Identification of genes affecting flowering time variation in *Brassica* species

O.A.Shavorskaya
Department of Plant Biology and Forest Genetics
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2004

Acta Universitatis Agriculturae Sueciae

Agraria 453

ISSN 1401-6249
ISBN 91-576-6467-6

Tryck: SLU Service/Repro, Uppsala 2004

Abstract

Shavorskaya, O. 2004. Identification of genes affecting flowering time variation in *Brassica* species.
Doctor's dissertation
ISSN1401-6249, ISBN 91-576-6467-6

One prerequisite for a deeper understanding of genetics of adaptive variation is the identification and analyses of genes affecting such variation. In this thesis, I have searched for genes affecting variation in the adaptive trait flowering time in different Brassica species.

Comparative QTL mapping of flowering time indicated that duplicated copies of one ancestral gene, possibly *CONSTANS* (*CO*) might regulate flowering in several Brassica species. A region homologous to the *CO* region of *Arabidopsis thaliana* is present in three copies in diploid, and six copies in polyploid Brassica species, and QTL mapped to two out of three regions in the diploids, and three out of six in the polyploid.
The two QTL identified in *B. nigra* were studied further. *CO* alleles from early and late flowering plants were isolated from both QTL regions (Bni *COa* and Bni *COb*). No effect of allelic variation at Bni *COa* on flowering time was detected when comparing *A. thaliana co* mutants transformed with the two *B.nigra* alleles. However, allelic variation at Bni *CONSTANS LIKE1 (*Bni *COL1)* located 3kb upstream of Bni *COa* showed association with flowering time in natural populations of *B. nigra*. Specifically, an 18 bp indel in Bni *COL1* showed significant association with flowering time in several populations. These results and more thorough studies support a hypothesis where a naturally occurring quantitative trait nucleotide (QTN) for flowering time affects the function of Bni *COL1* or the expression of Bni *COa* through variation in the intergenic region between *COL1* and *COa*.

Analysis of variation at Bni *COb* suggested that it has been turned into a pseudogene although some aspects of the data are difficult to reconcile with the loss of function.

Key words: *Brassica*, flowering time, comparative mapping, *CONSTANS*, duplication, adaptation, genetic variation, indel polymorphism, molecular evolution, silencing.

Author's address: Oksana Shavorskaya, Department of Plant Biology and Forest Genetics, Swedish University of Agricultural Science. Box 7080, SE-750 07 Uppsala, Sweden
E-mail: Oksana.Shavorskaya@vbsg.slu.se

Appendix

Papers I-IV

This thesis is based on the following papers, which will be referred to as their Roman numerals.

1) Axelsson, T., Shavorskaya, O., Lagercrantz, U. (2001) Multiple flowering time QTLs within several Brassica species could be the result of duplicated copies of one ancestral gene. *Genome* Oct;44(5):856-64.
2) Osterberg, M.K., Shavorskaya, O., Lascoux, M., Lagercrantz, U. (2002). Naturally occurring indel variation in the Brassica nigra COL1 gene is associated with variation inflowering time. *Genetics* May;161(1):299-306.
3) Shavorskaya, O., Lagercrantz, U. (2004). Sequence divergence at the putative flowering time locus *COL1* in Brassicaceae (manuscript).
4) Shavorskaya, O., Hedman, H., Lascoux, M., Lagercrantz, U. (2004). Molecular evolution of a recently duplicated flowering time gene in *Brassica nigra* (submitted).

Contents

Introduction **7**

The Genetic Basis of Adaptation **7**

Theoretical models 7
Empirical data 8

Gene duplication and adaptation **10**

Origin of duplicated genes 11
Gene duplication 11
Polyploidization 11
Fate of gene duplication 12

Methods for identification of genes effecting adaptive traits **13**

Quantitative trait locus (QTL) mapping 13
Experimental design 13
Fine mapping 14
Linkage disequilibrium mapping 14
Positional cloning and comparative mapping 15
Analysis of DNA sequence variation 15
Frequency-distribution tests 16
Tests based on comparing variability in different loci or different classes of mutations – heterogeneity tests 16

Timing of flowering is an adaptive trait **17**

Genetic basis of flowering time control 18
Photoperiod pathway 19
The connection between the circadian clock and flowering time 21
Vernalization pathway 22
Autonomous pathway 23
Gibberellin pathway 23
Interaction of photoperiod, vernalization and gibberellin pathways 23
Genes affecting natural genetic variation in flowering in Arabidopsis 25
Flowering time genes in other species 25

The Brassica species **26**

Genome structure of diploid Brassica species 27
B.nigra CO genes as candidates for the control of flowering time variation 28

Aims of the study **30**

Specific aims **30**

Results and discussion **30**

Multiple flowering time QTLs map to homologous positions within and between several Brassica species (I) 30

Flowering time variation and sequence divergence at the putative flowering time locus COL1 - CO*a* (II and III) 31

Molecular evolution of a recently duplicated flowering time gene in *Brassica nigra* (IV) 34

Future perspectives **35**

References **37**

Acknowledgements **43**

Introduction

One of the greatest challenges in evolutionary biology is to understand the origin of interspecific differences and the relationship between variation in natural populations and species divergence (Lynch and Walsh 1998). To achieve this goal, it is necessary to characterise the patterns of genetic variation within and among species and to elucidate the evolutionary forces acting on this variation (Kopp et al 2003).

Many characters of evolutionary and agricultural importance, such as timing of flowering, are quantitative in nature. Variation in these traits can be partitioned into environmental and genetic components, the genetic component presumed to be due to the segregation of alleles at a number of loci that affect the trait (Falconer and Mackay 1996). As plant varieties spread, there is strong selective pressure for changes in flowering time that confers an advantage in that new environment. There are several questions that needs to be answered in order to explain the nature of standing variation in quantitative traits and how the adaptations that distinguish closely related species become established. These questions include the number of loci contributing to quantitative variation, the number and effects of alleles at these loci, the interaction between loci and with the environment, and the molecular nature of the variants that give rise to quantitative variation.

It is known that polyploidisation has occured often in the evolutionary history of plants, and polyploidy might promote adaptive evolutionary change. *Brassica* is one of the highly polyploid genomes. The questions of the evolution of duplicated genes and the extent of involving homologous genes in controlling flowering time variation within and between species will be considered.

The Genetic Basis of Adaptation

Adaptation is a mechanism by which populations gradually approach a phenotypic optimum via the stepwise substitution of favourable mutations. Classical genetic models of adaptation suggest that it depends primarily on mutations with small phenotypic effects. However, recent empirical data indicate that major gene effects may underlie many of the important differences observed not only between species but also within species.

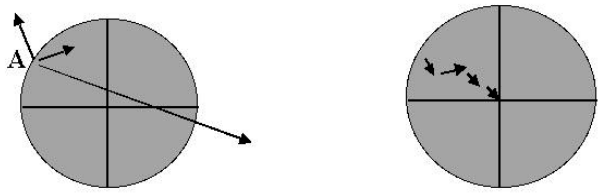
Theoretical models

On of the first theoretical models of adaptation was formulated by Fisher (1930). His classical theory of quantitative genetics predicts that adaptive variation is mostly based on a large number of loci with small effects. He showed that while very small mutations enjoy a 50% chance of being advantageous, larger mutations suffer a rapidly decreasing probability of being favourable. An illustrative example

of Fisher's model is shown in Figure 1. If the population, (point A), in which each organism is composed of just two traits (represented by the two coordinate axis), is away from the optimum (O), any random mutation (arrows) that moves the population inside the circle improves adaptation, whereas any mutation moving the population outside the circle worsens it. Any mutation that is larger than the circle's diameter overshoots the optimum, inevitably landing outside the circle. (Orr 1998). The conclusion was that factors of large phenotypic effect play little or no role in adaptation.

A second model of adaptation was suggested by Kimura (1983). He pointed out that Fisher's calculation ignored the probability of fixation of favorable mutations. Fisher correctly derived the probability that a random mutation of a given size will be favorable. However, he overlooked the fact that large mutations, while less likely to be advantageous, enjoy higher probability of fixation when favorable. So, Kimura supposed that, contrary to Fisher's views, mutations of intermediate phenotypic effect are the most likely to play a role in adaptation. However, neither Fisher nor Kimura provided the size distribution of factors fixed over the entire approach to the optimum. This is important because during a bout of adaptive evolution, a population gradually closes in on a phenotypic optimum. Orr (1998) derived the size distribution of mutations fixed in such a stepwise approach to the optimum.

Orr (1998) showed that the distribution of sizes fixed during adaptation is roughly exponential. Furthermore, he also showed that the expected size of the largest factor can be fairly large (Fig 1b).



A. B.

Figure 1. A two-dimensional (two-character) organism. Each character is represented by a coordinate axis. The optimal phenotype sits at the origin *O* and the population's present position is given by point A. All phenotypes inside the circle are better adapted than the present population. The arrows represent random mutations having different magnitudes and directions. A. Fisher's model. B. Orr's model: an adaptive "walk" to the optimum in two-character species. The length of each arrow represents the "size" of a mutation.

Empirical data

Results from quantitative trait locus analysis (QTL analysis, detailed description in later chapter) appear to be inconsistent with the Fisherian view too. QTL studies allow two of the most fundamental questions in the genetics of adaptation to be addressed: 1) how many genes are involved in the evolution of adaptation and 2) what is the distribution of phenotypic effects among the alleles fixed during adaptation (Orr 1998)?

Identification of the genes underlying adaptation has been difficult for a number of reasons. It is necessary to understand how ecologically important traits affect fitness in different environments. Furthermore, many fitness-related traits are quantitative and are unlikely to have a simple genetic basis.

Interspecific examples. One of the best known examples of a major gene affecting an adaptive trait involves the evolution of a derived life cycle mode in the Mexican Salamander (*Ambystoma mexicanum*). Using an interspecific cross (between *A. tigrinum tigrinum* and *A. mexicanum*) and genetic linkage analysis, a major quantitative trait locus for expression of metamorphosis was identified. Results from this study suggest that paedomorphosis (an adaptation for an entirely aquatic life cycle, i.e. metamorphic failure) in *A. mexicanum* evolved via a major gene. This recessive, major gene yields a phenotype that is clearly adaptive; metamorphic failure or paedomorphosis alters the entire life history of the organism because individuals remain in the aquatic habitat throughout the life cycle (Voss *et al*.1997).

Two sister species, pink-flowered *Mimulus lewisii*, which is pollinated by bumblebees and red-flowered *M. cardinalis*, which is pollinated by hummingbirds, were used in field studies to determine the adaptive value of alternative alleles at a major quantitative trait locus, YELLOW UPPER (YUP) (Bradshaw et al. 2003). This locus controls the presence or absence of yellow carotenoid pigments in the petals of both species. YUP was identified in a QTL mapping experiment based on a cross between the two species. The QTL study showed that most floral traits had at least one QTL with large effect. The YUP allele from each species was substituted into the other in near-isogenic lines (NILs). *M. cardinalis* NILs with the *M. lewisii* YUP allele had dark pink flowers and received 74-fold more bee visits than the wild type, whereas *M. lewisii* NILs with the *M. cardinalis yup* allele had yellow-orange flowers and received 68-fold more hummingbird visits than the wild type. Thus, a single major mutation in a flower colour locus produced a large pollinator shift.

The genetics of morphological differences between cultivated maize (*Zea mays L. Ssp.mays*) and its wild progenitor teosinte *(Z.mays ssp. parviglumis*) is also an interesting example, indicating that genes of large effect can be involved in species differentiation. QTL experiments identified 5 regions, with important effect on morphology (Doebley *et al*. 1990). Using complementation tests it was demonstrated that one of the QTLs is the locus for the known maize mutant *teosinte (tb*1). *tb1* is hypothesized to be a locus involved in the plastic response of the teosinte plant to its local environment. Genetic changes at *tb1* also results in a major change in plant architecture observed between maize and teosinte.

These examples provide evidence that single genes can induce major morphological shifts between species.

Intraspecific examples. The rock pocket mouse, *Chaetodipus intermedius*, provides a useful system for studying the genetics of adaptation. Classic studies (Benson. S.B. 1933 and Dice, L.R. *et al*. 1937) in the 1930s revealed a strong correlation between the color of the dorsal pelage and the colour of the substrate on which *C. intermedius* live. Recently, association studies were conducted using

markers in candidate pigmentation genes, and four mutations in the melanocortin-1-receptor gene, Mc1r, that seem to be responsible for adaptive melanism in one population of lava-dwelling pocket mice, were discovered. Interestingly, another melanic population of these mice on a different lava flow shows no association with Mc1r mutations, indicating that adaptive dark colour has evolved independently in different populations of this species through changes at different genes (Nachman M.W. *et al*. 2002).

