Functional and Evolutionary Analysis of Flowering Time Genes in *Brassica nigra* and *Physcomitrella patens*

Harald Hedman
Faculty of Natural Resources and Agricultural Sciences
Department of Plant Biology and Forest Genetics
Uppsala

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2008

Acta Universitatis Agriculturae Sueciae
2008:55

Front cover: *Physcomitrella patens,* leaf and filament (photo: H. Hedman)

ISSN 1652-6880
ISBN 978- 91-85913-88-6

Tryck: SLU Service/Repro, Uppsala 2008

Abstract

Hedman, H. 2008. *Functional and evolutionary analysis of flowering time genes in Brassica nigra and Physcomitrella patens*. Doctor's dissertation. ISSN 1652-6880, ISBN 978-91-85913-88-6.

Plant genomes harbour many gene families reminiscent of previous duplication events. In this thesis, the molecular evolution of duplicated genes is analysed with special emphasis on flowering-time genes. In *Brassica nigra*, a close relative to *Arabidopsis*, three COL genes have been associated with flowering time (*BnCOa, BnCOb* and *BnCOL1*). The molecular evolution of these genes have been studied. All aspects of the data can not easily be explained by demography, indicating that selection has played a role in shaping the variation at these genes. *BnCOb* appear to be in the early process of pseudogenization.

A functional study of COL genes in the moss *Physcomitrella patens* shows that in *Ppcol3* knockout mutants, the generation of gametophores (leafy shoots) was diminished under blue light. Indicating that *PpCOL3* is required for gametophore development Stem elongation was affected in white light in *Ppcol2-Ppcol3* double mutants, but not in a *Ppcol3* single mutant, suggesting that *PpCOL2* might be involved in growth regulation of the gametophore.

A phylogenetic study of plant PEBP genes show that three major clades exists in land plants: the basal MFT clade and FT and TFL1 clades. An analysis of positively selected sites in the tree branches identified six putatively selected sites. A detailed study of the basal MFT clade, together with an expression analysis in *Physcomitrella,* shows that moss MFT-like genes are light-induced and appear to be associated with development of reproductive structures.

Keywords: Flowering time, Molecular evolution, CONSTANS, FLOWERING LOCUS T, *Brassica nigra*, *Physcomitrella patens*.

Author's address: Harald Hedman, Department of Plant Biology and Forest Genetics, Swedish University of Agricultural Sciences, Box 7080, SE-750 07 Uppsala, Sweden, *E-mail:* Harald.Hedman@ebc.uu.se

Contents

Appendix

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Sjödin, P., Hedman, H., Kruskopf Österberg, M., Gustafsson, S., Lagercrantz, U., Lascoux, M. (2008). Polymorphism and divergence at three duplicate genes in *Brassica nigra*. *Journal of Molecular Evolution 66 (6), 581-590*.

II Sjödin, P., Hedman, H., Shavorskaya, O., Finet, C., Lascoux, M., Lagercrantz, U. (2007). Recent degeneration of an old duplicated flowering time gene in *Brassica nigra*. *Heredity 98 (6),* 375-384.

III Hedman, H., Thelander, M., Ronne, H., Lagercrantz, U. (2008). PpCOL2 and PpCOL3 are involved in growth and development in *Physcomitrella patens*. (Manuscript).

IV Hedman, H., Källman, T., Moore, D., Lascoux, M., Lagercrantz, U., Gyllenstrand, N. (2007). Molecular evolution of the PEBP gene family in plants: evidence of positive selection on amino acids involved in functional change (Submitted).

V Hedman, H., Källman, T., Lagercrantz, U. (2008).Early evolution of the MFT-like gene family in plants (Submitted).
Papers I-II are reproduced with the permission of the publishers.

Papers I-II are reproduced with the permission of the publishers.

Introduction

The expansion of gene families involved in plant development suggests that gene duplication may have played a significant role in the evolution of plants. This thesis investigates the evolution of duplicated genes that control flowering in higher plants. The first part aims to examine the molecular evolution of duplicated flowering time genes; the annual seed plant *Brassica nigra*. The second part investigates the evolutionary history of flowering time genes in a comparative approach using the moss *Physcomitrella patens*.

Evolution by gene duplication

Two years before Watson and Crick presented the structure of the DNA double helix, proposing a "possible copying mechanism for the genetic material" (Watson and Crick 1953), S.G. Stephens (1951) presented an article about "Possible Significance of Duplication in Evolution". Even though he based his ideas on the concept that a gene was a nucleoprotein, Stephens recognised that if a gene was to gain a new function by mutation this could only be at the expense of losing the old one, and in many cases the mutation would be detrimental. According to Stephens it was "difficult to regard such a mechanism [...] as an efficient method of effecting evolutionary progress from the simple to the complex", and he suggested that "another mechanism where new functions could be added to the former" would be of selective advantage. The mechanism that Stephens envisaged was an increase of genes either by the de novo synthesis or duplication of pre-existing genetic material. However, Stephens wasn't breaking entirely new grounds. Studies of the effects of polyploidy in plants were at that time a mature discipline. Haldane stated that hybridisation in plants could cause rapid evolutionary jumps and offer protection against deleterious mutants, and concluded that polyploidization could possibly be of an evolutionary advantage (Haldane 1933).

Even though the true nature of genes was not known until 1953, the association between chromosomes and heredity was established (Sutton 1903). It was also known that genes were linearly arranged on the chromosomes, and that chromosomal recombination occurred (Sturtevant et al 1919, Morgan et al 1920). Moreover, studies of the *bar* eye-mutant in *Drosophila* showed that a gene could be duplicated by unequal crossing over

(Muller 1936, Bridges 1936). Muller's statement: "Every gene from a pre-existing gene" (1936), clearly shows that evolution by gene duplication was certainly considered at this time. Another *Drosophila* geneticist, Serebrovsky, proposed in 1938 that duplication "should result in a specialization of genes, when each then fulfils only one function", a process today known as subfunctionalization (Taylor and Raes 2004). Thus the two major causes of gene duplication; polyploidization and tandem duplication had been already been observed and contemplated upon before Stephens article in 1951.

However, the idea of evolution by gene duplication did not receive much attention until the idea was reintroduced in Ohno's classic "Evolution by Gene Duplication" (1970). Today Ohno is much cited as the originator of the concept of gene duplication, possibly because he failed to recognise the previous publications concerning this matter as opposed to Mayo(1970). Nei (1969) also proposed that evolution at the molecular level was caused by the increase in the DNA content of germ cells; and he predicted that higher organisms "carried "a considerable number of non-functional genes"(pseudogenes), which today has been confirmed in many genomic studies. As we enter the age of genomics, I will let Ohno himself explain the concept of evolution by gene duplication:

"Only a redundant copy of an original gene created by the mechanism of gene duplication escapes from the stranglehold by natural selection, and while being ignored by natural selection, it is free to accumulate formerly forbidden mutations which change the active site. As a result, it may emerge as a new gene locus with a previously nonexistent function" (Ohno 1972)

Even though evolution by gene duplication offered an attractive explanation for the evolution of complex organisms, there was more to come. As soon as molecular genetics came of age it became apparent that morphological differences could not easily be explained by the molecular evolution of protein coding genes. In 1975 King and Wilson analysed electrophoretic differences between proteins in chimpanzees and humans and concluded: "It appears that molecular change has accumulated in the two lineages at approximately equal rates, despite a striking difference in rates of organismal evolution". The authors proposed that morphological differences between individuals, populations, or species are mainly caused by regulatory differences (King and Wilson 1975). However, this had already been put forward by Ohno and others: "we realize that major steps in vertebrate

evolution were more often accomplished by changes in regulation of already existing structural genes rather than by the acquisition of new structural genes" (Ohno 1972). However, the idea that major leaps in evolution are caused by regulatory differences is difficult to test, because regulatory DNA follows other evolutionary dynamics than protein-coding DNA.

Subfunctionalization

Gene duplication theory predicts that one gene in a duplicated pair will be reduced to a non-functional pseudogene, unless constrained by functional or dosage requirements. Only in rare occasions will both copies be retained in a functional state, this process can be either *neofunctionalization* where one copy gains a new function, or *subfunctionalization* where both copies are needed to perform the ancestral function. According to Nei (2005), morphological evolution will always occur through positive selection in coding sequences, however Nei does not discuss promoter evolution, and here other mechanisms may be of importance. Lynch and co-workers proposed a subfunctionalization process where duplicated genes can diverge in function without the need for positive selection (Force et al 1999, Lynch and Force 2000). In this model, *degenerative* mutations in regulatory regions can produce a situation where duplicated genes diverge in expression domains. Thus, the ancestral function becomes divided between two genes, allowing for a refinement in both the coding region and in the regulatory region. One important point is that since the expression of each gene is more restrained, the dependency of other genes is reduced. This will lead to a reduction in pleiotrophy and can fine-tune the regulatory cascades into more elaborate pathways. The essential point is that the subfunctionalization process is dependent on degenerative mutations in the regulatory region and not beneficial mutations in the coding region - which are much less frequent (Lynch and Force 2000).

