtacttgggcacgagcatgtgacacttgtcgttcagccgcctgcactgtctactgtgaagccgactctgcctacttgtgcaccacctgtgatgctcgagttcacgcagccaaccgtgttgcttctcgccatgaacg
tgttcgtgtctgtcagtcatgtgagtctgccccagctgcgtttttgtgtaaagcagacgctgcgtctctttgtacagcttgtgatgcagagattcattccgcaaacccactagctagacgtcatcagcgagttccgatt
cttccattatctgcaaactcttgcagctccatggccccaagtgagactgatgctgataacgatgaggatgatcgagaggttgcttcgtggttactccctaatccagggaaaaacatcggtaatcagaataatgggtt
cttgtttggtgttgagtaccttgatcttgtggattacagctccagcatggacaaccaattcgaagataatcagtacactcactaccagaggagctttggaggagatggagttgtgccgcttcaagttgaagaatcaa
caagtcacctgcaacaaagccaacaaaactttcagttgggtatcaactatggattctcctcaggagctcactacaacaacaactccttaaaggacctaaaccatagcgtaagctttttatacacttgccagctcttta
gttcacaataatcagtatgcatgctgtgtcatcacaagtatgcttcgctatacattgattctttgctcctataacctcggaagattgaaactgtctttgtctctgtcattaggcatccgtttcatcaatggacatcagtgttg
tgccagagtcaactgcaagtgacataacagtccaacacccaagaactactaaagagacgatagaccaactatctggtcctcctactcaggtggtgcaacagcttactccaatggagagggaggctagggtcct
gagatacagagagaagaagaagacaaggaaatttgacaagacaataagatatgcttcaagaaaggcatatgcagagataagaccacggatcaagggcagattcgcaaagaggatagaaactgaagctga
ggctgaagaaatcttctcaacatcgctaatgtctgagactggatatggaattgtcccttccttctaaactcctatggcgaagatagactagattgcatgctgtaaattacttctagtttggatttctacaagttattactgc
atatgggtccttagttttgtgacttttgtcaaactagtacgagtttaatcagatatgttcctatagctgcagttgtttaagcaacatctcgctacataattgaattggtgaaatcttatcatatggaacagtgcctatcatatc
taattgaattggtgccacatgaaagagatatgcatcactatgagtagctgtaagtagaaaactaaaagcggatttcaagttgtcacttcacaagaaagacaacgcctagtaggttcaaagtaaatacctttcgattat
tttctggctgtagaatcttctgtttcttgatactttccagatcacaagaacacgtaaatcgcaggtcaattccacatatgccaactcattccctgaggaaggattattcagaataacatgatttgtaatttaaacaattaaa
ctgataacatagaacaactttgtccacacaaacttaagaaactccaaaaaccaacatgcaaatcagcattaattgtcattttcggttaacgggaatcataagttaacgtattcagccaagtgggagtaagatagtga
agatactgagaaccctaatatacaccagcaactagcataggatgttgatgcctatgccaaattcaagtaaagaaggataattctttctctttaccttttggaataaaggactatactttctcactttttatggtatgtgctg
acagtaaaggatgatagagcaacgaaaaagaaagcacgacatttaaccaagccttactgtttcccatctggaaatgttcacagagctcaacattattatacatcacaacctattttttttttcttctcaattatacatcac
aacctatatgtgtagcaaaagtactgtatttatttcctaagttatgagagatctgaaaacagtgattacttaaccaaaataaagagcttacatcacaacctaaatttgtaacttgaaatgaaatgtagcaatagtgggac
aaattacgaaattctgcatcgtatacaattttcctataacatcataagatttatataaaagtgagtgatagtctaagtggaatcatagaaaccgtcagacatataagaagtaataaaaactaatataagacctagttgag
agttaatacagtatatactatatagtaatgggatcgagacctcacgtgaaatgtatcaaagtgacgaccctttccaaaaaggttactgcacttttcctgatcaacttaaaaagagtcccagccataaagttgagaag
ggatggtaaaagcgtcttccagactttttaaaaagcaaaaactcttttcatcagttgcatagagctcaagcaagccacactccactagagcgccatagattcctcttactatggaccacaaaacaggcggattttca
tcgagccactgatccaaaagtgtctaactgaagctacttcttagattatcagattaacgtgtttagagatcagggttgcaaatcaaagatattggcagttgcattcacataataatggcagatgaccaaagaagaaa
attgagagactaaaatcatgatgtagcacaaagtgagttatagagagataattaagaaacatatgccttaaaaatctcattcaaggaagcttttgtcaatcatatttttggatccctttacaattcctattatgccggtag
aaaacataaagttgtacatattaccttatcaatgagtcagccaagtcactccaagtgtatgcaaactctcag

>Aethionema arabicum CO
gaaacctatgatgtaaggaacctattatctttttctttagaaattaaatgataacggaacagttccataacattttacattacattcctcaaaaaaccagtaaattcgacaaagaggaatttgaaagttaaataccaattat
caaaaacaaaaaaaacttttaaaatttctatatcttcccttacggaacagaaaatgaaagagtaagagatacgaaaagtgtatcctaggaccagccttttatacaacaaactcacttttcagccgcagaggcataga
gttgaagtaaaccacatgaatttggcgactacagtgtcgaagcatcctgcgctttttggaccacaaaataggtatcatacatccactgaacttaaaagtgactggagcaagtactcaaacctaaagagtatataca
ctatagcttccaaataataaataattgtactccatccacacaataaaatttggcatgtcacttctaagtagtatgcattagatttctatgaaacatatacggctaccttgctacatactttgctactatgatggtgatcatga
gaaaaacaatatttctacgaaacatatacggctacttgatataactaaccttgttctcctttggactccaagatgctgttcagcttgtggctcaggtgcatcctaatctctaggctcaagactagagggcggatccaa

attgtgtcctagagaatccatcatgtctacgtttgccatgtcagcttcctaatacgaaaacactttccaacaaacaaaccatttcaaaatgtcaattacctaatgccctgtttcatgagtttctctatcaatgtaaatccag
tgctttggtgatgaccctggtagctgaaaccttacacgggtaaacctgaagtaaaattacaaccacacaatgttaaaaactaatattcaacagcataacctattgtggagccataaagattcaaaagtatctaactct
acttcactccatccgaaaatagtaaaatgtaaaacgaaaatagtaatttttatacgaagagactaaatatgtgtctgcaacacaatgacgttaaggggaacacttcctcttcagaatctccttgtgccacgtggcaaa
acctaagcaacctgtggataagtcccaccgaaaaatacataggaggagattttacactttgcactttgcactttacactttacacagagagaggcgtccgtaagtcttggtatggcttttttctctattgccaagatat
atgaaattgaagtttttcctaataattaataattccaatagccactgtggcaactgtggtccctatacttgagcttttctgaaaaactcagccaagccacaaggtcacccattaccttttataccccatcatctcaagtgt
cagctctcaggattctatctttcctctgtttattctcttccagcagatatagccttaattattagccccttccatcagatacaatttcctacaccattagactacttcactgagttataatgtcaaaagaagagagtaacgac
agaggtagtggagagaccaaccgggcacgcgcctgtgacacatgccagtcagctgcctgcaccgtgtactgccaggctgattctgcctacttgtgcacgagctgtgatactcaagttcattcagccaaccgcg
ttgcttcccggcatgaacgtgttcaggtttgcgagtcatgtgaacgtgccccggctgcctttttgtgtaaggcagacaacgcatctctatgcagtgcttgtgattcagaggttcattctgcaaacccacttgctagac
gccatcagcgaattccaattcagccaatatctggaaactcttgcagctctatggtcactcaccagcaaagagcgacaacagagaacagactggtggtgagtaaagaggtagcaggtgaagacgataatgaag
atgacagagtcgcagcttcgtggttgtgccctgatacggagagaaacagtgacaataacaataaccagaacaatgagttcttatatggtgaggagtatctagatcttgtggactacaactcgagtatggactaca
aatcgagtatggactacacaagtacaagtcaatacaatcaacaccaacaagactgtggcgtaccacagacgagctacggcggagatggagttgttcctcttcaattcgaagaatcaaaggccccccctgcgcca
tgagcaacagaactttcagttagatatcacatatggctcctcaagaaatcaatacagctactatgattccctaaatcataacgtgagtttttatatattatgttagcgattcaatttggcatcttccaaataggcagtacg
ctgcttctttcatcaacaaatccatgaacagttcatctacagtacataaactttttttgaccagaagtctataataaaaaattatgttaccctatgcatcgactctttctttagtcctcttaccaggaggattgagaatggcttt
gtttctgtcattaggcacgcattttatccatggaaactggagttgtaccagagccaacagccccacatccaagaacacccaaagggactatagaccaactaactggtccaattcagatgatagcccaactcactc
caatggacagggagggcaggctcctgagatacagagagaagaagaagacaaggaaatttgaaaagacaataaggtatgcctcaaggaaggcatatgcagagataagaccaaggattaatggccggtttgc
aaagagaagagaaaccgcagccgagaaccaagcgctctgcacaacactaatgtccgagacaggatatggaattgttccttcattctgatacttctagtcaaaagtaaaaactagattgcaagctgtaaaattaca
tatggtttgatttctacaaggtttaacaaagttcttggcttcaagggtcaatttattaccataggtacgcatatgagtctgtagttttgtcaactattctaattaatgaaaactagtccttttcagcagtagttaaacaacact
tctctgccttatacaattccacaatctgtgaaatacaccttctaagatattgccctcactctaattttacactgattgtacagatagtgatattcctttttgtgtagctgcaataagaaaagtaaatgcaggtttcacatacta
aatgggcggtgtcaaggttcataattgatagctttacagatcatatgaacacttggtaaactgcagaatcataaagataatggaataattcagcaattaagcagaatttatcaattccaaatttctaggtaatcacgatt
tacatcaaaaaaactctgtcattccgcaaaatcaatggccatagaaaatataccctaaagtttcaaaaaatcaagtaattagtgcttattatattatgtcattccaatccacaaatttacttcatcttccatgaatccacaa
atataatgttaacataatggacattgcaggtgattgcagtaggaatattcaataccgcttgcttatcgtatatatcactcaaaatatgtaaaagcatatggtcccttcatgtatgcaaagatgcttatatcatgaaacctat
ctcactcttatcttttagaataaaggggtatgtatacctcacttttacaaggtatatgaagcaagggaaaaaacgtgagattgtaactaaggattctttttttgccactatagagctcaacttatctatcactagcatacatc
cacaattgcacggttaagaaatgtgttcttcatgtaccaatcgattaacattgcaaaaaacaaaacaaaaagctcgctttaccacatggtatcagccaaagatctgagaacgatgttaatgcgccaaacaaaacct
ataactttcatatagtaagaacaaaaaacacttagttatctagtgaagaaatagcaacctagtaagaattgagtggataataatatctagtgatggattctagtcctatgaaaagtgcataaaactgacgaccttttgc
aaaagattgctgcactttttgcacttttacatatatctaaggatatttcagattgttttgacagccatggaaattgatccaaaacttataaccacaagctttacatggtccttgggacctcgtgcgatgctcaccctcaga
aaacaaaaaggctttcatatttattctaaaagacagaaaatgtctgcatctactgcatgaatggataagaaaatcttacaaataattgatactgtgatagcatgctgtatcagattcttctattgcttccggcaccacata
gacacaagtttcatgttatgtggggctatacactagggagaagtttctaaccatcattaaaatagattcacggattcgcaaacaatctttctttgaaccacaaatggaacagaccaacgcagatggctcatcatgaa
ctataaccactttaaattctccattctactttccagtctgagttctgtcagtgttatttagtttaccagagaaaggcttctaatgaattgatagccatcttaagtattaagcctagatgacacaaatatatccaatttctacta
ctgtatagacagaacactatggcaggatctgctgcacttatgttaacttatctgtggaacttaaaatccttctttcaatgaccatgaaacatttcatagcatttctcaatgccacccagaaacccagtaaacagatgaa
agaacctaatgaaaggtaacgagatatgaaaagtgcatcataggaccagaacaagtccagctctaacttctgaaattgcctttgtcatcctaaagacaagccctgaaaatttccacaccagatcctagagacatc
caaaaccctaatcttcaaataaaaaatgttaaaatcaaac

>Aethionema arabicum COL2
tatgtgtttctcttacagagaaataaacacgtttcttgttttctttggttcttccttttgcagcgaaacgtttacagaagctcttcgagctgctgaagttaaagctgagctagttatgtacaacggtaagacacatacggat
ctatttcttcaggtaagcgccgatacagaagctaccattcacagatcctatacgtaactcaatgtacactgtaacaaacaatggggaccacattgcattgcaggatcctttaagaggaggtaaagacgagttttttg
atcatatagtctctctgatacacgctgacgacaacgatgctctaacaagagacgccattgcaccaccaagaaaacgacttgttcccgagttcttacttaacctggctagtaaggtcagccctttctaaacacatgaa
acactaatctttgattcctctcttgtttcttcttcatttccattttatttgttgttgttactattctgtttctttattccaaatgttttatatataattacattaacgatataatatatagatatatgcacacatatttacttggcaaaacag
attttccgacaagcaatgttttattttatttaaaaacatggttaaggttttccatattaaatattaagacaatgggaacttaaaaatcctaatcttaacagagtcatccgttgaattggaccccatttaagccatttaaagtg
atcaggctttaacttaaacttgtggttcatctacattctaatcttcagactgaatgcttacctgtctgattctcattggttcctagaaatccaccgtgttacgcgtccgctacgtcactataatgtggaaacaaacaaagtt
agacaagtcggccctcatcttctacaaaaaccatctataagtcaaaaaggttttcactaaggtcaaacttttacagtctttttgtcagccacaagcattggcaacacatatacatccatatatatatatatatacacatat
atatagagatctcgtgatatacctggaccttaagaataaagcaaaagcaaatactaaaatagcacaagaaataaaaccaaaatagtaaaaagttaccaacttgacacactagacaattttcactcgttttaggcttc
actgcattctcgtgaacacttaatcagctcttgaggatctttttatggaagccacgtcagctcagatcctaaagcaaattgcagatggaccctacagagaaagtaacacacgaagagatgtttacacttactgagg
agttgtacatacacttggtgtagctttgttctcctttatcaaatgaataagatagactcggtcattctcttgatgcttaatgaatgattccaaaaagccaaagtagggtcgtggtccctaaactcattcagccacaagat
catctaacaacttttacacaccacgaattgtgtgccacgtgtccatcctctagattctatcttttccctctgctttttctctctctctcttaaatcagatatagtctcattagccgttctccctttaggtacaatcccccccctc
gaaactttaattagtaccaatatgttgaaagaagagagtaacgagagtggtggccgagagaacagctgggcacggacatgtgacacatgccgttctgcagcctgtactgtctattgtgaggctgattctgcttact
tgtgcactagctgtgatgctcaagttcacgcggctaatcgtgttgcttctcgtcatgaacgtgtccgggtttgtgagtcatgtgagtgtgccccagctgcctttttgtgtaaggcagatgcggcatctctctgcacag
cctgtgatgcagagattcattctgcaaatccattagctagacgtcatcaacgtgttccaattctcccaatatctggacactcttgcagttccatgaccactcaccacccaacaccaagtgagactgcagtcacggat
gcagataatggaatgatgatgggtcaagaggaaggtgaagacgaggatgaggacgaccgagaggtggcttcctggttattgcttaatccagggaaaaacagcggcaataacaataacaatcccaacaatgg
gttcttgtttggtgttgagtatcttgaccttgtggattacaactcgaatttggacaaccaatttgcagatcagtaccagcgagactgtggcgtgccgcagaggaactttgggggagacggagtcgtcccacttcaa
gttgaagaatcaaacggtcacctgcaccacaatcaacagaactttcagtcaactatcaactatacttgttcttcaggagctcaatacaactacaatagttccttaaaggacctaaaccataacgtaagttagctttctt
catgataattagtatgttgtctctttggctattaactttccaacagatgactgaaaatgtctttgtttctatcattaggtatccattttatcaatggacatcagtgttgtgccagagtcaacagcaagtgacataacagtccc
acacccaagaacacccaaagggacgacagaccaactatctggccctcctactcagatgatgcaccagcttacgccaatggacagggaggctagggtcctgagatacagagagaagaagaagacaaggaa
atttgagaagacaataaggtatgcttcaagaaaggcatacgcagagatacgaccacggatcaagggccggtttgcaaagaggacagacattgaagcagaggcagaacaattgttctccacaacgctaatgtc
cgacactggatatggaattgttccttcattttgaaaggcatatacggactagattgcataatgtaaattacttttagtttgaacctcagcaagttataaccatagttacgcatatggatccgcagttttgtgacttttagctt
attaattaaatatgttcctatagcagcagtagtctcctaacatggttgaaatccttaacatatgtgaaatattaccttccaaaacatttcctacgtcattatgagttgctgttaaaagaaaagtaaaaaaggggaatttca
ccttgtcatctcccaaggggtaccacttgaaaacttaaatcacaggcaattccagtcctgccaactcatgccactttttttaatcgtccctgaggaagaagttttctgaaaataagacttatgaattaaagatttaaagt
aaatgaaaacataaacaattctgtagaaactgtagaaactccacaaattcagcaaaatcaacattaatccattattttctactaatgagaaccattcctttaacttattcagcctagtgggagaaaggtcgtgtagaca
ccgagaaccccgagtggaccattctaatgtgaatggttctgtctaagcatattctttcttatgaatggcactggaggtccaaggccggtaccaaatattcattaccgcgtgctctctaagacttttatccctttcagtatt
gatgcctatgcgcaatccaatttagcatagatactcaattctttctctttttttatcttttcgaataaaggtctacactttctcactttctatagtataatgctgacgttaaagatgacggagcgatggaaaagaaagagactt
atggtgactgagccttacctttttttccacctggaattgttcacagaactcgccatatctaccactcgcatagcttcacacttctatggttaggatgatgacgactgtatacccgtaccactcaataactcaactgtataat
aagcagattcttctttagattttatctaccggagtatataacacagtgattagttaccaaaattttaaaatgtatctatgaaattaatgttgcattagtaagataaaaggccaaattatgcatcctagaaatgcgtcctgag
ttatccaacataatatttttattagtgataaataatccaggtgcaatattgaaaccatcacacatatgttttccatctgaaattctgcttcggagcctaaagatgcaaaccaaactaacttgagaactaattgactgatgat
actatacccactacccagatacttcacgtgaaagagtataaaagtaaccacactttccaaaaggttactgcacttttgcactttttgcactttatagatattaactatctaaaagctcacaaccacgaaagttgagaag
gcatccaacacttacccttaccaactttacatgatgctcgggtcctctgaataccattttatgcctttgttaaaagggattcattcaccagaaaaaagtttaaaaaaggaaaactacatgacttgttaaatattagaaag
aaatatgagataagagtaacgcaaccaatcaatttcttgttaccattcccttacttcagtgttgagagatcacagctactatacaagaatcctcgcacgccacataatctggacaaacacgtcgccctttgttagaag
agcaaacctcacaataaacactgaaacagtcaataaagaggcttaagcattcaacggaaatacatgaatttatccattaattaagtaaattcagctactcatcattcgattagtccgatcgtatcacctaatccttaga
tttcctaccaaatggaaatcaaaatcaaacaccacaaaaattcgagaatccggtaaaaaaatcgtacgtagcttcaaatttagacgattgcgagtagctaattataaaatcggtgatcacgcatcccttaaacccta

aaacatcttcactcaaaaggatcgaggtaatcatcttcgtcaaaactttttgaatattcaagtacatcagccccaaacaaaggcgtatttcacattgcgagatacaaatccgtaatcaaaaacagtcacctcctcagt
acgaatagccacgacgacgacgacgaacgatcgccgcgatttgattctccgtttaggtgggaaactccaatgccgcgtagcaaaatattgatgtagcaatcttaaactacacgcgcctaaacatatatgggcca
aatatccaacatgggcttttattataaaattataaagcccaattaaaaaaataactcaatagcaaagtaacaactagagtccgaagctaaggaaagaaacacaaactcgcagcgtggtcacgtcgttaacacgcg
ccaatcttgaaaatatagagaaagaaactcagttctacaacaacgttttaaaacattcaacaacaaaattactggtattaaaaataaagtctcgagatttttcttattgaaggagaacaataccacacaacgagagag