Studies of classic quantitative traits in *Drosophila melanogaster* have shown that much of the mutational and segregating variation affecting bristle number, a model quantitative trait, is attributable to alleles with large phenotype effects at a small number of candidate loci (Mackay 1995). The association between quantitative genetic variation in bristle number and molecular variation at a candidate neurogenic locus, scabrous, was examined in *D. melanogaster*. Approximately 32 percent of the genetic variation in abdominal bristle number (21 percent for sternopleural bristle number) among 47 second chromosomes from a natural population was correlated with DNA sequence polymorphisms at this locus. Several polymorphic sites associated with large phenotypic effects occurred at intermediate frequency (Lai *et al*. 1994).

Also in Drosophila, QTL responsible for intraspecific variation in several sexually dimorphic morphological traits were identified (Kopp A. *et.al*. 2003). The traits included female abdominal pigmentation and the number of ventral abdominal bristles in males and females. To capture the pattern of genetic variation present in the wild, a panel of recombinant inbred lines (RILs) was created from two heterozygous flies taken directly from nature. The large number of RILs and the dense spacing of cytogenetic markers used in the study allowed mapping the QTL with high resolution, opening the way for candidate gene analysis. With the help of quantitative complementation tests, *bric a brac* gene (*bab*) was identified as the locus responsible for most of the variation in sexually dimorphic pigmentation but not in bristle numbers. The three strongest QTL account for >60% of variation in the number of ventral abdominal bristles, and *bab* alone accounts for 60% of variation in female pigmentation. Since the flies used to establish the RILs were taken directly from nature and did not undergo any inbreeding or artificial selection, the results suggest that QTL of large effect are present, and possibly common, in natural populations (Kopp *et.al*. 2003).

All these examples show that major genes can be an important force in evolution. Major genes can be involved in species differentiation, but also in within species variability. However, more data, in particular within species, is needed to clarify the generality of these results.

Gene duplication and adaptation

Gene duplication has long been thought to play a major role in evolution and the adaptation of organisms to the environment, as duplicated genes may be free to evolve and acquire new functions (Ohno, 1970). Ohno (1970) maintained that the evolution of new genes and novel biochemical processes could arise only via gene

duplication. Although other mechanisms such as alternative splicing, post-transcriptional and post-translational modifications, and regulatory mutations among others can serve to increase the functional diversity of a gene without duplication, the pervasive role of gene duplication in the generation of genomic complexity cannot be denied. Gene duplication in conjunction with domain shuffling has frequently been suggested to play an important role in the origin of novel genes (Long et al. 1993; Nurminsky et al. 1998).

Recent genome sequencing efforts have confirmed that gene and genome duplications have been prevalent in the evolution of most species (Lynch and Conery 2000). Estimates of the fraction of genes that represent recent duplication events range from 11 % in *Haemophilus influenzae* to 65 % in *Arabidopsis thaliana* (AGI 2000). In *A. thaliana,* 17 % were found in tandem arrays.

Origin of duplicated genes

Genes have replicated both through duplication of relatively small regions of DNA and through whole genome duplication events, i.e., polyploidy.

Gene duplication

Shorter stretches of DNA can duplicate in several ways, but mechanisms involving transposable elements and unequal crossing over are the most common. Transposons may occasionally replicate not only themselves, but also affect neighboring genes. Unequal homologous recombination may occur between repeated elements, resulting in duplication or loss of intervening sequences.

Polyploidization

The evolution of plants in particular, seems to be characterised by recurrent polyploidization (Wendel 2000). Even the small *Arabidopsis thaliana* genome contains numerous large duplicated chromosomal segments (Blanc et al. 2000), suggesting at least two large-scale duplication events (Blanc et al. 2003).

Similarly, genomes of related Brassica species have experienced additional rounds of polyploidization after the divergence from *Arabidopsis*. Diploid Brassica spp., such as *B. nigra, B. oleracea and B. rapa* contain three copies of a genome similar to that of Arabidopsis (see below; Lagercrantz and Lydiate 1996).

Polyploidy can occur as a result of errors at three points: mitosis, meiosis or fertilisation (polyspermy) (Stebbins 1980, Ramsey and Schemske 1998) Failure of reductive divisions during meiosis and errors in fertilisation are the most common source of polyploids.

One of the biggest stumbling blocks to the successful establishment of polyploidy in sexual species is the requirement for a genetically compatible mate. Polyploidy is more often found among perennial plants (Stebbins 1938), probably because having a long life span increases the chances that rare events will occur (e.g. polyploidisation following hybridization), and allows for mating between polyploids and their offspring (Otto 2000). Polyploid populations more often have

multiple origins than single origins (Soltis et al. 1999). Polyploidy is far more rare in animals than in plants, partly because polyploidization would interfere with sex determination in animals. Shifts in flowering time and ecological habitat are frequently found among plants of different ploidy level (Thompson et al. 1992), which would increase the mating isolation between diploids and polyploids. In contrast, animal mobility would reduce the chance that newly formed polyploids are spatially isolated.

The fate of duplicated genes

Theory suggest three alternative fates for duplicate genes (Otto 2000):

1) Maintenance of both copies performing the exact same function.
2) Inactivation or loss of one copy by the accumulation of mutations that disrupt its function.
3) Divergence in function through neo-functionalization, specialization or subfunctionalization.

In the case of neo-functionalization the two copies may diverge and take on novel functions. If two copies specialize in function, each becoming better at performing one of the original functions of the single-copy gene, it may led to specialization. Two copies might also accumulate deleterious mutations disrupting different sub-functions, causing both copies to be maintained in order to perform all the functions of the original single-copy gene. This process has been coined sub-functionalization.

Most empirical data on the likelihood of each of these fates comes from studies of ancient polyploids. The data suggest that a surprisingly high proportion of genes have been retained over long periods of time (Otto and Yong 2002). Based on genome sequence data, Lynch and Connery (2000) estimated the half-life of duplicated genes to about 4 million years. This estimate indicates a higher silencing rate than that obtained from studies of polyploids.

Several theoretical models of the fate of duplicated genes have been developed (Watterson 1983, Walsh 1995, Force et al. 1999). Fisher (1935) showed that if two copies are completely redundant, they are very unlikely to be maintained, and the most common immediate fate of duplicated genes is silencing. Inactivated pseudogenes have been observed in many gene families (nematode worm *Caenorhabditis elegans*: Mounsey et al. 2002; bacterial genomes: Mira A. et al. 2001; 9; *Drosophila melanogaster*: Robin et al. 2000). Such pseudogenes are characterized by a high rate of nonsynonymous nucleotide substitutions (causing amino acid replacement), suggesting that purifying selection is weak or absent (Otto 2000).

Long term maintenance of duplicates is more likely if mutations are deleterious even when they occur in one gene only. This could occur in a polyploid, where loss of a duplicted gene could cause a deficit of its gene product relative to those with which it interacts (that are still produced from duplicated copies). This could explain why duplicated genes derived from polyploidization seems to be

maintained for longer periods of time than those that arose by other means (Lynch and Conery 2000).

Even though silencing is likely for most duplicated genes, a large number of genes have been retained and diversified in function. Hughes (1994) reviewed examples where positive selection is indicated in diversification of gene duplicates. Force et al. (1999) showed that diversification could occur as a result of inactivation of sub-functions. This process of subfunctionalization preserves duplicates over longer period of evolutionary time, also allowing for more adaptive mutations.

Methods for identification of genes effecting adaptive trait

Quantitative trait locus (QTL) mapping

Variation in qualitative, major gene traits, are due to allelic differences at just one or a couple of genes. For qualitative traits, the effects of allelic differences on the phenotype are sufficiently larger than the effect on quantitative traits due to the environment to produce discrete classes between genotypes.

However, most traits cannot be put in this category. Rather, most traits show continuous variation, and these characters are called quantitative or metric, because their study depends on measurement instead of on counting. A quantitative trait locus (QTL) is a region in the genome, containing one or several genes affecting a quantitative trait. QTL, like other genes, can be mapped, and the effect of individual QTL can be estimated.

Experimental designs for mapping of QTL are based on linkage disequilibrium between alleles at marker locus and alleles at the linked QTL. One of the requirements for mapping QTL is a linkage map of polymorphic marker loci that adequately covers the genome. Marker loci should preferably be highly polymorphic, abundant, neutral and co-dominant. Restriction fragment length polymorphisms (RFLPs), variable number of tandem repeat (VNTR), randomly amplified polymorphic DNA (RAPD), simple sequence repeat (SSR) and single nucleotide polymorphisms (SNPs) are all suitable genetic markers. Generally, to identify a QTL, we need to search for statistical association between marker genotypes and trait phenotype.

Experimental design

Most commonly, parental inbred populations are crossed to produce the F_1 generation, which is then either backrossed to one or both parental lines (the BC design), or crossed *inter se* to produce the F_2 generation (the F_2 design). In the simplest case (*single-marker analysis*), the mapping population is divided based on the genotype at a marker locus, and the phenotypic values are compared statistically between genotypes at the marker locus. This step is repeated for each marker locus on the map.

If QTL analysis identifies the association between a marker genotypes and phenotype, one can assume that QTL is located near the marker. However, it is

difficult to distinguish the case where the detected effect, which is due to tight linkage between the marker and a QTL with a small effect or loose linkage to QTL with a large effect. In this case *interval mapping* can be used to estimate both the QTL effects and their map position. In interval mapping, the genotype at a large number of positions along the chromosome is predicted based on the observed genotypes of markers flanking an interval and the distance between markers. This makes it possible to make independent estimates of location and effect of the QTL. *Composite interval mapping,* that include additional markers as cofactors, can also be used in order to reduce variation that is associated with other QTL in the genome.

Fine mapping

Linkage analysis utilizes information on recombination events that occur between genetic markers when gametes are formed and transmitted from parents to offspring. The low number of recombination events between closely linked markers and environmental variation restrict the precision of quantitative trait loci (QTL) mapping. In genome scans a QTL can typically be mapped to a chromosomal region of 10–30 cM. To utilize QTL in selective breeding, or to identify genes that define QTL, a higher level of resolution is required.

Three approaches can be applied to refine the map position of a QTL: increase in marker density, increase in crossover density, and improvement of the ability to deduce QTL genotype from phenotype.

Increasing marker density can still be time-consuming in many organisms but is conceptually the simplest remedy. Two options are available to increase the local crossover density: breed recombinants de novo or exploit historical recombination events; i.e., use linkage disequilibrium (LD) in populations. The former approach is generally used with model organisms that have a short generation interval (Darvasi 1998), while the latter is the only practical alternative when working with human or large livestock species. The ability to deduce QTL genotype from phenotype can be improved by using "clones" (e.g., recombinant inbred lines; Darvasi1998), by means of progeny testing (Georges et al. 1995), or by marker-assisted segregation analysis (Riquet et al. 1999).

The isolation of tightly linked markers flanking a QTL is just the initial step to isolate the gene for the QTL. Once such a region is believed to be isolated, it is important to actually confirm that it contains a QTL, for example by using markers to introgress the region into a standardized background and directly testing its effect. If the region has an effect on the character, the next step may be to produce near-isogenic lines (NILs) (Lynch 1998). NILs are constructed by first crossing a donor parent to an inbred line (the recurrent parent) to form an F_1. The resulting offspring are then backcrossed to the recurrent parent for several generations, selection for the QTL each generation.

Linkage disequilibrium mapping

Linkage disequilibrium is the nonrandom allelic association of genetic markers and phenotypes in a population. Disequilibrium can arise from deliberate crosses

of divergent lines, intermixture of populations with different gene frequencies, or by chance in small populations. Disequilibrium can also be produced, and maintained, by selection favouring one combination of alleles over another. Linkage disequilibrium is commonly found in natural populations between closely linked loci for which recombination has not had sufficient time to dissipate the initial disequilibrium.

Linkage disequilibrium mapping uses all recombinations in the population since the mutation occurred, which increases the precision of the estimate of the position. Linkage disequilibrium mapping methods seem, therefore, more useful for precise estimation of QTL positions, while classical mapping is more useful for a genome-wide scans for QTL.