Today studies of duplicated genes are rapidly growing and the sequencing of whole genomes reveals that organisms have experienced several large-scale duplications. Another finding is the prevalence of large gene families with several homologs of high similarity, for example the HOX gene families in animals, and the MADS gene families in plants, thus Ohno's statement that "The creation of a new gene from a redundant copy is the most important role that gene duplication played in evolution" (Ohno 1970) seems more and more true.

The neutral theory

At the beginning of the 70's another forceful idea had been introduced by Kimura (1968). This was the hypothesis of neutral evolution, which initially challenged the idea of Darwinian selection, because genetic variation was postulated to be gained from the random accumulation of neutral mutations, as opposed to the prevailing idea that genetic variation was due to the maintenance of beneficial mutations by balancing selection (Nei 2005, Hughes 2007). The neutral theory is still debated, even though supporting evidence is rapidly accumulating. In the neutral view, genes are accumulating mutations at a steady pace. If a site has a selective value harmful mutations will be selected against (purifying, or negative, selection), and beneficial mutations will be fixed (positive selection). The result is that most *fixed* mutations will be either neutral or nearly neutral, this is because deleterious mutations are removed by purifying selection and beneficial mutations are very rare. Purifying selection is more efficient in large populations than in small ones, thus in small populations genetic drift will fix even slightly deleterious alleles (Nielsen 2005). Neutral theory also emphasises that gene duplication must occur before the emergence of a new gene function (Kimura and Ohta 1974).

In coding DNA, the genetic code is arranged in triples, or codons, where each triplet codes for an amino acid. The first two positions are conservative, while the third position is allowed to toggle for many amino acids. This means that mutations have different effects depending on their position. Mutations that result in an amino acid replacement are called non-synonymous (dN), and mutations that have no effects are called synonymous (dS). Neutral theory predicts that selectively *neutral* amino acids (or whole pseudogenes) will have a ratio of non-synonymous (dN) to synonymous (dS) mutations close to 1 ($dN \approx dS$). However, this ratio becomes skewed when selection is acting. For functionally *important* parts of the protein, amino acid substitutions are selected against ($dN < dS$). For sites where amino acid replacements are beneficial, substitutions are favoured ($dN > dS$) (Yang and Bielawski 2000).

As mentioned previously, most sites in a gene will be under either neutral or purifying selection. The result from purifying selection is the conservation of functional domains, and this is rather evident in alignments of distantly related genes. It is from these conserved domains that genes can be grouped

into gene families or functional classes. If a conserved domain is functionally described, this can help in understanding functions of an unknown protein carrying the same domain, for example a nuclear localisation signal or a DNA binding motif. However, conserved sites cannot explain phenotype variation. Thus much evolutionary research is focused on detecting positively selected sites, because these sites can tell us something about ongoing evolution. However, examples of coding regions under positive selection are rare. In many cases where positive selection is invoked, the genes are selected for amino replacements in variable domains (Hughes 2007). Some examples are MHC and other immune-system proteins, and proteins involved in plant self-incompatibility systems (Nei 2005).

Reports of genes under positive selection are growing, but according to Hughes (2007), some claims based on the McDonald-Kreitman (MK) test may be artefacts due to the random fixation of alleles in small populations, it is also possible that some sites are in the process of being deleted by purifying selection. Hughes also warns to rely too much on codon based methods for detecting positive selection (like PAML) because detected sites may be under relaxed purifying selection, for example after a duplication. In studies of positive selection without an *a priori* hypothesis, there is also a risk for the formulation of just so stories to account for the observed pattern. According to Hughes, positive selection is most likely to occur in proteins that are involved in protein-protein interaction and host-parasite evolution, but for other types of proteins the available methods may be inappropriate, for example, if a protein is evolving by deletion of previously important domains. Positively selected genes tend to reduce the variation in closely linked genes. This is called a selective sweep. Selective sweeps can be used for the detection of selected alleles, for example by genotyping large population samples by single nucleotide polymorphisms (SNP)(Nielsen 2005). However, the ongoing search for adaptive variation in coding regions, may have to be re-directed to regulatory regions, since many gradual phenotypes like flowering time are likely to be caused by expression differences (Hughes 2007).

Plant polyploids

Polyploids are most often formed from unreduced gametes (2n), this can happen within a species (autopolyploidy) or between related species (allopolyploidy). Both types of polyploids will benefit from the increase in

genetic variation unless there is a gene dosage problem (Birchler et al 2001). However, in an autopolyploid, the high similarity between chromosomes can cause unequal segregation of sister chromatids during cell division (Comai 2005). There are several potential advantages in polyploidy, like protection from inbreeding or deleterious recessive alleles. Another potential advantage may be a loss of self-incompatibility and finally, the potential advantage of having a duplicated set of genes on which evolution can act (Wendel 2000, Comai 2005). Polyploidization is far more common in vascular plants than in animals, this has been explained by the fact that many plants lack sex chromosomes; polyploidy in animals almost always leads to sterility. Another fact is that plant pollen is constantly spread across species boundaries. However, in most cases hybridization is hindered by pollen incompatibility. Polyploidization is believed to fuel plant speciation by introducing sexual isolation between polyploids and parent species. As more sequence data is added to the databases, the emerging pattern is that almost all plants have been experienced several rounds of polyploidization and local gene duplications. Thus it is "unlikely that the pure diploid plant genome exists" (Gale and Devos 1998).

Local duplications

Some gene families are evolving by local duplications. Occasionally tandem repeats are produced by unequal crossover in meiosis, this will in turn increase the probability of subsequent duplications, thus tandem repeats can expand to tens and hundreds of genes as seen in ribosomal genes and polyubiquitin. Some gene families appear to have evolved by tandem duplication followed by polyploidization, resulting in a rapid expansion of the family. One example is the HOX gene family in animals (Ohta 2003). Tandem repeats are often highly conserved. Previously this has been explained by concerted evolution, where duplicates are homogenised due to DNA repair mechanisms (gene conversion). This idea has been challenged by Nei et al (2000) that have investigated the evolution of the polyubiquitin gene family. According to these findings, the similarities are caused by purifying selection on duplicates subject to birth-and-death evolution and not by gene conversion.

Transposable elements

Another potentially important case of local gene duplication is the occasional insertion of transposon-mediated reverse transcribed mRNA in the gametes. The inserted gene lacks introns and also promoter, thus the expression will be different from the original gene. Transposons of class I involves an RNA intermediate while class II jumps directly within the chromosomes, both classes are extremely abundant in plants. When transposons jump, repetitive retroelements are left in the original position. In angiosperms with haploid genomes above 2000 Mb, over 50% of the nuclear DNA consists of LTR retrotransposons or other repeats, and as much as 70% of the maize genome consist of retroelements (Bennetzen 2005a, 2005b). Transposons have been regarded as parasitic DNA, but this idea is challenged by Brosius (1999), that postulates a more complex evolutionary dynamic: "Retroposition is an efficient route to move coding regions around the genome 'in search' of novel regulatory elements and to shotgun regulatory elements into the genome 'in search' of new target genes". Transposable elements can rapidly induce changes in gene expression, and together with polyploidization this could lead to a rapid subfunctionalization (Force et al 1999, Freeling and Thomas 2006). Transposons may also cause major chromosomal rearrangements by introducing double breaks (Feshotte and Pritham). Moreover, plants may be more protected against negative effects by transposons because of purifying selection of dominant lethal mutations in somatic cell lineages, and a quality check of lethals in the haploid gametophyte (Walbot and Evans 2003).

The genetic control of flowering

Before setting out an investigation about the evolution of flowering, it is necessary to have a thorough understanding about the genetic control of plant reproduction. Much understanding has been gained about the control of flowering in *Arabidopsis thaliana* and other seed plants, much less is known about the genetic control of reproduction in bryophytes (mosses, hornworts, liverworts) and less advanced tracheophytes (lycopods, ferns, horsetails, gymnosperms). The elucidation of flowering pathways in *Arabidopsis* has been extremely valuable and the knowledge is already used in comparative studies. A few functional studies of putative flowering time homologs have been made for the moss model organism *Physcomitrella patens* (Cove 2005), these include LFY (Tanahashi et al 2005), CO (Shimizu et al 2004, Zobell

et al 2005), MFT (Hedman et al 2008), and the MIKC MADS-box genes (Singer et al 2007, Quodt et al 2007). The evolution of gene families involved in the flowering pathway has also been investigated, although much focus has been on seed plants. However more species will be added to the comparisons due to the rapid expansion of expression libraries and genome sequences. Functional studies of flowering time homologs in the morphologically simple *Physcomitrella* (Reski 1997, Schaefer 2002) may shed some light on the more complex pathways in *Arabidopsis*. Together, these two model organisms will cross-fertilise the research in plant development. However, the large evolutionary distance between mosses and seed plants, makes it necessary to add more species to the functional comparisons, Phylogenetic studies may work as a framework, but functional data will be needed to test evolutionary hypotheses. Thus functional studies of crucial "missing link taxa" will increase the resolution in our understanding of plant evolution, and also bring forward new theories about the role of gene duplication in evolution (Soltis et al 2002). The understanding of the genetic control of flowering in *Arabidopsis* is the fundament on which comparative studies in plants are based. Below is a brief summary of some of the flowering pathways in *Arabidopsis*.

