# Genetic control of flowering in the diploid wild strawberry *Fragaria vesca*

Elli Koskela
Kasvinjalostus
Soveltavan biologian laitos
Maatalous–metsätieteellinen tiedekunta
Helsingin yliopisto
Toukokuu 2009

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

| Tiedekunta/Osasto — Fakultet/Sektion — Faculty | Laitos — Institution — Department |
|---|---|
| Faculty of Agriculture and Forestry | Department of Applied Biology |

Tekijä — Författare — Author
Elli Koskela

Työn nimi — Arbetets titel — Title
Genetic control of flowering in the diploid strawberry *Fragaria vesca*

Oppiaine —Läroämne — Subject
Plant breeding

| Työn laji — Arbetets art — Level | Aika — Datum — Month and year | Sivumäärä — Sidoantal — Number of pages |
|---|---|---|
| Pro gradu | 05/09 | 48+ 3 pages appendices |

Tiivistelmä – Referat –Abstract

Strawberry (*Fragaria* × *ananassa*) is the most important berry crop cultivated in Finland. Due to the species' economic importance, there is a national breeding programme aimed at extending the cropping season from the current one month to up to three months. This could be achieved by growing cultivars which would initiate flowers throughout the summer months, without the requirement of a period of short days as is the case with currently grown cultivars.

The cultivated strawberry is an octoploid and therefore has complex patterns of inheritance. It is desirable to study the genetic mechanisms of flowering in the closely related but diploid species *F. vesca* (L). In the diploid *Fragaria*, a mutation in a single locus, namely the *SEASONAL FLOWERING LOCUS* (*Sfl)*, changes the flowering phenotype from seasonal to perpetual flowering. There is also an array of genetic tools available for *F. vesca*, which facilitate genetic studies at molecular level.

Experiments described here aimed at elucidating the identity of the gene which confers perpetual flowering in *F. vesca* by exploring the flowering characteristics and genotypes of five $F_2$ populations (crosses between seasonal × perpetual flowering cultivars). The study took advantage of a genetic map for diploid *Fragaria*, publicly available EST and genomic *Fragaria* sequences and a recently developed BAC library. Sequence information was used for designing gene–specific primers for a host of flowering–related candidate genes, which were subsequently mapped on the diploid *Fragaria* genetic map. BAC library was screened with molecular markers supposedly located close to the *Sfl*, with the aim of positionally cloning the *Sfl*.

Segregation of flowering phenotypes in the five $F_2$ populations showed, that the *Sfl* indeed controls flowering in all the tested cultivars. A genetic map was constructed of the chromosome with the *Sfl*, and a positional cloning attempt was initiated with the closest flanking markers. 45 gene–specific primers pairs were designed for 21 flowering–related genes, and eight genes were successfully mapped on the diploid *Fragaria* map. One of the mapped genes, namely *PRR7*, located very close to the *Sfl*, and is a potential candidate for the gene that has evaded identification so far.

Avainsanat — Nyckelord — Keywords
*Fragaria*, genetic control, flowering, *SFL*, mapping, BAC library, candidate genes

Säilytyspaikka — Förvaringsställe — Where deposited
Viikki Science Library

Muita tietoja — Övriga uppgifter — Further information
Supervisors MMT Timo Hytönen (Dept. of Applied Biology) & Dr. Daniel J Sargent (East Malling Research)

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

| Tiedekunta/Osasto — Fakultet/Sektion — Faculty | Laitos — Institution — Department |
|---|---|
| Maatalous – metsätieteellinen tiedekunta | Soveltavan biologian laitos |

Tekijä — Författare — Author
Elli Koskela

Työn nimi — Arbetets titel — Title
Genetic control of flowering in the diploid strawberry *Fragaria vesca*

Oppiaine —Läroämne — Subject
Kasvinjalostus

| Työn laji — Arbetets art — Level | Aika — Datum — Month and year | Sivumäärä — Sidoantal — Number of pages |
|---|---|---|
| Pro gradu | 05/09 | 48 + 3 sivua liitteet |

Tiivistelmä – Referat –Abstract
Mansikka (*Fragaria* × *ananassa*) on tärkein Suomessa viljelty marjakasvi. Nykyisin mansikan satokausi kestää vain noin kuukauden. Satokautta pidentämällä viljelijän olisi mahdollista tasata tuotantohuippuja ja saada sesongin ulkopuolisesta tuotteesta parempaa hintaa. Mansikalle onkin perustettu jalostusohjelma, joka tähtää jatkuvasatoisten, Suomen oloihin soveltuvien lajikkeiden jalostamiseen. Jatkuvasatoiset lajikkeet pystyvät muodostamaan kukka–aiheita läpi kesän, toisin kuin nykysin viljellyt lajikkeet, jotka tarvitsevat kukka–aiheiden muodostamiseen lyhyenpäivän oloja.

Viljelty mansikka on oktoploidi, jonka periytyminen on monimutkaista. Tämän takia on kannattavaa tutkia kukkimisen geneettisiä mekanismeja lähisukuisessa diploidissa lajissa, ahomansikassa (*F. vesca*). Ahomansikassa yhden lokuksen, niin kutsutun *SEASONAL FLOWERING*–lokuksen (*Sfl*), mutaatio saa aikaan jatkuvan kukkimisen. Ahomansikkaa on tutkittu paljon ja lajille on kehitetty monia geneettisiä työkaluja, jotka mahdollistavat molekyylitason geneettisten tutkimusten teon.

Tässä työssä pyrittiin selvittämään *Sfl:*ssa sijaitsevan geenin identitetti tutkimalla viiden ahomansikan $F_2$–populaation kukintaa ja genotyyppejä. Tutkimus hyödynsi diploidille mansikalle kehitettyä geneettistä karttaa, julkisia EST– ja genomisia mansikan sekvenssejä sekä äskettäin kehitettyä BAC–kirjastoa. Sekvenssitietoja käytettiin tunnetuille kukintageeneille spesifisten alukeparien suunnitteluun. BAC–kirjastoa seulomalla yritettiin löytää klooni(t), jotka sisältäisivät *Sfl:*n.

Kukintafenotyyppien segregaatio viidessä $F_2$–populaatiossa osoitti, että *Sfl* kontrolloi kukintaa kaikissa kokeessa mukana olleissa lajikkeissa. Kromosomille, jossa *Sfl* sijaitsee, kehitettiin geneettinen kartta. *Sfl:*n lähelle sijoittuvien molekyylimerkkien avulla aloitettiin BAC–kirjaston seulonta, jonka tarkoituksena oli *Sfl:*n kloonaaminen. 21 kukintageenille suunniteltiin yhteensä 45 geenispesifistä alukeparia. Kokeessa onnistuttiin sijoittamaan kahdeksan kukintageeniä geneettiselle kartalle. Yksi kartalle sijoitetuista geeneistä, *PRR7*, on hyvin lähellä *Sfl:*a, ja on hyvä kandidaatti ahomansikan keskeiseksi kukintageeniksi.