>Tarenaya hassleriana COL (BAC Cs1 Gene Bank ID: DQ415922)
tttacttgaatccgcgtttagaggttgaagaagaacatcttcaggaagttaccttgaagcggttcgttccagagttcttgctgaatctggctcgtgaggccagccccttctgaatagtctccgatcctcttgcagtgtg
gatcatgaatcagtagttatagcatcctgatctgcatttctcctgttaatattgttgtttatgggtttttttgccctctgtccttttcttgttcatatgtagaaactcccatcccacctgttgataatgaacgaaacatggtttcc
ttcagctcccgcatcactcaaaatctgaaataaacacatcccacgagttaaaaaaagcataatgtagttgtactctactgcttataagttggataaactcttctattttgtagtcattgtacttccttcggggaaaggag
caaaaaaaaaatatgcaaaaagatgattcaacacacataataaaacaaaatttaaatccagttcattaacataaatcatatagaaaaaaagtggagctttttgatggttggggatcaagaaaagacgattcggctaaa
tcaggagagagacaagagagagatgatactaacctctcgtttaggaaccaacctcgaagagatcaggtttggcaacttgtggctcagttccatcctaatctcttgactacatcttacttggcagtttccgattggct
gtcaggagcatcttaggctttgtcctcgtcagctttctctccttccatgtatatcggcatgttccacgtcgtccctgatgtggcaaaagcaaacaaagtgaaacaagtcagcttcatctacatccacacacagggac
caccttaacaccaaagatcaccggctaataacattatgatcttcaaggatatgaaaccaaaaaaaaaaatagtttaaaaaagtaaaaaatcatgccagagaaattcaaaacttcgcaactttatacacaaactaatat
ggaagcgaggtttatattaccacatcgaggctgtccactgagtcggcaacttcaggcgttacaacacttggaaactttttgtctgccacgtgtcggacactgagcaacttacggaagcccaatgaaggggaaca
taggaagagattttatttttccttattacacacgagagtagcatagtgtagtgtgtgatctcttctcttataccaaagtgtaacaagaaaccaagtattttgctggtgtttaatgattccaaaaggccaaagcagcaaaa
aaactgtggtccctttgctcaaaactccacaaaaactccgccagccacaagatgacccaacaacttttgtatgccactagttccacaccactagctcccgccatgtgtcagtatggccctcagaggattctatcctt
ccttcctctgctcatattttcacaacatttatagccctttccctccctccaacttcccctaccatcagacctcacattgagacacagagaggggtaggaacatagttgattgattatgttgaaggaagagcgaaccag
cggtggagaaactggagagaacaactgggcacgcatctgtgacacatgccgttcagccgcctgcacggtgtattgccgggctgattctgcatacttgtgcacgagctgtgatgcccgtgttcacgctgctaatc
atgtggcctcccgccacgagcgcgtctgggtctgtgagtcttgtgagcgtgccccggctgccttcttgtgcaaggcagatgccgcatctctgtgcgcagcctgtgatgcagagatccattccgctaacccactc
gctagacgccatcatcgagtcccaattctgccaatatctggaagcatgtctggtcccatggccaatcaccaccccagtgagacagccatgacagatacagagaacgacatggtggttggtcgagaggaagct
gaagatgaggacgaggatgatgaagaggcggcctcatggctgttgcttaatccagggaagaacagcgggaacaacaataatcagaacaacgggtctttttttgatggagaggccgatgagtatctggacctt
gtggagtacaattcaagcatggagaaccaattttcagatcaatacagtcaatcaatcatcaagattgtggggtacctcagaagagctttggcggagatggagttgttccgcttcaggttgaagaatccagaggtcaa
ctgcaccacgaacaacaaagctttcagttggctatcacctatggctccccaggggcactgtacggcagctacaacggctccatgaaccacagcgtgagcttttctggattttcatcaataatctatcaccaaaga
gctcaatattttaattaggaagtctagacaaataagtaaatttccatgttcaatgattctctcttgtcttcccaagtttaattgttggaatttttgtttctgtgattaggtatcaatgtcatccatggacatcgttgttgttccaga
atctacagcaagtgatatggcagtagtctcacagttaagagcacccaaagggacaacagatctactaattggtcctccaattcagatgatgccccagcttagtccaatggacagggaggcaagggtcctgagat
acagagagaagaagaagacaaggaaatttgagaagacaataagatatgcatcaagaaaggcttatgcagagacaagaccacggatcaagggcaggtttgcaaagaggacggacattgaagccgaggtgg
atcaagcgttctctacaacgctaatgcaagaatctggatatggaattgttccttcattttaaaatcccatgacatataaggacctaagtaaatgctgtaaatttcttatggtttggatttatgcaggttgttacggtagctat
gcatataggtctgtatttttgcctaactattaagaatcaaaggagtatatcattcggtctgctacaatagttaaaaataaaacatcccatgacatgttaaatttctatcatatttaaccttctggccaactactttcacactggt
gtcatgggagagatgtatgccttgctataactggccataaagagcaaagtaaaagcagatttcaccatgcttaccagaccagaggaaccacagaggtggaaagaaaatacctttgatccttggcaatagatttt
tgtgctacttccaaatcacttgaaagcgtacacgctggaagattccacttatgccgcctcttggcataatcttttatcgcctctgagaaagaaaaaggttatgctttgcaagatcttgatttgtttgaggcaaacatcat
agagctattatctctcatccattcccatgagtagcagaaaacacttgcattaagataaaagtcatacattatccgtgcatttgtgttaataggaatcatcccgtgacatactcagacaagtggaagtaaggtagtaac
aacactgagaatcctgaatggaccacactttaatgaaaatggttaatccacgcatgaattttgtggtaagcatacggagatgttactaatattttgtccttaggcatcttgctgacagtaaactactggcagaagatg
ttaattctccgatccatcaaagttcacaaatgacattcagcttgtgaaacgtatttgtcacaacatcccttttgagaaaagggctatactttctcacattttctgaggaatacgcttacataaaagagtgacgaggcaag
ggggaagagcaagtgacaatgtactcaggcttaccattttcgacacctgaaagaaccatatagctcagcatatctaccacttgcacaaaacaggtggctcagctgatgaattcctttcttcacacagatattatcttc
taaagatctgaaagctgccgataagctaccaaaataaatcacgtataccataaccattatgcatatcttacaaacagaaatcatgtaatagtagcaaaaaagaaacaaattgtgcatcgtaagttcgtaacacaaaa
gttctctaaggtatacaatatacatcagaaattccatgtaagagtaacatgctgggcacgttacagaaaccatttgacataaatactgaatcatccaactatgcattctatttgaaactctgcttcgggattttgagatg
caatggaraccaattctattcctaactgggtgttaatatctagtgctatatctgagacctcaagtgaaaagttcatcgaggtgatgactttagcaaaaggtcactgcactttgcacttttcaagagatatatgtagata
ttttagagaatttctagagccatggaaagtggagaatcttcaaacttacattgataacttttaactgatcctcgggccctatagattagattcaatttctgtcattgctaggatagatgcctaaatctttcaccaaaaaag
aagagaaatcattctcaacttctaaatgaattgttatagttaaaatgaacaagtgatggaagcatcataacaccaaagttcagatttcagagtacccaccatatatcaaaatggcaatttcctctacatcacccagac
aaaagaacagttttactagggggctataatgtatgtcagagaagctcctaattaactactgacgagaaactcgcagattgtctgaaatcaacaagcaaagacttctcccccaaaacataaaacagaaccctcatg
gacttgaatctcggatggttwtcaccccrrrtcaysgtagtcgaattctctttagtaaataagtcaccagaaccaaaaagacagattaagaaagggagaagtatagatcaaggtctatgttcaatcaatcgtctttg
gagcaaacatataaaagttgacatggctcattcagttactacttttgtacagacatcaactaagacgccccggagttctgaagatgttacacatttaatggtaccaaaagatgttctttagccatttgatcggatccc
acataataaccattctaagggatcaatagcgattattcttcaaagatctcaatggtcatggagctatttctacaagcctcacaaaaacagatcataacgaaaagggcaacaaacagatcaattaccagcttgcaattt
cagttttcctaagtccaatactaaccatacaccgaaatgagatagaagagctaagctactggagaaaacgaaagaagatgaaagaaggaaaacgtatcttcgaacaaacctttatgaaagcatactcgcttttct
gctttggcgctaaagtaagctacgaagtggaaagagcgttgcaaaattcattccttatggaccacaaaaatggatgtccactcactccaaaaaatgtaccattgaaatttgtcccgcaaccccaaagcgtatccgt
tactgcaatc

Promoter sequences used for *cis*-element comparison (4kb):

>Medicago Truncatula pCOLa (TC86982)
gattctaaagactatttcgctgcctttacaatctttgtctaaattcacttgttcgaatatcaaagctccactttcgttttatgaattgatatttgagatgatggataaacaattatcaaaccctccactttcgtttccacagtttg
ataatggttgatccatgttaaagcggtgaatttggattagagattagggttacataagattaatgtttaaagcttggtttgattaggattatgtcattagacttatccaacaatcccaagacaagagaaatctactcaata
acgtgcattgtagacatagagaagaagagagaaagcataaaagtaaataacaagaaagtaaaggtgaacttttattcaaagaaacagttgtcacttattatttccgagaacaaaatatgaatatttgtgatggctag
ctacgctatttatagtttacattcaacttaaaatctaagcaatcactagaatgttccacaagtaaactaaaacagagtagcttaaaatctaagcaaaagcagaaaaagtagttctagagtgtggcacggccgtgcta
agcctagcacgggccgtgccaagtttatgacttgagggagatatattttgtctccgaatcttcatattttcataccaaatcagctcccaaaaccccttctttgctccaagactcaatccatcaaatattcattcctgaaa
tataacaaaatgcaattgtagtaaaatagttcaaaaaaaaaaaatattaatataaaataaatttaaatctaattaaaacgaaactaaataatatattaaaataaataataaatgacccatcatctatctacacaaaaaaaa
tcaataaaacattatttccaaatttaggcatcattgaacctcgaaacgtatcaaacataaattgttagctaataaccagattttgagattgtgctgtcccattgaatgcacctaagatgttatttaatatgttgtcacatcctt
cagtacttactaagatgttaagtagcataaagatctggaggtaagttcgttgctgtttctccaactcctggtccttttattgaatgtaatgtgcctatgatggtacaatgacattttatctaaaaattgttgtcccacctcat
atgttgatacttacattcacacaatatgctccacaatatttagcagttcccccaatttcagataggtattgtttcattcctttgatcattgttgaaaactattgcagcatatgcatccaacattttttttgggcgttagccctcc
gattctcaggggaaatgaacctttgtaattcagagtttgaccgtgaaataaagatattctgatcgaccaataattattctaccaaagaatggaactcgagattaaaatattacctcctaacaataaagcggagggctt
cacacgatccccgatctatcccgaaagcatatctcactctgttcgatcgaccttacccacgtatgagtaactatgatcaataatcaataacttacctaacatatacacattttgaagctaacattagcccaactgccaa
ctcccaactatactatgcttaccccgttttggaattacctattgttagttactaacggcataggttctactaggtcataaatgcataccactttttgtctaatataatatgtaccaaatggaaaatttcgattatgaaggtact
acttttatattaggcatggattgtggtctaattgaaaccagatttcacaatcaatttatatatgtatacaagacttattgatgtatatatatatagtatcactttttagcattaagataagacaatgctccctaaacagaatac
acacaaacaatataataaaattgagttaaatcactagcgttaggtacataaaaaagttaataacttcttattcttatgattaaaaagtggaagtcaacaaaaagtagtcactttcatgcattagaaaaatgaaacctagt
attaaaatcatttttaaggtcggataaagtaaatatccaaaagcaagaaaatgtaatagattaaactaaagcagagccaagatccattatatttataatcacgaaataatccatccgattttggttatatatataaacaa
ctaaatggaaacatttttaggtgaaagttaatatcatatgttaaatccataatgcaaaacaaaggagaaatagaaggatatattaatcatcgaattcaaattgaataatttaaatggacgagggccgcagacgttttag
gtggtgatagaatcagacatgacaaaactatacatgatataattagaaatgtccgtgttagagagagattggctacgaattaacattttgcctactattagataaaattgcatataaaattgtttgcttgagattgactag

gcattagctacaaaaaagtaatattttttagttgttacaaattagactctcaagggttactaaaatcatttagtaagacgggatttacctaacatttatcaaatgttacaagatttatatagtaacattgtttattatttagttaa
aattttcaagtattactaaatctcatatatgttattgtgatgatagaatattactacgtcgtttatctttgtagtcaactctacttaatgaaataaagcttgattattttatttgtaactattcgtcaaccgtagtgtccaagtttta
aaaaagactagagtcaaggagggaggggagtattttgttgcataggaataaaaagggtcttacaaaaaaagttaattatattgtaatagttgaatattttaaaaatgacatcattaaatagatataaaaaataactaaactc
tgttgaaaagaaaagttttgtttatatttaagatcaataaatatgtatatttgtttaaatgtaggatatcgttcccttgtcatttaggtaagctctaaagtgttatccaccacatattttctccaattcacaactacaactaatttg
aagtcaagattgctagcctaccacataataaaggcttaaagaaggaaagagtagcacttgcacatggtcttatcatatatatatcttcatttttatagctggaattaaagtcacccctccaagttaatgaattatgacttt
aaattcttagtagggtatggatctctcatctatattggtgtgatgaaagatcaataatataacatctcaaatatcaagttaattaggataatatataacatctcaaatatcaagttaattaggataatttatgcagcaaagc
gtaacaatataacagaagataaactaaataaatctagtggatgcatgagctatattaggctgaggttaaccactgcatatatatgatttatcagtacttacggaattgaataggaaacaccttattattgcaattgcagt
cgcatatgtggttacgctacaaacctttatattgtggtataacgcagataaatatggttgattctgtcacaattgcgattgtaatgttgtaaagactataaaactgttatattacaacctcaatcgcgattacggctcacat
tttaaaaccacgacgagaaattcaaagggtgaaacaaagaaaaaagagataaaaaaacacaagaaagacatcataggacaaatttaacaagcatatactataatattatttttaacatgacattttaagatataaat
aaaccattattagtagtttgtttcttctactcatgattgttaagcacattaaaaatttacaatctcaatcctctgaccccaaaacttgtccctaccacaaagcatagccacccacaagatagataatgtcacaactttccc
acaccactagttttctgccacgtgtcatacacagattctctttatactcccaacaccaccaaaaacagcaagaacaatcactgcaaatacaaacttcacctatttgaacttccttaacactgtctgtccaaaactgaca
aacaaaccattcatcactccctctctccacaaatg

>Tarenaya hassleriana pCOL (BAC Cs1 Gene Bank ID: DQ415922)
ctcggccgagacatgttgtcccgttgtgattgtccgaacttcttctccataaaatcgccaaatataataatcttcgacaaatccccttcgaaatagatgtttacaaacatcttctgtactgcgatatttaaaattgcgata
acttgtacatggacatagcatttttccacgttcttgacacaacggctgcatagaagcaaaattcataaactgttccaaaccgttaatgaattcttgcgttacctcttttgtaatcgaatcaactctttcgtatatccactgtc
gggacgaatgaatattcggatttgttgaagaattcatttttaggagaatttttggtaaaagaatagtataagagtgacttagagagtggtttgtgcgaattatggccgggtgctttaaatagagaaagtttccaacgga
tatccgttggaaatccgttgaaaagttcgttaccgttacacttaccaacggatatccgttggaatatttcgaacgttataacgttcgactatttccaacggatatccgtcggaattattgccgcgtttcataacgttcgaa
taatttcaacggattgccgttgaaattcgtcgaacgttatttccaacggaaatccgttggaatatacgtggaaaatttccaactcaaatatccgttgatatgccgttggaaatttccaacggatttctaacggacgaatc
cgttagacgtccgttggaaatccgttagaaatgcgttggaattattccaacggtataccgttggaaatccgttggaaaaattacaacggattaatccgttggaaaaaatccgttggaatttcgtattttttttgtagtgtag
ataacacaacttggactgccgaagatcgagctccgggaaaattccagtttgaacctctaacagccttagcgacagcagcatataggggaattctgagagcacgagggcctgcagccccaatgcgctctataa
gtggccccatccctgtccagtcatcgtgccaaaaagttgttgtcttcccattgccgagtacacaactgagaaattcccttgcatggtccttgtacccaagaagttttctgagatggcaagactgtcgctgtgaacgc
aaagctccccagaatcccctcttcttaaagacgttttccttcaaccaagccacccatagagatcccgattctttaaagatcatccaaatactttttaacctgaagatggaactgaagtcctccaatctcctgatgccca
gtcctccttcttctttcggttggcatactagcttccatgcaattctcgctttgctcctgacttcaatcgagttcttccacagaaaggttgaacacaacctgtcaatctcctttaatagatgttttgggagcataaacactcta
ctccaagcgttcaccatcccgtaaatgactgatgagattaacaagatcctagttgggtatgatagaaacttataggtccacgatgtaatcctcttcttgaccctctcaagtagcggcatgaaatcctgtcgacgcagt
ttccatgggcctaagggaatccccaagtatcggactggcatcagtccaaggttaaaacccactgaggagctgatgtccatagcttcctgttgcgttagtcccccaaaaataactcagttttgctgtgatttagggt
tagtcccgagtattccccaaaatcccataacagctccttgacatgcatgattgcatcactttctcccttgacaaagaccatcaaatcatccacaaaagcaaggtgagagataagcggagaaccacacctcggatg
cgatgatagccttccttggagcatggcaaagtccaaggatcttgacagtgcctccatcgagataatgaacaagtatggcgagatggggtctccttaccggagacctcttttgcccctgaaaaagccagccaattc
tccattaattgcaaccgagaaggaagggctggtaatacactccctgatccatccaatgaagacgaggggaaggtccaaggctcttagtatatccagaatgaaatcccagctaattgagtcaaacgccttcttgat
atcgatttttagcattgcacttggctcggccgagggtctattgtagtcctttaccagctccgctgcaagcataatgtgatcacccatagaacgtcctttttggaatgcagcctgattccctgagattaaccaaggaagt
agaagttgcagccggttagcaatgattttggaaataaccttgtatagaaaattacaacaactgattggtctaaagtccccaaggttaatggcttctgggatcttcggaaccaatgtgatgatggtcgcattgagctct
tttagtagttttccattgcggaagaactccatggttgcttcaattaggtgatccccgactgtatctcaagctgaaacaaagaactccttcgggtaaccgtctgggcccggctctaagtttgtctcaccctaatcatattc
tgactttacttgaatccgcgtttagaggttgaagaagaacatcttcaggaagttaccttgaagcggttcgttccagagttcttgctgaatctggctcgtgaggccagccccttctgaatagtctccgatcctcttgcag
tgtggatcatgaatcagtagttatagcatcctgatctgcatttctcctgttaatattgttgtttatgggtttttttgccctctgtccttttctttgttcatatgtagaaactcccatcccacctgttgataatgaacgaaacatgg
tttccttcagctcccgcatcactcaaaatctgaaataaacacatcccacgagttaaaaaaagcataatgtagttgtactctactgcttataagttggataaactcttctattttgtagtcattgtacttccttcggggaaag
gagcaaaaaaaaaatatgcaaaaagatgattcaacacacataataaaacaaaatttaaatccagttcattaacataaatcatatagaaaaaaagtggagctttttgatggttggggatcaagaaaagacgattcggc
taaatcaggagagagacaagagagagatgatactaacctctcgtttaggaaccaacctcgaagagatcaggtttggcaacttgtggctcagttccatcctaatctcttgactacatcttacttggcagtttccgattg
gctgtcaggagcatcttaggctttgtcctcgtcagctttctctccttccatgtatatcggcatgttccacgtcgtccctgatgtggcaaaagcaaacaaagtgaaacaagtcagcttcatctacatccacacacaggg
accaccttaacaccaaagatcaccggctaataacattatgatcttcaaggatatgaaaccaaaaaaaaaaatagtttaaaaaagtaaaaaatcatgccagagaaattcaaaacttcgcaactttatacacaaactaat
atggaagcgaggtttatattaccacatcgaggctgtccactgagtcggcaacttcaggcgttacaacacttggaaactttttgtctgccacgtgtcggacactgagcaacttacggaagcccaatgaaggggaa
cataggaagagattttatttttccttattacacacgagagtagcatagtgtagtgtgtgatctcttctcttataccaaagtgtaacaagaaaccaagtattttgctggtgtttaatgattccaaaaggccaaagcagcaa
aaaaactgtggtccctttgctcaaaactccacaaaaactccgccagccacaagatgacccaacaacttttgtatgccactagttccacaccactagctcccgccatgtgtcagtatggccctcagaggattctatc
cttccttcctctgctcatattttcacaacatttatagccctttccctccctccaacttcccctaccatcagacctcacattgagacacagagaggggtaggaacatagttgattgattatg

>Populus Trichocarpa pCO1 (POPTR 0017s14410.1)
caattgaaattgagagagacaacgtgcttcttctttttttcatgtttttaaaaataaaaaatcactaaattattcaagaaataaatttattttatatttatattcctccaaccaaacatatgtatatttgtattattatttctattcgt
tttgtatttatcttttctcttaatctacgttatgtctaattcatcaccaccttgattttgttcgattacataatatacagacaaagggattgtttaaggatcaatttatcttgtagttaaatcttaaatttccaacctagaaaatatc
agatttctatgatcataagaattcttttactaatgggtttttttttcccttttattttctgtgaattattgcaataagctggccgtttaattcaacatcttggctgccaagtgggttttaagagattgggagacctaaaaagagtt
gtagataattaagtactcaatgaatcttagaataaactaatttaatctgaagattgaattaaaaaacagaggaaaaaacagaaaaggtactgaattaggacttcatgtaacaaatatgcaatcctcacacggcaaat
attctttttgaaggcatagaaaaaaaaaaattcttttttttatatgataaagttgtcgtggaaaacattttaaggttgtatttgtttttgcggtttaattgtactttaaataaatttatattttttttcttcaagttaatttttttttggtgtt
tttggattgttttaatgtactgatgtcaaaaataaaaaaaaatattatttgaatgtatttttaaataaaaaaatattttaaaaaataaattatactacaattccaaacatgctccacaaaacactaatttggtatctcaaaatctac
acacgtaaaaagtagactaattcttttagaaaaattactacacacaagattaagattattattatatagtgtaatgataaaattgctctcttaactattaaactttttagctagggttttagggtaagatgattattattatata
gtgtaatgataaaattgctctcttaactatgacacaattatttttacaaattatatgaggttagattgatcattttgtctttttttatatagtaaaattactaatttaaccttgaaatagaatatataatgtctactactcatgaac
atttttttgtctttttttttattaatgtaattactttgatgtccttaaagaaaaaaaccaatgtaactcttacctaatagctttttagttattttacattggttttaccgtaaattttaggggtactcaaggacattcttgtaaattaaatt
tattaaatataatttaaaaaactttgttggaacgagccaacatgtggatggcctttatgtgtgggattgtccggccaccaaacatcgccttctaggacggcgtttgaaagctctcgatgctcctaaaacaatatggtg
gcacatgcccttaatgcgcattgagtttgacctcaataaccatttttttttcatttttttttctcatctcacctcgaatttcacgtcagcctctgcaaaacaacataccaacacctacaatttctttccccttcatatttggtcctta
tcctttctattactatttttttttatttttgactaagtcatgaaactacaaatattttttcaatttcacattccttttatttttttaatttttttagatttggtcctcattcttctgattactattttttttttaataagctataaaattacaattgtttttca
atttcaccccttttattttttttatctttcaaatttggtccccattattttgtttgggatctttttttagtttttttttttttacaatttcatcatccatgagttttttttcctttcaaatttgattttcattctttttattgctattttttttttttgtaagttttttta
aattgatatatttttttataggattagtctcaggtccatgatcagagttactagtttcaaaggttaatgtaggttgacttttttttatgtcctttaaaaaaaaagatttaacatttttttaaaatgatttattttaatttcatcattcaacaatt
aatttatttgaaattcatctttctattttttttttctattgggttatcttgatcctatgcttatgattttggggtttcggagttaacctgtgtcaactaatatttgtttttttgttgctatgagattaatcctaggtttatgactatagtcatc
ggttcaaaaggttaaggtaaatttatttttttgggccttttaaaaattaaattatttcaatttcatcctttaatatttgatttattggaaagtgattttcatattttttttcttcatttttttttctattgggttatctcgatcattttattttttta
caatttcatcctccatgagtttttacttcctatcaaatttgatccttatttttttttattgttatcttttttttttttctttggccagttttctaaattgatatttttcatgatttcacccttcaaaattaaattggttgggaattaagctttttaatt
gaactcggtttaaggatttcatgatttccatggtcatgaggtttgcagattaatctggttgactcatatttttttttcattgcttttataaattaaatttctttttttcatttttgtccttcaacattaaattattcaataattaatcttcatg
atattatttaatttttttttcgaatgcaattaccctgatatcacgattgggtcagagattttgcagacccaagtgggctcagatcgagttttttttttttaccttttaattttttttttcttaattttcatcgttcaacattttattttctaaag
attgagtttcaccttttttccttttttatttttttgggagctatcatgttctcatttgatttattagaaattatttttttaaactttttttcctgcgtttctttatatgatattatcccaatattatgttcattacaacaaagtttgcgtgttaacc
cggtttgactgagtttttttatgtttttttaattttttattcaatttcaccatttatcattgaattatttcataatccgggttgactcaaatcaagttttttttatattttttttattgtcatttttttcttatatattttttttgttgttgtattgttttaatc
atattatttaaattacaaattaaaccaattacttgagtattaattcttttttaaatattttttatattaaaaaaatattttatttagcctgcgggatactgtctatcacaaatctggcctgtttgtttgtatgtttcaaaaacgttgttgaa

aaaaattaaatttttttattttttcttactttaaattaatattttatattatttgatatattaatacaaaaaattattttaaaaaatatattatttaaataaaaaatattttaaattacattataaaacattcctttttatcagctaccag
aacttcaacgttacaacgtcaaaatttgcatctgaagaaagctgaaaaccccaccaaatgttgaacccattacactccacgtgttattcacatgttacaacttccaaaaaaacatgaaaactaatactctctgactga
cacgtgtacaggcagaatcttacccattacagagcggaccaacaagggcttgactgtgacaagtgtcttcataagcaagcgagtatcgcaatctttctctattccccgcagtgtactccaaaccccaaaatggttc
taactaagagccatagcagaaaaccatgtagtaactctctctcaactgggacccactccaacactcaaacagccacaagataatgaaacaacttctcagccacttaattgatgccacgtgtcacatccttgatact
ctcctccccacttgtgtttttcctacaggtatccctcatccttgcaactaatgctcactcactctcatacatagctaaagacagtataactgagattggtattgggagagaaagagaggaagtgatg