Flowering time pathways in *Arabidopsis*

Arabidopsis thaliana is a facultative long day annual found in temperate regions worldwide. The adaptation to temperate regions is reflected by the fact that flowering is accelerated under long day (LD) conditions, and in some cases if the plants have been exposed to a long period of cold (vernalization), the later response is found in the accessions that are winter annuals and flower the next spring after germination (Mouradov et al 2002). After germination a variable number of rosette leaves are produced before flowering is initiated. Flowering can occur within 3-4 weeks after seed germination, and is manifested by the rapid growth of a main stem carrying one central and several axillary inflorescences. Flowering time can be measured as the number of leaves produced before flowering occurs, which is strongly depending on environmental and genetic factors. *Arabidopsis* populations differ in flowering time; in general, southern populations flower earlier than northern ones, and southern populations respond very little to vernalization while some northern populations are very responsive. However, flowering occurs normally in all accessions reflecting the annual weed habit (Koornneef et al 1991).

Flowering time in *Arabidopsis* is affected by many different external factors like day-length, nutrient availability, ambient temperature, stress and neighbour competition (Blazquez 2000). Flowering can be also be promoted by applying the plant hormone gibberellin, most likely mimicking an internal hormonal response to flower-promoting conditions (Bernier and Perilleux 2005). Flowering mutants in *Arabidopsis* have been available since 1962 when Reideri identified *constans (co)*, *gigantea (gi),* and *luminidependens (ld)* (Kobayashi and Weigel 2007). However, the progress in understanding the genetic control of flowering was slow until researchers joined forces and focused on flowering mutants in *Arabidopsis* (Meinke 1998). The cloning of flowering time mutants generated and described by Koornneef et al (1991) and others have resulted in a rapidly growing understanding about some of the key players in flower induction. Some of the flowering time mutants Koornneef identified are functionally linked e.g., *GIGANTEA (GI), CONSTANS (CO), FLOWERING LOCUS D (FD), FLOWERING LOCUS T (FT),* and *FLOWERING LOCUS C (FCA)* (Putterill et al 1995, Macknight 1997, Abe et al 2005, Wigge et al 2005, Kardailsky et al 1999, Kobayashi et al 1999, Fowler et al 1999). New genes involved in the control of flowering time are continuously discovered, and encode proteins involved in the light reception, circadian rhythm, protein degradation, chromatin remodelling and mRNA processing. The multitude of regulatory levels reveals that flowering is tightly controlled by a complex genetic machinery (Mouradov 2002, Putterill et al 2004).

The photoperiodic pathway

One of the earliest findings regarding photoperiod and flowering came from greenhouse experiments by Tournois in 1912 and Klebs in 1913. These experiments showed that given artificial light, flowers could develop even in winter, indicating that light duration was critical in flowering. Later, controlled day-length experiments by Garner and Allard [1920, 1923] (by carrying plants to a windowless shed) showed that a shortened day-length could accelerate flowering in soybeans and tobacco. This led to the classification of plants as (facultative or obligate) short day (SD), long day (LD), or day-neutral (Kobayashi and Weigel 2007).

The circadian clock

The duration of the day is one of the few predictable parameters to a plant; local growth conditions change constantly, but the seasonal light rhythms stay the same. By monitoring day-length and light quality a plant can adjust growth and reproduction to meet seasonal changes and avoid shading. Thus, the photoperiod pathway starts with genes involved in the perception of light and circadian rhythms. Plants have an internal diurnal genetic rhythm that is reset at dawn every day. This internal rhythm is provided by the circadian clock. A circadian clock is present in all organisms and is composed of genes connected by negative auto-regulatory loops. The circadian clock in *Arabidopsis* seems to involve three interconnected loops, one between *CCA1/LHY* and *ELF4*, one between *CCA1/LHY* and *ARR5/ARR7/ARR9,* and one between *GI* and *TOC1*. The clock also include several modifiers that affect the stability and light-sensitivity of the clock (Covington 2001, Locke et al 2006). One important clock-associated gene is ELF3, proposed to shelter the clock from transient light at dusk (Carre 2003). The clock is reset every morning and this entrainment is mainly caused by phytochromes and cryptochromes. One known mechanism is the binding of a PHYB-PIF3 complex to the promoters of CCA1 and LHY (Putterill 2004). Many genes in the photoperiod pathway, including the photoreceptors and many other genes as well, have a circadian expression that is controlled by the proteins encoded by the clock. The phase however can be shifted, and some genes have opposite expression rhythms.

The light coincidence model

Light enters the photoperiodic pathway by photoreceptors specialised to different wavelengths. In *Arabidopsis* red/far-red light is perceived by the five phytochromes (PHYA-PHYE), and blue light is perceived by cryptochromes (CRY1 and CRY2) and phototropins (Thomas 2006). Several of these photoreceptors are directly involved in the protein turnover of the key integrator of photoperiod in *Arabidopsis,* the CONSTANS protein. During the night and morning, PHYB and three clock-controlled genes *SUPPRESSOR OF PHYA-105-1 (SPA1, SPA3, SPA4)* degrade CO, while PHYA and the cryptochromes have an equally important function in stabilising CO during the day. Together these opposing forces causes CO protein to accumulate by the end of a long day and return to low levels at daybreak, despite high levels of CO mRNA during the night (Turck et al 2008).

The *CO* gene expression is controlled by the circadian clock and the rhythm is constant regardless of day-length. CO expression is light-induced and the mRNA starts to accumulate during the day. In short days, however, the accumulation is below threshold by the onset of darkness, while in long days, the accumulation is prolonged - leading to high mRNA levels by the end of a long day (Suarez-Lopez et al 2001). Moreover, CO mRNA abundance is affected by *FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 (FKF1)* protein, that together with GIGANTEA putatively degrades the *CO* repressor CYCLING DOF FACTOR 1 (CDF1) (Sawa et al 2007). FKF1 activity is blue-light dependent and necessary for the stabilisation of the CO mRNA during the day (Imaizumi et al 2003). Suarez-lopez et al (2001) proposed that the accumulation of CO mRNA caused the acceleration of flowering in long days. This is called the "external coincidence model", a mechanism originally proposed by Bünning 1936, where a genetic response is triggered if light coincides with high levels of a cycling internal protein (Samach and Coupland 2000). CO in turn activates FT, which promotes the developmental transition at the shoot meristem (Kobayashi et al 1999, Kardialsky et al 1999).

FT moves to the meristem

In the present model, the flowering time pathway in *Arabidopsis* starts in the leaf vascular tissue where CO protein is stabilised by light. High levels of CO promote the expression of *FT* in the leaf, and of the close homolog, *TWIN SISTER OF FT (TSF)* in the stem. Both proteins are then unloaded to the phloem and transported to the apical meristem (Turck et al 2008). In the meristem FT (and possibly TSF) interacts with locally expressed *FLOWERING LOCUS D* (*FD*) protein (Abe et al 2005), and promotes transcription of the MADS-box genes *APETALA 1 (AP1)* and FRUITFUL (FUL) (Conti and Bradley 2007). Once AP1 expression is stabilised FT is no longer required and the shoot meristem becomes totally committed to flower development. How light controls growth processes in roots stems and seeds is not known, but it is plausible that pathways similar to the one implicated in flowering are involved, possibly involving other members of the *CO* and *FT* gene families. For example, it has been shown that the CO homolog COL3 is involved in lateral root formation and accelerates flowering under SD (Datta 2006).

Other pathways

The photoperiodic pathway acts in parallel with several other pathways like age, stress and nutritional status that all converge at the floral integrator *LEAFY (LFY)* (Blazquez 2000). Hence, other factors will eventually induce flowering in *co* and *ft* mutants. Another important factor that control flowering in *Arabidopsis* is the MADS-gene *FLC*, which acts as a repressor at several levels in the flowering pathways, targeting SOC1, FT and also CRY2 which promotes CO transcription (Boss et al 2006). In the *Arabidopsis* winter annuals, FRIGIDA (FRI) up-regulates FLC and causes a strong vernalization requirement (Mouradov et al 2002). One point made by Bernier and Perilleux (2005) is that annuals like *Arabidopsis* flower once and then die, while perennials and trees delay flowering until the accumulation of resources have reached a certain threshold level. This difference in life history is likely to be reflected by the relative importance of flowering time pathways. Size is often a good determinant of flowering in perennials, and is essentially an integrator of growth-limiting factors like irradiance, water and mineral availability, ambient temperature and competition. Integration of size and age into the flowering pathways is especially pertinent to trees, and how this is controlled may have to be investigated in other species than *Arabidopsis* (Bernier and Perilleux 2005).