Avainsanat — Nyckelord — Keywords
*Fragaria*, genetiikka, kukkiminen, jalostus, *Sfl*, molekyylimerkki, kandidaattigeeni

Säilytyspaikka — Förvaringsställe — Where deposited
Viikin tiedekirjasto

Muita tietoja — Övriga uppgifter — Further information
Työn ohjaaja MMT Timo Hytönen (SBL) & Dr. Daniel J Sargent (East Malling Research)

# Table of Contents

List of abbreviations....5
1 Introduction....6
1.1 Background of the study....6
1.2 Photoperiodism in F. vesca....7
1.3 How to identify a gene?....9
1.3.1 Positional cloning....9
1.3.2 Candidate gene approach....12
2 Research aims....15
3 Materials & Methods....16
3.1 Determining the phenotype and genotype of F2 populations....16
3.1.1 Plant materials, growing conditions and phenotyping....16
3.1.2 DNA extraction....18
3.1.3 Primers & PCR conditions....19
3.2 Linkage map construction and BAC library screening....20
3.2.1 Constructing linkage maps of the linkage group six....20
3.2.2 BAC library screening....20
3.3 Designing and mapping gene–specific primer pairs....22
4 Results....25
4.1 Determining the phenotype and genotype of F2 populations....25
4.2 Linkage mapping and BAC library screening....27
4.3 Designing and mapping gene–specific primer pairs....29
5 Discussion....32
5.1 Single–gene control of flowering in F. vesca....32
5.2 Chromosome walking....34
5.3 Candidate gene approach....36
6 Conclusions....39
Acknowledgements....39
References....40
Appendices....47
Appendix 1. Markers used for genotyping and mapping the F2 populations....47
Appendix 2. An example of how to identify a positive colony. ....48
Appendix 3. Bin mapping; the first six sequences represent the bin set seedlings, the 7th is the F. vesca parent and the last the heterozygous F1....49

## *List of abbreviations*

| | |
|---|---|
| *AGL14* | *AGAMOUS-LIKE 14* |
| BAC | bacterial artificial chromosome |
| bp | base pairs |
| cM | centiMorgan |
| *CCA1* | *CIRCADIAN CLOCK ASSOCIATED 1* |
| *CO* | *CONSTANS* |
| *COP1* | *CONSTITUTIVE PHOTOMORPHOGENIC 1* |
| *CRY1* | *CRYPTOCHROME 1* |
| CTAB | cetyl trimethylammonium bromide |
| dNTP | deoxy-ribonucleotidetriphosphate |
| EDTA | ethylene diamine tetraacetic acid |
| *ELF3* | *EARLY FLOWERING 3* |
| *EMF2* | *EMBRYONIC FLOWER 2* |
| EST | expressed sequence tag |
| EtBr | ethidium bromide |
| *FLC* | *FLOWERING LOCUS C* |
| *FraPRR7* | *Fragaria PRR7* |
| *FT* | *FLOWERING LOCUS T* |
| *GA2OX* | *GIBBERELLIN-2-OXIDASE* |
| *GI* | *GIGANTEA* |
| GDR | Genome Database for Rosaceae |
| INDELs | insertions or deletions |
| LB | Luria Broth medium |
| LD | long day |
| LG | linkage group |
| *LHY* | *LATE ELONGATED HYPOCOTYL* |
| LOD | logarithm of odds |
| kb | kilobase |
| M | million |
| Mb | million base pairs |
| NCGR | The National Resource for Genome Resources |
| PCR | polymerase chain reaction |
| *PHYB* | *PHYTOCHROME B* |
| *PRR5* | *PSEUDO-RESPONSE REGULATOR 5* |
| *PRR7* | *PSEUDO-RESPONSE REGULATOR 7* |
| PVP | polyvinylpyrrolidone |
| rpm | rounds per minute |
| SCAR | sequence charaterised amplified region |
| SD | short day |
| *Sfl* | *Seasonal flowering locus* |
| SNPs | single nucleotide polymorphisms |
| SSR | simple sequence repeat |
| *TOC1* | *TIMING OF CAB EXPRESSION1* |
| *VIP3* | *VERNALISATION INDEPENDENCE 3* |
| *VRN1* | *VERNALISATION 1* |

# *1 Introduction*

## 1.1 Background of the study

Strawberries (*Fragaria*) are amongst the most economically significant berry crops. The annual production volume worldwide is nearly 4 Mt, while the yearly production volume in Finland amounts to 10 000 tons (FAO 2009) with an estimated value of 40 M euros (MMM 2007, Puutarhaliitto ry 2009). Due to the species' economic importance, many countries have national breeding programmes aimed at improving quality characteristics, disease resistance and yield. In Finland, interest lies in extending the production period from the current one month to up to three months. This goal could be achieved by growing cultivars which initiate flowers throughout the summer months, without the requirement of a period of short–days as is the case with currently grown cultivars.

Breeding of the cultivated strawberry (*Fragaria* × *ananassa*) has been hindered by the species' octoploid nature. Investigations aimed at elucidating the inheritance of the perpetual flowering trait have yielded notably conflicting results. Ahmadi *et al.* (1990) made several crosses between day neutral and June–bearing cultivars and used an extensive seedling population of some 28 000 individuals. They reached the conclusion that perpetual flowering in strawberries was controlled by a single dominant allele which is inherited in Mendelian fashion. More recent studies have, however, indicated that while it is probable that there is a major dominant gene conferring perpetual flowering in *F.* × *ananassa*, it is likely to be modified by a number of minor genes (Shaw & Famula 2005). In addition to complex patterns of inheritance, breeding for perpetual flowering is complicated by the strong effect the environment has on flowering characteristics. It is therefore conceivable that characterizing the gene(s) conferring perpetual flowering, and subsequently developing markers associated with the gene(s) could greatly benefit the breeding work for perpetual flowering in *F.* × *ananassa*.