Positional cloning and comparative mapping

If we can localize the position of a QTL to a sufficiently small region of DNA, it becomes possible to examine all of the genes in that region. The gene can be identified by:

1) Sequencing of the entire region.
2) Using sequence data from related species to isolate only those parts of the region of interest that are conserved between species, as these likely represent genes.
3) Exon amplification or exon trapping methods.
4) The timing and pattern of expression of genes localized within a region may indicate likely candidates (Lynch 1998).
5) Map position of homologous genes is partly conserved across species, providing an avenue by which results from one species can be used to identify genes affecting a QTL in a related species. For example, humans and mice show rather extensive conservation of gene order. A marker tightly linked to a QTL in mouse has a good chance of also being tightly linked to the homologous gene in humans (Nadeau 1989). Similar observations have been made for e.g., the mustards Brassica and Arabidopsis (Teutonico et al. 1994), potatoes (Solanum) and tomatoes (Lycopersicon) (Bonierbale et al. 1988).
6) If one or more candidate gene has been identified, the effect of the gene(s) on the trait has to be confirmed. To test the effect of different alleles on a particular trait, a transformation experiment, where different alleles of the candidate gene are put into a specific genotype, can be performed.

Analysis of DNA sequence variation

Analysis of DNA sequence variation between and within species can indicate whether selection has acted on a particular gene. This information can add important evidence to support or refute the hypothesis that variation in the genes is important for adaptation. The neutral theory of molecular evolution provides an essential framework in which functional sequences can be defined and functional changes can be identified. The neutral theory (Kimura 1983) proposes that mutations fall primarily into two fitness classes. A fraction of DNA changes are

strongly deleterious and are quickly eliminated from populations by natural selection. The vast majority of non-deleterious mutations have little or no effect on an organism's fitness (they are redundant with respect to physiology and function) and their evolutionary dynamics are governed solely by genetic drift.

Adaptive theories of molecular evolution allow for a large fraction of mutations with deleterious effects, but, of the remaining mutations, a relatively large proportion confers a fitness advantage to the organism. Under this model, positive selection plays an important role in DNA evolution (Akashi 1999). Since the introduction of the neutral theory, the relative contribution of positively selected and neutral mutations to DNA polymorphism and divergence has been intensely debated. Still, the neutral theory has provided an invaluable theoretical framework in which both neutral and selective models of molecular evolution can be tested.

Testing the neutral hypothesis has also been a major objective of population genetics. In recent years, the focus has shifted towards using the neutral theory as a null model against which specific types of selection may be detected, e.g. positive selection and selective sweeps. If regions that have experienced selection, or that are currently under selection, can be identified, these results could aid in the development of hypothesis about function from sequence data.

Statistical tests for detecting selection (or tests of neutrality) can be divided into two categories, frequency-distribution tests and heterogeneity tests.

Frequency-distribution tests

The *Tajima D-test* (Tajima 1989) is one of the most widely used neutrality test for nucleotide data. The test examines the relationship between two estimates of the population mutation rate, θ, based on the average pairwise differences, π, and the number of segregating sites, S. Tajima's D is defined as :

$$D = \frac{r - S/a_n}{\sqrt{Var(r - S/a_n)}} \qquad a_n = \sum_{i=1}^{n-1} 1/i$$

D will be negative if a sample has an excess of rare variants, and positive if there is an excess of intermediate frequency variants.

Deviation could be due to selection (e.g., selective sweeps, balancing selection), but also demographic events such as bottlenecks. The Fu and Li D test (Fu et al.1993) is based on the same idea but uses sites present only once in a sample (singletons). The test is more sensitive to certain population genetic scenarios, such as selective sweeps.

Tests based on comparing variability in different loci or different classes of mutations – heterogeneity tests

The tests discussed above considered diversity at a single locus, and that all sites were equivalent. An important class of tests use contrast between loci or classes of mutations. An important element in many such tests is the contrast between patterns of polymorphism and divergence.

The Hudson-Kreitman-Aguadé (HKA) test. In the HKA test (Hudson et al. 1987), variability within and between species is compared for two or more loci. In the absence of selection, the expected number of segregating sites within species and the expected number of fixed differences between species are both proportional to the mutation rate, so the ratio of the two should be constant among loci. If the variance among loci is too high, selection is inferred.

McDonald and Kreitman test (MK). The MK test tests for independence between the number of nonsynonymous (amino acid replacement, R) and synonymous (silent, S) polymorphic sites to the number of nonsynonymous and synonymous fixed differences between species (McDonald and Kreitman 1991). If polymorphism and divergence is affected only by mutation and drift, the ration should be the same for both replacement and silent sites.

Tests based on the dN/dS ratio. The ratio ω = dN/dS is usually used as a cumulative measure of selection pressure on protein-encoding homologous genes. $\omega = 1$ (i.e. amino acid replacement substitutions occur at the same rate as synonymous ones) corresponds to neutral evolution with relaxed selection. More often, genes are subjected to strong purifying selection resulting in a $\omega < 1$. The rare case of $\omega > 1$ is indicative of positive selection.

The dN/dS ratio test differs from other test, by providing direct evidence for positive selection if the ratio is significantly above one. However, the test has limitations, as it assumes no recombination. Furthermore, the power could be low if only a few sites are affected by positive selection.

Timing of flowering is an adaptive trait

Clearly, flowering time has a high adaptive value, as reproductive success in plants closely depend on the timing of flower initiation. A correct timing ensures that flowers develop at a time most favourable for fertilization and consequently leads to a sufficient production of seeds. Plants show diverse responses to environmental factors such as daylength and could be classified as short day (SD), long day (LD) or day neutral plants. SD plants (e.g., *Nicotiana tabacum*) are induced to flower when daylength is shorter than a particular duration, called the critical daylength (Cremer et al.2003). Conversly, LD plants (e.g., *Arabidopsis thaliana*) are induced to flower when daylength is longer than the critical daylength.

Many plants species occur over a wide range of latitudes, and within a species, accessions from different regions can differ substantially in flowering behaviour (Michaels et al.2003). Genetic variation for flowering time is important for the adaptation of plant species to different environments, and for the selection of crop plants that meet specific cultivation and consumer needs (Kole et al.2001).

Genetic basis of flowering time control

In this section I will focus mostly on *A. thaliana*. It is an excellent model system, because it responds to many of the environmental conditions that control flowering in other species, and genetic tools are well developed. Studies in Arabidopsis have led to the identification of several components within individual signaling pathways that affect flowering, and to their positioning within molecular hierarchies. Also, genetic analysis of *A. thaliana* varieties showing natural variation in flowering time has demonstrated how the activity of these pathways can be altered in nature and how balancing the effects of different environmental stimuli on flowering time is important in plants adapting to growth in different geographical locations.

Flowering is closely allied to seasonal progression. Naturally occurring Arabidopsis ecotypes have evolved two broadly different flowering strategies: rapid cycling (RC) and winter annual Arabidopsis (Simpson et al.2002).
RC accessions complete their life cycle in as little as 6 weeks and may complete more than one life cycle in a growing season. This is one property that has led to the widespread use of Arabidopsis as a model for plant biology, and the most commonly studied laboratory accessions. *Landsberg erecta* (*Ler*) and *Columbia* (*Col*), are in this class.

However, most naturally occurring Arabidopsis ecotypes are winter annuals. That is, they germinate before winter and flower in the favourable conditions of the following spring. Winter annuals therefore complete only one life cycle in a growing season. Unlike rapid cycling accessions, winter annuals require the long cold treatment of winter (vernalization) to accelerate flowering.

In plants, the transition from vegetative growth to flowering can be stimulated by both environmental and developmental signals. At least four different pathways that regulate the timing of flowering have been identified.

The photoperiod and vernalization pathways are both involved in interpreting environmental cues, while the gibberillin (GA) and autonomous pathways are relatively independent of environmental cues. The photoperiod pathway promotes flowering in response to daylength and the vernalization pathway promotes flowering in response to the prolonged exposure to cold temperature that occur in winter. The autonomous pathways may coordinate flowering with the developmental state (Simpson 2002). The GA pathway mediates the floral promotion effects of GA and is required for flowering in noninductive photoperiod (Blazquez et al.2000).

The identified flowering time pathways, some of the genes, and natural allelic variation, which is important for the adaptation of plant species to different natural and agricultural environments, are discussed below.

Photoperiod pathway

The duration of the daily light period, or photoperiod, is one of the most important factors controlling flowering time in temperate regions. The photoperiod pathway in Arabidopsis comprises three parts: photoreceptors, a circadian clock and an output pathway from the clock specific to flowering (Figure 2).

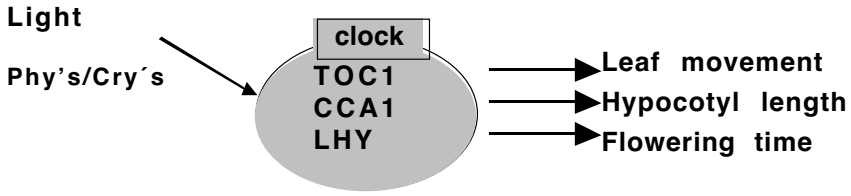


Figure 2. Model of a simple circadian system. Photoreceptors (the phytochromes and the cryptochromes) mediate the input pathway. Interaction among the components of the central oscillator creates the autoregulatory negative-feedback loop that generates the approximately 24-h oscillations. Three different output pathways arise from the central oscillator.

The primary photosensory receptors of higher plants are the red/far-red light receptors called phytochromes and the blue/UV-A light receptors called cryptochromes (Kendrik 1994). Arabidopsis has 5 phytochrome genes, *PHYA* to *PHYE*, which encode the apoprotein of PHYA to PHYE, respectively (Quail et al.1995). Different phytochromes regulate either distinct light responses or similar responses under different light conditions (light quantity, quality, and timing) (Lin 2000). For example, although *PHYA* and *PHYB* both mediate light inhibition of hypocotyl elongation, *PHYA* functions primarily in far-red light whereas *PHYB* acts mainly in red light and plays the most prominent role in shade-avoidance (Halliday et al.2003). The majority of Arabidopsis phytochromes play inhibitory roles in flowering (*PHYA* is an exception, *PHYA* signalling may suppress the biosynthesis of a floral suppressor) (Weller et al.1997).

Cryptochrome genes have been isolated not only from plant species and algae, but also animals including fruit fly, mouse, and human (Cashmore et al. 1999). Arabidopsis has two cryptochrome genes, *CRY1* and *CRY2*. *CRY* genes, encodes a protein associated with a flavin chromophore (FAD) that absorbs blue/UV-A light (Lin et al 1995). The function of *CRY1* in flowering seems complicated, although it may have a promotive effect (Lin et al.1998). *CRY2* promotes flowering since the *cry2* mutant is late flowering in long-day and transgenic plants over expressing *CRY2* are early flowering in short-day. These data also indicate that *CRY2* is involved in sensing photoperiod (Guo et al 1999).

It has been suggested that the different *PHY* and *CRY* photoreceptors show a complex interaction (Neff 1998). Some photoreceptor seems to interact antagonistically (e.g., *CRY2* & *PHYB*), while some show partly redundant function (e.g., *PHYB*, *PHYD* & *PHYE*).

The circadian clock is an endogenous timer that measures the duration of the day or night. This timer also controls daily rhythms of gene expression and patterns of behaviour such as leaf movement. Circadian rhythms, discovered 250 years ago, have a period length (the duration of one cycle) of ≈ 24 hours. Gene transcription, Calcium levels, and some enzyme activities are examples of intracellular processes under circadian clock regulation (Sweeney 1987).
The mechanism that generates these rhythms is often described in three interrelated section:

1) an input pathway that synchronizes the clock mechanism to daily cycles of light and dark,
2) a central oscillator that generates the 24-h time-keeping mechanism,
3) an output pathway that regulates particular processes, such as the control of flowering via *CONSTANS* (*CO*) and *FLOWERING LOCUS T* (*FT*) (Roenneberg et al. 2000).

In Arabidopsis, light signals (through e.g., *CRY1* and *PHYB*) entrain the circadian clock to the daily cycle of night and day. The oscillator generates the 24-hour rhythm, and is proposed to be a negative feedback loop (Samach et al.2000a). In Arabidopsis, like in other species, the expression of clock genes is promoted by "positive elements". As the abundance of clock gene products rises, they repress their own synthesis by antagonizing the function of the "positive elements". The abundance of clock gene products then falls and the function of the "positive elements" is restored. This regulation produces peaks in clock gene expression that are approximately 24 hours apart. The "positive elements" might also regulate the expression of clock-controlled genes that direct overt rhythms or output from the oscillator. In Arabidopsis, these include leaf movements, hypocotyl elongation, the expression of many genes and the photoperiodic control of flowering time (Samach 2000a).