Flower development is the product of a multitude of genes that ultimately converge at the meristem. Thus, Boss et al (2006) suggested that genes in flowering pathways could be classified as either promoting or enabling floral meristem development. One particular aspect of meristem development is that suppression is just as important as promotion. This is because once flowering is initiated it cannot be reversed. Many genes in the autonomous pathway are negative regulators of FLC. Hence, the enabling of the meristem response by the down-regulation of FLC may be a requisite for the flower-promoting pathways (Boss et al 2006).

Meristem phase changes

In *Arabidopsis*, apical growth occurs at the shoot meristem. At the core of the meristem lies a quiescent organising centre of a four cells that rarely divide, this core is surrounded by stem cells comprising the central zone. As the stem cells divide the apical daughter cells are displaced towards the flanks - keeping the apical layer at constant thickness. Further away from the centre

a peripheral zone is established where rapidly dividing cells initiate leaf primordia under the influence of auxin. It is believed that the primordial cells can differentiate because they are beyond the range of cytokinin, produced in the organising centre by the action of *WUSCHEL (WUS)*. *WUS* is only expressed in the organising centre and maintain the stem cell population by promoting the expression of *CLAVATA3 (CLV3)* (Sablowski 2007, Dinneny and Benfey, 2008, Dello Loyo et al 2008). During the vegetative phase, the meristem events result in a spiral arrangement of leaves and secondary shoots along the central stem. During floral transition, the meristem changes to an inflorescence meristem reiterated at the flanks; the secondary stems then produce the flowers (Krizek and Fletcher 2005). Two fundamental branching patterns are seen in plants. In the monopodial branching, as seen in *Arabidopsis*, only one round of secondary branches are allowed. In sympodial branching as seen in tomato, tertiary branches are allowed, some with vegetative growth, some carrying inflorescences (Lifschitz and Eshed 2006).

A putative TFL1-gradient

As previously mentioned, a cytokinin gradient originating from the meristem central region has been proposed to regulate differentiation of stem cells in *Arabidopsis*. Recently Conti and Bradley (2007) proposed a similar model, where a gradient of TFL1 protein originating from the same region is involved in flower development. In their model TFL1 represses LFY and AP1 in the central region, but is unable to repress LFY at the periphery, resulting in flower development at the flanks. This is supported by the fact that TFL1 mRNA is only found in a small region at the core of the inflorescent meristem. The protein however has a much wider distribution and is probably distributed through plasmodesmata, establishing a gradient of TFL1 with a maximal concentration in the central region (Conti and Bradley 2007, Blazquez et al 2006).

Meristem switching

In a hypothetical model, the TFL1-gradient could be involved in the switching between two semi-stable networks of MADS-box genes, one set controlling vegetative development and the other floral development; each set maintains its own state by positive feedback loops and at the same time repressing the opposite set. In this way the meristem is locked into either pathway, until one of the genes in the opposite set is up regulated. A similar regulatory network of MADS-genes is controlling the patterning of floral

organs in the *Arabidopsis* meristem. In the classical ABCE model, two or more MADS genes; expressed in different regions of the meristem, control the spatial arrangement of floral organs. Moreover, in the quartet model, these MADS genes are proposed to function as tetrameric complexes, explaining the functional redundancy and combinatorial requirements of these genes (Krizek and Fletcher 2005).

Before the ABCE expression domains are established, the flanking meristems must switch from vegetative development to floral development by expressing the floral meristem identity MADS genes AP1, CAL (Blazquez et al 2006). One question is how this switch is controlled, since the central meristem in *Arabidopsis* continues the vegetative growth while the flanking meristems embark on the flowering pathway. In the vegetative phase AP1 and CAL are repressed by the MADS genes FUL and AGL24. This repression might be controlled by TFL1. TFL1 and FUL expression coincides in the central region of the vegetative shoot before floral transition is initiated (Blazquez et al 2006). After flowering has been initiated, the repressors TFL1, FUL and AGL24 are replaced with AP1, CAL and LFY. This only occurs in the flanking meristems; at the centre of the shoot TFL1, FUL and AGL24 remain expressed and keep the central meristem in vegetative phase. Thus somehow a developmental patterning is imposed on the shoot, with a vegetative set of MADS genes expressed in the centre and flower promoting set at the flanks. One model to explain this is by a TFL1 gradient similar to the model proposed by Ratcliffe et al (1999).

Are FT and TFL1 competitors?

The photoperiod pathway results in the unloading of FT in the shoot, where FT bind to FD and up regulate AP1 and FUL (Turck et al 2008). However, AP1 is only expressed in the flower initials and FUL is only expressed in the central meristem (Blazquez et al 2006). Is it possible that this pattern is caused by opposing actions of FT and TFL1? These two proteins are structurally very similar (Hanzawa 2005), and it is possible that both compete in binding to FD. Thus, one model is that the vegetative phase is maintained by TFL1 repression on AP1. After flower induction, FT competes with TFL1 and up-regulates AP1 in the peripheral meristems where TFL1 concentration is low. Subsequently AP1, CAL and LFY up-regulate each other and repress FUL to establish a flower meristem. In the central core TFL1 maintains the inflorescence by promoting FUL permitting more rounds of flower initials. Eventually FT activates AP1 in

the core, resulting in a terminal flower. This hypothetical model could explain why TFL1 is central in maintaining the inflorescence.

The autonomous pathway

The autonomous pathway, is more complicated but may involve a LFY-gradient in the opposite direction to the TFL1-gradient. LFY may also be transported by plasmodesmata, since the LFY protein is small is found beyond its zone of expression (Weigel et al 1992) Blazquez et al 1997). During vegetative development, *LFY* is expressed at low levels at leaf primordia, but after floral transition *LFY* becomes highly expressed in floral primordia, but do not extend into the centre of the shoot (Blazquez et al 1997, Blazquez et al 2006). This could indicate that *TFL1* repress *LFY* in concentration-dependent manner, directly or indirectly, with a maximal repression at the centre of the inflorescence meristem. The gradual increase in *LFY* expression during plant development, possibly due to increasing levels of gibberellin, suggest that *LFY* acts an integrator of external factors (age, temperature, nutrients), comprising a light-independent flowering pathway (Mouradov et al 2002). This highly speculative view does not take into account the effects of many other regulatory genes, like FLC. However, to this initial model, additional levels of regulation can be added.

The CO gene family

The key photoperiod pathway integrator *CO* belongs a large gene family comprising 16 members in *Arabidopsis* (Robson et al 2001). In *Physcomitrella* the expansion is smaller with only three close homologs: *PpCOL1, PpCOL2 and PpCOL3* (Zobell et al 2005). The CO protein contains two conserved domains; one amino terminal domain with two Zn-finger motifs and one carboxy terminal domain containing a CCT motif, shared among several other proteins in *Arabidopsis* (Wenkel et al 2006) The CCT domain also has a nuclear localisation signal (Putterill 1995). CO is a transcriptional factor, but has not been shown to directly bind to DNA. Recently, it has been shown that CO can form a trimeric complex with *HAP3* and *HAP5* belonging to the *HEME ACTIVATOR PROTEIN (HAP)* family. The interaction is mediated through the CCT domain in CO, and the complex is believed to activate transcription by binding to a CCAAT box in the promoter (Wenkel et al 2006, Ben-Naim et al 2006). There are 35 members in the HAP gene family in *Arabidopsis* (Wenkel et al 2006). Moreover, Cai

et al (2007) has showed that a *hap3b* mutant delayed flowering under LD while overexpression of HAP3b promoted early flowering. These findings suggest that genes with CCT domains form trimeric complexes with HAP-genes, and regulate expression (Cai et al 2007)

The PEBP-like gene family

FT has received much attention lately because of its central role in the photoperiodic pathway as a candidate for the elusive florigen (Turck et al 2008). FT belongs to a gene family with five other members in *Arabidopsis*: *TERMINAL FLOWER 1 (TFL1)*, *TWIN SISTER OF FT (TSF), ARABIDOPSIS THALIANA CENTRORADIALIS (ATC), BROTHER OF FT AND TFL1 (BFT) AND MOTHER OF FT AND TFL1 (MFT) (*Kobayashi 1999). These genes are similar to Phosphatidylethanolamine binding proteins (PEBP) found in other multi-cellular organisms (Bradley et al 1997). In *Physcomitrella*, only MFT-like genes are present, thus the FT/TFL1 homologs have evolved in the tracheophytes (Hedman et al 2007). One interesting feature of *FT* is the close homology to *TFL1*. The proteins have similar structures, with a putative ligand-binding a pocket and an external loop that is assumed to interact with residues residing in the pocket (Banfield and Brady 2000, Ahn et al 2006). However, while FT is a promotor of flowering, TFL1 is a repressor and keeps the inflorescence in an indeterminate stage (Kobayashi 1999, Kardialsky 1999). The expression domains of these proteins are different; *FT* is highly expressed in the leaf mesophyll, and *TFL1* is highly expressed in the shoot apical meristem. By changing one single residue within this pocket Hanzawa et al. (2006) could switch functions between FT and TFL1. Thus it appears that one single mutation can produce a drastic functional exchange in these genes. We do not know much about the function MFT-like genes in plants, but in *Arabidopsis,* MFT expression is highest the seed and ovule which could indicate a role in seed development (Winter et al 2007). Yoo et al (2004) found no clear association with flowering time in the loss-of-function *mft* allele; however, over-expression of MFT caused a small decrease in flowering time suggesting a weak FT activity.