Given the complexity of inheritance in *F.* × *ananassa*, it is desirable to study the genetic mechanisms of trait(s) of interest in the closely related

but diploid species *F. vesca* (L). In *F. vesca* the inheritance of perpetual flowering seems to follow Mendelian segregation ratios of a single recessive gene (Brown & Wareing 1965, Sargent *et al.* 2004). In addition to the simpler patterns of inheritance of a diploid species, *F. vesca* possesses also other characteristics of a good model species; it has a short generation span, a genome size of 206 Mb per haploid genome (Folta & Davis 2006), which is comparable to that of *Arabidopsis* and there are a number of genetically divergent accessions readily available from gene banks. Furthermore, there is already a genetic map available for the species with almost 350 markers covering a map distance of ca. 570 cM (Sargent *et al.* 2008), as well as a bacterial artificial chromosome (BAC) library (Bonet *et al.* unpublished), both of which are necessary tools for the map–based cloning of genes. Utilising these resources, it should be possible to identify the gene conferring perpetual flowering habit in *F. vesca* and then to design gene–specific markers to permit marker–assisted selection in *F.* × *ananassa,* in which the everbearing habit appears to be a dominant trait (Ahmadi *et al.* 1990, Shaw & Famula 2005).

## 1.2 Photoperiodism in *F. vesca*

Growth and development of the wild type *F. vesca* are strongly controlled by both photoperiod and temperature, and can be divided into two distinctive phases, as described by Brown & Wareing (1965). During the long, warm days of summer, the plant is in a vegetative growth phase and the axillary buds develop mostly into stolons. The plant enters generative phase at the end of summer as the days get shorter and temperatures decrease. Runner production ceases, and axillary buds develop into lateral crowns. Flowers are initiated during short days in autumn, as the terminal apices of the main and lateral crowns produce inflorescences until growth and development cease for the winter. Growth is resumed in spring and the inflorescences complete their development, followed by flowering.

The perpetually flowering *F. vesca* f. s*emperflorens* differs notably from the wild type *F. vesca* ssp. *vesca* in terms of growth and development. The *semperflorens* type remains generative throughout the growing period;

the apex develops into an inflorescence after producing a few nodes and vegetative growth is resumed by the topmost axillary bud below the inflorescence. (Brown & Wareing 1965).

Since the classic study by Brown & Wareing (1965), the consensus has been that wild type *F. vesca* is a temperature–dependent short day (SD) plant, flowering only when the day length is shorter than a critical photoperiod, whereas *F. vesca* f. *semperflorens* is a true day neutral, flowering continuously independent of temperature or day length. However, the recent studies by Heide & Sønsteby (2007) and Sønsteby & Heide (2008) have shown that while there indeed is a strong interaction between temperature and day length in the control of flowering in *F. vesca* ssp. *vesca*, an interaction is found also in *F. vesca* f. s*emperflorens*. According to Heide & Sønsteby (2007), flower induction in *F. vesca* ssp. v*esca* takes place in both SD and LD at 9 ºC, whereas at 21 ºC no flowering was observed under any day length regime. Within the temperature range of 15 and 18 ºC and with increasing temperature, shorter day lengths were required to induce flowering. In the case of *F. vesca* f. *semperflorens*, flowering is promoted by long days across a wider temperature range, from 9 to 27 ºC, the requirement of LD being almost obligatory at 27 ºC. It is especially notable that 'Elan', a perpetually flowering octoploid cultivar included in the trial, had nearly identical responses to day length and temperature with the two *F. vesca* f. *semperflorens* cultivars (Sønsteby & Heide 2008). Despite the contradictory modes of inheritance of the everbearing trait in the diploid *F. vesca* and the octoploid *F.* × *ananassa*, the results of Sønsteby & Heide (2008) suggest that the genetics controlling flowering in diploid and octoploid strawberries may be similar, and also grant further support for using *F. vesca* as a model for studying flowering in the genus.

The locus which induces the change from a SD plant to a quantitative LD plant in *F. vesca* was first characterised by Brown & Wareing (1965), who denoted the locus *semperflorens* (*s*), but which was later termed *SEASONAL FLOWERING LOCUS* (*Sfl*) by Albani *et al.* (2001). The individuals carrying the dominant allele (*Sfl/–*) are induced to flower only under short days, whereas the plants homozygous for the recessive allele (*sfl/sfl*) flower perpetually. Also more recent investigations (e.g. Sargent *et*

*al.* 2004) have shown that in $F_2$ populations, the inheritance of perpetual flowering in *F. vesca* follows approximately the expected Mendelian segregation ratio of 3:1 of a single locus. However, there is some evidence to indicate that flowering may be controlled by different genes in divergent *F. vesca* accessions. Ahmadi *et al.* (1990) crossed European perpetually flowering *F. vesca* f. s*emperflorens* with a native Californian *F. vesca* accession, and found segregation ratios that closely fitted a three–gene model of inheritance. This result may be explained by segregation distortion rather than a multiple gene model of inheritance. According to Li *et al.* (2007), segregation distortion is quite a common phenomenon especially in interpspecific crosses, and it has been reported to occur often in both inter– and intraspecific crosses of diploid *Fragaria* species (Davis & Yu 1997, Sargent *et al.* 2004).

Considering the economic importance of strawberries, the seemingly simple inheritance of perpetual flowering in *F. vesca* and the status of *F. vesca* as a model species, it is rather surprising that no genes working on the flowering pathway in *Fragaria* have so far been characterised at the molecular level. However, as there is a vast array of genetic tools available for *F. vesca,* it seems to be only a matter of time before the genes behind strawberry flowering are elucidated.

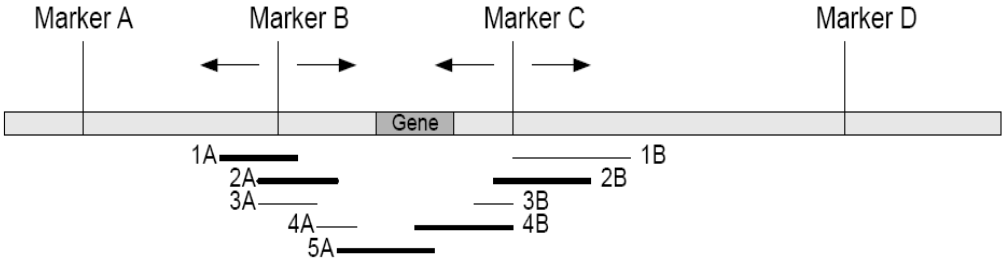
## 1.3 How to identify a gene?

### 1.3.1 Positional cloning

Positional cloning, or chromosome walking is a method for identifying gene(s) by the combined use of a mapping population and a large–insert genomic library such as BAC library. The first step in positional cloning is to determine flanking markers for the gene of interest, after which the genomic library is screened for positive clones. The positive BAC clones are end–sequenced and the sequence information is used for designing new primers. These primers are employed in the second round of screening. Successive screens identify further clones, which can essentially extend to four directions from the initial flanking markers. After a sufficient number

of “screening rounds”, the positive BAC clones can be sequenced and the sequences assembled into a single contig which spans the entire genomic region in between the initial flanking markers. (Figure 1).