In *Arabidopsis*, three likely components of the oscillator (*TOC1*, *CCA1* and *LHY*) are involved in the photoperiodic induction of flowering. *TOC1* (*TIMING OF CAB EXPRESSION1*) transcript level cycles in light/dark cycles with a peak in the evening and shows a circadian rhythm in continuous light (Strayer et al 2000). *CIRCADIAN CLOCK ASSOCIATED1* (*CCA1*) and *LATE ELONGATED HYPOCOTYL* (*LHY*) genes encode highly similar single Myb domain DNA binding proteins and are each regulated by the circadian clock, with a peak in their expression soon after dawn (Mouradov et al 2002). *TOC1*, *LHY* and *CCA1* may act in a negative feedback loop, in which *LHY* and *CCA1* expression rise in the morning and repress the expression of *TOC1*. According to this model, *LHY* and *CCA1* feed back to repress their own expression, and as their protein levels fall, the expression of *TOC1* rises. *TOC1*, in turn, promotes the expression of *LHY* and *CCA1*, so initiating another cycle (Alabadi et al.2001). This feedback loop may regulate flowering time by determining the time of day that flowering-time genes in the long-day pathway, such as *CONSTANS* (*CO*) and *GIGANTEA* (*GI*), are expressed (Mouradov et al. 2002) (see below and Fig.2).

The connection between the circadian clock and flowering time

Genes that are required for the photoperiodic response of Arabidopsis are circadian-regulated and their phase of expression changes in long days (LDs) to short days (SDs).

Mutations in several genes, e.g., *constans* (*co*), *gigantea* (*gi*), *flowering locus t* (*ft)*, disrupt photoperiodic responses. These mutations delay flowering under LDs but have no, or only slight, effects under SDs, and are considered as genes required to promote flowering in response to LDs (Samach et al. 2000b). A linear pathway has been established through the action of photoreceptors that entrain the circadian clock, the function of which affects *GI (GIGANTEA)*. *GI* controls the expression of *CO*, which is the latest acting of the known genes in the photoperiodic pathway. *CO* in turn activates the floral pathway integrators, *FT* and *SOC1*. This results in the activation of floral identity genes *AP1* and *LFY*, which promotes the floral transition (Fig.3) (see below) (Simpson 2003).

GI encodes a novel nuclear protein (Fowler et al. 1999, Park et al. 1999, Huq et al. 2000). *GI* mRNA expression exhibits a circadian rhythm and gi mutants exhibit defects in clock function (Park et al. 1999), indicating that GI is closely associated with the clock itself (Simpson et al. 2003). *FT* encodes a protein with homology to Raf kinase inhibitor protein (Kardailsky et al. 1999, Kobayashi et al. 1999). It is a powerful promoter of flowering, activating the floral meristem identity gene, *APETALA1* (*AP1*) (Kardailsky et al. 1999, Kobayashi et al. 1999) and is the target of several pathways. (Mouradov et al. 2000).
As *CONSTANS* homologues were identified as candidate genes in the control of flowering time variation in *B. nigra* (see results), *CO* and related genes are describe here in more detail.

The *CO* protein is 373 amino acids long, located in the nucleus and regulates transcription of a diverse set of target genes. *CO* has two domains essential for its function. These are zinc fingers of the B-box class located near to the amino terminus, and a domain of approximately 60 amino acids close to the carboxy-terminus. This combination of domains is plant-specific. The carboxy-terminal motif was dubbed the CCT domain because of its presence in *CO*, *COL1* and *TOC1*. The CCT domain of *CO* is implicated in nuclear localization, but it may also have a role in protein-protein interactions. B-box zinc fingers similar to those present in *COL* genes mediate protein-protein interactions in animal proteins, and this may be their function in *CO* (Robson et al. 2001).

Four early target genes of *CO* were identified using a steroid-inducible version of the protein: *FT* (described above), *SOC1*, a gene encoding a MADS-box transcription factor within the shoot apex that results in activation of floral meristem identity genes such as *LFY*, *ALS10*, which encodes an enzyme involved in ethylene biosynthesis, and *AtP5CS2*, which encodes an enzyme that catalyzes proline biosynthesis (Samach 2000).

Transgenic plants containing increased copy number of *CO* flower earlier than the wild type under long- and short-day conditions, suggesting that the amount of *CO* expression in wild-type plants limits flowering time. Transcriptional regulation of

CO is therefore an important determinant of the photoperiodic regulation of flowering time (Onouchi et al. 2000). Recent studies have shown that regulation of CO protein stability is central to the photoperiodic regulation of flowering (Valverde et al. 2004). Flowering is induced when *CO* mRNA expression coincides with the exposure of plants to light, that only occurs in long day photoperiods. Photoreceptors regulate *CO* stability so that *CO* protein abundance high enough to induce flowering only in long days.

Vernalization pathway

Long exposure to low temperature will often accelerate flowering. Such vernalization requirement prevents flowering before winter and promotes rapid flowering in the spring. The late flowering phenotype of winter-annual accessions of Arabidopsis is created by the interaction of two dominant genes, *FLOWERING LOCUS C* (*FLC*) and *FRIGIDA* (*FRI*) (Michaels et al. 2000). *FLC* encodes a MADS domain-containing transcription factor whose expression is sufficient to inhibit flowering (Sheldon et al. 1999). However, the native *FLC* gene is only expressed to high levels in the presence of *FRI*, which encodes a novel protein (biochemical function is unknown) with two coiled-coil domains that may be involved in protein-protein interaction (Johanson et al. 2000).

FLC expression is regulated by pathways that act antagonistically to either promote or down-regulate its expression (Sheldon et al. 1999). However, in winter-annual accessions, active *FRI* alleles override the activity of the autonomous pathway and promote elevated levels of *FLC* mRNA expression (Michaels et al.1999). This function of *FRI* is antagonized by the vernalization pathway, which acts to down-regulate *FLC* in response to the long cold temperature treatment of winter (Gendall et al.2001).

How cold results in low *FLC* expression is largely unknown. One gene that appear to regulate the response to cold is *HOS1*. *HOS1* regulates the expression of cold-induced genes such as *RD29*, and provides a link between vernalization and cold stress.

Genes involved in the stable maintenance of *FLC* repression have recently been isolated (Gendall et al.2001). *VRN1* encodes a DNA-binding protein and *VRN2* encodes a homologue to Polycomb-group proteins indicating that the repression of *FLC* by vernalization requires chromatin-remodelling factors. Vernalization causes changes in histone methylation within the *FLC* locus. Such changes result in silenced chromatin states in other organisms. Analysis of methylation patterns of *vrn* mutants, revealed that a ”histone code” that specify a silenced chromatin state mediates the memory of winter (Bastow. et al.2004).

Autonomous pathway

This pathway was identified via a group of mutations that are late-flowering under all photoperiods, and are highly responsive to vernalization (Martinez-Zapater et al.1990). These mutants (e.g., *fca*, *fy*, *fpa*, *ld*, *fve*) contain higher levels of *FLC* mRNA than do wild type plants or late-flowering mutants affected in the long-day or GA pathways (Koorneef et al.1998). These mutants delay flowering by increasing expression of FLC. Therefore, this pathway negatively regulates *FLC* expression in wild-type plants.

Several genes in the autonomous pathway have been cloned. *FCA* encodes a protein with two RNA recognition motifs. (RRM)-type RNA binding domains that can bind RNA in vitro (Macknight et al.1997; Quesada et al.2003) and a WW protein-protein interaction domain (Mcknight et al.1997). Interestingly, *FPA* also encodes an RNA binding protein, suggesting a general role for post-transcriptional regulation in the autonomous pathway. *LD* may encode a transcription factor. How these proteins regulate *FLC* expression is not known.

Gibberellin pathway

The growth regulator gibberellin (GA) has been known to induce flowering in many plants species (Bernier 1988). Flowering of Arabidopsis is also promoted by GA. Exogenous treatment of GA accelerates flowering of Arabidopsis particularly under short days (Chandler et al.1994), and strong GA biosynthesis mutants fails to flower under short days and shows a slight delay in flowering under long days (Wilson et al.1992). Thus, GA is absolutely required for flowering under short days in Arabidopsis.

Interaction of GA, photoperiod and vernalization pathways

Even though separate genetic pathways controls flowering in response to different environmental signals, they converge and regulate the expression of the same downstream genes (Fig.3).

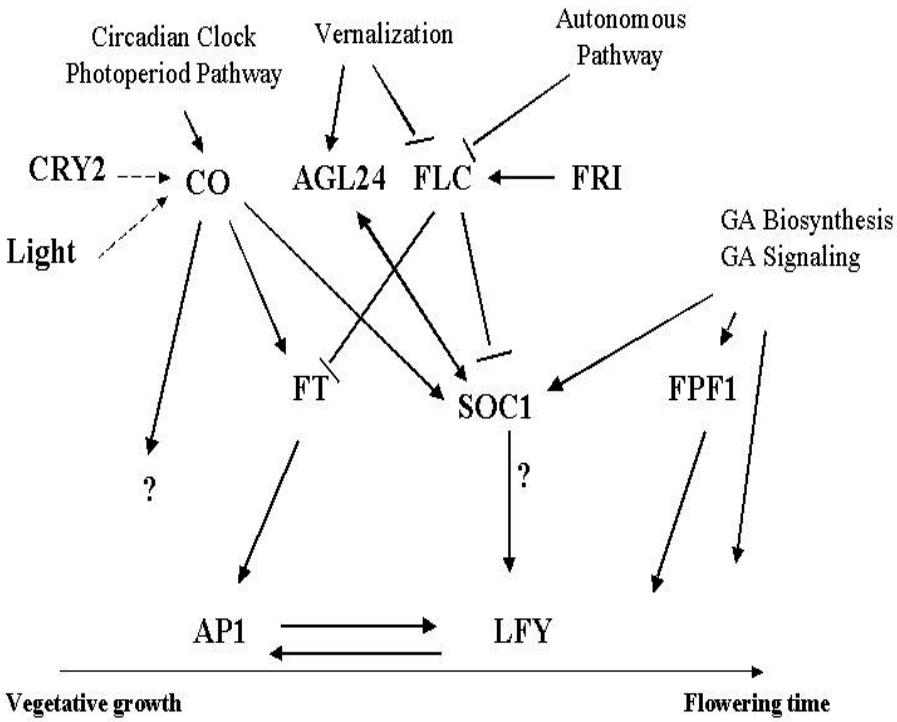


Figure 3. Overview of the relationships among *Arabidopsis* flowering pathways.

LFY is the earliest of the known floral identity genes to be expressed, and it directly activates at least one of the later genes, *AP1* (Wagner et al.1999). The *GA and long-day pathways* converge on the *LFY* promoter where separate promoter elements respond to the GA and photoperiod pathways both inducing *LFY* expression (Blazquez et al.2000) The *autonomous* and *long-day* pathways convergence on *FT* and *SOC1*. *CO* in the photoperiodic pathway, and *FLC* in the autonomous/vernalization pathways act antagonistically in the regulation of *FT* and *SOC1*. *CO* induces the expression of both *FT* and *SOC1* while *FLC* repress both genes. This antagonism represents an important aspect of the adaptation of flowering time to changing seasons. The signals from different environmental stimuli need to be balanced to produce an appropriate response. In winter annuals, the signals that promote flowering in long days need to be prevented in the first year. Molecularly, this adaptation is explained by the action of *FLC* to prevent the activation of *FT* and *SOC1* by *CO*.

FT might activate the expression of floral meristem identity genes such as *AP1* (Ruiz-Garcia et al.1997), while *SOC1* probably acts on *LFY*. The primary activation of *LFY* and *AP1* may occur by parallel pathways, and subsequent direct activation of *AP1* by *LFY* may be required to rapidly amplify floral meristem identity gene expression in young primordia (Mouradov et al.2002).

Genes affecting natural genetic variation in flowering in Arabidopsis

Allelic variation at the *FRI* and *FLC* loci was found to be a major determinants of flowering time in *Arabidopsis* accessions. Analysis of the allelic variation at *FRI* showed that the majority of early-flowering (rapid cycling) accessions carry *FRI* alleles with one of two deletions, both predicted to causes loss-of-function (Johanson et al.2000). This indicates that rapid cycling accessions have evolved independently from winter annual progenitors by losing *FRI* activity, and that *FRI* activity was lost multiple times during the course of evolution. In a study of 25 accessions from western Europe (Le Corre et al.2002) six novel loss-of-function *FRI* alleles and a high degree of polymorphism within exon one (which contains the first of the coiled-coil domains, thought to play a role in protein-protein or protein-nucleic acid interaction (Johanson et al.2000)) were identified (Le Corre et al.2002).