Aims of the study

This endeavour is an attempt to increase the understanding of the evolution of genes involved in the photoperiodic control of flowering. The photoperiodic pathway in *Arabidopsis thaliana* has been extensively studied, and the emerging picture is that the initial event where day-length is translated into a flowering signal involves the two genes *CONSTANS (CO)* and *FLOWERING LOCUS T (FT)*. Both *CO* and *FT* belong to families that have expanded by gene duplication during the evolution of land plants. The expansion of gene families involved in plant development, suggests that gene duplication may have played a significant role in the evolution of plants.

In this study we are examining *Brassica nigra,* which is a close relative to *Arabidopsis*. *B.nigra*, is a polyploid with essentially a triplicated ancestral *Arabidopsis* genome. Thus *B.nigra* offers a possibility to examine the evolution of recently duplicated flowering time genes, which have been extensively investigated in its sister species. We are focusing on three genes belonging to QTL linkage groups previously associated with flowering time in B.nigra (Lagercrantz et al 1996). These genes, *BnCOa*, *BnCOb* and *BnCOL1* have different evolutionary histories; *BnCOL1* probably originates from an ancestral tandem duplication of *BnCOa*. The triplicated genome suggests that both *BnCOa* and *BnCOL1* once existed in two more copies, but that all these copies have been lost, except *BnCOb*. The remaining genes are presumably retained due to functional constraints. This study aims specifically to examine the molecular evolution of these duplicates, but also to gain a deeper understanding of the evolution of flowering in the moss *Physcomitrella patens*.

Specific aims

I. To examine the molecular evolution of three closely related flowering time genes in Brassica nigra: *BnCOa, BnCOb* and *BnCOL1* (paper I).

II. A detailed study of the molecular evolution of *BnCOb* since this gene was indicated to be involved in flowering time in *Brassica nigra* (paper II).

III. To perform functional studies of CO-like genes in *Physcomitrella* by the generation of loss-of-function mutants (paper III).

IV. To study the phylogenetic history of PEBP-like (FT-like) genes in plants and test for site-specific positive selection (paper IV).

V. To examine the evolution of *MFT-like* genes in plants from phylogenetic reconstructions and expression studies in *Physcomitrella* (paper V).

Results and discussion

Molecular evolution of *COL* genes in *Brassica nigra* (Paper I and II)

The black mustard *Brassica nigra* (2n =16) is an out crossing annual and a close relative to *Arabidopsis thaliana*. *B. nigra* is believed to be an ancient hexaploid in having essentially a triplicated *Arabidopsis* genome. The genome is extensively rearranged, but linkage groups can be assembled into what is probably three ancient diploid genomes originating from an ancestor of *Arabidopsis* and *Brassica* (Lagercrantz 1998). Previous QTL studies of genetic variation affecting flowering time in *B. nigra* revealed two candidate chromosomal locations (Lagercrantz et al 1996). The first QTL mapped to a region with *BnCOL1* and *BnCOa*, and the second smaller QTL mapped close to the more recently duplicated *BnCOb*. These findings have spurred more investigations on the molecular evolution and genetic variation of these three genes.

In a study of flowering time in *B. nigra* populations, Kruskopf-Österberg et al (2002) found that variation in *BnCOa* coding region + 600 kb upstream sequence could not explain variation in flowering time. However indel variation in *BnCOL1* was correlated with flowering time, possibly by linkage disequilibrium with a downstream region. This downstream region is located in the intergenic region between *BnCOL1* and *BnCOa* (the genes are tandem duplicates). Thus, variation in this region could cause differences in *BnCOa expression* and be correlated with flowering time. The maintenance of high variation in this intergenic region was suggested to be caused by balancing selection on flowering time. A later extended study of the same populations could not detect any sign of positive selection in the *BnCOL1* coding region, again indicating that the flowering time is correlated with the *BnCOL1* - *BnCOa* intergenic region (Lagercrantz et al 2002). In *Arabidopsis*, overexpression of *COL1* had no effects on flowering time, suggesting that *COL1* is not redundant to *CO* (Ledger et al 2001).

In Paper I, we present a study of the molecular evolution of *BnCOa*, *BnCOb* and *BnCOL1* detected in the initial QTL study. The sequences of all three genes were analysed in a subset from the previously used accessions (five individuals from each population, Greece, France, Germany, Ethiopia

and India). First we show that the nucleotide diversity is different in the three genes, and that this could in part be caused by selection rather than random genetic drift. *BnCOa* has the lowest level of variation, *BnCOL1* has approximately a doubled variation, and *BnCOb* twice the variation of *BnCOL1*. Thus *BnCOa* appears to be most conserved and *BnCOb* least conserved. The ratio of non-synonymous to synonymous substitutions (dN/dS) was high for all three genes, confirming the previous finding that CONSTANS -like genes are evolving exceptionally fast in Brassicaceae (Lagercrantz and Axelsson 2000). We used the McDonald-Kreitman test to look for signs of positive selection in comparison with *Arabidopsis*, The results showed that *BnCOL1* had a significant excess of polymorphic *synonymous* mutations. This could indicate a stronger purifying selection in *BnCOL1* than in *Arabidopsis COL1*. No evidence of positive selection was found for *BnCOa* or in *BnCOb*. A possible scenario compatible with our data is that *COL1* went through a phase of relaxed constraint after the duplication from *CO*, followed by a recent increase in *COL1* in *B. nigra*. The results obtained in this study contradict a previous study (Lagercrantz et al 2002) that found no evidence of positive selection in a smaller subset of populations, our results from a larger dataset could therefore be caused by random fixation in subpopulations as suggested by Hughes (2007).

In paper II a more detailed study of *BnCOb* is undertaken since this gene was located close to a QTL for flowering time, and showed high insertion/deletion polymorphism (indels). In this study we show that although indels in many cases probably disrupt protein function, other features of molecular evolution show no signs of pseudogenization (the estimate of dN/dS ratio is smaller for *BnCOb* than for other *COL* genes in *B. nigra)*. These contradictory findings lead us to conclude that it is plausible that *BnCOb* is in an early stage of pseudogenization with loss of function indels. The small QTL for the *BnCOb* region indicate that *BnCOb* still might have a redundant effect on flowering time, and *BnCOb* is still transcribed. Thus, it is possible that *BnCOb* could have an effect on flowering time in some populations. However *Arabidopsis co* mutants transformed with putatively functional *BnCOb* alleles showed no acceleration in flowering time. We propose that *BnCOb* was retained for a long period after duplication, but a recent fixation of a detrimental mutation, possibly as an effect of a bottleneck, resulted in its non-functionalization. This event is then too recent to result in any effect on the dN/dS ratio.

Functional analysis of COL homologs in Physcomitrella (Paper III)

In paper III we investigate the functional role of two CONSTANS-like genes in *Physcomitrella*, *PpCOL2* and *PpCOL3*. The function of these genes is analysed by the generation of (putative) single *Ppcol3* and a double *Ppcol2-Ppcol3* mutants. The mutants show normal phenotypes under normal conditions, but in blue light the generation of gametophores (leafy shoots) is diminished in *Ppcol3* mutants, suggesting that *PpCOL3* is involved in gametophore development and receives input from the blue-light pathway. We also see indications that the gametophores of the double mutant are less elongated in blue light. This could indicate that *PpCOL2* is necessary for growth promotion, and could receive input from blue light photoreceptors. The analysis of the mutant genotypes reveals that the homologous recombination process had resulted in multiple insertions in the target sites. The insertion in *PpCOL2* appears to have resulted in the duplication of either the whole gene, or parts of the gene, however expression studies show that both *PpCOL2* and *PpCOL3* are silenced in the double mutants. One remaining question is if PpMFT-like genes are downstream targets of PpCOL genes, and this was tested, but no clear connection could be found in the mutants (data not shown). In order to clarify this, more experiments will be necessary, particularly under conditions where phenotypic effects are strong (blue light and low nutrient levels). This study is the first to demonstrate that PpCOL genes are involved in developmental transitions in *Physcomitrella*, and should be repeated as the results can give new knowledge about developmental regulation in mosses, and provide clues about the evolution of the CO-FT regulatory module in plants (Böhlenius et al 2006).