*Figure 1. Five successive rounds of BAC screening starting from the initial flanking markers B and C. Fingerprinting the clones has indicated, that the genomic area can be covered by sequencing the clones 1A, 2A, 5A, 4B and 2B.*




This should result in sequencing the desired gene, usually alongside with many other genes. (Tanksley *et al.* 1995, Bortiri *et al.* 2006). It should be noted that if the library has been created from a mutant form instead of the wild type, the gene of interest may elude all identification efforts. This is especially true if the mutant form has arisen from a deletion in the gene.

The number of successive screens required for covering the area between the flanking markers obviously depends on the mutual proximity of the markers. Fortunately, the genome of *F. vesca* is small compared to that of many other cultivated plant species, and there is a genetic map available with an average distance of ~1.3 cM between markers. The reference map is based on an interspecific cross between *F. vesca* and *F. nubicola* (later on referred to as FV × FN population), spanning 568.8 cM over seven linkage groups and comprising at present almost 350 molecular markers, most of which are transferable SSRs. (Sargent *et al.* 2008, Vilanova *et al.* 2008). Moreover, finding the closest flanking markers for a gene causing an observable phenotype such as seasonal vs. perpetual flowering does not present a problem, as the location of the gene can be determined by mapping the phenotype in relation to a suitable set of markers. In the case of *Sfl*, a set of three SCAR markers have been reported by Albani *et al.* (2004) to lie in the proximity of the *Sfl* in the linkage group six (LG6), one

of these markers being inseparable from the *Sfl*. Unfortunately, these markers have not been mapped in the reference population, and therefore it is difficult to estimate whether they indeed are the closest flanking markers to the *Sfl*. At present, there are 37 markers on the LG6 with an average distance of ~1.9 cM between markers (Sargent *et al.* 2007). It should be possible to find markers located close enough to the *Sfl* for initiating a chromosome walk.

Coverage of the BAC library, as well as the insert size affect the usefulness of the library. Positional cloning can be seriously compromised by presence of chimeric clones, i.e. clones that contain DNA from more than one genomic region, or by high frequency of repetitive DNA (Tanksley *et al.* 1995). The library should also be checked for and cleared of non–recombinant or cross–contaminated clones by, for example, employing a restriction–based BAC fingerprinting method (Han *et al.* 2007). According to Cai *et al.* (1995), clones with insert size much smaller than 100 kb are not desirable for positional cloning. Recently, a BAC library of the *F. vesca* cultivar 'Ali–Baba' has been characterized by Bonet *et al.* (unpublished). In the investigation, the library was estimated to cover ca. 8.8 x the *Fragaria* genome with an average insert size of 98 kb, and the occurrence of non–recombinant clones was assessed to be low.

Positive clones can be identified either by arraying the library onto filter and employing Southern hybridization protocol (e.g. Kaufmann *et al.* 2003) or by pooling the library in multiple dimensions and screening the pools by PCR (e.g. Xu *et al.* 1998). The former protocol is rather expensive and time–consuming and requires the use of radiolabelled isotopes. Moreover, hybridization tends to produce a notable number of false positives if the probes contain long tandem repeats, as is often the case with SSR markers (Bonet *et al.* unpublished). PCR–based screening of a pooled library is a rapid and cost–effective alternative, and has been proven a reliable method in a number of studies (Xu *et al.* 1998, Han *et al.* 2007). The 'Ali–Baba' BAC library of *F. vesca* consists of 18,432 clones, in forty–eight 384–well plates. After initially condensing each 384–well plate into a 96–well plate, the library was pooled in three dimensions, creating 48 plate pools, 48 column pools and 32 row pools. For prescreening the library, eight

superpools were created, each containing all clones for six 384–well plates. The entire BAC library of 18,432 clones was thus condensed into 80 BAC pools and eight superpools, which could be screened in just 38 PCRs. (Bonet *et al.* unpublished).

Although positional cloning is a robust and straight–forward method for identifying candidate genes and numerous attempts to positionally clone genes have been made, there are only few reports that have applied the approach successfully in other species than the model plants. Undoubtedly, the high cost of initiating a positional cloning attempt from a scratch including the creation of a mapping population and construction of a large–insert genomic library, has limited the number of plant species in which the approach can be employed. Well–established protocols for map–based cloning exist only for a few model species, such as *Arabidopsis* and rice, in which it is possible to locate a gene within a genetic map distance of 200–800 kb within 1–2 weeks (Peters *et al.* 2003). However, research in this field is intense within the Rosaceae scientific community, and physical maps have been recently developed for *Malus* (Han *et al.* 2007), *Prunus* (Zhebantyayeva *et al.* 2008) and *Rosa* (Kaufmann *et al.* 2003).

### 1.3.2 Candidate gene approach

The candidate gene approach aims to apply the knowledge gained from studies on model species to identifying genes controlling agriculturally interesting traits in cultivated crop species. For example, ESTs from the target species can be assigned a putative function based on their homology with *Arabidopsis* genes, given that the gene function in *Arabidopsis* is already known. *Arabidopsis* genomic sequence data can also be used for predicting intron–exon junctions within EST sequences of the target species, and this information can be further utilized for designing gene–specific primers. This approach has been used in eg. *Prunus* (Georgi *et al.* 2003) and *Malus* (Sargent *et al.* 2009).

There are five pathways in *Arabidopsis* that can initiate flowering either independently or by integration with each other (Izawa *et al.* 2003). In short, the five pathways are the gibberellin, photoperiod, light quality, vernalisation and autonomous pathways (Figure 2).