>PopulusTrichocarpa pCO2 (POPTR 0004s10800.1)
tggttgcatatataatgcagagacaaagggattgttcgaagttcaaacatataataccttactacttccaatatatactcacgagaatgagttaaatattgaattttttgatagttttaatctttccatttccttgttgtgaaa
tgtaataatggaaccattaattccttgtttggctgtctaatactaattcttaagttcgaagagtctttaatccgtaaaaactaagaaaagaaagaaagaaatacaaaaatcttgggttgccaaagcgggggtttaaaag
attgagagagctaaaaagttatagataagtactagatgaatcttatgataaacccaaattgttaatttaatttatatgtacccataattttggaaaaaataatttaatcttaagaatgaattaattaagaaaaaaagggag
caaaaatactgaattgaagacgtcatgtggaggtagtttatgactttatgcaacaaaaatgcaatccttaattatatggcaaacattctttctcaaaggcatggaaaacattttaatactcgcggaagtaactaatttgg
tataaaatcaaaattaaatttagtttcttcaatattaccagtttaagtattataaatttcaaaactactggaatcttatatagttgttaactttaggcttataaaattagttgaaatgcgcgtaagttgacccagatatccaca
ttaataaagaaaaataaatttggtttcataagtcaattaacatgtcaagatataattataatcaatagacgaaactagctagtctttgggcatcaatttaatttgattaatataagaatttacctataataaaatccattctag
caatattataattagacagtgttgattcacatttcattacaatcatgtaatgattgagcttgactactagttggtgagatgacaaggtgacatccacgagttttcctacaaaaagtaagtttagaatatcttttaatgtttaa
gtcagtaaatatcaatagaagataatatgttgagcaacagatgttaggagtaggagaaagaaatgtgtggaaaagagagtaggagcaaggggaaagcaatgtggctttgtgaccttgtgtggtctattataatgg
tcctaattaagttttctgcttgtgttgtgaaaacaagcttatgatttttatccggttataaaagatatttataaatcctaatcacatcagtttatttcatgtagagtgatataataataatagagaataattactcagtttacctta
tgtagagtgatataataataatggagaataattactttttcttgtaggtatctcatagagagtggagaaaattgattattggttgttatatttcgttgtaccattgatgggacggttggtgttcatgaataataagcaagaa
aggtaggtgtcattttaaagagacatgttgagaaaacatgacaagttgaatagatggagcaagatataaaaataaatatatctttaacacaattttagagtgaattcaaatatttctaaaaccaaagagaaaaagaag
gcatatctttgttatctcaattatctctatctcttgtatccaaaatagtaatgaaactttacttaattagcacatccaacataacattactacttctatactaaaaaacttgaaactcattataaaacttattgttcatataaataa
tgaagcttcttttttcttctttttttattttttattttcaatttagcacctcatttttcgatttcccccttttttatttcaatttttccttttatttctcaattttattcttatactttttttaattgtcatccaagttgccttgaattggccaaaa
aaaatagatcatattgagataattctcacataatttaaatttaaatccgagttaagcaaaaagcaagcccaagagattttctgtcaatttcatcataaaatttttttatatcaattaactaggttacctttagactagctaagtt
aactagatcgcgttagaataactctcacacaatttaattttaaacctaagctaagcaagaagttgagttagaaggttttttttatgttaatttgattatttttttaacttcatccttaaaattgttttactagtcagcttgaatatct
ttggatcgatcaagttgatctaatcatatcgaaaacatttctacataatttaattttaaactcgagttagataatgagttgaataaaaaaaaatttaatgctaactttctagatgaaatttaataaaaaatacaaaataattttata
taacttgaataattgtttttattattatgagataatcggaaactactttacatgtgagaaacttgaacatcatagttattaaacccagattggcctaatgggtcgacccgagatccggtcgatcaatggttggactggttc
gggtaaggaaaaagaccgaccggagtaaaaacctcatcaacccggcgggtcgacccatgacccgagccacccgggcgaacccggacgagacctttttttttcaaatatggtttttctcctagacctcttttttttta
tatttttttagttggttattagccctttcaaatttcattatataaataccagaagaatgttttatttttcaatatggaatttgaaaccctttagtatatatattctatgtttccaagaaaaaaaattatttttgaatgtgagatttgaa
accctttactatatatgctatatgttcataagaaaaaaaaattatttttcaatgtgggatttgaaaccctttagtatataaactctatgttcacaaggaaaaagttatgtttttcaatgtaggatctaaaaaaacttttttgattta
aatatttcaatttaaaaggataacatatagtatctttttaatatgagatttgaaaccctttagtatatatattttatgttcacaagaaaaaaaattattttgcaatgtggaatttgaaaccttttagtatatatactttatgttcaca
agaaaaaaaattatattttttcaatatgggataaaaataattttttgtttaaatacttcaacttaaaaggacaatataatatcttttcaatttaggataaattttttttatatatattcttttaaactttattatttacaacatatttaatcaat
atccacataaattatttttttaatttttttctataaaatattaaaattttaattttttttttaattttttttaaattaatctaaattgatctaagtttatctatataatctagaattcagctttttaagtagattatcctctaatcgggtttgacag
ctatgttgaaaaccccactaaagcttaagacctaatacagtacgacgtgttcctcacatgttacaaagttccttaaaaccataaaagatgacacgtctacatgaagaatcttatccattacagggtggacccaagat
cttgattatgacaagtgtcttcattttcaagcgagtatcacaatcttcctctacttccagcagtactctaaacactaaaatgattcataaagccaaatcagcccacaaccttatgtaccgcgtaaacaagtctttgaaatttt
ttgaatttgaattttatttttaatttttaaaatttttaaaattattgattttaaaaataaaaaataaaaataaataaattattttaataaaaacattttaaaaaataaaaaatattatttaatatattttttaataaaaaaaaactttaaaaa
taatatttactgtgtatgtactctcacttacaactgcgacctacctcaaaactcagccaatctcaagataatgaaataacttttctttaccactaaattgatgccatgtgtcgcactctcgattctctccaccccgcttgttt
atcctagaggtatatccctccctccttgcaactaaggctcactcatagctagctcaaaacacagtaaatttgagagagagaaagaaaggcaatatg

>Solanum Tuberosum pCO1 (Gene ID: PGSC0003DMG401010056)
ccggtgcgacgagctcgacccagaaggcaaacccccttgcttttgctggcaaggggtttcgaacttgagacctccaacatggaagtcccaagcccaaaccactgggccatccctaagggttcactttactttattc
ttttgtggatcaagaatcttaagaaaaagtgaaaagagcatattccatctcacaagaacattgaagttatcctggacctgttgtgttccacagaagtaaccagagaaaggctaaaatttcaagtttgagatccctcct
attccatgggtactttatgtatctatttctgctctgccatttgctgagtcacagcactttttctacagctcataaatgtcaagcaatccacttgtgattcctatgcttagaaaattactagtttttggaggtagagtcctgaga
aaaagtgaaaagagtaggttccacttaagtcacctgagcatagctaagccttttaagctttgaggtccctaagattacaagggaactttagctatctatttctgctctgccatttgctgagccgcaggtctttttccca
caactcataaatgctaagcaatccaccatggagcaattcttgtgatagctatgcttaaaaaattattaaacaaaacaaaacaatttccagtgattttctaattaatgcgaaaatatcgaacttcatttctggttgattacac
aacaatgtccagatacaacaatcttattattgtgagtgaacaactaatttcgagcaagttcgctggtcgagattgtgccttacagaaacacatagtgaggccaggctatgcagcttgattaagtcacaaggtaccta
ctacctccaaccggctctacaagtatcgaataattatgtccaccaaaatttaggcaaatgcgaaaaattctagtagcgtcacctctgttgggatttgtaccttggtacccaaggctattattccaagtcctcaactgct
aggggacacaggcaaatcatattgactaaagataattttactagctgttatttttcataatgtgcgaacgaaagaagcacacacacagacgtaggcaagatttcaacaaaaatgtttcatcttagattgtatcagttga
ctgtcaaagaatgcatacagaggctaatccaggctaatattttaaagcatataattaaaagacaatgtagagcttctttcttttaacttttgtggatgaagagtcctaccataaaggtgaaaagagtagattctatctca
gtgagaacattgaagttgttccacagaagtcattagagcctggctaagcctttcaagtttgaggtccctcttgatccatgacactttagtacctatttcttctcttccattcgctgaaaacctgaaactctgcactttttac
aaagctcataaatttcttagtttttgtgggtgtgaggccctttcaaatccttaggacttttatttatctatttccgcactgtcatttgccaaacatcattactttttgatctggttgacagccacaacaaagattctttcgtgc
catatcatcaaattttatccagtaatagcattaagaggtaaaattgtctaagcatttcaaatttgtgtatttagtggggaaaaaactgaaaaatgcgtttgtcctctttttttatataccaagatgcatttgtatgtagtgtaat
ccctttgtatataccaaacaaagattctttcatgccataccaagatgctctaccttagaggcaattgtccctttatagatataaaaaaataatgagaacaaacctagtttaatcccacaagtggggtctggaggggga
cagatgtatgcagaaatacctttgtgggatagagagattgtttctaacagacctcggctaaaaaagtgaagttattcggaaaaggattgcaggaagaaaaagaaatgacagcaaaatatcaagctacatgcaaa
atagaacaacagatagcaaaaacattgacaaaaaaaaaactgtgataactaaattaatactaataaatagggacaaggggatatgaggtttagccccctgccactcccaactaatgagcgaccacattcagctac
ctactaaccttctaccctaatcctcgattgccatacccttcttatctagggacatgtcctcaataagctgaagttgtgtcatgtcttgtcttgtctaatcaactcccccatttcgtttttggtcaccactacctctcctaactcc
tgttacaaccaacctcctaactagcgtatctgcacacctcctaactaatccacattcagctacctactaaccttctaccctaatcctcgatcgccatgccctcttatctagggacatgtcctcaataagctgaagttgtg
tcatgtcttgtcttgtctaatcaactccccccatttcgttttggtcaccactacctctcctaaatcctgttacaaccaacctcctaactagcgtatctgcacacctcctcttcacatgcccatgcccgaaccatctcagtc
tcacttctctcatcttgtccgtcatgaatgtcattcccaccttgccccatatagctttgttcctgatcatgtttctccctttatatataccatttctaaaaatatgtgtacctatatgaagcttgtattctttatcagaaaaagaca
tatttggaaaaagaaataaatccttaaaccatcacgtctcaaattcaaacagacttataagttgaaccataacaaaccgtaacaatagatgttagtatcatatgttcaccacaataaagcagaatatacgaaataatc
aggaaagaaaaatttaacgaaaaatacagaactgtgaattaattatccgagcccacagaattcgctatgtgtccttaataagccctagccccatataagagatctaaaaaggtcatttacttcttttcaatttgtacgt
ggtcataagagaccacttcctctctatcattagataaaagatacatgtcaatattcagacaccaaatcagttgcccaacatacattatttcaaatttgaaaaaagaactttcatcgtgtctgttgcagcaccaattatttc
ttttgatgtggtgctgagtttcttttcctaatgttagtagagcatcaaagtctcaagtctttcaaatgaggttttctatgacaaagttcctctatattttctctaactgctattttcttctccttcctacactactattttgaaatcat
agaaagggaaaacgtcaatactatagatgcttatcaaagtactgaactctgaagggatcttccaaacagtataaattcaacatgtgaagtaagcctacgtatgccaaactcagcaattcttacgatgaaacaaaaa
taatcttgctctccaagaacaagacttcctaattcttccactccactactatatatgacctttaaagcattcttcccgtttaacctgcttaggcatatgttttgatcatttggattctattatgagattcaacgagaatcacatc
acgagagtattctaaaaaccgaaattacaaaatcaatgttttcaataccataccattcaggtccgtaccatgcccacccctaacggtgaaaatctaataaactatgggaagtgtatcggtatagtccaccccttaaa
ggcaatctcttgaacaaagctctcccacctcaattgattccaaaagcccaaggcccctggacccacttgtttcaacctccttgtggcttgaccccaccttccggccccaccagaatcaccagccaccacaagata
atgaaaaaacttttcctacccattaacttccctccacatgtttcccttggattctctccccacttgatagttgcctactgttattttccctcctttccccccttcatattccacagtgccccataagtctgcagcatacctgcta
aacttcagtattcgaagtacaacagaaaatagagagcaagaagaaaaaaatagggccaaaaatagttcaatg

>Solanum Tuberosum pCO2 (pCO) (Gene ID: PGSC0003DMG402010056)
taattatttgttactgggcggtgccagtaacgcccaagtcttgcttggtctcatccttgccagcaccaaaaacagggcctcctgtcctcaagtatctcaatttgaataccgcattcatcttcacgatccattctaggctt
cttctctccctctttctcatcaactaactgattcatatgtagttggtgttaaacatccctttggctgtgcccctgatttactctgcttcgttttaatttacgtgatataatttgatttgttatagaatttaaaataataataaaaatt
gtaatcttcacagatgtgaacaaggagtaatatatgttcctagctcttaagtactacttgtatagtgcctatctaactttgagtaaacatggacgcttcaattgtcatcatggaatgactgaaaaggctgccacgtatcg
tacatgaaactcccatgtgactaatgtagacaagctccatttgtaccactccatagatatgcactgtcggttaggggcagagttagaacgttaggttcatcccaactcctctttatcaaaatttcttactctactattgct
ataaccatttcataaaatgggacatgaccgagttctgtaggctgtcactcgctgtgaaaaaatagaagatgagtaattactttacttagagcctgattgacttttaaaaaatacaatttttaagttaaaaataaaaaattt
aagtcaaaaaataaaataaaaagttgggggacacctactcttggtttttgacttatttttgaccttgtcaaacacttcctaacttgtttttaatcatttttacttattttaagttaatttttatatttttcaaacactttcaaaaaag
acctacttttaagccaatccaaagaccctcttagtttttttttttccttccctaggagctcccacccctttgctcccttgatgactcgaactcgcaaccttcgaagtgaaagtgagggtgcttaccatccgagcaactcc
ctctcgtcaaagaccctctgttgttacaatgattttatgtttttcgctagatttaagataatttggactaattaaaagaattaaatggtagttttattgtccaaataatttaaattatatgaaattattatcataattatctatatac
atttaaaaatcgcaatcaaataatactattgcttgtcgaaagctagtttgagcaagagaaagtatagtattttacacggtatctcaatgaaagatttattgaacttattactcagaaagctacaactacgtataaaagata
gctctacacattataacaccttaaaaattcagagaagaaaacaatgatgcatgtttgacgcaaaagttaaaaatgcctatttctccgagttttactatgatgccagtgccactggtcaacaataagaaggatgatagt
caaaactttggtcattgactgactgactctatgtatgctggtcatcaattgatttacgcaaagctatcaataactgacaaatgaaaatacaagtcattttgtggccaattgagataaaaggctggggatgaagataaat
actattccccttccatctcatattagttgatcatttcatattaattgtttaactcattaaatttaaaaaatattaattaaactttttctatattactctcacaatcagttctttttaaagtgttcaagttttatttataaaaagtgaatta
atgaaaaaagaaattgatcgactaatataaaacggaggggtagcttttaatttttcatgtctatcaacattgcgaagacagcaaatcgaaatctaataaatgcaaagctaaggatattagagtagtggccagtaatgt
ttggagttaggaaggccattatattgtcaagtttttgattgcggctttacctttcaactgggcccaaaatccatgtgaattaggccctgacttctaacccactaaaacttgtggttaagggtggatctcaatccagctcc
aattggcagcaacccattgggtgaaattcttacacaaaaactcatctgactcagctttactaattgtgttaatgtctgactgagcagcagccacgaggtctctcaacagagtacatttttgggtttatcctaaacaatgt
atgagtatttgaaaggacaagaaaaatgagtgatttgacgcagacgtttttttcctgtgcattaggaattggtaacagaacacacgtacgtagcatttgtagtcggtatcctgggaatgtgctctcaaaataatttaaa
ttaaacttctctcgtaaaaaaataaaaaagttgtaaagtttgagaagaattataactggactacaaaatcaattctctttaataaataaataagtgtgcattactgactattttcgaattcaagaatgacaacttccatgta
acactaaacagaactttttcacacgtaaaaaattcaatctctagaacaatgttttgaagaagaaaaataaacgggtgctatgactatttttcaattattttatattaattagcagtcttaaaagtggctaccactattaagat
ctccttccaaatatgatcgtaacacttagtcatgggtggtgtttgaggagacaaggtgctgctaaatgctaatggtggaagaagtgagtggagaggatttgcacaaataattaacggaggataaaattaacccattt
cacatagtaataaggcaaaattgaaccttatcccttccgaagatcggatgacttttaaaagtaaaaagggaaatcaagaatccagcttttcagtttcaaacttttcgaaatggatgtctcattccttaaatgatacttcct
ccatccacttttatttgtcatgttacgtttttcaaaagtcaatttgactaattttccaagtaaaattagattacattaatttaatattttaaacaaaaaaaaaatcagatattcaaaaactatacgaatagtactataaattgcattt
ttttgcatatcaatatcacgaaaaaatacatcgtaaaatgttaaccaaagttcttataatttgactctaaaaaaggaaaccgtggaccactctaaattttgaactgtaaatttatatacagtctcgtaattaacgttcccaa
tcaaaatatacaaaacgttctggcaaagtctcccagtaaaaataaacaaggtgaatttctcaaaatattactgcttccatgaattaatggcttagttgacttggattgtttatatcaatggagcattactctttcttttaaga
tcaaaatcttacacgagggcattgaaaatagaacaaggtgtattggcttattaagcttgtattcttaattgtcagataaaaaacatatttggaagatgtaattaatccttaaatcatcacgtctcaaacacaaaaatcact
tacaagcagaattgtgacattacgctagtactatctgaacaaaatcttgcatacaccactatgcatcagagtgcaccaatgttagtcagacctgataaacaatggaaagtgtgtcaataaagtcgatatattgaataa
aatatgaaaaatgttattgacataaattataataaaaatgcttaaatagagttatatgggtgtgattatctctctagtctcccccctaaaaggcaatctcttggacaaagctctcccacctcaattgattccaaaagccaaa
ggccccaggacccacttgttcaacctccctgttgcttgactccaccagaatcatcagccaccacaagataatgaaacaacttttcctacccattaacttgtctccacgtgtcacctgtggattctctccccacttggta
gttgcctagttgtcattttccgtcctttgccccttcacatttcacagtgccccataagtctccagcattcttgctcaacttcactactccaactacaagttttacaacacaaaatagagagcaagaaagaaaagaaaag
aaaagagaaaccaaaattagtcagtccaatg