All these facts lead to the view that *FRI* is under strong selection pressure and is a major target for natural selection of flowering time in Arabidopsis. However, some early-flowering accessions do not carry these deletion alleles. The *Shakhdara*, *Kondara* and *KZ-9* accessions carry a functional *FRI* allele with earliness being caused by allelic variation at other loci including *FLC* (Gazzani et al.2003). A novel non-autonomous Mutator-like transposon was found in the weak *FLC* allele in Landsberg erecta, positioned in the first intron, a region required for normal *FLC* regulation. The transposon could affect the processing or stability of the primary *FLC* transcript. However, this transposon was not present in *FLC* alleles of most other accessions including *Shakhdara*, *Kondara* or *KZ-9* In this three cases, other mechanisms have led to the weak *FLC* alleles (Gazzani et al.2003).

Flowering time genes in other species

Plants show diverse flowering responses to environmental cues. To what extent are the results from Arabidopsis applicable to other species? Recent result from rice and other species indicate that many of the genes identified in Arabidopsis are present in other angiosperms, and perform a similar function.

Flowering time in rice is promoted by short days, which can align flowering with the rainy season (Simpson 2003). Even though Arabidopsis and rice show opposite photoperiodic responses, several flowering time genes identified in Arabidopsis seems to have similar functions in rice.
Hd1 encodes a gene homologous to *CO* (Yano et al.2000), and Hd3a encodes a gene highly related to *FT* (Kojima et al.2002). *OsGI* a homologue to *GI* promoted the expression of *Hd1*, consistent with Arabidopsis. However, data suggests that *Hd1* negatively regulate *Hd3a*, the opposite of what happens in Arabidopsis (Simpson 2003). So, while the promotion of flowering in long-day in Arabidopsis results from *CO* activating *FT*, the delay in flowering in long-days in rice results from Hd1 repressing *Hd3a* (Simpson 2003):

Rice Long days⇒*OsGI* ⇒*Hd1*—| *Hd3a* **flowering delayed**
Arabidopsis Long days⇒*GI*⇒*CO*⇒*FT*⇒*AP1* **flowering promoted**

In addition to the *Hd1* repression of flowering in long-days, is seems that *Hd1* also promotes flowering in short-day condition (Yano et al.2000).
A rice QTL affecting photoperiod sensitivity, *Hd6,* encodes an alpha subunit of protein kinase *CK2*. Previous studies in Arabidopsis have implicated a role for Arabidopsis *CK2* in the photoperiod pathway (Takahashi et al.2001).

Diploid Brassica species are hypothesized to be ancient polyploids derived from three copies an *Arabidopsis thaliana* like genome (Lagercrantz 1998). The genome of one of these diploid Brassica species *B. rapa* contains at least four *FLC* homologues. Three of these copies co-segregate with flowering time loci and appear to interact in an additive manner to modulate flowering time (Schranz et al.2002).

The Brassica species

The family *Brassicaceae* (Cruciferae) comprises around 3000 species in 360 genera, organised into 13 tribes. The family's major centers of diversity are southwestern and central Asia and the Mediterranean region. The most economically important species are in the genus *Brassica* (tribe Brassiceae), which includes 6 species that are grown worldwide for a variety of uses: *B. rapa*, *B. nigra*, *B. oleracea*, *B. juncea*, *B. napus* and *B. carinata* (Price et al. 1994) (Fig. 4). Some of them are important sources of edible and industrial oils, vegetables, condiments, fodder and forage. Three of the species, *B. rapa*, *B. nigra* and *B. oleracea* are diploids and the other three species, *B. juncea*, *B. napus* and *B. carinata* are amphidiploids. The amphidiploid Brassica species are thought to contain the whole nuclear genome of both diploid ancestors (Axelsson et al. 2000).

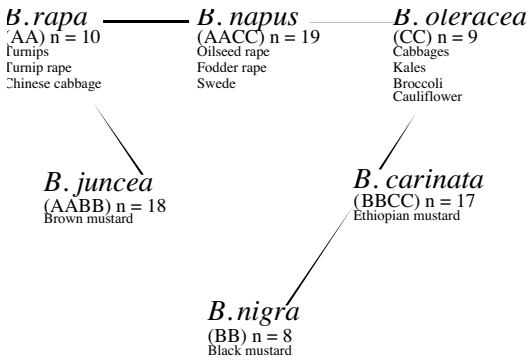


Figure 4. The relationships between the Brassica crop species, arranged according to U, (1935).

Genome structure of Brassica species

The genomes of the diploids *B. nigra, B. oleracea* and *B. rapa* have a recognizably triplicated structure (Figure 5; Lagercrantz and Lydiate 1996).

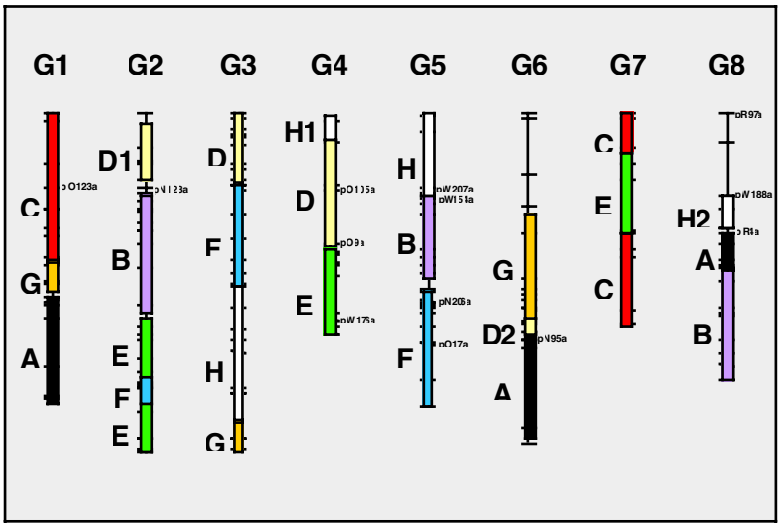


Figure 5. Genetic map of *B.nigra* showing the distribution of the sets of triplicated chromosomal segments. Chromosomal segments with the same colour share common sets of homologus loci. Three copies of the unit genome are rearranged with respect to one another (modified with permission from Lagercrantz and Lydiate 1996).

This triplicated nature of the diploid Brassica genomes suggests that these species have all descend from a hexaploid ancestor. Comparative mapping between *B. nigra* and *A. thaliana* indicated that the *B. nigra* genome contains three copies of a genome similar in complexity and structure to that of *A. thaliana* (Lagercrantz 1998). These data suggests that the lineage leading to Brassica went through a triplication event after the divergence of the lineage leading to Arabidopsis. The polyploidization in the Brassica lineage has probably been followed by an extensive reduction in chromosome number through chromosome fusion events and rearrangements (Fig. 6).

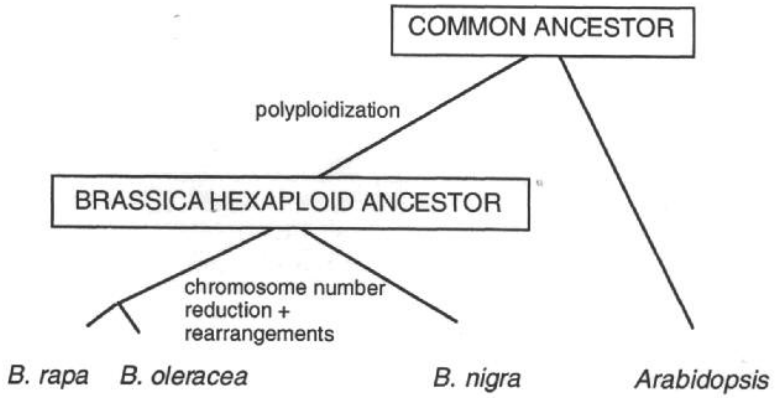


Figure 6. Hypothetical scheme for genome evolution of *Brassica* and *Arabidopsis*.

Comparative maps may provide a conduit for communication, permitting information gathered during study of one species to be quickly and efficiently applied to related species.

The Arabidopsis - Brassica comparative mapping data showed that the Brassica genomes have experienced a large number of chromosomal rearrangements (Lagercrantz 1998). Even though the rearrangements have been frequent since the divergence of Arabidopsis and Brassica, the average length of conserved segments between *A. thaliana* and *B. nigra* was estimated at 8 cM. Thus, mapping a Brassica gene to an interval of less that 10 cM is often likely to allow the identification of the homologous collinear region in *A. thaliana.* Thus, comparative mapping may be one alternative to identify candidate genes for QTL identified in Brassica species utilizing the knowledge gained about gene function in the model species *A. thaliana*.

B. nigra CO genes as candidates for the control of FT variation

Lagercrantz et al. (1996) detected QTL for flowering time in two out of three replicated chromosomal segments in *B. nigra* (Fig. 7). The triplicated segments were homologous to a region on *A. thaliana* chromosome V, where two key genes of the photoperiod and vernalisation pathways, respectively are located (*CO* and FLC). *B. nigra CO* homologues mapped close to both *B. nigra* QTL. These results suggested *B. nigra CO* homologues as possible candidate gene for the flowering time QTL identified in crosses between early and late flowering *B. nigra* genotypes.

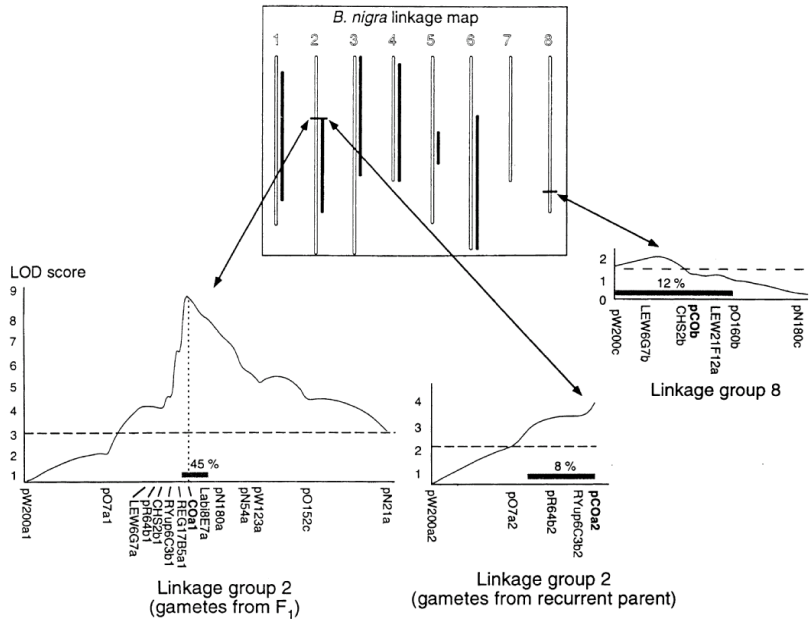


Figure 7. Chromosomal localization of QTL controlling flowering time in *B.nigra*

Aims of the study

The overall aim of the thesis was to identify genes affecting variation in flowering time within different Brassica species, and to study the molecular evolution of such genes.

Specific aims:

1.To study to what extent flowering time in different Brassica species are controlled by homologous genes (paper I).

2.To test identified *CO* homologues as candidates for the control of flowering time variation in *B. nigra* (paper II)

3.To provide information on the molecular evolution of *B. nigra COL1*. Does sequence data reveal any effects of selection on the putative candidate gene for FT variation? (paper III).

4. To investigate if variation in a duplicated *CO* homologue in *B. nigra* (*COb*) affects flowering time (paper IV).

Results and discussion

Multiple flowering time QTLs map to homologous positions within and between several Brassica species (I)

Comparative mapping studies between species have revealed an unexpected conservation of genome structure (Gale et al.1998). However, it is largely unknown to what extent this conservation also reflects a conservation of function of individual genes. Furthermore, if general function is conserved, to what extent is phenotypic variation for a particular trait in two species caused by allelic variation in homologous genes?

Most adaptive traits are controlled by several genes. Thus, for a particular trait, there is more than one gene that could exhibit allelic variation within species or diverge between species. Brassica species also have highly replicated genomes probably as a result of ancient polyploidization events. The resulting duplication of genes could potentially contribute to the extent of multiple QTL for particular traits.

Variation in flowering time is frequent both within and between Brassica species. Previous mapping efforts have identified QTL for parts of this variation in several Brassica species (Lagercrantz et al.1996, Osborn. et al.1997, Bohuon. et al. 1998, Butruille et al.1999, Schranz et al.2002). The various QTL studies have been performed with different mapping populations and genetic markers. Thus, it is difficult to know to what extent QTL in different species might represent the same locus.