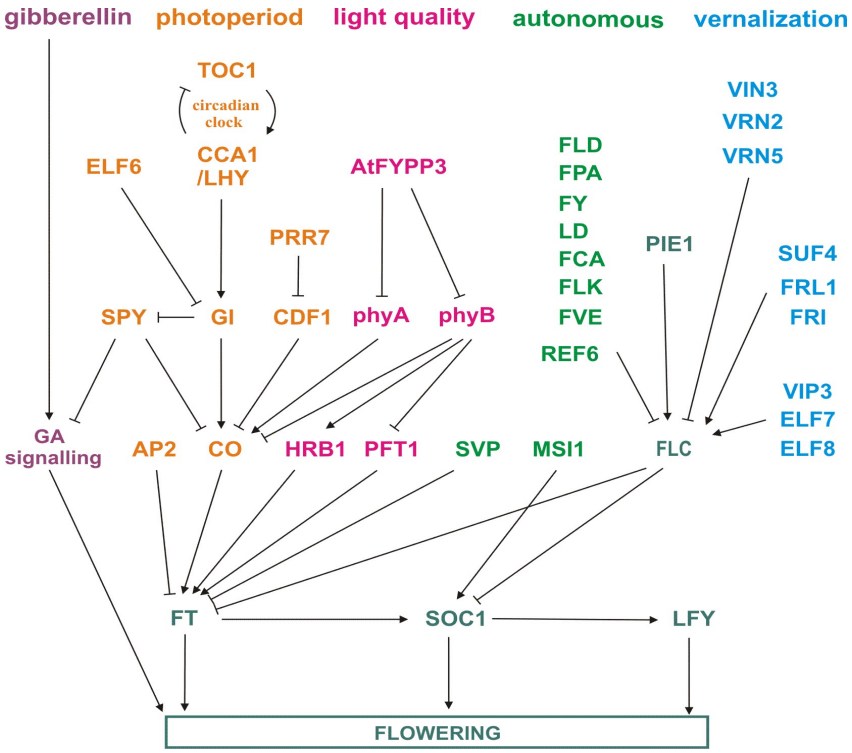



Figure 2. Flowering–related genes of the pathways controlling flowering in *Arabidopsis* (Mouhu *et al.* manuscript in preparation).

Both the autonomous and the vernalisation pathways activate flowering by repressing *FLOWERING LOCUS C* (*FLC*), a transcription factor that strongly represses flowering (Simpson *et al.* 1999). Light and light quality are perceived mainly by phytochromes and cryptochromes (Mockler *et al.* 2003). The photoreceptors control the expression of *CONSTANS* (*CO*) through the circadian clock and by affecting the stability of CO protein. The circadian clock is a complex time–keeping device which contains several interlocking feedback loops (reviewed in e.g. Gardner *et al.* 2006). Some of the genes associated with control of flowering in *Arabidopsis* are presented in Figure 2. The gibberellic acid pathway in *Arabidopsis* promotes flowering even under noninductive short days (Taiz & Zeiger 2006). Mutations of gene(s) on any of these pathways have been shown to result in altered flowering patterns in *Arabidopsis* (e.g. Gardner et al. 2006; Thomas & Sun 2004, Michaels & Amasino 2000). The genetic research conducted with *Arabidopsis* has identified dozens of flowering time–related genes

(McClung 2001) that can be utilised in flowering gene studies in less–well characterized species.

The candidate gene approach is based on the assumption that a mutation within the candidate gene sequence, or change in the gene's expression can be correlated to an observable phenotype. Mutations, such as single nucleotide polymorphisms (SNPs) or polymorphisms caused by insertions or deletions (INDELs), can be scored in a suitable mapping progeny. Evidence for the candidate gene being associated with the expression of the trait is obtained if the phenotype and the candidate gene co–segregate in the mapping population. (Silva *et al.* 2005).

A suitable mapping population is obviously one where the trait of interest has been scored and mapped as accurately as possible. On the other hand, only those candidate genes that map into the same linkage group as the trait need to be fine–mapped. For initial mapping, i.e. assigning candidate genes into linkage groups, it is not plausible to perform costly mapping in the entire mapping population. Instead, a more cost–effective approach, namely bin mapping, can be adopted. A bin–set is a set of selected seedlings which represent a maximum number of genotypes in a mapping population (Vision *et al.* 2000). Recently, a bin–set of six seedlings representing 46 bins was selected from the reference FV × FN population by Sargent *et al.* (2008). Using this bin set, it is possible to genotype and map candidate genes with SNPs or INDELs at a greatly reduced cost.

The success of the candidate gene approach to trait characterisation depends on the choice of the candidate genes. Although function conservation of the central circadian oscillator genes such as *GI*, *CCA1*, *TOC1* and *ELF*3 has been observed between *Arabidopsis* and rice (Murakami *et al.* 2006), it has been shown that the photoperiodic pathway–related genes may have additional and/or different functions in divergent species. According to Böhlenius *et al.* (2006), the *CO/FT* regulatory module of aspen controls not only photoperiodic flowering, but also short day–induced growth cessation and bud set. In apple, Hättasch *et al.* (2008) discovered no clear expression patterns of the apple homologue of *CONSTANS*. Thus, it is also possible that the gene controlling

seasonal/perpetual flowering in *F. vesca* is not present in the *Arabidopsis* genome, or has a different function in that organism. Even with this limitation, the method offers a solid foundation for initiating studies on the genetic control of the flowering pathway in species where these issues have not previously been characterised to a great extent.

## *2 Research aims*

This Master's thesis aimed at elucidating the gene conferring the change from seasonal to perpetual flowering in *F. vesca*. The work was divided into three sub–goals, namely 1) finding out whether the same gene is responsible for perpetual flowering in *F. vesca* accessions of different origins by determining the phenotype and genotype of several $F_2$ crossing populations; 2) identifying markers flanking the *Sfl* and employing these markers for initiating an attempt to positionally clone the *Sfl*; and 3) designing gene–specific primers for a group of flowering–related genes in *F. vesca* and finding out whether one of these genes would co–segregate with the *Sfl*.

# *3 Materials & Methods*

## 3.1 Determining the phenotype and genotype of $F_2$ populations

### 3.1.1 Plant materials, growing conditions and phenotyping

The trial included five perpetually flowering *F. vesca* f. *semperflorens* cultivars and one seasonally flowering Finnish genotype of *F. vesca*. The exact origins of some of these varieties are unknown, but the collection should represent *F. vesca* accessions with different flowering characteristics from Europe and USA (Table 1). Seeds of cultivars and the information on the origins of the *F. vesca* f. *semperflorens* cultivars were obtained from NCGR–Corvallis (http://www.ars–grin.gov/cor/catalogs/fraeverbear.html).

**Table 1. *Fragaria* accessions included in the study, their characteristics and geographical origins.**

| Plant material | Variety / accession | NCGR accession number | Flowering | Runnering | Origin |
|---|---|---|---|---|---|
| *F. vesca* f. *semperflorens* | Alexandria | PI 551826 | perpetual | no | old European Alpine cultivar |
| *F. vesca* f. *semperflorens* | Baron Solemacher | PI 551507 | perpetual | no | old German Alpine cultivar |
| *F. vesca* f. *semperflorens* | Hawaii–4 | PI 551572 | perpetual | yes | collected in Hawaii, but not native |
| *F. vesca* f. *semperflorens* | Piikkiö | – | perpetual | no | unknown |
| *F. vesca* f. *semperflorens* | Yellow Wonder | PI 551827 | perpetual | no | US, probably a seedling selection of *F. vesca* |
| *F. vesca* ssp. *vesca* | – | PI 551792 | seasonal | yes | collected in Punkaharju, Finland |

The $F_1$ populations were created by crossing *F. vesca* ssp. *vesca* with all the perpetually flowering cultivars, producing five cross combinations (Table 2).