The phylogeny of PEPB-like genes in plants (Paper IV)

In Paper IV we investigate the phylogenetic history of PEBP-like genes in plants. This study confirms previous findings that plant PEBP genes have expanded during land plant evolution, and that MFT is the basal clade with representatives from all extant plant phyla (Izawa et al 2002). We find no evidence for a TFL1 like gene in gymnosperms, even though our data is limited to EST libraries. One gymnosperm gene PaFT4 have been reported to exhibit FT-like properties in having an expression which was positively correlated to bud burst and bud set under SD conditions. In the same study, three other FT-like homologs were not found to be correlated with bud burst or bud set (Gyllenstrand et al 2007). Still the possibility that a TFL1

function exists in gymnosperms should be considered, given the essential role for TFL1 in specifying the vegetative meristem in *Arabidopsis* (Conti and Bradley 2007).

The plant PEPB genes are highly conserved throughout the whole protein, suggesting an important role in plant development, and also functional constraints regarding size and activity. The finding that FT acts through its counterpart FD highlights that FT interacts with other proteins in order to promote transcription (Abe et al 2005). Another interesting feature is that TFL1 can be converted from a repressor of flowering to a promoter of flowering with FT-like properties by a change in one single amino acid (Hanzawa et al 2005). This implies that the evolution of TFL1 could involve positive selection at this site. This was tested by PAML analysis on the phylogenetic tree of the PEPB gene family. Indeed, the branch leading from the MFT clade showed sign of positive selection in six sites on the proteins, one of which was this crucial amino acid (Tyr85/His88). The PAML test has been criticized for generating a high rate of false positives (Nei 2005). However the high number of samples and high overall sequence conservation provide some support for our results. One particular problem is the old age of the nodes in the phylogeny. With an estimated divergence of vascular plants and bryophytes of approximately 450 Myr, it is likely that silent sites are saturated, resulting in an under estimation of silent site substitutions. Therefore, more conservative tests should be made gain more support for the suggestion that the Tyr85/His88 residue was selected in the TFL1/FT branch (Hughes 2007).

The early evolution of MFT-genes in plants (Paper V)

In paper V the phylogeny of *MFT* in all major plant lineages was reconstructed in a thorough analysis including the cloning of *MFT* members from *Physcomitrella patens* and *Selaginella sp.* (three species). In addition the expression of the *Physcomitrella* genes was analysed in different photoperiods and in different tissues. The emerging picture is that the *MFT* genes are duplicated within *Physcomitrella* (four genes) and that higher vascular plants generally have two MFT-like homologs. In our study we also report that a new *MFT* subclade is found in angiosperms, the function of this extra copy is unknown. We have examined available expression data for angiosperm MFT homologs and find support for the idea that *MFT* may be involved in seed development, an idea put forward by Danilevskaya et al 2008. The presence of two MFT-like genes in most vascular plants and four genes in

Physcomitrella, lends support to the hypothesis that the ancestral species had two MFT-like genes. Still, there is no evidence for this. In the basal liverwort *Marchantia polymorpha* we have at present only evidence for one MFT-like gene, and in *Selaginella moellendorphii* there is only one MFT-like homolog, but in this species there is one additional partially MFT-like gene which has diverged considerably. If the FT/TFL1 functional dichotomy is crucial for the determination of the shoot meristem, we may expect to find both FT and TFL1-like genes in all vascular plants.

The expression analysis reveals that *Physcomitrella MFT* genes are induced by light, and that the expression increases as the moss develops, with a maximal expression in late developmental stages. These findings suggest an involvement in the development of reproductive tissues in the moss. If the MFT-like genes in *Physcomitrella* also have a similar role as TFL1, in maintaining vegetative growth, or if these genes have a role in the induction of reproduction, is still unclear. It will be interesting to examine the functional role of the four genes in the moss, as these will shed some light on the emergence of TFL1-like properties in vascular plants.

Future perspectives

Plant reproduction has been subject to extensive research during decades but our knowledge of the genetic pathways involved is relatively new. The analysis of flowering time mutants in *Arabidopsis* has pushed open the door for much exciting research. Now, as the genomic sequences are being reported, we are in the position to ask how plant reproduction have evolved, and at the same time test hypotheses about neutral evolution and the role of gene duplication in the evolution of plants. The rise of *Physcomitrella* as model species is in its infancy, and many exciting news will come from this beautiful and simple plant. By comparing the genetic control of reproduction in this bryophyte with the control of flowering in seed plants, we may be in a position to both make and test hypotheses about the evolution of plants.

Functional genomics is also comparative genomics, and it is clear that knowledge acquired from *Arabidopsis* can be applied in other model organisms, even in those very distant, like bryophytes. It is also clear that functional studies in early lineages can shed light on the functional evolution of regulatory pathways in later lineages; this can be very valuable when it comes to the studies of gymnosperms that have very large genomes and also long generation times. In this thesis, I have studied the evolution of two key genes involved in seed plant reproduction, *CONSTANS (CO)* and *FLOWERING LOCUS T (FT).* From functional studies in *Arabidopsis,* the *a priori* hypothesis is that these two genes are involved in the initial stages of plant reproduction. However, the function of *TFL1* in *Arabidopsis* indicates that *PpMFT* genes also may be involved in establishing a developmental pattern in the gametophore. The evolution of genes is intimately connected to their function, and as we gain more knowledge about their ancient function we will also learn about their evolution. Therefore I would like to proceed with the analysis of the functional roles of the *PpCOL* and *PpMFT* genes in *Physcomitrella*, in order to be able to make educated guesses about the evolution of vascular plants from an ancient bryophyte. It will also be interesting to follow the search for a plant sister group in algae, as an algae ancestor would immensely improve the understanding of early land plant evolution.

For both *PpCOL* and *PpMFT* genes, the main goal will be to generate new loss-of-function mutants (single and multiple) and overexpressors.

Much has been learned from the previous studies, as to where and when phenotypes may appear, and this is very valuable. The second step would be to study the regulation of these genes and also identify their downstream targets. Indications of the upstream regulators may be found in functional studies. Downstream targets may be identified either by large scale microarray approaches or candidate gene approaches. One particularly appealing approach is to study conditional over expressors in a mutant background. One promising conditional promoter is the heat shock promoter reported by Saidi et al (2005).

From the *Arabidopsis* studies, some questions are evident, for example is the CO-FT pathway conserved in plants? Both MFT-like and CO-like genes exist in *Physcomitrella* but if these are functionally connected is unknown. What are the downstream targets of MFT-like genes in *Physcomitrella*? It will be interesting to examine if PpLFY or MADS genes are affected in a PpMFT knockout. Moreover, there are several FD-like genes in *Physcomitrella*, and whether any of these genes interact with PpMFT will be an interesting study. It will also be necessary to analyse tissue expression patterns on a smaller scale, in order to make a detailed description of the expression domains of the *PpCOL* and *PpMFT* genes in *Physcomitrella*. Furthermore, the protein distribution patterns will also have to be examined, since *PpMFT* protein may be transported and *PpCOL* genes may be tightly regulated at the protein level. As more is learnt from studies in *Physcomitrella*, some hypotheses about conserved function have to be tested in *Arabidopsis*. Some questions are already quite pertinent, like the investigation of the function of the second MFT-like gene in seed plants (Paper V), and studies of a potential MFT function the seed and ovule (Paper V), and finally more investigations about the function of the close *Physcomitrella* homologs, COL3, COL4 and COL5 in *Arabidopsis* (Paper IV).

References

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science*, 309 (5737), 1052-1056.

Ahn, J., Miller, D., Winter, V., Banfield, M., Lee, J., Yoo, S., Henz, S., Brady, R., Weigel, D. (2006). A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *Embo Journal*, 25 (3), 605-614.

Banfield, M., Brady, R. (2000). The structure of *Antirrhinum centroradialis* protein (CEN) suggests a role as a kinase regulator. *Journal of molecular biology*, 297 (5), 1159-70.

Ben-Naim, O., Eshed, R., Parnis, A., Teper-Bamnolker, P., Shalit, A., Coupland, G., Samach, A., Lifschitz, E. (2006). The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant Journal*, 46 (3), 462-476.