As a first attempt to assess to what extent multiple QTL within the extensively replicated Brassica genomes are the result of duplicated genes, and if those QTL

are conserved in related species, we conducted a comparative QTL mapping study in *B. juncea*, *B. nigra*, *B. oleracea* and *B. rapa*. Our approach was to produce mapping populations in each species based on comparable plant materials. To this end, we used rapid-cycling lines as one parent in each cross. Each of these lines is derived from a large number of early flowering genotypes in each species, that were crossed and recurrently selected for early flowering. Artificial selection has probably fixed naturally occurring alleles for early flowering in each species. One of the parents in each mapping cross had previously been used in the production of other maps. All the different maps were also produced with the same set of RFLP probes. These procedures allowed alignment of homologous chromosomal segments based on a comparative map of the A, B and C genomes, present in *B. rapa*, *B. nigra* and *B. olereacea*, respectively (Lagercrantz and Lydiate 1996).

The analysis focused on duplicated chromosomal regions corresponding to a region of *A. thaliana* chromosome five, where at least two potential candidate genes for flowering time variation (*CO* and *FLC*) are located. In each of the diploid species, *B. nigra*, *B. oleracea* and *B. rapa*, three *CO–FLC* regions were identified, and six in the allopolyploid *B. juncea*. This degree of duplication is in agreement with previous RFLP mapping data (Lagercrantz and Lydiate 1996, Lagercrantz 1998).

Major QTL controlling flowering time were detected in two out of three regions in *B. nigra, B. rapa* and *B. oleracea*, and three out of six in *B. juncea*. The identified QTL explained a considerable part of the variation in flowering time in each species, from 21 to 56 %. The results suggest that flowering time in *B. nigra*, *B. oleracea*, *B. rapa* and *B. juncea* may in part be controlled by genes resulting from ancient and more recent polyploidization events. All detected QTL were located close to Brassica homologues to *A. thaliana CO* and *FLC*. The mapping data favoured *CO* over *FLC* as the most likely candidate gene for several of the QTL. Thus, duplicate copies of one ancestral gene, possibly *CO*, could regulate flowering time in several Brassica species. Further studies are needed to test this tentative hypothesis.

Flowering time variation and sequence divergence at the putative flowering time locus COL1 - CO*a* (II and III)

Previous QTL mapping has identified two homologues to *Arabidopsis thaliana CO* as candidates potentially affecting naturally occurring flowering time variation in *B. nigra* (Lagercrantz. 1996).

We cloned alleles of the *B. nigra CO* homologue Bni *COa* at one of the QTL from early and late flowering plants. To test if the two alleles had any differential effect on flowering time, they were transformed into an *A. thaliana co* mutant. These transformation experiments did not show any difference in flowering time between the two alleles. The DNA sequence of the coding region of the two alleles was identical. Therefore, we examined allelic variation around *COa*. About 3.5 kb upstream of *COa* we detected a related gene, Bni *COL1,* that probably is the result of a tandem duplication of *COa*. The corresponding gene pair is also present in *A.*

thaliana, so the duplication probably occurred at least 20-30 million years ago (mya).

The Bni *COL1* gene displayed highly diverged alleles in the early and late flowering plants. The coding region in the two alleles differed by 16 nucleotide substitutions and two in frame insertion/deletion (indels: *Ind1* and *Ind2*). *Ind1* is part of an AAC repeat coding for a run of asparagine residues, and the allele from the early plant contained 6 additional repeats. The 18 nucleotides longer *Ind1* was associated with a deletion of 18 nucleotides at *Ind2* in the «early» allele. The majority of substitutions (9) were non-synonymous, and were concentrated around the indels and the C-terminal end of the protein.

Studies of association between *Ind1* and *Ind2* polymorphisms and flowering time were conducted in natural populations. These studies detected a significant association for *Ind2* in several *B. nigra* populations, indicating that nucleotide variation within or close to Bni *COL1* affects flowering time.

We found that allelic diversity also extended into the non-coding region between Bni *COL1* and Bni *COa*. The intergenic region was highly diverged between alleles from early and late flowering plants (Fig. 8).

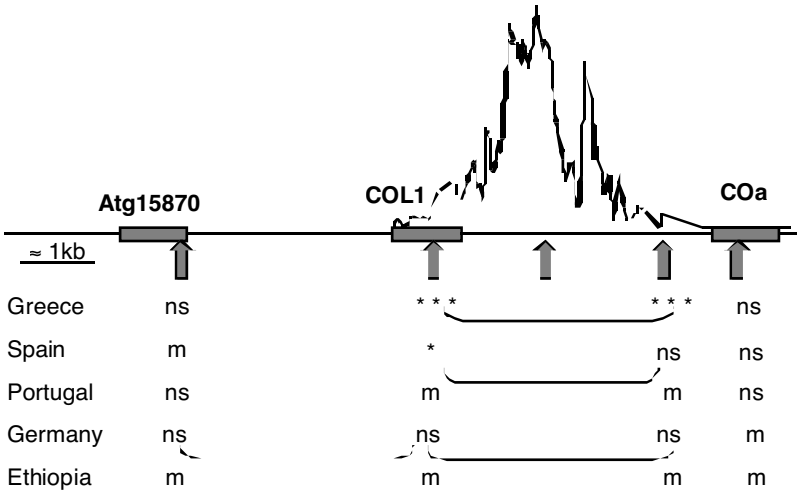


Figure 8. The *B. nigra COL1 – COa* region. The diagram on top illustrates the silent nucleotide diversity between *COL1* and *COa*, and below are results from association studies between markers (arrows) and flowering time in five populations. *: $P < 0.05$, ***: $P < 0.001$, ns: no significant association, m: monomorphic locus. The brackets indicate significant linkage disequilibrium between markers. *GH*, *COL1* and *COa* are *B. nigra* homologues to *A. thaliana* At5g15870, At5g15850, At5g15840, respectively.

To further delimit the nucleotides that affect flowering time we assessed the association between flowering time and genotype at several genetic markers in the *Bni COL1* - *COa* region in natural populations (Fig. 8, unpublished). These data suggest that the flowering time QTN is located upstream of Bni *COa* and downstream of gene *Bi GH*, a *B. nigra* homologue to *At5g15870* (Fig. 8). These data are consistent with our previous transformation experiments, which imply that flowering time QTN is outside the Bni *COa* coding region.

The intergenic region could contain regulatory elements for the downstream Bni *COa* gene. Thus, our data are consistent with a hypothesis where a naturally occurring quantitative trait nucleotide (QTN) for flowering time affects the function or expression of either Bni *COL1* or Bni *COa*.

Preliminary sequencing of *COL1* from closely related Brassica species indicated that the *Ind2* polymorphism is old as expected for a balanced polymorphism. Aligment of a part of the *COL1* gene around the *B. nigra* Ind2 indel *from A. thaliana, B.juncea, B.rapa* showed that the insertion was present in *A. thaliana* but absent in alleles sampled from *B. oleracea*, *B. rapa* and *B. juncea*.

The association between allelic variation at *Ind2* and flowering, and apparently unusual distribution of *Ind2* in the Brassicaceae family prompted us to perform a more detailed study of the evolution of *COL1* in the Brassicaceae family (III). We studied the molecular evolution of both nucleotide substitutions and indels, with particular emphasis on *Ind2*. Forty-two *COL1* sequences corresponding to 27 haplotypes were obtained from 11 Brassicaceae species.

A. lyrata and *B. vulgaris* COL1 sequences share the presence of a single intron with *A. thaliana* COL1, in contrast to other COL1 genes sequences. RT-PCR products of the *B. nigra* COL1 gene showed that the intron is absent from this gene, suggesting that the intron in COL1 was lost in the lineage leading to Brassica, Raphanus and Sinapis.

A total of 24 indels were found in the 600bp of the middle variable region of the gene (Figure 1, III). Such an unusual high frequency could be explained by a high rate of mutations or be the result of positive selection. Currently however, we have no means to determine rigorously whether the high frequency of indels in *COL1* is the result of selection or relaxed constraint. Several small regions of conservation have been identified in the variable middle region of gene in the *CONSTANS LIKE* family (Griffith. et al.2003), indicating that at least parts of the variable region have important functions. Genomic sequence data from closely related plant species could be used to estimate the neutral rate of indel substitution.

Sequence data showed that *Ind2* resulted from a deletion after the Arabidopsis lineage diverged from the lineage leading to Brassica, Raphanus and Sinapis. Data also indicate that the *Ind2* polymorphism present in *B. nigra* has existed more than 8 million years (III). The persistence of a neutral polymorphism for such a long time seems unlikely.

It could be hypothesized that the stark divergence observed in the variable region could be an effect of positive selection resulting in novel protein function. To test for evidence of a ratio of non-synonymous to synonymous substitutions greater

than one, we used the codon substitution models developed by Yang and colleagues (III). The analysis of the different codon substitution models suggests that purifying selection is relaxed in the variable middle region of *COL1*, but that positive selection has not played a major role in shaping the diversity in the *COL1* gene. The results do not exclude that a small number of codons are subjected to divergent selection, but go undetected due to limited power.

In conclusion, we find an association between flowering time and *Ind2* in *B. nigra*, and an unexpected distribution of *Ind2* in related species, both suggesting that selection has affected *Ind2*. However, analysis of the nucleotide sequence of *COL1* shows no signs of selection.

Molecular evolution of a recently duplicated flowering time gene in *Brassica nigra* (IV).

Two homologues of the *A. thaliana CO* gene (Bni *COa* and Bni *COb*) have been identified in *B. nigra*. QTL for flowering time have previously been mapped to both of them (Lagercrantz. 1996), indicating that both genes might still be functional and involved in the control of naturally occurring flowering time variation.

Analysis of DNA sequence variation at *COb*, revealed a number of apparently deleterious mutations that occur at a high frequency in several natural populations of *B. nigra* (IV, Table 1). These mutations include premature stop codons, deletions of apparently important protein domains and frame-shifts resulting from small insertion or deletions (IV, Figure 1). These data suggests that *COb* might be a pseudogene. However, some of the *COb* alleles were free of apparently deleterious mutations, indicating that at lest some alleles could still be functional.

As a flowering time QTL was previous identified close to *COb*, an important question was whether the QTL could be due to segregation of functional and non-functional alleles at *COb.* We tested this hypothesis using a complementation test that we previously used to check *Bni COa* alleles (Österberg, et al.2002). Two apparently functional and two non-functional *COb* alleles were transformed into an *A. thaliana co* mutant. Transformed lines with potentially functional and non-functional alleles showed almost identical flowering times (IV, Table 2). These data therefore does not support the hypothesis that the QTL at *COb* is the result of segregation of functional and non-functional *COb* alleles. Rather, these data indicate that *COb* is a pseudogene.

These results prompted us to check if *COb* was expressed. RT-PCR experiments showed low levels of expression of *COb* both in *B. nigra* and in the transformed *A. thaliana* lines. The low level of expression was similar to the level of the functional Bni *COa* gene. However, a 12 bp deletion in sequenced *COb* promoters removed a putative GATA-1 box identified in functional *CO* genes. GATA-1 boxes are important in the activation of genes by light and by the circadian clock. A combination of light and circadian clock regulation is important for the induction of flowering by *CO* in *A. thaliana* (Mouradov et al. 2001).

To further examine if *COb* is likely to be functional or has degenerated into a pseudogene, we studied nucleotide variation in a large sample of *COb* alleles. The variability at *COb* was high, with 182 segregating sites, but ratio of replacement to silent substitutions (within species as well as between species) was relatively low compared to other *CO* genes indicating that *COb* has not experienced relaxed selective constraint. Frequency distribution tests (Fu and Li D* and F*) were significantly negative, more likely due to population growth rather than selection.

Thus, part of our data strongly suggests that *COb* is a pseudogene whereas other aspects seem difficult to reconcile with the loss of function. The large number of apparently deleterious mutations, the lack of effect in transformation experiments and the deletion in the *COb* promoter of part of a putative GATA-1 box identified in *A. thaliana CO* are the main arguments for the loss of function. On the other hand, *COb* is expressed and sequence variation indicates that the gene is under selective constraint. The apparently conflicting data could partly be explained if *COb* has only recently become a pseudogene.

The large-scale duplications in the *B. nigra* genome are likely an effect of polyploidy, and initially there were probably three copies of most genes in the polyploid ancestor of *B. nigra* (Lagercrantz 1998). However, extensive screening of genomic libraries only identified two paralogues of *A. thaliana CO* (*COa* and *COb*, Lagercrantz and Axelsson 2000), suggesting that a third copy has been deleted or has diverged to an extent that it is no longer recognized in a Southern hybridization experiment. Thus, it seems likely that at least one and possibly two of the three replicated *CO* genes were silenced in a few million years.