**Table 2. The cross combinations and their abbreviations.**

| Cross ♀ / ♂ | Referred to as |
|---|---|
| *F. vesca* × 'Alexandria' | FV × AL |
| *F. vesca* × 'Baron Solemacher' | FV × BS |
| *F. vesca* × 'Hawaii – 4' | FV × H4 |
| *F. vesca* × 'Piikkiö' | FV × PI |
| 'Yellow Wonder' × *F. vesca* | YW × FV |

For observing the flowering trait in the $F_1$ generations, 14–23 plants per cross (and 18 *F. vesca* plants as a control) were planted in 8 x 8 cm pots in commercial potting medium (Kekkilä, Tuusula, Finland). The plants were grown under LD conditions for 12 weeks, after which half of the plants of each cross combinations were moved to SD conditions while the other half was kept under long day. After 4 weeks the SD treated plants were moved to LD conditions, and flowering was observed for 20 weeks. As the SD treated plants started flowering, four plants of each cross were isolated for production of $F_2$ seeds. The selected plants were placed into fabric–covered cubes in an otherwise empty greenhouse compartment. Before the isolation, all the open flowers and developing berries were removed. After self pollination had resulted in new berries, seeds were collected by slicing a mature strawberry into two halves, pressing the halves onto tissue paper and letting the seeds dry for a week.

$F_2$ seeds were sown at the end of May 2008 in LD. In mid–June 2008, 110 plants from each cross (except for the cross FV × AL ,which consisted of only 94 plants) were transplanted into 8 x 8 cm pots. Flowering was observed and recorded twice a week for six weeks, and thereafter twice a month until the end of the experiment (November 2008). Phenotypic segregation ratios were analysed statistically with $\chi^2$ goodness–of–fit test by Microsoft Excel (Microsoft Corp.).

All the plants were fertigated by ebb–and–flow with Marja–Superex (Kekkilä, Tuusula, Finland). The relative humidity was set to 65 %. The LD and the SD treatments had day lengths of 18 and 12 hours, respectively. Lighting was provided by natural daylight and 400 W SON–T lamps

(Airam, Kerava, Finland). The set day/night temperatures were 20/18 ºC for both treatments.

### 3.1.2 DNA extraction

DNA was extracted from young unexpanded leaves. A modified Doyle & Doyle (1990) CTAB protocol was used, scaled down to fit in a 1.5 ml Eppendorf tube. A leaf sample was placed in a 1.5 ml Eppendorf tube, which was dipped in liquid nitrogen. The sample was ground into fine powder by a small pestle. CTAB (Table 3) was heated to 60 ºC and β–mercaptoethanol to concentration of 0.2 % was added. 0.5 ml of CTAB was added to the sample, the sample was stirred shortly and incubated at 65 ºC for 30 minutes. The tube was stirred once during the incubation.

**Table 3. The reagents and their concentrations in CTAB buffer.**

| Concentration in the CTAB buffer | Concentration of the reagent (M) | Measured amount |
|---|---|---|
| 2 % CTAB | | 5.0 g |
| 1.4 M NaCl | 5 M | 70 ml |
| 20 mM EDTA | 0.5 M | 10 ml |
| 100 mM Tris pH8 | 1 M | 25 ml |
| 1.0 % PVP–K40 | | 2.5 g |
| | | ad. 250 ml |

After the incubation, 500 µl chloroform:isoamyl alcohol (24:1 v/v) was added, the sample was stirred gently and centrifuged (14 000 rpm) for five minutes. The upper aqueous phase was pipetted into a fresh Eppendorf tube and the chloroform:isoamyl alcohol extraction was repeated. The upper aqueous phase was pipetted into a fresh tube and the DNA was precipitated with cold (–20 ºC) isopropanol (2/3 of the sample volume). The sample was centrifuged at 4 ºC (15 000 rpm) for 10 minutes. The pellet was washed with 1 ml 70 % ethanol and centrifuged at 4 ºC (15 000 rpm) for 10 minutes. The pellet was dried and resuspended into 50 µl TE buffer. After measuring the absorbance (A260/A280 nm) with GeneQuant 1300 spectrophotometer (GE Healthcare, Chalfont St. Giles, UK), 0.2 µl RNAse (10 mg/ml) was added to each sample. Before running PCR, the samples were diluted to the concentration of approximately 4 ng/µl.

### 3.1.3 Primers and PCR conditions

The initial genotyping was performed with markers SCAR1–3 originally developed by Albani *et al*. (2001) with the forward primers labelled with 6–FAM fluorescent dye. SCAR primers were ordered from Oligomer Oy (Helsinki, Finland). Per each population, 46 or 48 individual plants were genotyped. Half of the genotyped plants represented the seasonal flowering phenotype, and half were perpetually flowering.

The PCR reactions for SCAR markers were performed in 20 µl reaction volume. The reaction mixtures contained 1 ng/µl template DNA, 1 x HF buffer (Finnzymes Oy, Finland), 200 µM dNTPs (Fermentas, Ontario, Canada), 0.5 µM each primer and 0.25 U/µl Phusion Hot Start DNA polymerase (Finnzymes Oy, Finland). The PCR reactions were performed in a Mastergradient thermocycler (Eppendorf, Hamburg, Germany).

The optimal PCR conditions for the marker SCAR2 were determined experimentally by gradient PCR and were: 30 seconds at 98 ºC, followed by four cycles of (10 seconds at 98 ºC, 20 seconds at 61 ºC and 15 seconds at 72 ºC with the annealing temperature decreasing 1 ºC per cycle) and 45 cycles of (10 seconds at 98 ºC, 20 seconds at 57 ºC and 15 seconds at 72 ºC) with the final extension of 4 minutes at 72 ºC.

The PCR fragments produced by SCAR markers were separated on a 2 % agar gel in 0.5 x TBE buffer with 0.1 µg/ml ethidium bromide (EtBr). Electrophoresis buffer contained 0.1 µg/ml EtBr. The PCR products were fractionated at 5.9 V/cm for 1 hour. Gels were imaged with Bio–Rad Chemidoc (Bio–Rad Laboratories Inc., Hercules, CA, USA) and the approximate sizes of the amplification products were estimated visually. For more accurate determination of the amplification product size, the fragments were separated on contract by capillary electrophoresis through an ABI 3130 XL genetic analyzer (Applied Biosystems, CA, USA) at the Institute of Biotechnology, Helsinki. Data from the capillary electrophoresis were collected and analysed using the software PeakScanner 1.0 (Applied Biosystems, CA, USA).