>Vitis Vinifera pCO (CAO49246)
tagtatttcatgcccatgactaactcattgcaacttagactctttatggttcacaagttcaaagttctcatttatgcatatttaacctcttctatttccttggttgagaatgacttttgcaacatttttttcatcccttaaacccata
attaggctcaaaatagaagttaaaatccataactaaggtctttgcatcaatttaagttaaatgatttgagctttttatattacataaatcatatcatatatgtgccattggaccacatattttgactctagacaatcaattttta
atataaagtagtcctataatatgagaaaaaaacctaaaagaatggatttagactgtaagagagatgttcttttggatctaattgggctttgggattcaagagctctagggttggattctccttgagaaatagtggaaatt
ctttgttatcttctcttgtgaatgtaagttttagacctaggtggtgtttgttttttggctttttgttgaaaaccatttagttttagaatttaggttgtttgtttttttacttttttcatgacttattataaactttttactaaatagaaaaag
ccagaatatgtggtttttctaaatagaaaaaataacacaatggtttttttacttttaatacttaatagaaataaaatactacaaaaacaaacaacctaatatttaatactattaagtattaaggttttatttagaattaagta
aaaaaacaaacaccaccctaatcttataaattattgcatgctttatgtgccttctatttttatttttattttattttattttgctcctattattgttctttgttatgatttatcattaaaaaaaaaatggtattagaattgaggttgttaga
ctgacaattaatggccaagttcgaagttgagaagtttgattaaaaaaatgattttaatttattgcacactaagatgcatagtatgcttgtttaacaaaagttgtattaagcattgaaaggcagggaggctctattgaaaa
caatgtttgacatacgaaagaatgaacttattgaaaatatgtatgatgctattttgttgttcttatatcttctaaaaaaaaaaatacaactatcgagtgataattgaagctgaaaagtctatatgatgaaatttcttacaaatg
acctttgtatttgaagaagcaattatttactctttatatgaaggaaggtgcactcataaagatgatcttaatgagttcaatagaactattgtagacataagggatattgatgttaaaataaataatgaagattaagccaca
attttgatgtgttttctatcaaatcttgaagtactaccccttattgcatgtcatgtgtcctacttttggtacctaattgtgatcttttgttgtactaaaatactaaacttaattatgtgataaccctaaccaagaaaaaccatag
aaatgattatcacacaacttgtagtattcttggtattgttctcaaggatagacctataagatttgtcaatacttctctacaatgatgactcaaactaaatactaaagtgaaaatattgaatggatcaactttaatgaattaag
attaacttaattaaaaactaattaactaatgaaatgattattgatttctaactcaattgattcaagtttgggacatttacaatccaacttaaaaaaaaatagaaattgaagcgaataagggtctctcatttgtagtgatcttag
ggttcttaggatccttgatcaatcgagattaatttgagttctaacactcaaagtttcatttgaaacctaattttgtccttctagtctagattcctaaatactattaagtgaatgaaattcgttttaggtttaagtcttagatttttct
atcctcactccttttacatatttgccttaggatggataaagataggaaaaaccatgatagttagtggtcaagaggtgctaagaatctctagtgttcgatagataactatcgtaccatctcttaagtcaaactagaatgga
gagcatgttaatcctatcaatttacgccttaaattattcaacaaatccattgttaaatcttatttacccatttttcctcctagcctttaaccctaggtcttcatatttcacatcataggatgacttttaatttctcgtggaggtgtt
gtcactcacataatttgtaagagaaggaacatcaatatgaataaaaagaaactaaaaataataataaaattgaatgtataaagaacaaacctaaaaatcttcaaagcaatctttgacatagtctctccaatcttctaaa
tgaagtttatatattttaaaacttaaaatcagtttacatagtaatgtaaaacctaatacacatgacataaccttattagccaaaatcatccaaaaaaaaacacaattattatcccaaataatcataaaaattacaaattttga
gtgaattagattcattttgaactaggttggtgtttgagaaatccaaaattcgaattggaagtgttctatgtgaaaattggaaccctaattcaactagttgagggttagttcaaaattccattgtgtaattccaaaatttgaa
attgaatttttgccaaagctttgctagttcaaaatccttatggatttcaaattcaagttttatctacccctcttgctcaacttagttcaaaattggtgatagaattggaactcaaacttttgatttttggtttctttgcttgattgag
ttttaatctctttccaaatttggaatcaagcttttaagttgatgacttcttttgcttccaaagtggttggatactctgtagctctttgggcaatcttcaattactctttcttttgcttcatgctccttatgcttttgcttggattctaatt
catgttaagttggatgatttaaaatttacatcatgtagatgatcagattgtttatatgctattggactacaatgagtaccctatttcaaaattagttcatgaaatcaaattggtggagaagaaaatgggttttgttgtatggg
tagggatgcaccagccattatgtcacgtgggaactccatgaaggaggggtttgggaacactaggtgtaggacacatgcatcagtggcgatgaggacaggccgagaaacagacaatcaaagtggaaagcac
ttgtggctggctggtggctgggggcagcagaccgactcggccactgttggcacatcagcataaaagcaggtctttccgttggacccattaccccattcgagaggatgagcctcacctcgctctcccccactctc
cctccttccctcgtgtcaccaatatgtggtgctcctcaatcctaatctgaactccttcccatgtggcaccatcccattggctgcaagaattccctcccgcgttgtcccagtcagcttcatatggcgtagacatcgactc
tacccgcttccacatcagctctagctaatccttttgttctaccgtaactacgtccgatgatgacatgttggcctctggtgagattacccatattcaatcttatcacacacaatcccctaccaattccaatttcatgaaggc
ccttaaaccaatttcatcaatttagtgtgattcattacgaaagttgatagctactgttcatcaacttccgagtccccattctaggtatatgtcaatgtcttttcctaatcaattcgttaacatagatcacctgcatctatttcct
caccacccaaatttattattattattatttttcttttacacgctgtcatatctgatgatgatgtgggttgtgggaggaggataaaggccgaaagccctttttgttttgggcttcccatcaatcatcatctctttaatcttcctttt
aatgaatcaaaaccaaagggcccttctttactttaaactacacccgtggcaaaatcttggaacctattataaatttatacatcaaaagtgcaaatattaagcaactgggggtgggttgtcacatg