Bennetzen, J. (2005). Transposable elements, gene creation and genome rearrangement in flowering plants. *Current Opinion in Genetics & Development*, 15 (6), 621-627.

Bennetzen, J., Ma, J., Devos, K. (2005). Mechanisms of recent genome size variation in flowering plants. *Annals of botany*, 95 (1), 127-132.

Bernier, G., Perilleux, C. (2005). A physiological overview of the genetics of flowering time control. *Plant Biotechnology Journal*, 3 (1), 3-16.

Birchler, J., Bhadra, U., Bhadra, M., Auger, D. (2001). Dosage-dependent gene regulation in multicellular eukaryotes: Implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Developmental biology*, 234 (2), 275-288.

Blazquez, M. (2000). Flower development pathways. *Journal Of Cell Science*, 113 (20), 3547-3548.

Blazquez, M., Weigel, D. (2000). Integration of floral inductive signals in *Arabidopsis*. *Nature*, 404 (6780), 889-892.

Blazquez, M., Ferrandiz, C., Madueno, F., Parcy, F. (2006). How floral meristems are built. *Plant molecular biology*, 60 (6), 855-870.

Blazquez, M., Soowal, L., Lee, I., Weigel, D. (1997). LEAFY expression and flower initiation in *Arabidopsis*. *Development*, 124 (19), 3835-3844.

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

Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R., Coen, E. (1997). Inflorescence commitment and architecture in *Arabidopsis*. *Science*, 275 (5296), 80-83.

Bridges, C. (1936). The Bar "gene" - A duplication. 83, 210-211.

Brosius, J. (1999). Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica*, 107 (1-3), 209-238.

Böhlenius, H., Huang, T,, Charbonnel-Campaa, L., Brunne,r A., Jansson, S., Strauss, S., Nilsson, O. (2006). CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science*, 312 (5776) pp. 1040-1043.

Cai, X., Ballif, J., Endo, S., Davis, E., Liang, M., Chen, D., DeWald, D., Kreps, J., Zhu, T., Wu, Y. (2007). A putative CCAAT-binding transcription factor is a regulator of flowering timing in *Arabidopsis*. *Plant Physiology*, 145 (1), 98-105.

Carre, I. (2002). ELF3: a circadian safeguard to buffer effects of light. *Trends in Plant Science*, 7 (1), 4-6.

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, 6 (11), 836-846.

Conti, L., Bradley, D. (2007). TERMINAL FLOWER1 is a mobile signal controlling *Arabidopsis* architecture. *Plant Cell*, 19 (3), 767-778.

Cove, D. (2005). The moss *Physcomitrella patens*. *Annual review of genetics*, 39, 339-358.

Covington, M., Panda, S., Liu, X., Strayer, C., Wagner, D., Kay, S. (2001). ELF3 modulates resetting of the circadian clock in *Arabidopsis*. *Plant Cell*, 13 (6), 1305-1315.

Danilevskaya, O., Meng, X., Hou, Z., Ananiev, E., Simmons, C. (2008). A genomic and expression compendium of the expanded PEBP gene family from maize. *Plant physiology*, 146 (1), 250-64.

Datta, S., Hettiarachchi, G., Deng, X., Holm, M. (2006). *Arabidopsis* CONSTANS-LIKE3 is a positive regulator of red light signaling and root growth. *Plant Cell*, 18 (1), 70-84.

Dello Loio, R., Linhares, F., Sabatini, S. (2008). Emerging role of cytokinin as a regulator of cellular differentiation. *Current opinion in plant biology*, 11 (1), 23-27.

Dinneny, J., Benfey, P. (2008). Plant stem cell niches: Standing the test of time. *Cell*, 132 (4), 553-557.

Feschotte, C., Pritham, E. (2007). DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet, 41, 331-368

Force, A., Lynch, M., Pickett, F., Amores, A., Yan, Y., Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151 (4), 1531-1545.

Fowler, S., Lee, K., Onouchi, H., Samach, A., Richardson, K., Coupland, G., Putterill, J. (1999). GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. *Embo Journal*, 18 (17), 4679-4688.

Freeling, M., Thomas, B. (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome research*, 16 (7), 805-814.

Gale, M., Devos, K. (1998). Plant comparative genetics after 10 years. *Science*, 282 (5389), 656-9.

Gyllenstrand, N., Clapham, D., Kallman, T., Lagercrantz, U. (2007). A Norway spruce FLOWERING LOCUS T homolog is implicated in control of growth rhythm in conifers. *Plant physiology*, 144 (1), 248-257.

Haldane, J. (1933). The part played by recurrent mutation in evolution. *The American naturalist*, 67, 5-19.

Hanzawa, Y., Money, T., Bradley, D. (2005). A single amino acid converts a repressor to an activator of flowering. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (21), 7748-53.

Hughes, A. (2007). Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity*, 99 (4), 364-373.

Imaizumi, T., Tran, H., Swartz, T., Briggs, W., Kay, S. (2003). FKF1 is essential for photoperiodic-specific light signalling in *Arabidopsis*. *Nature*, 426 (6964), 302-306.

Izawa, T., Oikawa, T., Sugiyama, N., Tanisaka, T., Yano, M., Shimamoto, K. (2002). Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice. *Genes & development*, 16 (15), 2006-20.

Kardailsky, I., Shukla, V., Ahn, J., Dagenais, N., Christensen, S., Nguyen, J., Chory, J., Harrison, M., Weigel, D. (1999). Activation tagging of the floral inducer FT. *Science*, 286 (5446), 1962-5.

Kenrick, P., Crane, P. (1997). The origin and early evolution of plants on land. *Nature*, 389 (6646), 33-39.

Kimura, M. (1968). Evolutionary rate at molecular level. *Nature*, 217 (5129), 624-&.

Kimura, M., Ohta, T. (1974). Some principles governing molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 71 (7), 2848-2852.

King, M., Wilson, A. (1975). Evolution at 2 levels in humans and chimpanzees. *Science*, 188 (4184), 107-116.

Kobayashi, Y., Weigel, D. (2007). Move on up, it's time for change - mobile signals controlling photoperiod-dependent flowering. *Genes & development*, 21 (19), 2371-2384.

Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., Araki, T. (1999). A pair of related genes with antagonistic roles in mediating flowering signals. *Science*, 286 (5446), 1960-1962.

Koornneef, M., Hanhart, C., van der Veen, J. (1991). A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Molecular & general genetics : MGG*, 229 (1), 57-66.

Krizek, B., Fletcher, J. (2005). Molecular mechanisms of flower development: an armchair guide. *Nature Reviews Genetics*, 6 (9), 688-98.

Lagercrantz, U. (1998). Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that Brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics*, 150 (3), 1217-1228.

Lagercrantz, U., Axelsson, T. (2000). Rapid evolution of the family of CONSTANS LIKE genes in plants. *Molecular Biology and Evolution*, 17 (10), 1499-1507.

Lagercrantz, U., Osterberg, M., Lascoux, M. (2002). Sequence variation and haplotype structure at the putative flowering-time locus COL1 of *Brassica nigra*. *Molecular Biology and Evolution*, 19 (9), 1474-1482.

Lagercrantz, U., Putterill, J., Coupland, G., Lydiate, D. (1996). Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. *Plant Journal*, 9 (1), 13-20.

Ledger, S., Strayer, C., Ashton, F., Kay, S., Putterill, J. (2001). Analysis of the function of two circadian-regulated CONSTANS-LIKE genes. *Plant Journal*, 26 (1), 14-22.

Lifschitz, E., Eshed, Y. (2006). Universal florigenic signals triggered by FT homologues regulate growth and flowering cycles in perennial day-neutral tomato. *Journal of experimental botany*, 57 (13), 3405-3414.

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

Lynch, M., Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics*, 154 (1), 459-473.

Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C., Dean, C. (1997). FCA, a gene controlling flowering time in Arabidopsis, encodes a protein containing RNA-binding domains. *Cell*, 89 (5), 737-745.

Mayo, O. (1970). The role of duplications in evolution. Heredity, 25 (4), 543-53.

Meinke, D., Cherry, J., Dean, C., Rounsley, S., Koornneef, M. (1998). *Arabidopsis thaliana*: A model plant for genome analysis. *Science*, 282 (5389), 662-+.

Morgan, T., Sturtevant, A., Bridges, C. (1920). The evidence for the linear order of the genes. *Proceedings of the National Academy of Sciences of the United States of America*, 6, 162-164.

Mouradov, A., Cremer, F., Coupland, G. (2002). Control of flowering time: Interacting pathways as a basis for diversity. *Plant Cell*, 14, S111-S130.

Muller, H. (1936). Bar duplication. *Science*, 83, 528-530.

Nei, M. (1969). Gene duplication and nucleotide substitution in evolution. *Nature*, 221 (5175), 40-2.