## 3.2 Linkage map construction and BAC library screening

### 3.2.1 Constructing linkage maps of the linkage group six

15 selected markers (see Appendix 1 and Table 6) of the *Fragaria* linkage group six on the reference map of Sargent *et al.* (2007) were checked for polymorphisms in the parents AL and BS and the $F_1$ hybrids of the mapping populations (FV × H4 & YW × FV). Six markers were chosen for genotyping the FV × H4 population and seven markers the YW × FV population, as these markers could be conveniently genotyped in a single run. Later on, six markers were scored in the FV × AL and FV × BS populations.

PCR reactions were performed in a final volume of 13 µl containing approximately 1 ng template DNA, 1 x PCR buffer, 2 mM $Mg^{2+}$, 0.05 U Taq DNA polymerase (all Qiagen, CA, USA), 200 µM dNTPs and 0.2 µM each primer. Reactions were carried out according to the touchdown protocol described by Sargent *et al.* (2003). Prior to genotyping, the PCR products were diluted 1:10. All seven markers were genotyped simultaneously on an ABI 3100 prism genetic analyser (Applied Biosystems, CA, USA). Data were collected and analysed using the Genescan and Genotyper software (Applied Biosystems). Linkage maps were constructed with JoinMap 4.0 software (Van Ooijen 2006), using Kosambi's mapping function. Minimum LOD score was set to 3.0, recombination frequency threshold to 0.35, jump threshold to 3.0 and ripple value to 1.0. The maps presented were constructed using MapChart 2.1 (Voorrips 2002).

### 3.2.2 BAC library screening

Screening the BAC library was initiated with the three markers that mapped the closest to the *Sfl* in the YW × FV mapping population. Colony–PCR was first performed on the eight superpools as described by Bonet *et al.* (unpublished), and subsequently the column, row and plate pools included in the positive superpools were screened. After identifying the columns, rows and plates which contained positive colonies, the four adjacent colonies on the 384–plates (see Appendix 2) were toothpicked and cultured

overnight in 1.2 ml LB medium (Sigma–Aldrich Inc., MO, USA) with 200 µg/ml chloramphenicol. Colony–PCR was performed on the cultured colonies; the PCR mastermix was prepared as described in 3.2.1, except 2,2 µl template DNA was used. PCR reactions were carried out according to the touchdown protocol by Sargent *et al.* (2003). Products were fractionated on 1.2 % agarose gel in TAE buffer (5.2 V/cm for 2 h) which was stained with EtBr for 20 min. Gels were imaged with Bio–Rad Chemidoc (Bio–Rad Laboratories Inc., Hercules, CA, USA). Positive clones were cultured for 16 hours in 10 ml LB with 200 µg/ml chloramphenicol and the bacteria were spun down, supernatant was discarded and the bacteria were sent to Qiagen (Germany) for end sequencing. The end sequences were utilised for designing new primers with Primer3 software (Rozen & Skaletsky 2000). The design criteria were a primer length of 20 to 24 bp (optimum 22 bp), $T_m$ of 55 to 65 ºC (optimum 60 ºC), a minimum GC content of 45 %, a 2 bp GC–clamp at the 3' end and expected product length of 270–400 bp Table 4). Primers were synthesised on contract by VH Bio (Gateshead, UK) and a new round of BAC library screening was initiated using new primers.

**Table 4. Markers designed from BAC end sequences of positive clones and the predicted sizes of the amplification fragments.**

| Marker | F primer | R primer | Size (nt) |
|---|---|---|---|
| 3D05–P1 | GAGGATGTTGCATTCTATGACC | GTTTGGTTGCATGTGATATTGC | 344 |
| 3D05–P2 | TGTTGGACTATGCTTCTTCACC | ATGACAATCGGGAGATTAGACC | 376 |
| 3D05–P3 | CATGCAACCAAACATCGTTC | CTTCCTACTCCTCTCGGTCTCC | 340 |
| 3D05–P4 | TTTCAAACTCCGCAAACTGG | TCTGCGGGTCTTCTCTCTCC | 400 |
| 15L02–P1 | GAGGATTTCCTTCCAACAAAGC | TGCTTCTAAATTTGGGTTCAGC | 368 |
| 15L02–P2 | GGTCTTCAACCACAAAATGAGC | CCTTGCAAAGTGCTTATGTCC | 335 |
| 19G01–P1 | GATTTCCGGCCAACACTACTC | AAAACCGACGACATGTTTTTC | 276 |
| 05K19–P1 | ATTAGCAATGGACGTAGTGAGC | TAAATTGTGGGGTAGGAACG | 379 |
| 05K19–P2 | TGAGACCTTGTGCTTGTTGC | GAGCTCATACGGTCCTTTGC | 294 |
| 05K19–P3 | TGAGGGAACAGCATATTACAGG | AATTAGTGAAGGGGGCATAACC | 478 |
| 05K19–P4 | TCCAGTTTTCACCAGGATTAGG | ATCAATCAGAACCCGCAAGG | 465 |
| 29F06–P1 | GCTTCATCATTTTGCCATCC | GTCATGGCAAACAAACATACCC | 413 |
| 29F06–P2 | GGTATGTTTGTTTGCCATGACG | GTTGGGACTTAAGGTTTTCACC | 387 |
| 29F06–P3 | GAAAACCTTAAGTCCCAACTCC | CACCTCATGTCGTGTATTCTGG | 435 |
| 31B24–P1 | CTCGAGTTTGCACGGTTAGG | TCTTTGAACTTCTCCAGAGTCG | 447 |
| 31B24–P2 | GACTCTGGAGAAGTTCAAAGACC | AGGACCAGAAGATCCTCAAAGC | 423 |
| 31B24–P3 | AGCCGAATGTACAGCTTCTAGC | GTGCAGAAGAAAAGAGCTTCG | 531 |

## 3.3 Designing and mapping gene–specific primer pairs

The EST sequences used in this study were previously identified as homologues to *Arabidopsis* flowering genes by Mouhu *et al.* (manuscript in preparation). Most EST sequences were derived from *Fragaria* 454 sequencing project of the berry research group of the Department of Applied Biology (University of Helsinki), and some were retrieved from the Rosaceous EST databank of Genome Database for Rosaceae (GDR). 91 EST sequences that showed significant homology to 30 flowering–related genes in *Arabidopsis* were first assembled into larger contigs using SeqMan 6.1 (DNAStar Inc., WI, USA) and then blasted against *Fragaria vesca* genome sequence assembly version 6 (Vladimir Shulaev, unpublished). Genomic contigs were retrieved from the *Fragaria vesca* genome sequence assembly version 6 (Vladimir Shulaev, unpublished) and aligned with corresponding ESTs using MegAlign 6.1 (DNAStar Inc.) to identify putative intron–exon junctions. The gap penalty was set to 20.0 and the gap length penalty to 0.