Future perspectives

The results in this thesis lend support to an hypothesis where a naturally occurring quantitative trait nucleotide (QTN) for flowering time affects the function of Bni *COL1* or the expression Bni *COa* through variation in the intergenic region between *COL1* and *COa*. Obviously, these hypotheses needs further testing. Two avenues seem particularly important. Association studies where a number of markers in the region around Bni *COL1* and Bni *COa* are assayed for association with flowering time in a larger number of populations, may refine the position of the putative QTN. Such studies should also reveal the extent of linkage disequilibrium in the region, and the thereby mapping resolution that is possible to achieve. These association studies should be complemented with new transformation experiments, with different constructs comprising the *COL1 – COa* region. Initial transformation experiments should include *COL1*, *COa* and the intergenic region, from late and early flowering *B. nigra* plants, respectively. If a difference in flowering time is detected between *A. thaliana co* transformed with the two types of alleles, constructs with different parts of the region from early and late flowering plants should be tested in *A. thaliana*. A potential difference in the expression of *COa* between different alleles could also be tested with detailed gene expression studies. This type of studies is complicated by the complex expression pattern of *CO* genes.

An important question is whether flowering time variation in different Brassica species are controlled by homologous genes, as suggested by our comparative QTL mapping data. The map positions of the different QTL need to be refined. First, backcrossing of individual QTL alleles into a common genetic background may result in monogenic segregation for individual QTL. Screening a large population for recombinants in the QTL region with flanking markers will identify plants with recombination in the region of interest. These plants can then be assayed for a large number of markers within the region.

Even though much of our data indicated that *COb* has become pseudogene, we cannot exclude that *COb* still has a function or that the flowering time QTL close to *COb* could be the result of variation at *COb.* No clear difference between *COb* alleles could be detected in our transformation experiments, but the use of other *COb* alleles or *A. thaliana* genotypes might. The *COb* alleles were introduced into a mutant lacking a functional *CO* gene. It could be envisioned that the *B. nigra COb* alleles could interact differentially with a normal *CO* gene product. The *COb* allele products could for example compete for binding with functional *CO* gene products, and different *COb* allele products could have different affinity for binding. Furthermore, we only introduced two types of *COb* alleles into *A. thaliana*. Perhaps other types of alleles would reveal a differential function. Additional transformation experiment might resolve these questions. A 12 bp deletion in the promoter of all sequenced *COb* haplotypes obliterated a putative GATA-1 box. Deleting the corresponding region from a functional *CO* gene, and testing the effect of this deletion by introducing the modified gene into an *A. thaliana co* mutant, should tell us if this segment is necessary for proper *CO* function.

It would also be very interesting to study DNA sequence variation at *COb* homologues in other related species. Are these homologues functional or non-functional? When did the deleterious mutations occur? Are some of the mutations segregating in *B. nigra* also segregating also in other species? Maybe other *CO* homologues (e.g., those corresponding to *COa* in *B. nigra*) are silenced in other species. We could also study sequence variation at *COa* in *B. nigra*. Are all *COa* alleles functional in *B. nigra*?

The unusual distribution of indels in *B. nigra COL1*, and in particular *Ind2* needs further study. Most work on DNA sequence variation has so far be devoted to studies of nucleotide substitutions and multiallelic microsatellite indels. Relatively little is know about other predominantly diallelic indels. Recent studies have shown that indel formation is probably the most rapid form of mutation in eukaryotic evolution (Weber et al. 2002; Britten et al. 2003). The ratio of unpaired nucleotides attributable to indels to those attributable to substitutions was estimated from around three in Primates up to 51 in *A. thaliana*. One case of positive selection acting on indel substitutions was also recently reported (Podlaha and Zhang 2003). In this context, there are two types of experiment that are of interest to pursue. To evaluate if the high rate of indels observed in *COL1* is the result of a high mutation rates, or could be the result of positive selection, we could estimate the occurrence of indels in nongenic DNA. This could be achieved by comparing noncoding DNA from Brassica species and A. thaliana, and estimate

indel frequency. Some genomic sequencing data is already available for Brassica, and additional Brassica genome sequencing efforts are underway.

We should also study the distribution of *Ind2* in more detail. Our previous studies included a very limited number of individuals in each species. Screening specifically for the *Ind2* polymorphism could be done in large scale using a simple gel based screen on PCR amplified DNA from *COL1*. Such studies should reveal if *Ind2* is polymorphic also in other species besides *B. nigra*, and if this polymorphism is associated with flowering time. We might also get a better picture of the evolution of *Ind2* in particular, and indels in general.

References

Akashi, H. 1999. Within- and between-species DNA sequence variation and the `footprint' of natural selection. *Gene.* Volume 238, Issue 1 , 30 September, Pages 39-51.

Alabadi, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Mas, P., Kay, S.A. 2001. Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science* 3; 293(5531):880-3.

Axelsson, T., Bowman, C.M., Sharpe, A.G., Lydiate, D.J., Lagercrantz, U. 2000. Amphidiploid Brassica juncea contains conserved progenitor genomes. *Genome.* 43(4):679-88.

Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., Dean, C. 2004. Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature.* Jan 8;427(6970):164-7.

Benson, S.B. 1933 *Univ. Calif. Publ. Zool*. 40, 1-70.

Bernier, G. 1988. The control of floweral evocation and morphogenesis. *Annu. Rev.Plant Physiol*. 39: 175-219.

Blanc, G., Barakat, A., Guyot, R., Cooke, R., Delseny, M. 2000. Extensive duplication and reshuffling in the Arabidopsis genome. *Plant Cell*. Jul;12(7):1093-101.

Blanc, G., Hokamp, K., Wolfe K.H. 2003. A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. *Genome Res*. Feb;13(2):137-44.

Blazquez, M.A., Weigel, D.2000. Integration of floral inductive signals in Arabidopsis. *Nature* 20;404(6780):889-92.

Bohuon, E.J., Ramsay, L.D., Craft, J.A., Arthur, A.E., Marshall, D.F., Lydiate, D.J., Kearsey, M.J. 1998.The association of flowering time quantitative trait loci with duplicated regions and candidate loci in Brassica oleracea. *Genetics*. Sep;150(1):393-401.

Bonierbale, M.W., Plaisted, R.L., and Tanksley, S.D. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120: 1095-1103.

Bradshaw, H.D. and Schemske D.W. 2003. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. Nature. Vol. 426: 176-178.

Butruille, D.V., Guries, R.P., Osborn, T.C. 1999. Linkage analysis of molecular markers and quantitative trait loci in populations of inbredbackcross lines of Brassica napus L. *Genetics*. Oct;153(2):949-64.

Cashmore, A.R., Jarillo, J.A., Wu, Y.J., Liu, D.1999. Cryptochromes: blue light receptors for plants and animals. *Science.* 30; 284(5415):760-5. Review.

Chandler, J., Dean, C. 1994. Factors influencing the vernalization response and flowering mutants of *Arabidopsis thaliana* (*L.*) *Heynh. J.Exp.Bot*. 45: 1279-88.

Cremer, F. and Coupland, G., 2003. Distinct photoperiodic responses are conferred by the same genetic pathway in *Arabidopsis* and in rice. *TRENDS in plant Science*. Vol.8 No.9: 405-407.

Darvasi, A.1998. Experimental strategies for the genetic dissection of complex traits in animal models. *Nat Genet*. Jan;18(1):19-24.
Dice, L.R.. and Blossom, P.M. 1937. *Carnegie Inst. Wash. Publ.* 485, 1-129.
Doebley, J., Stec A., Wendel J., and Edwards M., 1990. Genetic and morphological analysis of maize-teosinte F2 population: implications for the origin of maize. *Proc. Natl.Acad.Sci.USA* 87: 9888-9892.
Falconer, D.S. and Mackay, T.F.C. 1996. Introduction to quantitative genetics. Fourth edition.
Fisher, R.A, 1930. The genetical theory of natural selection. *Oxford Univ. Press, Oxford*.
Fisher, R.A.,1935. The sheltering of lethals. *Am.Nat.* 69:446-55.
Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531-45.
Fowler, S., Lee, K., Onouchi, H., Samach, A., Richardson, K., Morris, B., 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 J*. Sep 1;18(17):4679-88.
Fu, Y.-X. and Li, W.-H. 1993 Statistical tests of neutrality of mutations. *Genetics* 133: 693-709.
Gale, M.D., Devos, K.M.1998. Plant comparative genetics after 10 years. *Science.* Oct 23;282(5389):656-9.
Gazzani, S., Gendall, A.R., Lister, C., Dean, C. 2003. Analysis of the molecular basis of flowering time variation in Arabidopsis accessions. *Plant Physiol.* Jun;132(2):1107-14. Epub May 22.
Gendall, A.R., Levy, Y.Y., Wilson, A., Dean, C. 2001. The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. *Cell*. Nov 16;107(4):525-35.
Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto,R., Pasquino, A.T., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X., et al. 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing.*Genetics*. Feb;139(2):907-20.
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.
Guo, H., Duong, H., Ma, N., Lin, C. 1999. The Arabidopsis blue light receptor cryptochrome 2 is a nuclear protein regulated by a blue light-dependent post-transcriptional mechanism. *Plant J.* 19(3):279-87.
Halliday, K.J., Salter, M.G., Thingnaes, E., Whitelam, G.C. 2003. Phytochrome control of flowering is temperature sensitive and correlates with expression of the floral integrator FT. *Plant J.* 2003 Mar;33(5):875-85.
Hudson, R.R. Kreitman M., and Aguade, M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116: 152-159.
Hughes, A.L. 1994. The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B Biol Sci* 256, 119-24.
Huq, E., Teppermanm, J.M., Quail, P.H. 2000. GIGANTEA is a nuclear protein involved in phytochrome signaling in Arabidopsis. *Proc Natl Acad Sci U S A.* Aug 15;97(17):9789-94.
Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., Dean, C. 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science.* Oct 13;290(5490):344-7.
Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., Weigel, D. 1999. Activation tagging of the floral inducer FT. *Science*. Dec 3;286(5446):1962-5.
Kendrik, R.E., Kronenberg, G.H.M., 1994. *Photomorphogenesis in Plants*, Ed.2 Kluwer Academic Publishers, Dordrecht, The Netherland.
Kimura, M. 1983. *The neutral theary of molecular evolution*. Cambridge Iniversity Press, Cambridge, U.K.

Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., Araki, T.1999. A pair of related genes with antagonistic roles in mediating flowering signals. *Science.* Dec 3; 286(5446):1960-2.

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T., Yano, M. 2002. Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant Cell Physiol*. Oct;43(10):1096-105.

Kole, C., P.Quijada, S.D., Michaels, S.D., Amasino, R.M., Osborn, T.C. 2001. Evidence for homology of flowering-time genes *VFR2* from *Brassica rapa* and *FLC* from *Arabidopsis thaliana*. *Theor Appl Genet* 102:425-430.

Koorneef, M., Alonso-Blanco, C., Peeters, A.J.M., and Soppe, M. 1998. Genetic control of flowering time in Arabidopsis. *Annu.Plant Physio.Plant Mol.Biol.*49, 345-370.

Kopp, A., Graze, R.M., Xu, S., Carroll, S.B., Nuzhdin, S.V. 2003. Quantitative trait loci responsible for variation in sexually dimorphic traits in *Drosophila melanogaster*. *Genetics.* Feb;163(2):771-87.

Lagercrantz, U., Lydiate, D.J., 1996 Comparative genome mapping in Brassica. *Genetics.* Dec;144(4):1903-10.

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

Lagercrantz, U.1998. Comparative mapping between Arabidopsis thaliana and Brassica nigra indicates that Brassicagenomes have evolved through extensive genome replication accompanied by chromosome and frequent rearrangements. *Genetics*. Nov;150(3):1217-28.

Lagercrantz U, Axelsson T.2000. Rapid evolution of the family of CONSTANS LIKE genes in plants. *Mol Biol Evol.* Oct;17(10):1499-507.

Lai, C., Lyman, R.F., Long, A.D., Langley, C.H. and Mackay, T.F. 1994. Naturally occurring variation in bristle number and DNA polymorphism at the scabrous öocus of Drosophila melanogaster. *Science.* Dec.9; 266 (5191): 1697-702.

Le Corre, V., Roux, F., Reboud, X. 2002. DNA polymorphism at the FRIGIDA gene in Arabidopsis thaliana: extensive non-synonymous variation is consistent with local selection for flowering time. *Mol Biol Evol.* Aug;19(8):1261-71.

Lin, C. 2000. Photoreceptors and regulation of flowering time. *Plant Physiology*. Vol. 123, pp.39-50.