Nei, M. (2005). Selectionism and neutralism in molecular evolution. *Molecular Biology and Evolution*, 22 (12), 2318-23442.

Nei, M., Rogozin, I., Piontkivska, H. (2000). Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proceedings of the National Academy of Sciences of the United States of America*, 97 (20), 10866-71.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual review of genetics*, 39, 197-218.

Ohno, S. (1970). Evolution by Gene Duplication. Springer Verlag. Berlin

Ohno, S. (1972). Gene duplication, mutation load, and mammalian genetic regulatory systems. *Journal Of Medical Genetics*, 9 (3), 254-&.

Ohta, T. (2003). Evolution by gene duplication revisited: differentiation of regulatory elements versus proteins. *Genetica*, 118 (2-3), 209-216.

Osterberg, M., Shavorskaya, O., Lascoux, M., Lagercrantz, U. (2002). Naturally occurring indel variation in the *Brassica nigra* COL1 gene is associated with variation in flowering time. *Genetics*, 161 (1), 299-306.

Putterill, J., Robson, F., Lee, K., Simon, R., Coupland, G. (1995). The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc-finger transcription factors. *Cell*, 80 (6), 847-857.

Quodt, V., Faigl, W., Saedler, H., Münster, T. (2007). The MADS-domain protein PPM2 preferentially occurs in gametangia and sporophytes of the moss *Physcomitrella patens*. *Gene*, 400 (1-2), 25-34.

Ratcliffe, O., Bradley, D., Coen, E. (1999). Separation of shoot and floral identity in *Arabidopsis*. *Development*, 126 (6), 1109-1120.

Reski, R. (1998). Development, genetics and molecular biology of mosses. *Botanica Acta*, 111 (1), 1-15.

Robson, F., Costa, M., Hepworth, S., Vizir, I., Pineiro, M., Reeves, P., Putterill, J., Coupland, G. (2001). Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. *Plant Journal*, 28 (6), 619-631.

Sablowski, R. (2007). The dynamic plant stem cell niches. *Current opinion in plant biology*, 10 (6), 639-644.

Saidi, Y., Finka, A., Chakhporanian, M., Zrÿd, J.-P., Schaefer, D., Goloubinoff, P. (2005). Controlled expression of recombinant proteins in *Physcomitrella patens* by a conditional heat-shock promoter: a tool for plant research and biotechnology. *Plant molecular biology*, 59 (5), 697-711.

Samach, A., Coupland, G. (2000). Time measurement and the control of flowering in plants. *BioEssays*, 22 (1), 38-47.

Sawa, M., Nusinow, D., Kay, S., Imaizumi, T. (2007). FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. *Science*, 318 (5848), 261-265.

Schaefer, D. (2002). A NEW MOSS GENETICS: Targeted Mutagenesis in *Physcomitrella patens*. *Annu. Rev. Plant Biol*, 53, 477-501.

Shimizu, M., Ichikawa, K., Aoki, S. (2004). Photoperiod-regulated expression of the PpCOL1 gene encoding a homolog of CO/COL proteins in the moss *Physcomitrella patens*. *Biochemical and biophysical research communications*, 324 (4), 1296-301.

Singer, S., Krogan, N., Ashton, N. (2007). Clues about the ancestral roles of plant MADS-box genes from a functional analysis of moss homologues. *Plant Cell reports*, 26 (8), 1155-1169.

Soltis, D., Soltis, P., Albert, V., Oppenheimer, D., dePamphilis, C., Ma, H., Frohlich, M., Theissen, G. (2002). Missing links: the genetic architecture of flower and floral diversification. *Trends in Plant Science*, 7 (1), 22-31.

Stephens, S. (1951). Possible significance of duplication in evolution. *Advances In Genetics Incorporating Molecular Genetic Medicine*, 4, 247-265.

Sturtevant, A., Bridges, C., Morgan, T. (1919). The spatial relations of genes. *Proceedings of the National Academy of Sciences of the United States of America*, 5, 168-173.

Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., Coupland, G. (2001). CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature*, 410 (6832), 1116-1120.

Sutton, W. (1903). The chromosomes in heredity. *Biological Bulletin*, 4 (5), 231-251.

Tanahashi, T., Sumikawa, N., Kato, M., Hasebe, M. (2005). Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic *Cell* division in the moss *Physcomitrella patens*. *Development*, 132 (7), 1727-36.

Taylor, J., Raes, J. (2004). Duplication and divergence: The evolution of new genes and old ideas. *Annual review of genetics*, 38, 615-643.

Thomas, B. (2006). Light signals and flowering. *Journal of experimental botany*, 57 (13), 3387-3393.

Turck, F., Fornara, F., Coupland, G. (2008). Regulation and Identity of Florigen: FLOWERING LOCUS T Moves Center Stage. *Annual review of plant biology*, 59, 573-94.

Walbot, V., Evans, M. (2003). Unique features of the plant life cycle and their consequences. *Nature Reviews Genetics*, 4 (5), 369-79.

Watson, J., Crick, F. (1953). Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature*, 171 (4356), 737-738.

Weigel, D., Alvarez, J., Smyth, D., Yanofsky, M., Meyerowitz, E. (1992). LEAFY controls floral meristem identity in *Arabidopsis*. *Cell*, 69 (5), 843-859.

Wendel, J. (2000). Genome evolution in polyploids. *Plant molecular biology*, 42 (1), 225-249.

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., Coupland, G. (2006). CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell*, 18 (11), 2971-2984.

Wigge, P., Kim, M., Jaeger, K., Busch, W., Schmid, M., Lohmann, J., Weigel, D. (2005). Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science*, 309 (5737), 1056-1059.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G., Provart, N., Baxter, I. (2007). An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLoS ONE*, 2 (8), 718.

Yang, Z., Bielawski, J. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, 15 (12), 496-503.

Yoo, S., Kardailsky, I., Lee, J., Weigel, D., Ahn, J. (2004). Acceleration of flowering by overexpression of MFT (MOTHER OF FT AND TFL1). *Molecules and Cells*, 17 (1), 95-101.

Zobell, Coupland, G., Reiss, B. (2005). The family of CONSTANS-like genes in Physcomitrella patens. *Plant Biology*, 7 (3), 266-275.

Acknowledgements

This work has been carried out at the Department of Evolutionary Functional Genetics at Uppsala University, and the Department of Plant Biology and Forest Genetics at SLU, Uppsala, Sweden. The research was financed by The AgriFunGen programme for functional genomics at SLU. Additional funding was also received from the Nilsson-Ehle foundation.

Most people start their reading here and I will begin my thesis by writing this. It has been a long journey and it was as rewarding and exciting as it has been difficult and frustrating. Therefore it is with great pleasure I here present my gratitude to many of those that have helped my on my way onward.

First I will give my deepest and warmest thanks to my supervisor Ulf Lagercrantz. Without whom, this thesis would not have been possible. Thank you for being patient and understanding. I'm especially grateful for the times I expected criticism, but received none. My deepest respect for your profound understanding of science. Finally, thank you for your kindness. For always keeping an open door to your office, and your efforts in helping me finish.

I would like to thank all my colleagues at the Department of Evolutionary Functional Genetics, EBC, especially my department brethren Thomas and Niklas, the tag team of spruce genetics, and Karl, my office wingman, and Kerstin and Susanne for your kind help and all the laughs in the lab.

I began my post-graduate studies at the Department of Plant Biology and Forest Genetics at SLU: I would like to than all the friendly people at Genetikcentrum, especially Jan, Mats, and Jenni. Jenny and Monica.

Thanks to the Uppsala moss community, Hans, Mattias, Tina and Anders. I think we can all agree that *Physcomitrella* rules.

Thanks to my friends. Terje, my first choice if I would ever take the Long Way Round. Thanks for days in Skyttorp. Svante and Anneli for all help in Uppsala and for expedition Fyrisån. Emil and Carina for keeping an open outpost on my long journeys back to Umeå. Friddi, my Icelandic friend, I

finally made it! Wish you were here. Thanks to Klas for your long lasting friendship ever since the early years in Umeå.

A deep and warm thanks to my mother, Vanja. You are the essence of motherhood. Thanks for all your letters with various food supplies! To my father Ola for your support. To my sister Torun and her family for all help in Umeå. Finally, to my great big brother Rikard. Thanks for all reports from your adventures when I was here in exile, and thanks to your family for all kindness back home in Arjeplog.

My girlfriend and beautiful companion Marika Nordström. Your love and endurance carried me to the finish. Now I stand here not only with a thesis but also with two wonderful children, giving me a whole new perspective on what's really important in life. You know that words can never be enough for your efforts. Thank you for walking the distance alongside with me. To you I owe everything and to your family I send my warmest thoughts.

To my wonderful kids Linn and Jack. Forgive me for being away from you so often. I love you.