CO Promoter alignment from ten Brassicaceae species as displayed in Fig. 11:

```
A.thaliana    A--------TG-----------------------------------------------------------CTGCTATGGCAAAACTAGACTGCATGCTTGTAAATTACTTCTGTTTTGAAATTCCTGCAAGTTATTACCACGGTTTTAT   82
A.lyrata      ---------TG-----------------------------------------------------------CTGCTATGGCAAGACTAGACTGCATGCTTGTAAATTACTTCTGGTTTGAAATTCCTGCAAGTTGTTACCACGG--TTAT   79
C.rubella     G--------TG-----------------------------------------------------------CTCCTATGGCAAAACTAGATTGCATGATT-----------------------------AGTTATTACCACTGT--AAT   50
N.officinale  ---------TG-----------------------------------------------------------CTCCTATGTCTAAACTAGACTGCATGTTT-----------------------------AGTTATTA-------------   39
A.alpina      TA-------TG-----------------------------------------------------------TTCCTATAGCAAAACCAGATTGCTTGCTTGTAAATTACTTATGGTTTG-AATTCCTGCAAGTTATTACCATGGTT--AT   80
B.napus       ------TAATA-----------------------------------------------------------CTCGTATGGTTTAACTAGATTCCATGCTTGT----------------------------AGTTATCACCACTGTAATGC   56
B.oleracea    TTCTGATAATA-----------------------------------------------------------CTCCTATGGTTCAACTAGATTCCATGCTTGTAAAT-------------GCATTACTGGAAGTTATGACCACTGTAATGC   77
S.irio        -------TATA-----------------------------------------------------------CCGCTATGGCAGAACTAGATTGCATGCTAGTAAATCACTTACTGGTTCGAATTCCTGCAAGTTATTACCGCTGTAATGC   83
S.officinale  -------TATA-----------------------------------------------------------TCCCTACGGCAGAACTATATTGCTTACGTGTAAATCACTA---ATTTCGAATACCTGCAAGTTATAACCGCTGTAATGC   80
A.arabicum    ----------AAGCATTCCCTACAATGCTAATCTTCGACACCGGATATAGAATTGTTCCTACATTTTGATGATCCTGTAATAAAACTAGACTGTATGCTTGTAAATTAATTATGATTT-GAATTCCTACAAGTTTTAAGACAGTTATGC  139

A.thaliana    GCATATAGGGTCTGTAGTTTACTAGTTTTGCCTAAC--------TAGTACGAGTTAATTAG-ATATACACTCATACTTAGCTGCTGTAAGTT-----AATACAAATTTTGCCCACTTGGTAGAAGTCCTCTT-ATCTATGA---------  208
A.lyrata      GCATATA-GGCCTG--------TAGTTTTGTCTAAC--------TAGTACGAGTTAATTAG-ATATATATTCATATTTAGCTAC--TAAGTT-----AAAACAAAATTTTCCCACTTGGTAGAAATCCTCTTCATTTATGAAATATTACC  204
C.rubella     GCATATAGGTCTA---------TAGTTTTGTCTAACTAGTACCGTAGTACGGATTAATGAG-ACAT--ATTCATAATTTGCTGCTGT-AGCT-----GAAACAAAATTTTCCCACCTGGTAGAAATCGTCTTATTTGTAAAATATTTACC  182
N.officinale  -------------------------------------------------------------------------------------------------------------------------------------------------TTATC   44
A.alpina      GCAATATAGGTCTGTA-------ATTTTTGTCTAA----------------ACTAATTAG-ATATA--TTCATATTTTGCTGCTCTA-GTT-----AAAACAAAATTTGCCAACTTGGTAGAAATTCTAT-------TGAAATATCACC  190
B.napus       --AAATAGGTCTC---------TATTTTTGTCT--------AACTAGTACGAATTAATT---AGATATATTTTTATTTTGCTGTTGT------------------------------------------------------------------  121
B.oleracea    --ATATAGGTCTC---------TATTTTTGTCT--------AACTAGTACCAATTAATT---AGATATATTTATGTTTTTCTGTTGTATTT------AAAACAGAATTTTCACAC---------------------------------------  164
S.irio        ATATATAGGTCAG---------TATTTTTGTCT--------ATATACTACGGATTAAGT-CGATATATATTCATAGTTAGGTAAAACAC------TAACACCAAATCT-ATATACTCTGATATGTG-------------AAATGAT-ACC  194
S.officinale  ATATATACGCCAT---------TATTTTTGTCTG-------ATGTACTACCGGATAATTACTATACATATTCATATTTTGCTGCGGTAACTTAGGTAAAAGCAAATCTAACATCCTCCGATATGTG-------------AAATGATCATC  201
A.arabicum    A--TATAGGTCTGTA----------GTTTTGTCTAACA--------AATTGAAATTAATGA---GATATGTTAATATTTTGTTGCAGTTTTTAA-------ACAAACTATGTTTAACTGGTAGAAATC-----------------------  237

A.thaliana    ------------------------------AATACTGCCCTACAGTCTA-AATTCAAACTG---ATGCTACAAAGAGATATGC-----------------ATTAGTATAGCTAGCTGTACCAAGAACAGACAACAC---------------  293
A.lyrata      ACCACATA---------------------GAATATTGCCCTACAGTCTA-AATTCAAACTG---ATGCTACAAAGAGATATGC-----------------ATTAGTTTAGCTAGCTGTACCAAGAACAGTCAACAC---------------  298
C.rubella     ACCACATG---------------------TAATACTGCTCTTCAGTCTA-AATTCAATTTTTGATGCTATAGAGAGATATCG-----------------TTTACTTTAGCTAGCTGTACCAAGAACAGTTAACAC---------------  279
N.officinale  AC----TG---------------------TAATTCTGCTCTTCCGTCCAGAACTCAGATTT---ATGCTATGAAGAGAAATGG-----------------TTTACTTGAGCCAGCTATACCAAGAACAGCTAACAC---------------  135
A.alpina      ACCACATA---------------------GAATA---------AACCTA-ATTTCAAATTG---ATGCTATAATGAG--ATGA-----------------ATCACTTTA-------TAACAAGAACAGTTAACA----------------  264
B.napus       -----------------------------AGTTACTATCCTTAAGTCTA-AATTCAAACCG---ATGCTATAGAGATATGTGC-----------------GTTACTTTAGCTAGATGTACCAAGAACAGTTAACAATGAAGCCTGTTCTTT  222
B.oleracea    -----------------------------AGTTACTATCCTTCAATCTA-AGTTCACACC-------------GATATGTGC-----------------GTTACTTTAGCTAGATGTACCAAGAACAGTTAACAA---------------  239
S.irio        ACCACATA---------------------GAATACTACCCTTCATTCTT-AGTTCAAATTG---ATACTATAAATATATATACA----------------ATTACTTT-------AATACCAAGAACAGTTAACAC---------------  282
S.officinale  ACCATATAT--------------------TAATACTACCCTTCCTTCTT-ATTTCCAATTG---ATACTATCAATATACATGCA----------------ATTACATT-------AATGTTAAAAACAGTTAACAC---------------  290
A.arabicum    --TACCTATTTGTGAAATATTTCCTATTAAAACTCCGTCCTAATGTCTA-ATTGCAAATTGATGCTACAGGAAAGAGATATGCATTACTCTGTTCTGTCCTAATGTCTAATGGCTGTAACAAGAAAACTAAAAATG--------------  370

A.thaliana    ------------------------------------------------------------------------------------------------------------------------------------------------------  293
A.lyrata      ------------------------------------------------------------------------------------------------------------------------------------------------------  298
C.rubella     ------------------------------------------------------------------------------------------------------------------------------------------------------  279
N.officinale  ------------------------------------------------------------------------------------------------------------------------------------------------------  135
A.alpina      ------------------------------------------------------------------------------------------------------------------------------------------------------  264
B.napus       AATGGTCGCTGATCGCAGCGGTTGCGTCAACATAAAATCACTATGACAGAAACAAATGAATGGATTTTGAAGATGAAATCAGTCGCCAACAAGTTTCCGGAGACAAAAATATCGGATGCCGCTGTCGCTGCTTTTATGGGCGTCAAATTT  372
B.oleracea    ------------------------------------------------------------------------------------------------------------------------------------------------------  239
S.irio        ------------------------------------------------------------------------------------------------------------------------------------------------------  282
S.officinale  ------------------------------------------------------------------------------------------------------------------------------------------------------  290
A.arabicum    ------------------------------------------------------------------------------------------------------------------------------------------------------  370

A.thaliana    ----------------------------------------------------------------------------------------------------------------------------------GTAACTCACCTTGCCTAA-G  312
A.lyrata      ----------------------------------------------------------------------------------------------------------------------------------GTAACTCACCTTGCCTAA-G  317
C.rubella     ----------------------------------------------------------------------------------------------------------------------------------GGAACTCACCTTGCCAAAGG  299
N.officinale  ----------------------------------------------------------------------------------------------------------------------------------GGAACTCACCTTTCCAAAAG  155
A.alpina      ----------------------------------------------------------------------------------------------------------------------------------------CACCTTGCCT-AGG  277
B.napus       TTCTTTTATTTTCACCGCAGCGACTGAGAAAGCAAATTGATATTGCTCCGCCAGCAGTGTATTTGCCGTGGTGAGAGATGTAGTTTTCGGTCGCTCCCGCTAATCGCGGCGTCTATGAAACGAACACGCCTGAACTCACCTTGCAAAAAG  522
B.oleracea    ----------------------------------------------------------------------------------------------------------------------------------TGAACTCACCTTGCAAAAAG  259
S.irio        ----------------------------------------------------------------------------------------------------------------------------------GGAACTCACCTTGTGAAAGG  302
S.officinale  ----------------------------------------------------------------------------------------------------------------------------------GGAGCTCACCTTGTGAAAGG  310
A.arabicum    -----------------------------------------------------------------------------------------------------------------------------------TACTTCACCTTG-------  382
```

```
                *
A.thaliana     TTCAAAG-----------------TTAATG------------ATTCCTTTCTTGGTGACACTTTCCACATCA----CTTTATCACGTAGAAGATAAATGATTC----------CACTTATGCCGCCCCAAAGACAATTTCGCACATCCCTG   420
A.lyrata       TTCAAAG-----------------CAAATAATTCCACTGTTCATTCCTTTCTTGGTGACACTTTCCACATCACTGTCTTTAT--CGTAGAAGATAAATGATTC----------CACTTATGCCGCCCC-ATGACATTTTCATACCTCCCTG   438
C.rubella      TTCATAG-----------------AAAATAA-----CTACTGATTCGTTTCTTGGTGACACCTTCCACATCACT----TGAACACCTCAAAGCTAAATGATTC----------CACTTATGCCGCCG-TTGGACATTTTCATACATCCCCT   413
N.officinale   TTCAAAA-----------------GA---------------------------------------------------------------------------------------------------------------------------   164
A.alpina       TTCCAAG-----------------CAACT---------ACTGATTCCTTTCTTAGTGACCATTGCTACATCGCTT----GAACAAGTTGAAGATAAATGATTC----------CACTTATGTCATCC----------------------CT   366
B.napus        GTTCAGA-----------------CAAGTAGT---------GATTGCTTTCTTGGTGAAAC-TTCCACATCAGCTTGATAAACAAGTAGAAGATAAATGATTC----------CACTTGTGCCTCCC-----------------------T   613
B.oleracea     GTTCAGA-----------------CAAGTAGT---------GATTGCTTTCTTGGTGACAC-TTCCACACCACTTTGATAAACAAGTAGAAGATAAATGATTC----------CACTTGTGTCTCCC-----------------------T   350
S.irio         TTCAAAG-----------------CAAACAA-------------------CCTATTGA----TTCCACATCACTTGG----AAACGTAGAAGATAAATGATTC----------CACTTGTTCCGCCC-----------------------T   376
S.officinale   TTCAAAG-----------------CAAAT---------------------------------ATCACT-----------------------ATCGATTC----------CACTTGTGCCGCCC-----------------------T   351
A.arabicum     -TCAATGTTCAATGTCGTTAGCTATTAGTGAGAT---------------------------TCCACATCACTTGAACACTT---TAGAAGATAGAAAATTTCCAAGAACATTTTCATGCAACCCTGAGGAA------------------   482

                                                         *  *  *    *                                                          ***         *  *  *    *
A.thaliana     AGA-AAAGAAAGTTGTACTCAA---TATGATTTTATACCAA-AAGTAGATTACAGAGACAAG-TGTAAACATG-TAA------AACTCC-A----------------------------ATAA-----AATCACACACTGAATAG-----   518
A.lyrata       AG--AAAGAAGGTTTTAATCAATAGTATGATTTTATAACCA-AACTAGATTACAGAGACTAG-TGTGAACATG-TAA------AACTTA-AGAGATCCAAC----------AAAATCAACATAA-----AACCACACACTGAATAGTCTTT   562
C.rubella      AAGAAAATAAGGTTTTACTCAAT---ACGATTTTATAATCATAAATAGATTACAGAGACGAG-TA--AACATG-TAA------AACTTT-GGAAATC-------------------CAATAA-----AATCACACACTGAATAG-TTAT   523
N.officinale   ------------------------------------------AAATAAATAACATAGACTAA-C-------TG-TAA------AACTCT-AGAAATCAATA---------AAGCTCAAAATAA-----AATCACATACTGAATAG-TCAT   241
A.alpina       GAGAAAAGAAGGTTATACACTTAT----GATTTTCTAATCAAAAATAGATTGCGTAGTCCAATTGTAAACATG-TAAATAGAGAAATTT-AA-----ACCA---------AAAGCAGACATAA-----TATCACACATTAAATAG-TCAT   490
B.napus        GAGAAAATAAGGTTTTACACA--AATATGATTTTCCAAACAAAAGTAGACTACATAGTCTAACTGTACACTTA-TAA------AACTTGAGCAAACCCAAT-------AAAAAAACATTGTAAC----CATCACACATTGA---------   734
B.oleracea     GAGAAAATAAGGTTTTACACA--AATATGATTTTCCAAACA-------ATTACATAGTTTAACTTTAGACATA-TAC------AACTTG-GAAAATTCAATTTAAAAAAAAAAGACATCATAA-----CATCACACATTGAATAG-TCAT   477
S.irio         GAGAAAATAAGGTTTTACAC---AATACGATTTTCCAAACA-AAGTAGACTACATAATCGAATTGTAAGCATT-TAA------AACTTG-AGAAGTTCAAT--------AAAAAACAACATAA-----CACCACACAGTGAATAG-TCAT   500
S.officinale   GA-AAAACAAGGTTTTACAC---AATATGATTTTCCAAACA-AAGAACATTACATAATCGAATTGTAAACATTGTAA------AACTTG-AGAAATTCAAT--------ACAAAACAACATAA-----CATCACACATTGAATTG-TCAT   475
A.arabicum     -----AACAAGATTAGACA---------------ACAA-----AGTGGATTACATAAACCA------------------------------------------------------ATAGTAAATAAGTATACTTTAGAAATCTAATA----AA   552

                                                                           ***                 ** *   *       * ****        *****
A.thaliana     ----------------------------------TCAGGTAACAAATAACCAAGAATCCTGAGTGAAC-AATGCTAATAGGTACAATCACATC--AACGCATATTCTTGTACTGATTAGTGATTAGCATTGGT------------------   614
A.lyrata       TTTCGGTAATGTACTCAGATAAGTGGGAC----TCAGATAACACGCAACCAAGAACCCTGAGTGAAC-AATGCTAATGGGTATAATTACATC--GACGCATATTCTTGTTCC------TGATTAGCATTGGT------------------   681
C.rubella      TTTCGGTAATGTAC-------------------TCAGATAGCAAATAACCAAGAACCCAGAGTGAAC-ATTGTTACTAGGTATATTTACATC--AATGCATATTCTCTTTCCT------GATTAACATTGAT-------------------   627
N.officinale   TTTCACGAATGTACTCAGATAAGTGGGAGTCAGTCAGATAGTAAATAATCAAGAACTCTGAGTGAAC-AATTCTACTAGGTATAGTTACATT--GGTGCATATTCTTGTTC---------------------------------------   349
A.alpina       TTTCAGTAACATACTCAGATAAGTGGCAGT----CATATAGCAAACAACCGAGAACCATGAGTAAAC-AATGCTACTTGGTATAATTACATC--AAAGCATATTCTTGTTCCTG------ATTAGCATTGGAG--CATGGTTGCGTT---   622
B.napus        --------------------------------------------------------CTTGAGTGAACAAATGCTACTATGTAAAATTACATCAAAAAACATATTCTTGTTTCTG------GTTAGCATTGGAGAGCATGGCTGTGTAACC   822
B.oleracea     TTCCTGTAACATACTTAGAAAAGTGGCAGTC----AGATAGCAAACAACCAAGAACCCGGAGTGAACAAATGCTACTCTGTAAAATTACATC-AAAAACATATTCTTGTTTCTG------GTTAGCATTGGAGAGCATGGTTGTGTAACC   616
S.irio         TTTCAGTAACATACTCAGATAAATGGCAGTC----AGATAGCAAGCAACCGAGAACCTTGAGTGAAC-AATGCTACTAGGTCTAATTACATT-AAGAGCATATTGCTGTTCCTG------ATTAGCATTG--GAGCATGGTTTCGGTACT   636
S.officinale   TTTCAGTAACATACTCAGATAAATGGCAGTC----AGATAGCAATCAACCGANAACATTGAGTGAAC-AATGCTACTAGGTATAGTTACATC-AAAAGCATATTCTTGTTCTTG------ATTANAATTG--GAGCATGGTTTCGGTACT   611
A.arabicum     TCTCTTTAAACCATCCACTAAATGATCAATTG-------------TTGTCACAAACTCAGAG----------GTTATTATGAATGATTACATC---GAACATATTCTCAC---------------------------------------   636

                                       **   *   *                                                    *  * ***   **       *  ** ** *    *      **  ****      *           *   *
A.thaliana     -GCTAAATCCATTCCCACGAGGTTGTCTCGGATAACAAG----------------------------------ATTCTATCACTCACAACCAC-TAC-TAGCAGAATATGTC----AATGCCTATGGACCATCCAAATGAACAAAGAT   721
A.lyrata       -GCCAAATCCATTCCCACGTGGTTGTCTCGGATGACAAG----------------------------------ATTCTATCACTCACAACCAAGTGC-TAACAGAATATGTC----AATGCCTATGGACCATCCTAATGAACAAAGAT   789
C.rubella      -GCCAAATCCATTCCCACGTGGCTCTCTCTGATAAACAG----------------------------------ATTCTATTACTCACAACCAAGTGC-TAGCAGAATATGTC----AATGCCTATGGACCATCCAAATGAACAAACAT   735
N.officinale   ----------------------------CTGATAACAAG----------------------------------ATTCTATCACTCACAGCCAAGTAC-TAGCAGAATATAAC----AATCCCTATGGACCATCCAAATGAACAAAGAT   430
A.alpina       -ACCAAACCATTTCCCACGTGGCTCTC--TGATAACAAAGATTCTAACACCTTTGTCGCTCAC----------ATTATATCACTAACAACCAAGTTC-TAGCAGAATATGTC----GATGCCTATGGACCATCCAAGTAAGCAAAGAT   752
B.napus        AAACCCATGCATTCCCACGTGGCTCTCTCTGATAAACAAGAATCTAAACACCTTTGTCCCATCCAGTATATTAA-CACTCATAAGTAACAAGCAACCAC-TAGCAGAATATGTC----AATGCCTATGTTTAGTTACTTGCTCTAAATCT   966
B.oleracea     AAACCCATGCATTCCCACCTGGCTCTCTCTGATAAACAAGAATCTAAACACCTTTGTCCCATCCAGTATATTAA-CACTCATAAGTAACAAGCAACCAC-TGGCAGAATATGTC----AATGCCTATGTTTAGTTACTTGCTCTAAATCT   760
S.irio         AAGCCAATCCATTCCCACGTGGTTCTCTCTGAGAAGCAAGATTCT-AACACCTTTGTCGCATCCAGTATATTATCCACTCACAAGTCACAACCAACTAC-TAGCAGAAAATGTC----AATGCCTATGTTTACTTGCTTT-----AATAT   775
S.officinale   AAACCAATGCATTCCCACNTGAATCTCTCTGANAAGCAAGATTTT-AACACCTT----------AGTATATTAT-CACTCACAAGTCACAACCAACTAC-TAGCAGAATATGTCAATGAATGCCTATGTTTACTTGCTCT-----AGTAT   743
A.arabicum     -----------------------------GATAAG-AAGATTCTAACACCTTTTTCCCTTATCACTATAT-----------TACTCACAACCAAGTAAGTAGCACAATACTTC----AATGTCTATGGACCATTTGTATACATAACGAT   740

                                                                             **            *        ****   *     *  *  *   *** *   *
A.thaliana     ATTCAGTGTGAGAAATCTATTTCTTGGTCTATCTTTTAGA----------ATAATAGCTCACTTTTT-------------CCAAGGT--ATATACTGGCACTAAAGAATAACGACGCAAC-------------------------CAAAAG   822
A.lyrata       ATTCAGTGTGAGAAATCTTTTTCTTGCTCTATCTTTTAGA----------ATAATAGCT--CTTTTT-------------TCCAATT--ATATGCTGGCACTAACAAATAACAAGGCAA--------------------------GAAAAG   887
C.rubella      ATTCAGTTTGAGAAATTCATATCTTTCTCTATCTTTTGGA----------AGGATAGCTTCACTTCT-------------CCAAGGT--ATATGCTGGCATTAAAGAATAACGAGCCAA--------------------------GAAAAG   835
N.officinale   ATTCAGTTTGAGAAATCTATATCTAGCTCTAACTTTTAGAAT--------ATAATAAC-TCACTTTT-------------CCAAGGT--ATATGCTGGCATTAACGAATAACGAGGCAA--------------------------GGAAAG   531
A.alpina       ATTCAGTTTGAGAAATCTTTTTCTGGCTATATCTTTTAGAATAAAGAAAAATAAAAGCTCC-CTTTT-------------CCAGGGT--ATATGCTAGCAATAAAGAATAATGAGGC----------------------------------   853
B.napus        ATTAGAATAAAGGGCTAT--------------------------------AGTGTGTC---ACCTTT-------------CCGAGGA--ATATAATAGCATTAAAGAATAAGGAAGC--------------------------AAAGAAA  1040
B.oleracea     ATTAGAATAAAGGGCTAT--------------------------------AGTGTGTC---ACCTTT-------------CCAAGGA--ATATAATAGCATTAAAGAATAAGGAAGC--------------------------AAAGAAA   834
S.irio         TTTA----------------------------------------------ATTAGCTC---AATTTT-------------CCAAGGA--ATATGCTGCCATTTAAGAATAAGAAAACAAAAGAGAAAAAAAAGCAAAGGAAACAAAGAAA   861
S.officinale   TTTAGAATAAAGAGCTATAG------------------------------AGTAGCTC---ACTTTT-------------CCAAGGCATATATGCTGGCATTAAAGAATAAGGAAA--------------------------AAAGAAA   820
A.arabicum     AC----TTCATGAAACCTATTTTTTGCTATTTCTTTTAGAATAAAG-----------------TTTATGCTTTGCTTTACCAAGGT--ATATGCTAAGATTAAAGAATGACGAGGCA---------------------------AGAAA   839
```

```
A.thaliana     AAAAAGGGTGACACTGTAAAG---CCTTCAC--TTTTTGCCAC-CTGGAAAGGGCCAGAGA---TCAACTTATTTACCACTAGGATAGCTTCCCACTTCAACGG----TTAAGCTGATGAATTCATGTACC----------------A--   941
A.lyrata       AAAAAGGGTGACACTGTAAAG---CCTTCAC--TTTCTGCCAC-CTGGAAAGGGCCAGATC---TCAACTTATCTACCACTAGGATAGCTTCCCACTTCCATGG----TTAAGCTGATGAATTCCTGTACC----------------A--  1006
C.rubella      AAAAAGGGTGAAACTGTAACGAAGCCTTCAC--TTTTTGCCGC-CTGAAAAGGGCCATAGATC-TCAACTCAG-TACCTCTACTATAGCTTCCCACTTCCATAGTTAATTAAGCTGATGAATTTCTGTACC----------------AAT   964
N.officinale   AAAAAGGGTGATACTGTA------CCTTCAC--TTTTTGCCAC-CT-AGAAAGGCCATAGAGCTTTAACTTATCTACCTCTACGA-AGCTTTCCACTTCCATGG----TTAAGCTGATGATTTCCTGTACC----------------AGT   650
A.alpina       -----------ACTGTAACTAATCCTTCAC--TTTTTGCCAC-CTGGAAAGGGCCATAGAGC-TCAACTTATCTACCACTTGGATGGCTTCCCA-------------TTAAGCTCATGAATTCCTATACC----------------AAT   958
B.napus        AATAGAAGTGACACTGTAACTAAGACTTCAC-TTTTTTGCCAC-CTGTAAAAGGCCATAGAGCCTCAGCT---------------TATCTACCACTTTCCATGGG---TTAAGCTGATGAATTCCTGCACC----------------AAC  1154
B.oleracea     AATAGAAGTGACACTGTAACTAAGCCTTCACTTTTTTTGCCCCACTGTAA------CTAAGCCCTC-------------------------ACCACTTTCCATGGG---TTAAGCTGATGA------GCACC----------------AAC   929
S.irio         AATTAAGGTGACACTGTAACTAAGCCTTCAC--TTCTCGCCAC-CTGGAAAAGGCCGTAGA-GCTCAGCTTA-----------TATATCTACCACTTTCCAC-GG---TGAAGCTGATGAATTCCTGTACC----------------AAT   976
S.officinale   AACTAAGGTGACACGGTAACTAAGCCTTCAC--TTTTCGCCAC-CTGGAAAAGGCC------ACTCAGCTTA-----------TATATCTACCACTTTCCAT-GG---TTAAGCTGATGAATTCCTGTACC----------------AAT   930
A.arabicum     AAAAACAATGACACTGTACCTAAGACTTCAC--TTTTTGCCAC-CTGGATAAGGCCCTAG------AGTTCAACTTAGCAAA----------ACATTTCCATGGTT---TAAGCAGCTGAATTCCTGAACCATTCAATAAACATTGGAAT   967

A.thaliana     -AATCAATACTTTTTAGCCATAAATGAGTCAGTTTTAG-TATCCACATGAATTTACCTACCAGAGTG-TTG---TAAATTATGTTCTTTTTGGGCCACTTACATGGATCTCATTCATTCACTGCAGCGA--------------------  1064
A.lyrata       -AATCAATATTTTTTAGCCATAACTGATTCAGTTTTAA-TACCCA----------------AGTGTG-TTG---CAA-------------------------------------------------------------------------  1061
C.rubella      CAAGATATTCGTTTCAGCCATAAGTGAGTGAGTGTTAA-TATCCATATGAATTTAGCCACTAGAGTG-TCA---GAA-----GTATTTTTTGGACCTCATAAATGGATCTCATTCATCCACTGAAGCCAGAAGTGT------------CC  1092
N.officinale   CAAAAGATTCGTTTCAGCCATAACTGAGAAAGTGATAATTAACAATATGAATTTAGCCACTAAAGT--TCA---AAACACATGTTATTTTTGGACA-CATAATTGAACCTCTTCATCTACTGAAGCCAGAAGTGT------------CC   782
A.alpina       CAAAAGATACGTTT-AGACATAACTCAGCGAGTGTTAG-TATTCATGTAAATATAGCCACCAAAGTG-TCA---TGATATG--TCCTTTTTGGACCACATAAATGGATCTCATACATCCACTGAAGCCAGAAGTGT------------CC  1088
B.napus        CAAAAGATTCGTTTTAGCCATCACTGAGTGAGTGTTGA-TATCCATACGAATTAAGCCACTAGATTGAGTG---TAAAATATGTTAT-TTTGGACCACATAAATGAATCTCATTCATCCACTGAAACCAGAAGTGTGACAA-GCGCCACC  1298
B.oleracea     CAAAAGATTCGTTTTGGCCATCACTGAGTGAGTGTTGA-TATCCATATGAATTTAGCCACTAGATTGAGTG---TAAAATATGTTCTCTTTGGACCACATAAATGAATCTCATTCATCCACTGAAACCAGAAGTGTGACAACCCGCCACC  1075
S.irio         CAAAAGATTCGCTTTGGCCATCACTGAGTGAGTGTTAA-TGTCCATATAAATTTAGCCACTAGATTGAGTG---TAAAATAAGTTATTTTTGGACCATATAAATGGATCTCATTCATCCACTGAAACCACAACTGTGACAA-GTGTCACC  1121
S.officinale   CAAAAGATTCGCTTTGGCCATCACTGAGTGAGTGTTAA-TATCCATATGAATTTAGCCACTAGATTGAGTG---TAAAATATGTTCTTTTTGGACCACATAAATGGATATCATTCATCCACTGAAACCAGAAGAGTGACAA-GTGTCACC  1075
A.arabicum     CAAAAGATTAGTTTCTGCTAGAGCTTAA-----------TATCCACATGAATTTTACCACTAGAT----TGTCTTAAAATCTGTTCTTTTTGGACCACTTTAACTGACCTCATCCATCCACTGATCCCAGAAGTGCC------------T  1090

A.thaliana     ---------------------GTTCTCAGACCACCAGAAAATTTATTCAGTGATCTG----TTTTGATCA-------TGCAACATAAACTTATAAGCCACACAAG-CAAAACAAAGATATCCCATGTTGC-ATATAATACGAGCTAGCATAT  1182
A.lyrata       -------------------------------------------AATCAGTGATCTG----TTTTCATCA-------TGCAGCATGAATTTATAAGCCACACAAG-CAAACCAAAGATGAGCCCTGTTGC-ATATAGTACTAGCTAGCATAT  1156
C.rubella      ACT--------GCAGTCAAGTTATCAGACCACCAGAAATTCTAGTCAGTGATCTG----GTTTGATCA-------TGCTGTACAAACTTATAAACCACGGAAG-CAAAACAAAGATATTCCCGGTTGC-ATATAATA----GTAGCATAT  1217
N.officinale   ACTAGAACACAATAACAATGTTCTCAGATCACCAGAAAATCTAGTCAGTGATCCC----GTTGGATCA-------CATAACACAAACTTATAAACCACACAAG-CAAATCAAAGATACCCCAGGTTCC-AT----------------ATAT   904
A.alpina       ACTGCAGCAAAGT-------TCTCAAGACCCCGAGAAAATCAAGTCAGTGATCTG----GTTTGATCA-------CGCAGCACAAAATTATAAACCACGCAAG-AAAATCGAAGATACCCCTGGTTGG-ATATAATAC----TAGCATAT  1214
B.napus        ACTGCA--------------TTCAGAACCCATCAGAAAATCTAATCTGTGATCTGTTTAATTTGATCA-------TGTAGCACAAACTTATAAACCACGCAAG-AAAATGG-------------------ATATGATAT----TAGCATA-  1403
B.oleracea     ACTGCA--------------TTCAAAACCCACCAGAAAATCTAGTCTGTGATCTG----GTTTGATCATGTAGGCTGTAGCACAAACTTATAAACTACTCAAG-AAAATGGACGATATCTCTGGTTAC-ATATGATAT----TAGCATAT  1201
S.irio         ACTG--------------------------CAG-AAATTTAGTCTGTGATCTG----ATTCGATAA-------TATAGAACAAACTTATAAACCACGCAAGAAAAATCGAAGATATCTCTGGTTTC-ATATGATAT----TAGCGTAT  1226
S.officinale   ACTG--------------------------CAGAAAATCCAGTCTGTGATCTG----AGTCGATCA-------TATAGCACAAACTTAAAAACCACGCAAG-AAAGTCGAAGATATCTCTGGTTTC-ATATGATAT----TAGCATAT  1180
A.arabicum     ACTGCAGCAAGC---------TCTCAGACCACAAGAAAAT-TAGTCAGCAATTTG----GTTTGTTCA-------TGGAATCCAGCTGAATACGTCAAGTAAA-TAAATCGCAGATTTCCATGGCCTGTACATATTACT----AGAATAT  1214

A.thaliana     CATAAAGAAGGAAACTTGAAGTAGCAAAG---TTTC---------------------------------------------------------------------------------------------------------------TACT  1219
A.lyrata       C-------AGGAAACTTGAAGTAGTAAAG---TTTC---------------------------------------------------------------------------------------------------------------TACT  1186
C.rubella      CATCAATATTGAAACTTGAAGCAGTAGAG---TTTC---------------------------------------------------------------------------------------------------------------TACT  1254
N.officinale   CATCATCAAGAAAACTTGAAGTAGTTGAG---TTCC---------------------------------------------------------------------------------------------------------------TACT   941
A.alpina       CATCAACAAGGAAACTTGAAGTACCAGAAAGTTTGT---------------------------------------------------------------------------------------------------------------TACT  1254
B.napus        ------------------------TAGAGAGTTTGATCTAGATATCTTTCCAGGAAATTTAAAGATATGCAATAACAATTTTAG--TAAATCCTAACGAGTGTAAAGAGCAAAATGTTAATAAAATGAATGTTAACATTTAGTTTGCTCT  1527
B.oleracea     CATTATCAACAAGG---AAGCTAGTAGAGAG-TTGATCTAGATATCTTTCCAGGAAATTTAAAGATATGCAATAACAATTTTAGGCTAAATCCTAACGAGTGTAAAGAGC----------CAATTGAGTGTTAACATTTAGTTTGCTCT  1336
S.irio         CATCATCAACA-------------------------------------AGAAAGCTTGAAG---------------------------------------------------------------------TAGTAGAGACTTTGTTCT  1268
S.officinale   CATCATCAACA-------------------------------------AGAAAGCTTGAAA---------------------------------------------------------------------TAGTAGAGAGTTTGTTCT  1222
A.arabicum     CATCAACAAGGAAATTTGAAGTAGTAGAGT---TCCC--------------------------------------------------------------------------------------------------------------ACT  1251

A.thaliana     AAATTTCTTGT----CAGGAATT-TTTAAAATGCAATGACAACC-------------ACTTGGAGCACTA---TGAGTTTCAGAGCCAATAGAATGTTACTATT---TGGTG----TGGATTCGAGCTAGCACGTGAAAGTGCATAAAA  1340
A.lyrata       AAATTTCTTGT----CAAGAAAT-TTTAAGATGCAATGACAACC-------------ACTTGGAGCACTAAC-AGAGTTTAAGCGCCAATAGAATGTTACTATT---TGATG---ATGGATTCGAGATATCACGTGAAAGTGCATAAAA  1310