Lin, C., Robertson, D.E., Ahmad, M., Raibekas, A.A., Jorns, M.S., Dutton, P.L., Cashmore, A.R. 1995. Association of flavin adenine dinucleotide with the Arabidopsis blue light receptor CRY1. *Science.* 18; 269(5226):968-70.

Lin, C., Yang, H., Guo, H., Mockler, T., Chen, J., Cashmore, A.R. 1998. Enhancement of blue-light sensitivity of Arabidopsis seedlings by a blue light receptor cryptochrome 2. *Proc Natl Acad Sci U S A*. 3; 95(5):2686-90.

Long, M. and Langley, C.H. 1993. Natural selection and the origin of jingwei, a chimeric processed functional gene in Drosophila. *Science.* Apr 2;260(5104):91-5.

Lynch, M., and B. Walsh, 1998 Genetics and Analysis of Quantitative Traits. Sinauer, Sunderland, MA.

Lynch, M. and Conery, J.S 2000. The evolutionary Fate and Consequences of Duplicate Genes. *Science* Vol.290 1151-1155.

Lynch, M. and Walsh, B.1998. *Genetics and Analysis of Quantitative traits.*

Mackay, T.F. 1995. The genetic basis of quantitative variation: number of sensory bristles of Drosophila melanogaster as a model system. *Trends Genetics*. Dec.11 (12): 464-470.

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*. May 30;89(5):737-45.

Martinez-Zapater, J.M., and Somerville, C.R. 1990. Effect of light quality and vernalization on late-flowering mutants of Arabidopsis thaliana. *Plant Physiology*. 92, 770-776.

McDonald, J.H., and Kreitman, M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila. Nature* 351: 652-654.
Michaels, S.D., Amasino, R.M. 1999. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell.* May;11(5):949-56.
Michaels, S.D., and Amazino, R.M. 2000. Memories of winter: Vernalization and the competence to flower. *Plant Cell Environ.* 23, 1145-1153.
Michaels, S.D., He, Y., Scortecci, K.C., Amasino, R.M. 2003. Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. *Proc Natl Acad Sci U S A*. Aug 19;100(17):10102-7.
Mira, A., Ochman, H., Moran, N.A. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet*. 17(10):589-96.
Mounsey, A., Bauer, P., and Hope, I.A., 2002. Evidence Suggesting That a Fifth of Annotated Caenorhabditis elegans Genes May Be Pseudogenes. *Genome research*. Vol. 12, Issue 5: 770-775.
Mouradov, A., Cremer, F., Coupland, G. 2002. Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell*. 14 Suppl:S111-30. Review. No abstract available
Nachman, M.W., Hoekstra, H.E. and D'Agostino, S.L. 2002. The genetic basis of adaptive melanism in pocket mice. *Proc. Natl.Acad.Sci.USA* Vol. 100 No. 9: 5268-5273.
Nadeau, J.H.1989. Maps of linkage and synteny homologies between mouse and man. *Trends Genet*. 5:82-86.
Neff, M.M., Chory, J. 1998. Genetic interactions between phytochrome A, phytochrome B, and cryptochrome 1 during Arabidopsis development. Plant Physiol. Sep;118(1):27-35.
Nurminsky, D.I., Nurminskaya, M.V., De Aguiar, D., Hartl, D.L. 1998. Selective sweep of a newly evolved sperm-specific gene in Drosophila. *Nature*. Dec 10;396(6711):572-5.
Ohno ,S. 1970. Evolution by Gene Duplication. *New York: Springer-Verlag.*
Onouchi, H., Igeno, M.I., Perilleux, C., Graves, K., Coupland, G. 2000. Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among Arabidopsis flowering-time genes. *Plant Cell*. Jun;12(6):885-900.
Orr ,A. 1998 The population Genetics of Adaptation: The Distribution of Factors Fixed during Adaptive Evolution. *Evolution.* Vol. 52 No 4: 935-949.
Osborn, T.C., Kole, C., Parkin, I.A., Sharpe, A.G., Kuiper, M., Lydiate, D.J., Trick, M. 1997. Comparison of flowering time genes in Brassica rapa, B. napus and Arabidopsis thaliana. *Genetics*. Jul;146(3):1123-9.
Osterberg, M.K., 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*. May;161(1):299-306.
Otto, S. 2000. Polyploid Incidence and Evolution. *Annu.Rev.Genet*. 34:401-37.
Otto, S.P. and Yong, P. 2002. The evolution of gene duplicates. *Adv Genet*. 46:451-83.
Park, D.H., Somers, D.E., Kim, Y.S., Choy, Y.H., Lim, H.K., Soh, M.S., Kim, H.J., Kay, S.A., Nam, H.G. 1999. Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. *Science.* Sep 3;285(5433):1579-82.
Price, R., Palmer, J. and AL-Shehbaz, I.1994 Systematic relationships of Arabidopsis: a molecular and morphological perspective, pp. 7–19 in Arabidopsis, edited by Meyerowitz, E. and Somerville, C. *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, NY.
Quail, P.H., Boylan, M.T., Parks, B.M., Short, T.W., Xu, Y., Wagner, D. 1995. Phytochromes: photosensory perception and signal transduction. *Science.* May 5;268(5211):675-80. Review.
Quesada, V, Macknight, R, Dean, C, Simpson GG. 2003 Autoregulation of FCA pre-mRNA processing controls Arabidopsis flowering time. *EMBO J.* Jun 16; 22(12):3142-52.
Ramsey, J., Schemske, D.W. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol*. Syst.29: 467-501.

Riquet, J., Coppieters, W., Cambisano, N., Arranz, J.J., Berzi,. P., Davis, S.K., Grisart, B., Farnir, F., Karim, L., Mni, M., Simon, P., Taylor, J.F., Vanmanshoven, P., Wagenaar, D., Womack, J.E., Georges, M. 1999. Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. *Proc Natl Acad Sci USA.* Aug 3;96(16):9252-7.

Robin, G.C., Russell, R.J., Cutler, D.J., Oakeshott ,J.G. 2000. The evolution of an alpha-esterase pseudogene inactivated in the Drosophila melanogaster lineage. *Mol. Biol Evol.* Apr;17(4):563-75.

Robson, F., Costa, M.M., Hepworth, S.R., Vizir, I., Pineiro, M., Reeves, P.H. Putterill, J., Coupland, G.. 2001. Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic . *Plant J.* Dec;28(6):619-31.

Roenneberg,T., Merrow, M. 2000 Circadian clocks: Omnes viae Romam ducunt. *Curr Biol.* 19;10(20):R742-5. Review.

Ruiz-Garcia, L., Madueno, F., Wilkinson, M., Haughn, G., Salinas, J., Martinez-Zapater, J.M. 1997. Different roles of flowering-time genes in the activation of floral initiation genes in Arabidopsis. *Plant Cell*. Nov;9(11):1921-34.

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

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., Coupland, G. 2000b. Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. *Science*. Jun 2; 288(5471):1613-6.

Schranz, M.E., Quijada, P., Sung, S.B., Lukens, L., Amasino, R., Osborn, T.C. 2002. Characterization and effects of the replicated flowering time gene FLC in Brassica rapa. *Genetics*. Nov;162(3):1457-68.

Simpson, G.G., Dean, C. 2002. Arabidopsis, the Rosetta stone of flowering time? *Science* 12;296(5566):285-9.

Simpson, G.G. 2003. Evolution of flowering in response to day length: flipping the CONSTANS switch. *Bioessays*. 25(9):829-32. Review.

Sheldon, C.C., Burn, J.E., Perez, P.P., Metzger, J., Edwards, J.A., Peacock, W.J., Dennis, E.S.. 1999. The FLF MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization andmethylation. *Plant Cell*. Mar;11(3):445-58.

Soltis, D.E., Soltis, P.S., 1999. Polyploidy: reccurent formation and genome evolution. *Trends Ecol. Ev*ol. 14:348-52.

Stebbins, G.L. 1938. Cytological characteristics associated with the different Groth habits in the dicotyledons. *Am.J.Bot*.25:189-98.

Stebbins, G.L.1980. Polyploidy in plants: unsolved problems and prospects. *New York: Plenum*, pp.495-520.

Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Mas, P., Panda, S., Kreps, J.A., Kay, S.A. 2000. Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog. *Science* 4; 289(5480):768-71.

Sweeney, B.M. 1987. Rhythmic Phenomena in Plants. *Academic Press, San Diego*.

Tajima , F.1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.

Takahashi,Y., Shomura, A., Sasaki, T., Yano, M. 2001. Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit ofprotein kinase CK2. *Proc Natl Acad Sci U S A*. 2001 Jul 3;98(14):7922-7. Epub Jun 19.

Teutinico, R.A., and Osborn, T.C. 1994. Mapping of RFLP and qualitative trait loci in *Brassica rapa* and comparison to the linkage maps of *B.napus, B. oleracea*, and *Arabidopsis thaliana*. *Theor.Appl.Genet*. 89: 885-894.

Thompson J.D. and Lumaret, R. 1992. The evolutionary dynamics of polyploid plants: origins, establishment and persistence. *Trends Ecol.Evol*. 7:302-7

U.N. 1935. Genome analysis in Brassica with specieal refernce to the experimental formation of B.napus and peculiar mode of fertalization, *Japan J.Bot*.7: 389-452.

Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., Coupland G. 2004. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science*. 2004 Feb 13;303(5660):1003-6.

Voss, S.R. and Shaffer, H.B. 1997. Adaptive evolution via a major gene effect: Paedomorphosis in the Mexican axolotl. *Proc. Natl.Acad.Sci.USA*. Vol. 94, pp.14185-14189.

Wagner, D., Sablowski, R.W., Meyerowitz, E.M. 1999. Transcriptional activation of APETALA1 by LEAFY. *Science.* Jul 23;285(5427):582-4.

Walsh ,J.B. 1995. How often do duplicated genes evolve new functions? *Genetics* 139: 421-428.

Watterson, G.A. 1983. On the time for gene silencing at duplicate loci. *Genetics* 105: 745-766.

Weller, J.L., Murfet, I.C., Reid, J.B. 1997. Pea mutants with reduced sensitivity to far-red light define an important role for phytochrome A in day-length detection. *Plant physiology* 114: 1225-1236.

Wendel ,J.F. 2000. Genome evolution in polyploids. *Plant Mol Biol*. Jan;42(1):225-49.

Wilson, R.N., Heckman, J.W., Somerville, C.R. 1992. Gibberellin is requared for flowering but not for senescence in Arabidopsis thaliana under short days. *Plant Physiol*. 100:403-8.

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

Acknowledgements

First and foremost, I should like to offer my sincere gratitude and deep admiration to my supervisor, Ulf Lagercrantz, for all the pleasure his work has brought me and for all the beauty he has captured for me to enjoy. I am also very grateful to Ulf for spending a great deal of time reviewing my thesis and for his valuable input.

Tomas Axelsson deserves a big thank you for the many opportunities he sent my way in the very beginning of my study as a PhD student. Tomas, such wonderful care you took of me during my first months in the lab!

Dear Maj-Britt, heartfelt thanks go out to you! You provided invaluable assistance, support and encouragement in my Swedish life!!!

I should like to acknowledge Harold for his generosity and helpful discussions.

I very much appreciate Dr.Martin Lascoux for his comments and suggestions for improving my thesis and manuscripts. Special thanks are reserved for Martin's group: Marita, Per!

Sincere thanks to Professor Kristina G., and the head of the department Per B. for the warm atmosphere at the department.

Let me thank Urban G., Anki R., Berit S. for the unforgettable teaching experience.

I wish to thank Ingrid E., Birgitta E., Astrid A. for always taking the time to answer questions, provide contact information.

Many thanks to Gunilla S., Gun R., Per Linden and Bertil Schelander for the helping with my transformation experiments and monitoring of the growth conditions.

I express my gratitude to all people who helped me during my PhD study. I should be remiss, if I did not thank all of the folks at GC, who are a great bunch of people to work with!

It was charming just to be with my Russian and foreign friends: Zhenya S., Anja G., Lada F., Petya B., Masha, Vlad, Igor I., Volodya S., Volodya B., Nadiya K., Sasha P., Andrei Z., Marina Tsh., Richard R., Mitya U., Nadya Ya., Lena M., Anton P., Kristina K., Mari B., Rita, Leticia, Vasilios, Rachel, Gosia, Natalie, Malin.

I am very thankful to Ira, Henrik, Sergei (b), Sergei (m) for lending me inspiration.

I should like to thank my parents for their love, and for the grace with which they raised and educated me!

I deeply appreciate the Nillson-Ehle foundation for its financial support.

.