C.rubella      AAATTTCTTGT----CAGGAAAT-TTTAAGATGCAATGATAATC-------------ACTTGGAGCACTAA--CTAATTTAAGAGGCAATACAATGTTAATATT---TGGTG---ATGGATTCGAGATAGCACGTGAAAGTGCATAAAA  1377
N.officinale   AAATTTCTTTTTTCCCAGGAAA--TTTAAGATGCAATGAAAACA-------------AGTTGAAGCACTAA--CGAATTTAAGA--CAGTAGAATGTTAAAATT---TAGTG---ATGGATTCGAGATATCACGTGAAAGTGCATAAAA  1065
A.alpina       AGATTTCTTTT----CAGGAAAT--TTAAGATGCAATGACAACT-------------AGTTTGAGTACCAACGAATGTATTGGAGTCAACTAAATGTTAATGTTC--TAGTA---ATGGATTCGAGATATCACTAGAAAGTGCATTAAA  1379
B.napus        AGATTTCTTTT----CAGGAAA--ATGAAGATACAGTAACAATT-------------AGTTGAAGTATACTAGCAAATGTAAGAGCCACTTGAATGCTAATCTT---TAGTG---ATGGATTTGAGACAC---GTGAAAGTGCATAAAA  1648
B.oleracea     AGATTTCTTTT----CAGGAAA--ATGAAGATGCAATAACAATT-------------AGTTTAGGTATACTAGCAAATGTAAGACCCACTTGAATGCTAATATT---TAGTG---ATGGATTCGAAACAC---GTGAAAGTGCATAAAA  1457
S.irio         AGATTTCTTTT----GAGGAAATTTTAAAGATGCAATAACAATTTA-----------ATTTTAAGTAGG--AACGAGTGTAAGAGCCAACTGAACGTTAATAAT---TATAGCCAATTGATTGCTAATAT---TTACGGATGCATAAAA  1394
S.officinale   AGATTTATTTT----CAGGAAA--TTGAAGATGCAATAACAATT-------------AGTATACGTACGCTAACGAATGTAAGAGCCAACTGAATACTAATATT---TAGGG---ATGGATTCGAGATATGGCGTGAAAGTGCATAAAA  1346
A.arabicum     GGAATTTTCTTCA---GGAAA---CTTAAAATGCAATAACAAGCACTTAAACTAAAATACTTAAGG----------ATCTAGGAGCCAAATCAATGATAATAGTTAGTGATGTT-----ATACCAAAAG-----TGAAAGTGCATAAAA  1374
```

```
A.thaliana     GTGATTACCTTTTGCCAAAGGTCACTGCAC--------TTTTCCTC----AGATA--GTTTCTCACAGCCATGGAAAGTGGAGAATCCGCAT-AAACGTACAATTACAAGCTTTATATGGTCCCTCGACTCT---------TATTCTCTT  1466
A.lyrata       GTGATTACCTTTTGCCAAAGGTCACTGCACTTTTGCATTTTTCCTC----AGATA--GTTTCTCACAGCCATGGAAAGTGGAGAATCCGCACAAAACGTACAATTAC-AGCTTTATATGGTCCCTCGACTCT---------TATTCTCTT  1444
C.rubella      GTGATTACCTTTTGCCGAAGGTCTCTGCACTTTTGCACTTTTCCTC----AAATATAGTTTCTCACAGCCATGGAAAGTGGAGAATCCGCAC-AAACGTACAATTACAAGCTTTATATGGTCCCTCGACTCT---------TGTTCTCTT  1513
N.officinale   GTGCTTACCTTTTGCCAAAAGTCACAGCACTTTTGCTCTTTTCTTC----AGATAATTTCTCTCACAGCCATGGAAAGTGGAGAATCCGCAT-AAA---ACTATTACAAGCTTTATATGGTCCCTGGACTCT---------TGTTCTCTT  1198
A.alpina       GTAATTACCTTTTGCCAAAGGTCACTGCACTTTTGTACTTTTCCTT----AGATAG------TTACAGCCATGGAAAGTGGAGAATCCGCACAATACTTACAATTACACGCTTCATATGGTCCATGGACTCT---------TGTTCTCTT  1510
B.napus        GTGATTACCTTTTGCCAAAGGTCTCTGCTCTTTTGCACTTTTCCTT----AGATAGT-TTTTTTACAGCCATATAAAGTGGAGAATCCTCACATAGCTTACAATTGCACGCTTTGTATGGTCCCTAGACTCTAAGATTCTAGGTTCTCTT  1793
B.oleracea     GTGATTACCTTTTGCCAAAGGTCTCTGCACTTTTGCACTTTTCCTC----AGATAGT-TTTCTCACAGCCATATAAAGTGGAGAATCCT-ACATAGCTTACAATTGCAAGCTTTGTATGGTCCATGGACTCTAAGATTCTAGGTTCTCTT  1601
S.irio         GTAATTACCTTTTGCCAATGGTGGCTGCACTTTTGCACTTTTTCTC----AGATAG--TTTCTCACAGCCATGGAAAGTGGAGAATCCGCACAAAACTTACAATTCCAAGCTTTATATGGTCCCTGGACTCT---------TGTTCTCGA  1529
S.officinale   GTGATTACCTTTTGCCAAAGGTCGCTGCACTTTTGCACTTTTCCTC----AGATAG--TTTCTCACAGCCATGGAAAGTGGAGAATCCGCACAAAACTT-CCATTACAAGCTTTATATGGTCCCTGGAGTCT---------TGTTCTCGT  1480
A.arabicum     GTGATTACCTTTTGCAAAAGTTTGCTGAACTTTTGCACTTTTCCACGTGTAG-------TTCTCACAGCCCTGCAAAGTGGAGAACCCATATAAAACCTACAA-TACAAGCTTTACGTGGTCCCTATACCCTT---------ATTCTCTT  1507

A.thaliana     CTCAGTCTTTG-----CAACTAAATAGGGTTT--TCGT-TAATCTGAAAGAAGCAAAG--------------------------------------------------------------TATTCGCAACCACGGAAACCT---------------  1537
A.lyrata       CTCAGTCTTTG-----CAGCTAAATACGCTTT--TCGTAAAATCTGAAAGAAGGAAAG--------------------------------------------------------------TATACGCAA-CACGGAAAACT--------------  1515
C.rubella      CTCAGTCTTTG-----CAAGGAAATAGGCTTT--TCCTTAAATCTGAATGCGGGAAAG--------------------------------------------------------------TATTCGCAACCACAAAGAAAT--------------  1585
N.officinale   CTCACTCTTTA-----CAAGAAAATGGGCTTTTCTCCTTAAACCTGACAGAAGGAAAG--------------------------------------------------------------TATTTGCAATCACGGAAAACT--------------  1272
A.alpina       CTCTGCCTTTT-----TAAGAAAAGAAGCTTTTCACCTTAAATACGAAAGAAGGAAAG--------------------------------------------------------------TATTCACAACCACGGTAAACT--------------  1584
B.napus        CTCAGTCCTTT-----CAAGAAAAGAAGGTTTTCTATTTAAATCTGAAAGAAGGAAAG--------------------------------------------------------------TATTCGCAACACAGAAGCT----------------  1865
B.oleracea     ATCAGTCTTTT-----CAAGAAAAGAAGGTTTTCTATTTAAATCTGAAAGCTGAAAAAAATAATAATGAAACAAATAAGTGACATGTTATAAAAGTAATAGAAAAGTCTATCTTTATTCATAACATAGAGGTTCTTTATATAGGAGATTA  1746
S.irio         CTCAACCTTTT-----CTAAAAAAGAAGCTTTTCTCCTTAAATCTGAAAGAAGGAAAG--------------------------------------------------------------TATTCGCAACCACGGAAACT---------------  1602
S.officinale   CTCAGTCTTTTTA---TTAAAAAAGAAGCTTTTCTCCTTAAATCTGAAAGAAGGAAAG--------------------------------------------------------------TATTCGCAAA-----------------------  1545
A.arabicum     TTTAGTCTTCGGCAGGAAAAAAAACAAATTTCT--CCTTAATTGTAGATGA-------------------------------------------------------------------TATTTGAA-------------------------  1564

A.thaliana     -----------------GATAAAGAATGAAAACA------------------------------------AATAAGCAATAGTGTTTTCTTG----A--AAATCTCGATGCAACTTTCAGG--------------ATATTGTT-AC  1609
A.lyrata       -----------------GATAACTAATCAAAACA------------------------------------AATAAGCAATATTGTTTTCTTTCCTCA--ATCTCTAGAATTGAGTCACAGTGGAAGAAAAAAATGCATTTTCCT-CC  1606
C.rubella      -----------------GATAAACAATGAAAATA----------------------------------------------AGCGATATAGTTTTCTTTCCTCA--ATCTCTAGAACTGAGTCACAATAG--ATCAAAAATGCAGCTTTCG-CC  1670
N.officinale   -----------------GGAAAATAATGAAAATA------------------------------------AATTAGCGATATAGGTTTGTCTCCTCA--ATCTCTAGAATTGAGT---------ATCAAAAATGCATTTAACT-CC  1353
A.alpina       -----------------GATAAATAATGAAAACA------------------------------------GATAAACCATATAGTTTTTTT--GTCA--ATATCTAAAATTGAGTCACAACAGA---TCAAAATACATTTTCTT-CC  1670
B.napus        -----------------GAAAAATAATGAAA-CA------------------------------------AATAAGTGACATAGTTTTCTTTCCATAATATCTCTAAAACTGAGTTTAATTGG--CTTAAAATTGAATTTTCCT-TC  1955
B.oleracea     CATCGTCATAGATAAATAGAAAGATTACAAAT-CATAATCTCTTGGTTATGAGCCATCCACAATCTGGTTCATAATA--TCACATAGTTTTTTTCCCTTAATATCTCTAAAACTGAGTCTAATTGG--CATAGAATTGAATTTTCCT-TC  1890
S.irio         ------------------GAGAATAATGAAATCA------------------------------------AATATGTGACATAGTTTTCTTTCCTTA--ATCTCTAAA----------ATTGG--ATTGAAAATGTATTTTCCTACC  1681
S.officinale   -------------------------------CT-------------------------------------GATAAGTGACATATTTTTCTTTCCTTA--ATCTCTAAAATTGAGTCGCATTGG--ATTGATAATGCATTTTCCT-CT  1619
A.arabicum     ---------------------------------------------------------------------------------------------------------------------------------------------------  1564

A.thaliana     ATATGATCTA------TTACTCGTAACA--------GTTATCCGAAGGCCTACACATGTGAGAGAAGTTCCAAACCGC--TACAACAATAAACTTAATTAGAGACTGTCAACGAGCAATAATAAGCAAAACTACTTTTTTCTTGAGCTAC  1743
A.lyrata       ACATAATCTC------AATGTCAT------------GTTATCTGAAGGCCTACACATGTGAGAGAAGTTCCAAACCGC--TACAACAATAAACTTAC----AGATTGTCAAGGAGCAATAATAAGCCAAA-TACCTTTTTCTTGAACAAC  1731
C.rubella      TTATAATCTC-ATGCAACTTTCAT------------GTTATGTGGAGGACTACACATGCGAGAGAACTTTCAAACAGC--AACAACAATAAATTCAC----AGAT--TCAAAGAGCAATAATAAGCGAAA-AACTAATTTCTTGAACTAC  1798
N.officinale   ACATAATCTCGATGCAACTTTCAT------------GTTATATGGAGGCCTACATATGCGAAAGAAGTTACAAACCGCTAACAAAAAATAAACTCAC----AGATTGTCAAAGAGCAATAATAAGCAAAA-TAGTTATTTCTTGAACTAC  1486
A.alpina       ATTTAATTCCAATGCAACTTTCATGTAAGATGGAGGTTTATACAGAGGTCTACACATGTGAGAGAAGTTGCAAACCGC--TACAACAATAAACTCAC----AGATTGGCAAAGAGCAACAATAAACAAAA-TACTTCTTTCTTGAACTA-  1812
B.napus        ACATAATTCCGATGCAACTTCCAT------------GTAAGGTGGAGGTCTACACTTGTGAATGAAGTTCCGAACCAC---------------------AGATACACAAAGAGAAATAATAAGCAAAA-CTA-TATTTCTTGAACTAC  2068
B.oleracea     ACATAATTCCAATGCAACTTCCAT------------GTAAGGTGGAGGTCTACACTTGTGAATGAAGTTCCAAACCAC---------------------AGATAGGCAAAGAGAAATAATAAGCAAAA-CTA-TATTTCTTGAACTAC  2003
S.irio         ACAACAAACCGATGCAACTTTCAT------------TTAAGGTGGAGGACTACACATGTGACAGAAGTTCCAAACCAC---------------------AGATAGGCAAACAGAAATAATAAGCAAAC-CTATTCTTTCTTGAACTAC  1795
S.officinale   ACAAAATTCCCATGTAACTTTCAT------------GTAAG--------------------AGAAGTTCTAAACCAC---------------------AGATAGGCAAACAGAAATAATAAGCAAAC-CAATTCTTTCTTGAACGAC  1712
A.arabicum     -------------------------------------TAAAATGATCATCTA------------AAGCACCAAAGC------------------------------------------------------------------CTTGCTCAAATTAT  1605

A.thaliana     AAGTG----AAAAGGCCAATACACA-------ATTTACTCTTCATGAACTCGAACCACGTTACAATCTCCAAA-AAATTTCATCACCAAAGCAC-TAAAAGCCAAAGATGCCTCAACT-TATCCAAC----------------------  1856
A.lyrata       AAGTGA---AAAAGGCCAATACACA-------ATTCACTCTTCATGAACTCAAA-CGCGTTATTTCTCCAAAA-AAATTTCATAACAAAAGCAC-TAAAAGCCAAAGATGCCTCAACT-TATCCAAC----------------------  1844
C.rubella      AAGTGAA-AAAAAGGCCAATACACAGAACACAGTTTACTCTTCATGAACTTAAACCACCTTAAAAA------A-AAATTTCATCATTAAATCAC-TAAAAGCC-AAGATGCCTCAAAT-TATCCAATTACTTC-GTGGCATAGACATCAA  1936
N.officinale   TAGTG---AAAAAGGCCAACACAC----------TTACTCTTCATGAACTTGAACCACCTTAAAACCTT---C-GTTTTTCATCATTACAGCAC-TAAAAGCC-AAGATGTC--------------------------------------  1579
A.alpina       ------------------------------------------------GAACCACTTTAAAACCTCCA-T-AATTTTGACCATTAATGC-C-CAAACGCC-AAGATGCCTCAAATATATCCAATTACTTCTATGGTTTAGACACCAA  1906
B.napus        AAGTGATAAAA-GGGTCAATACACA---------A--CTCTTCATGAGCTAGAACCACTATAAAATCCCCACT-TTTTCATATCATTAATGT---AATGAACCTAGAATCCTTCTCTTCTATAAGATTTA-------------------  2182
B.oleracea     AAGTGATAAAAGGGGCCAATACACA---------ACTCTCTTCATGAGCTAGAACCACTATAAAATCCCCACT-TTTTCATATCATTAATGT---AATGAACCTAGAATTCTTCTCCTTCATGAGATTTC------------------  2120
S.irio         AATTGAAAAAG---GCCTATACACA---------TTACTCTTCATGAACTAGAACCACTTCAAAATCTCCACTATTTTCATATCATTAAAGCGCTAAAAAAGCAGAGACACCTCAGAGTTATCCAATTACTT-------CTAAACATCAA  1926
S.officinale   AAGTGAAAAAG---GCCTATACACA---------TTGCTCTTCATGAACTAGAACCACTTCAAAATCTCCACTATTTTCATATCACTAAAGCGC-AAAAATGCAGAGATGCCTCGAATTTATCCAATTACTT-------CTAGACATCAA  1842
A.arabicum     ---------------CCAAT---------------TACTCCT-ATGGATTAGA---------------------------------CATCAAAC---------------AATGGCTAGCT-----------------------------  1646
```

```
                                                                                                                     * ***  *     ****  * *
A.thaliana    -------TTGGCAGGATAAAGATCTCCAAAAATGCTTA-----CTAAAGAACCTAGAAT-C-TTTTCTTTAGAATTCAATGATCATATAACCATTTCATAACAATTCTAAATGCCATTACATTCATCGTA-AAACCAGTAAATA-AC---   1987
A.lyrata      -------TTGGCAGGATAAAGATCTCCAAATATGCATAGTAATGTAAGGAACCTAGAAT-CTTTTTCTTCAGAATTCAATGATCATGAAACCATTTCATAACAATTCTACATGCCATTACATTCCTCATA-AAACCAGAAAGTA-AC---   1981
C.rubella     ACTTT--AGCGAGATATAAAGATCTCCAAATATGCTTACTAATGTAAGGAACCTAGTATCATTTTTCTTTAGAGTTCAATGATCATGAAACCATTTCATCACATTTCTCCATGCCATTACATTCCTCATA-AAACCAGTAAATA-AC---   2079
N.officinale  ----------------------------------------------------------------TNNNTTAGAGTTCAAAGATCATGAAACCATTTCGACACATTTCTCCATGCCGTTACATTCNTAATA-AAACCAGTAAACG-AC---   1660
A.alpina      ACTT----TGATGACATAAAGATCTCCAAAGATGCTTACTAATGTAATAAACCTAGATT--TTTTTTCTTAGAGTTCAGTGATGATGAAACCATTTCATAACATTTCTGCATACTATTACATTCCTCATACAAACAGGTAAACA-AT---   2046
B.napus       ----------------------------------------------------------------------------------------------------------------TCCATTACATTCATC-AT-AAACCAGTAA----AT---   2211
B.oleracea    ----------------------------------------------------------------------------------------------------------------TCCGTTACATTCATCAAA-AAACCAGTAA----AT---   2150
S.irio        ACTTTGGCGGG----ATAAAGATCTCCAAATATGTTTACTAT----------------------------------TAATG-------------------------------TCAATGACATTCCTCATA-AAACCAGTATATATAT---   2003
S.officinale  ACTTTGACGGGATAAATAAAGATCTCCAGATATGTTTACTAGTGTAAGGAACCTAGAATA---TTTTTATTAAGTTCAATGATCATG-AACAATTTCATAAGATTTC-----TCCATGACAATGCTCATA-AAACCAGTAAATATAT---   1979
A.arabicum    ----------------AAAACATTTGTAAAGAAACCTAT-GATGTAAGGAACCTATTATCTTTTTCTTTAGAAATTAAATGATAACGGAACAGTTCCATAACATTTTA-------CATTACATTCCTCAAA-AAACCAGTAAAT---TCGA   1769

  **    *              * * *     **  *** **                                                **** *   **  *                        * *
A.thaliana    --AA---GAACTTGTAT-GTT--ACGTT-----CCAATTACCAAGCA---AAAAAAA-------------------------------CCTTTTCAAAGTTTAAAG-------TT----CAAAATGGGAAAGAGA-AGTGCGGTGTAAG-   2077
A.lyrata      --AA---GAACTTAAAT-GTT--AAGTT-----CCAATTACCAAGC----AGAAAAA-------------------------------CCTTTTCAAAGTTTAAAT-------TTA--CTCATATAGGAAAGAGA-AGTGCGGTGTAAG-   2072
C.rubella     --AAGAAGAGCTTAAA----T--AAGTT-----CCAATTACCAAGC----AAAAAACTTATAAGTTCCAATTACCAAACAAAAAAAAACCTTTTAAAAGTTTAAATA------TTTTCCC-TTGTGGGAAAGAGA-AGTATAGTGTAAG-   2202
N.officinale  --AAGACGAACTTAAAA-GTT--ACGTT-----CCAATTACCAAGA----AAAAAAC-------------------------------CTTTTGAAAGTTTAAATT-------TTTTTCCGTTGTTGGAAAGGAG-AATGTGGTGTAAGA   1757
A.alpina      --AAA---AACCTAAAA-GTTTTAAGTT-----CCAATTACCACGAA---AAGAAAA-------------------------------CATTTTCAAAGTTTAAAACCCTCCCT--------TATTGGAAAGAAA-AGTGTGGAGTAAG-   2141
B.napus       --AACAACAACTGGAA--GTT--AAGCT-----CCAATTACCAAG-----AAATTTG-------------------------------T-TTTTGAGAGTTTAAAA-------TCGT-CCCTTACTAGAAAAAAATAGTGTGGTGTAAG-   2304
B.oleracea    --AACAAGAACTTGAA--GTT--AAGCT-----CCAGTTACCAAG-----AAAAAAA-------------------------------CATTTTGAGAGTTTAAAG-------TCGT-CCCTTACTAGAAAAAAA----ATAGTGTAAA-   2240
S.irio        --AACAGCAACTTAAAACGTT--AAGCT-----CCAATTACCAAGAAAAAAAAAAAA-------------------------------TCTTTTCAAAGTTTGAAA-------TCTT-CCC-TCTGAGAAAGAAA-AGTGTAGTGCAAG-   2102
S.officinale  --AACAGCAACTTAAAAAGTT--AAGTT-----CC-ATTACCAAG----AAAAAAAA-------------------------------CCTTTTCAACGTTTAAAA-------TCTT-CCCTTCTGAGAAAGAAA-AGTGTGCTGTAAG-   2074
A.arabicum    CAAAGAGGAATTTGA--------AAGTTAAATACCAATTATCAAAAAC---AAAAAAAA-------------------------------CTTTTAAAATTTCTATATCTTCC--------CTTACGGAACAGAAA---------------   1855

     ********  *******  * **********  *********  * ********  *              * ********      *    *******   ***    *                    ** ****           *
A.thaliana    CAAATATGAAAGAGGAAGAGATGCGAAAAGTGTATCCTAGGACCAG-CATTTTATACAAAAAAAAAACTCACTTTTCAGCTCTTAAGGCATAGAG-TGAAGGTAGCCATATGAA--TTTGGCCACTAGAGCGTCCGTCAAATCTCATTC-   2222
A.lyrata      CAAAAATGAAAGAGAAAGAGATGCGAAAAGTGTATCCTAGGACCAG-CATTTTATACAAAAAAAAAACTCACTTTTCAGCTCTTGAGGCATAGAG-TGAAGGAA----------------GCCACTAGAGCGTCCGTCAAATCTCATTC-   2203
C.rubella     CAAAAATGAAAGAGGAAGAGATGCGAAAAGTGTATCCTAGGACCAG-CATTTTATACAAGA--AAAAATCACTTTTCAGCGGTAGAGGCATAGAG-CGAAGGAAGCCACATGAA--TTTGGCCACTAGAGTGTCCATTAAGTCTCGTTC-   2345
N.officinale  AAAAAATGAAAGAAGAAGAGATGCCAAAAGTGTATCCTAGGACCAG-CATTTTATACTAG----AAAATCACTTTTCAGCGGTAGAGTCATAGAG-AGAAGGAAGCCACATGAA--ATTGGCCACTAGAGTGTCCATTAAATCTTCTTC-   1898
A.alpina      CAGAAATGAAAGAGGAAGAGATGCGAAAAGTGTATCCTAGGACCAGCCATTTTATACAAGTA---AACTCACTTTTCAGCTATAGAGGCATAGAGTTGAAGGAAGCCACATGAA--TTTGGCCACTA-----------ACATCTCGTTC-   2274
B.napus       AAATAATGAAAGAGGAAGAGATGCGAAAAGTGTATCTTAGGACCAGCCTTTTTATACAAGCA---AGCTCACTTTTCAGCTGTAGACTCATAGAGTTGAAGGAAGCCACATGAA--TTTGGCCACTAGGGTCTTC-TAA-----------   2437
B.oleracea    TA--AATGAAAGAGGAAGAGATGCGAAAAGTGTATACTAGGACCAGCCTTTTTATACAAGCA---AGCTCAGTTTTCAGCTGTAGACTCATAGAGTTGAAGGAAGCCACATGAA--TTTGGCCACTAGGGTCTTC-TAA-----------   2371
S.irio        CAGTAATGAAAGAGGAAGAGATGCGAAAAGTGTATCCTAGGACCAGCCTTTTTATACAAGCA---AGCGAACTTTTCAGCTGTTGAGGCATAGAGTTGAAAAAAGCCACATGAATTTTTCGCCACTAGGGTCTCC-TAAAATCTCGTTC-   2247
S.officinale  CAGAAATGAAAGAGGAAGAGATGCGAAAAGTGTATCCTAGGACCAGCCTTTTTATACAAGCA---AACTAACTTTTCAGCTGTTGAGGCATAGAGTTGAAGAAAGCCACATGAATTTTTTGGCCACTAGGGTCTCC-TAAAATCTCGTTC-   2219
A.arabicum    -----ATGAAAGAGTAAGAGATACGAAAAGTGTATCCTAGGACCAGCC-TTTTATACAACAAA----CTCACTTTTCAGCCGCAGAGGCATAGAGTTGAAGTAAACCACATGAAT--TTGGCGACTACAGTGTC---GAAGCATCCTGCG   1990

   ******   **    *    ****        ****  *  *  *  **   *****                      *   **  *******         *                   *    *  *  *   **  *    *   **
A.thaliana    -TTTTTGGAC--CAC-ATAATGGGTATCA----TACATTCACTGGACCCAAAAGCGTAACTGGAGCTAGTCCTC-AAACCTAGAGAGTAT------CGTATC------CTGTAGC-TTCCACATAGTAAA-CATTATGAGCATAACACCA   2349
A.lyrata      -TTTTTGGAC--CAC-AAAATGGGTATC-----TACATCCACTGGACCCAAAAGCGCAACTGGAGCTAGTCCTA-AAACCTAGAGAGTAT------CATATC------CTGTAGC-TTCCACATAGTAAA-CATTATGAGCATAACAACA   2329
C.rubella     -TTTTTGGAC--CAC-AAAGTGGGTATAA----TACATCCACTGAACCCAAAAGCGTCACTGGAG-CTAGTCGT-CAACTTAGAGAGTAT------CATATC------ATGTAGC-TTCCACGTAGTTAATTATTATTAGCATATCACTA   2472
N.officinale  -TTTTTGGAC--CAC-AAAACGGGTATCA----TACATCCGCCGAACCCAAAAGTGTCCTTCGAACAAAGTCCT-CAACCTAGAGAGTAT------CACATC------TTGTAGG-TTCCACACAGTAAAGTATTATTAGCATATCACTA   2026
A.alpina      -TTTTTGGAC--CAC-AAAATGGGTATCA----TACATCCACTGAACACAAAAGCGTAATTGGAGATAGTCCTCAAAACCTACAGAGTAT------CATATC------CTGTAGC-TTCCGCATAGTAAA-CATTGTTAGCATATCACGA   2402
B.napus       -CCTTTGGAC--CACATAAATGGGTATCATTAGTACATCCACTGAACCCAAAAGTGTCCTCA-------------AACTCTAGAGAGTAT------CATATTCATACACTGTAGCTTTCCACATAGTAAA-CATTATTAACATATCACCA   2564
B.oleracea    -CCTTTGGAC--CACATAAATGGGTATCATTAGTACATCCACCGAACCCAAAAGTGTCCTCA-------------AACTCTAAAGAGTAT------CATATTCATATACTGTAGCTTTCCACTTAGTAAA-CATTATTAACATATCACCA   2498
S.irio        -CTTTTGGAC--CACAAAAATGGGTATCA----GACATCCACTGAACCCAAAAGTGCCCTCC-------------AAACCTAGAGAGTAT------CATATTCATATACTGCAGC-TTCCACATAGTAAA-CATTATTAGCATATCACCA   2369
S.officinale  -CTTTTGGAC--CAC-AAAACGGGTACCA----TACATCCACTGAACTCAAAAGTCTCCTCA-------------AAACCTAGAGAGTAT------CGTATT-----------------CACATAGTAAA-CATTATTAGCATATCACCA   2324
A.arabicum    CTTTTTGGACCACA---AAATAGGTATCAT----ACATCCACTGAACTTAAAAGTGACTGGAGCAAGTACTC---AAACCTAAAGAGTATATACACTATAGCT--------------TCCAAATAATAA---ATAATT-GTACTCCATCC   2112

 **    *                                     ***      * *                                                                *
A.thaliana    ACAAGGCAACTCCAAGTACTAATG----GTTATTAGTACAGG--AAAACCCACGATGCTAAA------CACATGAATG---------------------------------GGTCACCAAATAGAGTGAAGATGGTTA------------   2442
A.lyrata      ACAAGGCAACTCAAAGTACTAATG----GTTACTAGTACAAG--CAAACCCAAGATGCTGAA------CACATGAATG---------------------------------GGTCACCAAATAGCTTGCATTTGCTAACCAAGAAACTAG   2434
C.rubella     CCAAGGCAACTCGAAGTACTAGTGGTTAATTACTAGTACAAG--CAAACCAAAGATGCCGAA------CACATGAATA---------------------------------GGTCACCAAATAGCTTGCATTTTCTAACCAAGAACTAG   2581
N.officinale  ACACGGCAACTCAAAGTACTGGAG----CTTACTAGTACAAG--CAAACTCAAGATGCTCAA------CACATGATTG---------------------------------GCTCACCAAATAGCTTTCTTTTACCAACCAAGAAACTAT   2131
A.alpina      CCAAGACAACTCTAAGTACTGGTG----GCTACTAGTACAAG--CAAACCCAAGAGGCTGAA------TAAAAGAACG---------------------------------AATCACCAAATGG--------------------AAATAG   2487
B.napus       ACAAGGCAACTCAAAGCACTGGTG----GTTATTAGTACAAAAGCAAACCCAAGATGATGGATATACCGTACACGAACGGTCATCAGATAGTTTTTTTTTTTTTGCTAAATGGTCATCAGATAGTTTTGTAGCTGCTAACCAAGAAAT-AG   2709
B.oleracea    ACAAGGCAACTCAAAGCACTGGTG----GTTACTAGTACAAAAGCAAACCCAAGATGATGGATATACCGTACACGAAC---------------------------------TGTCATCAGATAGTTTGCAGCTGCTAACCAAGAAAT-AG   2610
S.irio        ACAAGGCAACTCAACGGCCTGGTG----GTTACTAGTACAAA--TAAACCCAAGATGATCGA------TACACTTAT----------------------------------GGTCTCCAG----------TCTGCTAACCAAGAAACAAG   2463
S.officinale  ACAAGGCAACTCAAAGT---------------ATAGTATTAC--TAAACCCAAGATGATTGA------TACACGAAT----------------------------------GGTCACCAGATAGTTTGCATCTGCTAACCAAGAAACAAG   2417
A.arabicum    ACACAATA-------------------------------------AATTTGGCATGTC------------------------------------------------------ACTTCTAAGTAGTATGCATTAGATTTCTATGAAA----   2168
```

```
A.thaliana     --------------------AATTGCATCTATGGATCATGTGGACTAGTAAATGAGTG--TAGCAGAAAACTT-CACAATTACCTCTGTGATCTT--AGAAACATGTCCTGAAAA--TTCC------------------------------  2536
A.lyrata       C-------------------AATGGTGCTTAAAGATACTGTGGACTAGTAAACGAGTG--TAGCACAAAACTT-TACAATTACCTCTGTGATCCT--ATAAACAAGTTCCTGAAATTTTCC------------------------------  2531
C.rubella      CAAGCGAAGGTGCTTTTAAAATTTGCATCTACTGATCCTGTGGACTAGTAAACGAGTG--GAGCTCAAAATTG-TACAATTACCTCTGTTATCCT--AAAGACAAGTCCT-GAAATTTTCC------------------------------  2696
N.officinale   CAAGCAAAGGTGCTTCAACAATTTGCATCTACTGATC---------AGTA---------------------------AATTACCTCTA--------------------------------------------------------------  2183
A.alpina       -------------------------------------------------ACATGGGCG--CGAGACAACTCTC-GTTAATGACAT----------------------------------------------------------------  2520
B.napus        TAAACGAAGGTGCTTTTAA-ATCTGCATCTACTGACCATATGGACTAGTAAACGAGAG--TAGCTAAAAACGC-TACAATTTCCTCCGTGATCCTATAAAGATAAGTCATGTAACTT-CCATCTCTGGCCTGAGACATAGCTGAAGTGTA  2854
B.oleracea     TAAACAAAGGTGCTTTTAA-ATATGCATCTACTGATCCTATGG---AGTAAACGAGAGCATAGCTCAAAACGC-TAC--TTTTCTCTGTGATCCTATAAAGACAAGTCATGAAACTTCCCATATCTGGCCTGAGACATAA---CAGTATA  2750
S.irio         TAAACAAAGGTGCTTTGAA-ATTTGCATCTACTGAT---------TAGTAAACGAACG--TAGCTCAAAACTCTTACAATTTCCTCTGTGATCT--TGAAGACAAGTCATGAAACGTTTTATC--TGACCTTAAGCA------CAGAATA  2591
S.officinale   TAAACAAAGGTGCTTTAAA-ATTTGCATCTACTGATCATATGGGCTAGTAAACGAACG--TAGCTCAAAACTCTTACAATTTCCTCTGTGATCC--TGAAGACAAATCATGAAACGTTTTATA--TGACCCGAGGCA------CAGAGTA  2554
A.arabicum     CATATACGGCTACCTTG-----CTACATACTTTGCTACTATGAT------------------------------------------------------------------------------------------------------------  2207


A.thaliana     -----ATACAAGT------------------------GTCGTTTGTATTAGATTACTTCCACAGGTTGAGATCTA-ATAAAGCTACAATA-----AATAGTATAGAGTATCATCATAAACCCAAA----------------------  2626
A.lyrata       -----AAACAGGT------------------------GTAGTATGAATTAGATTACTTCCACAGGTTGAGATCTA-A-AAAGCTACAATA-----AATAATATAGAGTATCATCATAAACTCAAAATG-------------------  2623
C.rubella      -----ACACAAGT------------------------GTAGTACGTATTACATATCTTCCACAGGTTGAGATCTA-ATAAAGCTACAATATATAGTATAGTATAGAATAACAACATAAACTCAAAATG-------------------  2794
N.officinale   ---------------------------------------------------------------------------------------------------------------------------------------------------  2183
A.alpina       -----GCCAAAAC------------------------TTAGTATGTA---------TACCACATATTGAAATCTA-ATAAATTTACAATAA-----ATAATAAAGAGGGTGATCATATTCTCAAAATG-------------------  2604
B.napus        TAA--AAGCTAAT-----------------------------------------------------ATGGAAACATATATAAAGCTGCAATA-------AACAATATATTATCATCAAATTGTCAAATTGTTACAGAAAAAATTGTCAAA  2942
B.oleracea     AAATTAAGCTAAT-----------------------------------------------------ATGGAAACATATATACAGCTGCAGAA-------AACAATATAT--TCATCATATTGTCAAAT---------------------  2816
S.irio         TAA--ACGCTAATGGCAGAAATAACATTCCGAATTTGATTTAGAATGTGTTGGATTACTTTCACAGATTAAAATCTA-ATAAAGCTGCATTA-------AACAATAGAGTATCATCATATCCTCAAAAT-------------------  2710
S.officinale   TAA--ACGCTAATGGCAGAAATAACATGCCAAATTTTAGTT---ATGTATTGGATGACTGTCACGGGTTAAAATCTA-ATAAAGCTGCAACA-------AACAATAGAGTATTATCA--------------------------------  2658
A.arabicum     --------------------------------------------------------------------------------------------------GGTGATCATGAGAAAAACAATATTT------------------  2232


A.thaliana     ---TTACAGAGATGTGAAA-------ACACTCATGAGTCATGTTTTGTAACTACTTACTATAATGGTTA-CCAAGTGCAAATTTCTACATA----------CTATATATGATAAATCTAATTATTGCTCATGTGGACTCCAAAATGCCT  2754
A.lyrata       ---TTACAGAGATGTGACA-------ACACTCATGAGTCATGTTTTCTTACTACTTACTATGATGGGGA-TCAAGTGCAAATTTCTACATA----------CCATCTACGG-ACACCT-----TTGCTCATGTGGACTCCAACATGCC-  2744
C.rubella      ---TTACAGAGATGTGACA-------ACACTCATGAGTCATGCTATGTTACTAC-------GATGGTGA-TCAAGTGCAACTTTCTTCTACATACCATATACAGACATGTGATATAATTAACTTTTGCTCATTTGAACTCCAAAATGCCT  2926
N.officinale   ---------------------------------------------------------------------------------------AGAACTGTGAT--------------------------------------------  2194
A.alpina       ---TAAACGAGATGTGACA-------ATACTCATCACTGGAGCCATGTT--TTGTTACTATGATGGTGG-TCAAGTGCAAATGTCTACTTAACATACAC---AGACATGTAGTATAACTAACC-TTGCTCCTTTGGTCTTCAAAATGC--  2735
B.napus        CTGTTACAGAGACATTTGACA-----ACACCCATGAGTCATTCTTTA-------TTACCATGATGGTGA-TCAAGTGCAAATTTCTACATAAAATGTAGACATGTGATAC----AGACTAACCTAGCTCCTTTTGGACTACTAATTTGC-  3074
B.oleracea     -TGTTACAGAGACATTTGAC------ACACCCATGAGTCATTCTTTAT------TTACCATGATGGTGATTTAAGTTCAAATTTCTACATTAAATATAAACATGTGATATACTAACACTAACCTTGCTCCTTTTGGACTCCTAATTTTC-  2952
S.irio         --GTTACAGAGACATGTGACAATATAACACTCATGAACCATGTTTTA-------TTACTCTTATGTGAT-TCAAGTGCAATATTCTGCAT-AAATATACGGAGATATATGATAATAACTAACCTTTCTCCTTTTGGACTCCAAATTTCT-  2848
S.officinale   ----TACAGAGACTTGAGAC-----AACAGTCATGGGCCATGTTTTA-------TTACTCTAATGGTGA-TCAAGTGCAAATTTCTGCATAAAATTTACAGAGATG--TGAT-ATAACTAACCTTCCTCCTTTTGACTCCAAATTTCC-  2787
A.arabicum     ------------------------------CTACGAAACATAT-----------------------------ACGGCTACTTGATATA-------------------------ACTAACCTTGTTCTCCTTTGGACTCCAAGATGCTG  2296


A.thaliana     TTTAAGTTTTAACTTGTGCGTCAGGTAAATCCTAATTTGTAGTCTCAAGACTACTTGGCGGATTCGAGTTTGATCCTAGAAAATCCACCGTCTCTATGTTTTTCATGTCAC-------------------------------------  2865
A.lyrata       ------TTTCAACTTGTGCATCAGGTAAATCCCAATCTTTAGACTCAAGACTACTTGGCGGATTCGAGTTGGATCCTAGAACGCCCACAGTCTCTATGTTTTTTATGTCTTA------------------------------------  2850
C.rubella      -------TTCAACTTGCGGGTCAGGTAAATCCTAATCTTTAGACTCAAGACTACTTGGCGGATTCCAGT--AGTCCTAGAAAATCTACCATTTCTACTTTTTCGATGTCAG-------------------------------------  3028
N.officinale   -------------------------------------------ACTCAAGA-------------------------------------CAAGTCTAAGTTTTCCATGTCAG-------------------------------------  2225
A.alpina       ------TTTCAACTTGTAACTCAGGTAAA-------------------------------------------------ATACACCATGTATACGTTATCCGTGTCAG-------------------------------------  2787
B.napus        ------TTACCACTTGTGGTTCGGGTAACAGTTGGAGTTGG------------------------------------------------------------------------------------------------------  3109
B.oleracea     ------CTGCAACTTGTGGCCCAGGTAACAGTTGGTGATGGCTATGA----------------------------------GTAGGGATGTTACCCATGGGTTTACCCAAACCCACTAATAATGGGTTGGGTTTTTACCCAT  3054
S.irio         ------TTACAACTTGTTGCTCAAGTAACAGTTAG-------------------------------------------------------------------------------------------------------------  2877
S.officinale   ------TTACAACTTGTGGCTT-AGTAACAGTTAG-------------------------------------------------------------------------------------------------------------  2815
A.arabicum     T-------TCAGCTTGTGGCTCAGGTGCATCCTAATCTCTAGGCTCAAGACTAGAGGGCGGATCCAAATTGTGTCCTAGAGAATCCATCATGTCTACGTTTGCCATGTCAGCTTCCT-----------------------------------  2406


A.thaliana     ------------------------------------------------------------------------------------------------------------------------------------------------  2865
A.lyrata       ------------------------------------------------------------------------------------------------------------------------------------------------  2850
C.rubella      ------------------------------------------------------------------------------------------------------------------------------------------------  3028
N.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------  2225
A.alpina       ------------------------------------------------------------------------------------------------------------------------------------------------  2787
B.napus        ------------------------------------------------------------------------------------------------------------------------------------------------  3109
B.oleracea     TGGGTCAACCATGGGTATTAATTTTAAAACTCATATTTATGTTGGGTAAATACAAAAGGGTAATGGTGTACCCATGGGTTTTCAATTATATATATAGATTTTCAACATTTTAATTAAATTATATAAAATTTGCAAAAACATTTTTTTTT  3204
S.irio         ------------------------------------------------------------------------------------------------------------------------------------------------  2877
S.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------  2815
A.arabicum     ------------------------------------------------------------------------------------------------------------------------------------------------  2406
```

```
A.thaliana     ------------------------------------------------------------------------------------------------------------------------------------------------------ 2865
A.lyrata       ------------------------------------------------------------------------------------------------------------------------------------------------------ 2850
C.rubella      ------------------------------------------------------------------------------------------------------------------------------------------------------ 3028
N.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------ 2225
A.alpina       ------------------------------------------------------------------------------------------------------------------------------------------------------ 2787
B.napus        ------------------------------------------------------------------------------------------------------------------------------------------------------ 3109
B.oleracea     GTCAGAACCAAGAAAATAAAAATTTCCGCCAAAACCGTAAAATAAATTTTCTCACAGAAACCAAAAACGAGTTTTTCCACCAAAACCGAAAAATCGAGTTTTTCTGCCGAAACCACAAAACCGAGTTTTTCCGCCAAAACCGCAAAATAG 3354
S.irio         ------------------------------------------------------------------------------------------------------------------------------------------------------ 2877
S.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------ 2815
A.arabicum     ------------------------------------------------------------------------------------------------------------------------------------------------------ 2406

A.thaliana     ------------------------------------------------------------------------------------------------------------------------------------------------------ 2865
A.lyrata       ------------------------------------------------------------------------------------------------------------------------------------------------------ 2850
C.rubella      ------------------------------------------------------------------------------------------------------------------------------------------------------ 3028
N.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------ 2225
A.alpina       ------------------------------------------------------------------------------------------------------------------------------------------------------ 2787
B.napus        ------------------------------------------------------------------------------------------------------------------------------------------------------ 3109
B.oleracea     AATTTTTCCGCAAATACCGTATAATTGGGTTTTTTCGCTAAAATCGCAAAATCGAGTTTTTCGACTGAAAACCGTAAAATCGAGTATTCCCGCCAAAACCGCAAAATGTGTTTTTCCGCCAATACCGTAAAACCCAATTTTCCCGCCAAA 3504
S.irio         ------------------------------------------------------------------------------------------------------------------------------------------------------ 2877
S.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------ 2815
A.arabicum     ------------------------------------------------------------------------------------------------------------------------------------------------------ 2406

A.thaliana     ------------------------------------------------------------------------------------------------------------------------------------------------------ 2865
A.lyrata       ------------------------------------------------------------------------------------------------------------------------------------------------------ 2850
C.rubella      ------------------------------------------------------------------------------------------------------------------------------------------------------ 3028
N.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------ 2225
A.alpina       ------------------------------------------------------------------------------------------------------------------------------------------------------ 2787
B.napus        ------------------------------------------------------------------------------------------------------------------------------------------------------ 3109
B.oleracea     ACGATTTTTCCCGCCGAAACCGAAAAACCGAGTTTTCCCGCCAAAACCGAAAAATCGAGTTTTCCCGCCAAAACCGAAAAACCGAGTTTTCCCGTCAAAATAAAAACGAGTTTTCCCGCCAAAACCGAAACACCGAGTTTTTCCGCCAAA 3654
S.irio         ------------------------------------------------------------------------------------------------------------------------------------------------------ 2877
S.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------ 2815
A.arabicum     ------------------------------------------------------------------------------------------------------------------------------------------------------ 2406

A.thaliana     ------------------------------------------------------------------------------------------------------------------------------------------------------ 2865
A.lyrata       ------------------------------------------------------------------------------------------------------------------------------------------------------ 2850
C.rubella      ------------------------------------------------------------------------------------------------------------------------------------------------------ 3028
N.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------ 2225
A.alpina       ------------------------------------------------------------------------------------------------------------------------------------------------------ 2787
B.napus        ------------------------------------------------------------------------------------------------------------------------------------------------------ 3109
B.oleracea     ACCGCAAAACCGAGTTTTCCCGTCAAAATCAAAAACGAGTTTTCCCGCCAAAACCACGAAACCGAGTTTTCCTGCCAAAACCGCTAAACCGAATTTTCTCGCAAAAACCGTAAAACCGAGTTTTTCCGCCAAAAAAATCGCAAAAGATT 3804
S.irio         ------------------------------------------------------------------------------------------------------------------------------------------------------ 2877
S.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------ 2815
A.arabicum     ------------------------------------------------------------------------------------------------------------------------------------------------------ 2406

A.thaliana     ------------------------------------------------------------------------------------------------------------------------------------------------------ 2865
A.lyrata       ------------------------------------------------------------------------------------------------------------------------------------------------------ 2850
C.rubella      ------------------------------------------------------------------------------------------------------------------------------------------------------ 3028
N.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------ 2225
A.alpina       ------------------------------------------------------------------------------------------------------------------------------------------------------ 2787
B.napus        ------------------------------------------------------------------------------------------------------------------------------------------------------ 3109
B.oleracea     GTTCCCGCCAAAATCGAAAACAGGTTTCCCGCCAAAATCTAACAATTTTTTCCGTCAAAACCGTAAATACAAGTTTTTCCGCCAAAACCGCAAAAACGAATTTTCACGCCAAAACCAGAAAGATTATTTTTTCCGCAAACCGCAAAAATC 3954
S.irio         ------------------------------------------------------------------------------------------------------------------------------------------------------ 2877
S.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------ 2815
A.arabicum     ------------------------------------------------------------------------------------------------------------------------------------------------------ 2406
```

```
A.thaliana     ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  2865
A.lyrata       ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  2850
C.rubella      ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  3028
N.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  2225
A.alpina       ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  2787
B.napus        ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  3109
B.oleracea     GAGTTTTCCAACCAAAACCACAAAAAAAAAATTTACCGACCAAAAACCGCAAAAAGAAAATTTTCCGCTAAAACCACAAAAACAACTTAACAAGTTTTCTCGTTAAAACCATAAATGAGTTTTCCTGTCAAAACGATGTTTACTGCTAAA  4104
S.irio         ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  2877
S.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  2815
A.arabicum     ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  2406

A.thaliana     -------------TTTTCCGATATGATTCTCATTACCATGACTTTATGAACCAGATTAAACATTATAACACTTTTCATCAGAAAATCCTTCGAAAGTTTCAATTGCAAATCTTTCTAAATGATGCAGATGCATTCACAAATAATGGAA--  3000
A.lyrata       -------------TTTTCCGATATGCTTCTCATTGCCATTACTTTATGAACCAGACTAAACATTAAAACACTTTTGATCAGAAAATCATTAGGAAGTTCCAATTGCAAATCTTTCTAAATGATGCAGATGCAATCACAAATAATTG----  2983
C.rubella      -------------CGTTCCGATAGCTTTGACATTGCCATTAATTTATTAACCAGA-TACATACTAAAACACTTTTCATCAGAAAATCATTTGGAAG-TTCTATTGCAAA-CGTTCTAAGTGATGCAGAAGCAATCATAAA----------  3152
N.officinale   -------------CTTCCC-------TTCTCATTGCCATTACTTTATCAACCAGACTAAACACTAGAACAATTCTCAACAGGAAATCATTTGGAAGTTTCCATTGCAAATCTTTCTCAATGATGCAGATGCATTCATAAATAATGGAA--  2353
A.alpina       -------------CATCCCGATATGCTTCTCATTACCATTACTTTATTAGCCACACTAAATACTAATGCACATTTCATCAGAAAATCACTTCAAAGTTTCAATTGCAAATCATTCCAACTGATGCAGATGCAATCATCGATAACTGAA--  2922
B.napus        ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  3109
B.oleracea     ACCGCAAAATGAGTTTTCCGGTTAAAATTGCAACATAATTTTTTTTCCGAAAAAAACGAGTTTTTCCGCCAAAATCGTAAACGAGTTTTTGCG------CTAAAACAAGTTTTCTATAAAATTGCAAAATCTTGTTTATATATTAAAGCT  4248
S.irio         ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  2877
S.officinale   ------------------------------------------------------------------------------------------------------------------------------------------------------------------------  2815
A.arabicum     ------------------------------------------------------------------AATACGAAAACACTTTCCAACAAACAAACCAT-----------------------TTCAAAATGTCAATTACCTAATGCCCTGTT--------  2468

A.thaliana     ----------------------------------------------------------------------------------------------------------CAACAACTAT---------ACCATATTCACGAGTTTGTCTAACCTTTGTATA-GGTA  3047
A.lyrata       -----------------------------------------------------------------------------------------------------------AACAACTTT---------ACCATATTCACGAGTTTGTCTAACCTTC-TATA-GGTA  3028
C.rubella      --------------------------------------------------------------------------------------------------------------AATTCTTGTACTTACCCATTTTCACGAGTTCGTCTAGCCTTT-GATA-GGTA  3202
N.officinale   ----------------------------------------------------------------------------------------------------------CAACAACTCT--------ACCATTTTGACGAGTTTGTCTAACCTTT-GATAGGGTA  2400
A.alpina       ----------------------------------------------------------------------------------------------------------TAACAACTTT--------ACCAT-TTCACGAGTACGTCTAAACTTC--GTTGGGTA  2967
B.napus        ----------------------------------------------------------------------------------------------------------TGATGGCTCT--------GGAACTC------------------------------  3126
B.oleracea     CACATTAATGATATTAATATTTTACATGATCTTCAGAGTAAATAAGATAATATTTTGGGATATTGGGTTATTTTTTTGGTTTGGATTTCAACCTTGATGTTTCT--------GGGTTTTTGCTGGTTGGGTATAAACA---------GGG  4381
S.irio         ----------------------------------------------------------------------------------------------------------TGATGGCTCT--------GGGACTCGAAGGC-----------------------  2900
S.officinale   ----------------------------------------------------------------------------------------------------------TGATGGCTCTAA------GGGACTCGAAGGC-----------------------  2840
A.arabicum     ------------------------------------------------------------------------------------------------------------------------------TCATGAGTTTCTCTATCAAT-----------  2488

                                                                                                    *** * * **         *    *****       *  **            *  *  *  *   *  *  *** *
A.thaliana     GTCAACCCATAACAGTTGGTGATGGCT-----------CTGACACT--------------------------CGAAGCCTTAC-----TCGGAGAGATACCTGAACAGTAATCACA--------AGG-TTCAGGAT-GAATATTCAACCA  3145
A.lyrata       GTACACCCGTAACAGTTGGTGATGGCT-----------CTGAGACT--------------------------CGAAGCCTAAC-----TCAGAGATGTACCTGGACAGTAATCACA--------AGG-TTCAAAAT-GAATATTCAACTA  3126
C.rubella      GTGAACCCGTAACAGTTGGTGCTGGCTCTGATGAGAGACTGATACT--------------------------CGAAGCCTTAC-----TCGGAGATATACCTGAACAATTATCACA--------AGG-TTAAGGAT-GACTATTCAGCTA  3311
N.officinale   GTAAACCCGTAACAGTTGGTGAAGGCT-----------CTCAAACT--------------------------CGAAGCCTTAC-----TCGGAGATATACCTGGACAGTTATCACA--------AGGTTTCAAGAT-GAATATTCAACTA  2499
A.alpina       GAAACCCCATAACAGTTGGTGATGGTT-----------CTCTGACT--------------------------TGAAGCCTTAC-----TCGGAGATATACCTGGACAGTTAGCACA--------AGG-TTCAAGAA-GAATATTCAACTA  3065
B.napus        ------------------------------------------------------------------------TGAAGCCTTAC-----TCGGAGATATACCTGGACAGTTATCACA--------AAG-TTCAAGAAGGAATATTCAACTA  3190
B.oleracea     TTGGGTTATTTGTGGGTTGGGCTGGGCTGGGTTA-----TTTTACCACGTTAACATCCCTAGCTATGAGACATGAAACCTTAC-----TCGGAGATATACCTGGACAGTTATCACA--------CGG-TTCAAGAA-GAATACTCAACTA  4511
S.irio         ------------------------------------------------------------------------CGAAGCGTTACCCTGACCTGAGATATACCTGGCCAGTTATTAC---------GGG-TTGAAGAA-AAACATTCACCTA  2967
S.officinale   ------------------------------------------------------------------------CGAAGCGTTAC-----CCTGAGATATACCTGGCCAGTTATCACA--------AGG-TTGAAGAA-AAATATTCACCTA  2903
A.arabicum     GTAAATCCAGTGC-TTTGGTGATGACCCTGGTAGC-------------------------------------TGAAACCTTACAC------GGGTAAACCTGAAGTAAAATTACAACCACACAATG--TTAAAAACTAATATTCAACAG  2591

                *    *   *                                                   ** **                ** ******** **        ***                                  *                         *
A.thaliana     CTT--AAACTTTGT-----ATAAAG---CCAAAGAGATA-AAACG-AAT---------CTAGCTTTACTTTAAATAAAATGCATATGAAAATAGTAAAAGGTGATACGA-------------------AAAAATAGTAACAATTTGCCTGC  3256
A.lyrata       CTT--AAACTTCGT-----ATGAAG---CCAAAAAAAGGTAAACG-AAT---------CTAGCTTTACTTT---TAAAATTCATATGAAAATAGTAAAAGATGACACG-------------------TAAAATAGTAACAATGTGTCTGC  3234
C.rubella      CGT--AAACTTTGT-----GTGAAG---CCAAAGAGATGAAAGAA-TAT---------CTAGCTTTACTTTTA-ATAAATCCATATGAAAATAGTAAAAGGTGATACG-------------------AAAAATAGTAACGA--TGTCTGC  3419
N.officinale   TAT--AAACTTTGT-----GAGAAG---CCAAAGAGATAAAAGA---AT---------CTAGCTTTACGTTAA-ACAACTCCATATGAAAATAGTAAAAGGTTATACG-------------------AAAAATAGTAAATAAGTGTATGC  2607
A.alpina       CAT--GAACATTTA-----CTA-----------------------G-TAT---------CTAGCTTAAATTTAA-ATAAACCCATATGAAAATAGTAAAAGGCGATACG-------------------AAAAATAGTAACAATGTGTCTGC  3156
B.napus        TGT--AAACTTTTC-----CTGGAGCCACCACAGAGATAAAG----TAT---------CTAGCTTTCCTATAA-ATAAATCCAAGTGAAAATAGTAAAAGGTGGTACG-------------------AAAA---------------TAGT  3285
B.oleracea     TGT--AAACTTTTC-----CTGGAG---CCACATAGATAGAGTACCTAC---------CTAGCTTTACTATAA-ATAAATCCAAGTGAAAATAGTAAAAGGTGGCACG-------------------AATA---------------TAGT  4607
S.irio         CGTAAAAACCTTCCTCGAACCGAAG---CCAAAGAGATAAAC----TAT---------GTAGCTTTACTTTAA-ATAATTACAAATGAAAATAGCAAGTGGTGATACGT------------------AAAA---------------TGGT  3067
S.officinale   CGTAAAAAAATTTT-----CCAAAG---CCAGAGAGATAAAG----TAT---------CTAGCTTTACTTTAA-ATAATTCCAAATGAAAATAGTAAAAAGTGCTACGA-------------------AAAA---------------TAGT  2998
A.arabicum     CATA---ACCTATT-----GTGGAGC---CATAAAGAT----------TCAAAAGTATCTAACTCTACTTCA------CTCCATCCGAAAATAGTAAAATGTAAAACGAAAATAGTAATTTTTATACGAAGAGACTAAATATGTGTCTGC  2714
```

```
                *                              *  **   ******    *  ** *                  *** **   ***   *        ***
A.thaliana      AACACCATGGCATTATCCGGACCACTTCCTCTTGAGAATCTCAGTATGGCAAGTGGCAAAACC-------TAAGCAACTTGTG--AACGGGTCCCAACGAAGAAGTGCATAGGA--------------GGAGATGTTTACACTTTACACTT   3384
A.lyrata        ACAGAATTTTCAGTTTCCGGACCACTTCCTCTTGAGAATCTATGTATGGCAAGTGGCAAAACC-------TAAGCAACTTGTG--AACAGGTACCAATGAAGAAGTGCATAAGA--------------GGAGATG---------------TT   3348
C.rubella       AACACCATGGCTTCATCCAGACCACTTTCTCTTGAGAATCTCTGTATGGCAAGTGGCAAAACT-------TAAGCAACTTGTG--AACGGGTCCCAGTGAAGAAGTGCATAGGG--------------GGAGATGTTTACACTTT-CACTT   3546
N.officinale    AACACCATGGCTTTATCCGGACCACTTACCCTAGAGAATCTCTGTATAACATGTGGCAAAACT-------AAAGAAACGTGTGGTAAAGGGTCCCAATGAAGAAGTGTATAGGA--------------GGAGATGTCTA--------CACTT   2730
A.alpina        AACACCATGGCGTTGTCCAGACCACTTCCTCTTGAGAATCTCTGTATGGCAAGTGGCAAAACC-------AAAGCAACTTGTG--AACGGGGCCCATCCAAGAAATACATTGGA--------------GGAGATGTTTACACT--------T   3277
B.napus         AACACCATA-------TGGTGTTTTATCCTCTTGAGAATCTCTCAGTGGCAAG-----------------TAAGCAACGTGTG--AATAGTTCCCAA-----------------------------------------------------   3356
B.oleracea      AACACCACA-------TGGTGTTTTATCCTCTTGAGAATCTCTCAGTGGCAAGTAGCTGAAAATCTATAATAAGCAATGTGTG--AACCGTTCCCAA-----------------------------------------------------   4695
S.irio          AAC-----------ATCGGGACCACTTCCTCTTGGGAATCTCCGTATGGCAAGTGGCAAATCC-------TAAGCAACTAGTG--AACGGGTCCCAACGAAGGAGT----------------------AGAGATGTT---------------   3162
S.officinale    AACACCATGGTGTTATCGGGACCACTCTCTCTTGGGAATCTCCGTATGGCAAGTGACAAATCC-------TAAGCAACTAGTG--AACGGGTCCCAATGAAGGAGTACACATATACTACATAGGAGGAGAGATGTT---------------   3125
A.arabicum      AACACAATGACGTTAAGGGGAACACTTCCTCTTCAGAATCTCCTTGTGCCACGTGGCAAAACCT-------AAGCAACCTGTG--GATAAGTCCCACCGAA-AAATACATAGGAG--------------GAGATTTTA--------------   2828


A.thaliana      TACACTTTACACATAGGCCT------------------------------------------------------------------------------------------------------------------------------------   3404
A.lyrata        TACACTTTACACATAAGCCT------------------------------------------------------------------------------------------------------------------------------------   3368
C.rubella       TATACTTTACACATAAGCCA------------------------------------------------------------------------------------------------------------------------------------   3566
N.officinale    TATACTTTACACACAAGCCT------------------------------------------------------------------------------------------------------------------------------------   2750
A.alpina        TACACTTTACACATAAGCAT------------------------------------------------------------------------------------------------------------------------------------   3297
B.napus         --------------------------------------------------------------------------------------------------------------------------------------------------------   3356
B.oleracea      --------------------------------------------------------------------------------------------------------------------------------------------------------   4695
S.irio          TACACTTTACACATAAGTCT------------------------------------------------------------------------------------------------------------------------------------   3182
S.officinale    TTCACTTTACACATAAGCCT------------------------------------------------------------------------------------------------------------------------------------   3145
A.arabicum      --CACTTTGCACTTTGCACTTTACACTTTACACAGAGAGAGGCGTCCGTAAGTCTTGGTATGGCTTTTTTTCTCTATTGCCAAGATATATGAAATTGAAGTTTTTCCTAATAATTAATAATTTCCAATAGCCACTGTGGGCAACTGTGGT   2976


                                                                                  *  ** ***   *   * *  ****                ****  **   * *                  *
A.thaliana      ------------------TCCCAAAAGCTCAACTAGCTGCAAGAGGATCCAATAACATGTAAGAGCCACTAACGCTGTGCCACGTGTAGGCACTCAGGATTCGA---TCTCCCCTCTACTTATTCTCTCACACCAGATATAAGCTTTATT-   3533
A.lyrata        ------------------TCCCAAAAGCTCAACTAGCTGCAAGATGATCCAATAACATTTAAGAGCCACTAACTTTGTGCCACGTGTAGGCACTCAGGATTCGA---TCTACCCTCTGCTTATTCTCTCACACCAGATATAAGCTTTATT-   3497
C.rubella       ------------------TCCCAAAAGCTCAGCTAGCCGCAAGATGATCCAATCACATTTAAGAGCGACTAATTCTGTGCCACGTGTAGGGACTCAGGATTCCA---TCTCCCCTCTGCTTATTCTCTCAGAACAGATATAAGCTTTATT-   3695
N.officinale    ------------------TCCCAAAAACTCAGCTAGCCGCAAGATGATCCAATCACATTTAAGAGC--CTATCTTTGTGCCACGTGCAGGCACTCAGGATTCGG---TCTCCCCTCTGCTTATTCTCTCACAGCAGATAT-AGCTTTATT-   2876
A.alpina        ------------------TCCCAAAAGCTCACCCAGTCGAATAATGATCCAATAACATTTAAGAGCCAATAGATCTGTGCCACGTGTAAGCGCTCAATATTCGA---TCTCTCCTCTGCTTATTCTCTCACAACAGATA-----------T-   3416
B.napus         --------------------------------------------------------------------------CTAATAGCTCTGTGCCACGTGTAAGCACTGAGGATTCGG--------------------ATTCTCTCACAACAAGATAAAGCTTTATT-   3424
B.oleracea      --------------------------------------------------------------------------ATTATAGGTCTGTGCCACGTGTAAGCACTGAGGATTCGG--------------------ATTCCCTCACAACAAGATAAAGCTTTATT-   4763
S.irio          ------------------TCCTAAAAGCTGAGTTAGCCGCAAGATGGT------------------CCAATAACTCTGTGCCACGTGTAAGCGTTGATGATTCGAAATCTCTTCCTCTGCTTATTCTCTCACAAC-AGATATA-ATTTATT-   3295
S.officinale    ------------------TCCTAAAAGCTGA-------GCAAGATGAT------------------CCAATAACTCTGTTCCACGTGTAAGCGTTGATGATTCGAAATCTTTTCCTCCGCTTATTCTCTCACAAC-AGATATAGATTTATT-   3252
A.arabicum      CCCTATACTTGAGCTTTTCTGAAAAACTCAGCCAAGCCACAAGGTCACCCATTACCTTTTATA----CCCCATCATCTCAAGTGTCAGCTCTCAGGATTCTAT---CTTTCCTCTGTTTATTCTCTTCCAGCAGATATAGCCTTAATTA   3118


                  *******          ** *  ***  **     *** *  * *        ***          *       *  ***
A.thaliana      --AGCCCCTT-CTT--TCAGATACAAGCTCCC--ACACCATCAAACT----TACTACAT--CTGAGT-----TATTATG   3594
A.lyrata        --AGCCCCTT-CTT--TCAGATACAATCTCCC--ACACCATCAAACT----TACTACAC--CTGAGT-----TATTATG   3558
C.rubella       --AGCCCCTT-CTT--TCAGATACAATTTCCC--ACACCACCAAACT----TACTATAT--CTGAGT-----TAGTATG   3756
N.officinale    --AGCCCCTT-CTT--CCAAACACAATCTCCC--ACACAATCAAACT----CACTACAT--CTAAGT-----TATTATG   2937
A.alpina        --AGCCCCTT-CTT--TCAGATACAATCTCCC--ACACCATCAAACT----TACTACAT--TTGAGT-----AATTATG   3477
B.napus         --AGCCCCTGCTCT--TCAGACACAATTTCCCACACACCATCACACTTAATTACTACAT--CAGTGAGC---TAACATG   3494
B.oleracea      --AGCCCCTTCTC---TCAGACGCAA--TCTCACACACCACCAAAC--AA--ACTACAT--CAGTGAGCA--AAATATG   4827
S.irio          --AGCCCCTCACTT--TCAGAGCCAATCTCTC-CACACCATCAAACT-AGTTACTACAT--CAGAGTGTAAGCATTATG   3366
S.officinale    --AGCCCCTC-CTT--TCAGAGCCAATCTCTC-CACACCATCAAACT-AGTTACTACAT--CAGAGTGTTAGTATTATG   3322
A.arabicum      TTAGCCCCT---TCCATCAGATACAATTTCCTA--CACCATTAGA--------CTACTTCACTGAGTT-----ATAATG   3179
```

ABBREVIATIONS

General

°C	degree Celsius
µ	micro
3', 5'	3-prime, 5-prime
3-AT	3-Amino-1,2,4-triazole
3xHA	three times Hemagglutinin tag
A	adenine
amp	ampicillin
BAC	bacterial artificial chromosome
bHLH	basic helix-loop-helix
BLAST	basic local alignment search tool
bp	basepair
C	cytosine
$CaCl_2$	calcium chloride
CaMV	*Cauliflower mosaic virus*
CCT	CONSTANS, CONSTANS-like, TOC1
CDM	conserved distal motif
cDNA	complementary DNA
CDS	coding sequence
ChIP	Chromatin Immunoprecipitation
Chr	Chromosome
CMM	conserved middle motif
Col-0	Columbia
CPM	conserved proximal motif
CRM	conserved regulatory motif
DD	darkness
ddH_2O	ion exchange purified and sterilized water
DMSO	Dimethylsulfoxid
DNA	desoxyribonucleic acid
dNTP	deoxyribonucleic triphosphate
DOF	DNA-binding with one finger
EDTA	ethylendiaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
EtBR	ethidium bromide
EtOH	ethanol
g	gram
G	guanine
gent	gentamicin
GW	Gateway
h	hour
HA	Anti-Influenza, Hemagglutinin
HCl	hydrogen chloride
His	histidine
Kan	kanamycin
kb	kilobase
kDa	kilodalton
laemmli	laemmli buffer
LB	luria broth
LBS	LUX binding site
LD	long day
LED	light-emitting diode
L*er*	*Landsberg erecta*
Leu	leucine

LiAC	lithium acetate
LOV	LIGHT, OXYGEN, VOLTAGE
m	milli
M	molar (mol/l)
mg	milligram
min	minute
$MnCl_2$	manganese(II) chloride
mol	mole
mRNA	messenger RNA
MS	Murashige and Skoog
MTBS	tris buffered saline supplemented with milk powder
N	any base
NaF	sodium fluoride
NaOH	sodium hydroxide
O/N	over night
OD_{600}	optical density at 600nm
ORF	open reading frame
PBM	protein binding microarrays
pCO	promoter of *CONSTANS*
PCR	polymerase chain reaction
pH	negative logarithm of the proton concentration
PIC	protease inhibitor cocktail
pnos	minimal promoter of *NOPALINE SYNTHASE*
POL II	RNA polymerase II
Rif	Rifampicin
RNA	ribonucleic acid
RT	room temperature
SAM	shoot apical meristem
SD	short day
SDS	sodium dodecylsulfate
SNP	single nucleotide polymorphism
T	thymine
TAP	tandem affinity purification
taq	*Taq* polymerase
TEMED	N,N,N',N' tetramethylenethylendiamine
TF	transcription factor
Tris	Trishydroxymethylaminomethane
TS	transcriptional start
UTR	untranslated region
wt	wild-type
Y1H	Yeast One Hybrid
ZT	Zeitgeber time

Amino acids

Alanine	A
Cysteine	C
Aspartic acid	D
Glutamic acid	E
Phenylalanine	F
Glycine	G
Histidine	H
Isoleucine	I
Lysine	K
Leucine	L
Methionine	M

Asparagine	N
Proline	P
Glutamine	Q
Arginine	R
Serine	S
Threonine	T
Valine	V
Tryptophane	W

Genes

The nomenclature for plant genes follows the *Arabidopsis* standard:
GENES: upper case italics; *mutant genes*: lower case italics; PROTEINS: upper case regular letters

35S	35S RNA CaMV
AG	AGAMOUS
AGL24	AGAMOUS-LIKE 24
AP1	APETALA 1
CAL	CAULIFLOWER
CCA1	CIRCADIAN CLOCK REGULATED 1
CDF (1-5)	CYCLING DOF FACTOR (1-5)
CO	CONSTANS
COG1	COG1
COL (1-16)	CONSTANS-Like (1-16)
COP1	CONSTITUTIVE PHOTOMORPHOGENESIS 1
Cry1	Cryptochrome 1
Cry2	Cryptochrome 2
DNF	DAY NEUTRAL FLOWERING
ELF3	EARLY FLOWERING 3
FBH (1-4)	FLOWERING bHLH (1-4)
FD	FD
FKF1	FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 (ADO3)
FT	FLOWERING LOCUS T
FUL	FRUITFULL
GI	GIGANTEA
GUS	β-GLUCORONIDASE
H3	HISTONE3
HD1	HEADING DATE 1
HY5	ELONGATED HYPOCOTYL 5
LFY	LEAFY
LHY	LATE ELONGATED HYPOCOTYL
LKP2	LOV KELCH REPEAT PROTEIN 2 (ADO2)
LUC	LUCIFERASE
LUX	LUX ARRHYTHMO
NOS	NOPALINE SYNTHASE
PhyA	Phytochrome A
PhyB	Phytochrome B
PP2A	PROTEIN PHOSPHATASE 2A
PRR (3,5,9)	PSEUDO RESPONSE REGULATOR (3,5,9)
RFI2	RED AND FAR-RED INSENSITIVE 2
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1
SPA (1-4)	SUPPRESSOR OF PHYA-105- (1-4)
SPL (3-5)	SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (3-5)
SUC2	SUCROSE TRANSPORTER 2
TCP4	TEOSINTE BRANCHED/CYCLOIDEA/PCF 4
TFL2	TERMINAL FLOWER2
TOC1	TIMING OF CAB2 EXPRESSION 1

TSF	TWIN SISTER OF FT
ZTL	ZEITLUPE (ADO1) (LKP1)

Species

A. alpina	*Arabis alpina*
A. arabicum	*Aethionema arabicum*
A. lyrata	*Arabidopsis lyrata*
A. thaliana	*Arabidopsis thaliana*
A. tumefaciens	*Agrobacterium tumefaciens*
Aa	*Arabis alpina*
Ae	*Aethionema arabicum*
At	*Arabidopsis thaliana*
B. napus	*Brassica napus*
B. nigra	*Brassica nigra*
B. oleracea	*Brassica oleracea*
Bs	*Boechera stricta*
C. rubella	*Capsella rubella*
Cr	*Chlamydomonas reinhardtii*
E. coli	*Escherichia Coli*
Hv	*Hordeum vulgare*
Lj	*Lotus japonicas*
Mt	*Medicago truncatula*
N. officinale	*Nasturtium officinale*
Os	*Oryza sativa*
Pn	*Pharbitis nil*
Ps	*Pisum sativum*
Pt	*Populus trichocarpa*
Rl	*Reseda luteola*
S. irio	*Sisymbrium irio*
S. officinale	*Sisymbrium officinale*
St	*Solanum tuberosum*
T. hassleriana	*Tarenaya hassleriana* (*Cleome hassleriana*)
T. parvula	*Thellungiella parvula (Eutrema parvulum)*
Th	*Tarenaya hassleriana*
Vv	*Vitis vinifera*

ACKNOWLEDGEMENTS

I would like to thank…

Prof. George Coupland for making this PhD work possible. Thanks for finding the balance between good guidance and the freedom to develop my own ideas.

Prof. Ute Höcker for being my second supervisor as well as the support during my PhD work and in my graduate studies.

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

Dr. Franziska Turck for fruitful discussions over the years as well as being the fourth member of the committee.

all the member s of group Coupland and group Turck. You were a great help! Every Day!

especially Amaury de Montaigu for incredible intellectual support in the project.

Markus Berns, Fabio Fornara and Mark Rühl for help getting started with the project.

Stefan Wötzel und Christiane Kiefer für volle phylogenetische Unterstützung.

Tobias Gabriel and Rafael Martinez Gallegos for technical support.

Eva-Maria Willing, Vimal Ravat, Karl Nordström and Geo Velikkakam James for bioinformatic support.

the gardeners, specially Monika Bürgel and Heike Eichler for taken care of all these plants.

Isabell Witt and the International Graduate School for Health and Disease as well as the Max Planck Society for funding.

Michaela, Brigitte und Petra für eine tolle Zeit.

Maida, Liron, Loren, Juli, Liangyu and Theo for scientific but specially also for non-scientific support.

Ich möchte besonders meinen Eltern danken für all die Unterstützung in den letzten Jahren. Ein besonderer Dank gilt auch all den lieben Menschen die mich über die letzten knapp vier Jahre unterstützt haben.

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.

Teilpublikationen liegen nicht vor.

Köln, 18. November 2013 ___________________________

Samson Simon

LEBENSLAUF

Persönliche Angaben

Name	Samson Simon
Geburtsdatum	14.04.1983
Geburtsort	Köln
Nationalität	Deutsch
Familienstand	ledig

Ausbildung

2002	Abitur, Gymnasium der Stadt Kerpen, Deutschland
2002 – 2003	Zivildienst, St. Franziskus Hospital, Köln, Deutschland
2003 – 2009	Studium der Biologie, Universität zu Köln, Deutschland
	Hauptfach: Genetik
	Nebenfächer: Botanik, Biochemie
11 2009	Diplom in Biologie, Gesamtnote: sehr gut
	Titel der Diplomarbeit: Transcriptional regulation of the flowering-time gene *CONSTANS* in *Arabidopsis thaliana*
04 2010 -	Beginn der Doktorarbeit, Max-Planck-Institut für Pflanzenzüchtungsforschung, Köln
	Arbeitsgruppe von Prof. Dr. George Coupland