Primers were designed using Primer3 software (Rozen & Skaletsky 2000). The design criteria were a primer length of 20 to 24 bp (optimum 22 bp), $T_m$ of 55 to 65 ºC (optimum 60 ºC), a minimum GC content of 45 % and a 2 bp GC–clamp at the 3' end. Primers were designed to flank introns 150–800 base pairs in length, producing amplification fragments ranging from 190 to 820 bp (Table 5). Primers were ordered from VH Bio (Gateshead, UK).

Primer pairs were first tested in the $F_1$ hybrids of H4 × FV and FV × FN populations. PCR reactions were performed and the amplified fragments were visualised as described in section 3.2. Depending on the amplified fragment profiles, the markers were scored in the bin mapping population described by Sargent *et al.* (2008) by either scoring the mapping population on gel, or by capillary electrophoresis using fluorescently labelled forward primers. For the primer pairs that did not amplify polymorphic fragments readily scorable on gel, SNP mapping strategy was employed; eight PCR reactions for each primer pair with FV × FN $F_1$ DNA as template were performed (PCR reactions and protocols as in 3.2) and all

eight reactions were combined into a single sample of 100 µl, which was purified using QiaQuick PCR purification kit (Qiagen) according to manufacturer's protocol. The purified PCR products were sequenced by Qiagen (Germany). If SNPs were found in the sequenced PCR–product of the $F_1$, the marker was mapped in the bin set of six FV × FN seedlings. Additionally, one gene–specific marker, PRR7–i3, was mapped in FV × H4 $F_2$ population. Linkage maps were constructed as described in Chapter 3.2.

**Table 5. 45 gene–specific primer pairs designed for 21 flowering time–related genes . The –i followed by a number refers to the ordinal number of the intron which the primers were designed to flank.**

| Primer name | Forward primer sequence | Reverse primer sequence | Expected size |
|---|---|---|---|
| AGL14–i2 | TAGACCGTTTCCAAAAGAGAGC | ATTCTCATTGCCATGCTACTGG | 552 |
| AGL14–i4 | CATCAGTTGGAGAAGGATTTGG | CAGCAGTTTTAATTTACCTTTTCC | 240 |
| AGL22/SVP–i5 | TGAATTGGAACAAGAGAATAACC | CTGACAAGATGACACCATTTCC | 190 |
| AP1 2343–i3 | AAGATCTTGGCCCTTTGAGC | CTGAAGATCTGTGAGCTGATCC | 820 |
| AP1 2343–i6 | TTCTTCCAAGCCTTAGAGTGC | CTTGAATCGGATCATTCTGG | 211 |
| COP1–i1 | GAAAAAGCGGGTCCATGC | TTGCAGGTAACACTCTTGTAGC | 298 |
| COP1–i3 | TGCTGAACTTAGGCATGAGG | GGCGAACAACTCATAATCACG | 394 |
| COP1–i5 | CACGATCAAAGCTGAGTTGC | CCAAGCACGTTTCTCATGC | 283 |
| CRY1–i1 | TCAGGAGGACACCTTGATGC | ATCCAGAGGCACCACTTCTCC | 500 |
| CRY1–i2 | TGGCCCTGTGATCTCTAACC | CTCTTCTTGAGGTTTGTCTTGC | 594 |
| ELF7–i1 | ACCTCTTTTGCCTGATTTCG | ATTTCTGAATCAGCGGTTGG | 424 |
| ELF8–i1 | CTGATTTCCTTGACGGATGC | ACCATGGACCAACAAAGTCG | 554 |
| ELF8–i2 | AAGGCCATTGCACAATAAGG | GAGGCTCTTGTTGCTCTTGC | 241 |
| EMF2–i4 | TATGCTGATCCTGGTTGTATCC | GAGTGAAAGAATTGTCGCTTCC | 607 |
| EMF2–i5 | CAATTCTTTCACTCACATAGAGC | CGGACAAAACTTGTTCCAACG | 692 |
| EMF2–i6 | GAGGTTGATGATGATGTTGC | CTGTTGCTTTCTCACGAAGG | 234 |
| EMF2 CL7374–i1 | TTTGGTTTGGGACAGAGAGC | ACGGAATTGCAATCATAAGG | 328 |
| EMF2 CL7374–i2 | GGACAAGTGCCTTATGATTGC | CCTGCACTTGCTGGGATATG | 213 |
| FKF–i1 | CGATTGTCCCATCATTTACG | TGTCGTCTTTGAGCATGAGG | 817 |
| FVE MSI 4–i2 | AGCGGCTCTATCTCTCTGAGC | GCAATAACCAGAGTATTCGGAACC | 225 |
| FVE MSI 4–i4 | CCGTTTGTGAAGAAGTACAAGAC | TGCCACTATCTGAGTGTTCTGC | 290 |
| FVE MSI 4–i8 | GTGGCATTTTGCCCAACTAGG | CGTCACCAACACTACAGAACTCC | 284 |
| FVE MSI 4–i9 | ATACTATGGGATGCACGTGTGG | TTAGGATAAGGTTATCATCATGG | 593 |
| FVE MSI 4–i10 | GAATCCCCATGATGATAACC | ACTCCACCAGAAGTGAGATTCC | 226 |
| FVE MSI 4–i11 | ATCTCACTTCTGGTGGAGTTGG | CTCTGCAGAACTTCCAAAGACG | 239 |
| FVE MSI 4–i13 | CAGTAAAAAGGAGCGAACTTCC | ACAGTCCATGGGTCAGATGC | 221 |
| FVE MSI 4–i14 | TCATTGGAATGCATCTGACC | CTTCCAAAACCTCTTCTTCAGG | 372 |
| GA2OX–i1 | GTGGGCTGGGTTGAGTACC | CAAGCCATAGTCCTCACAGC | 248 |
| GA2OX–i2 | TGCCTCCAGACCAGTACTCC | GTGCCTCACACTTTGAAACC | 292 |
| GA3OX–i1 | CACCCTTGTTGGCTCACC | TCTTCAAAGCCACCTTTTGG | 498 |
| GI–i1 | GCACAAGTTAATTGCGTCTCC | GCTGTGGCTGCTTTAGTTGG | 796 |
| GI–i2 | CCAACTAAAGCAGCCACAGC | GATCATCATCTTTGGCAATCC | 590 |
| LHY–i1 | CGCAAAGCCTGAAATTGG | GCAGGAAGTTGTGTTTGAACC | 387 |
| PHYB CL 8836–i1 | ACGAGGAGCATAGAGCAAGG | CATGTTCAGCACCACACTGC | 230 |
| PHYB CL10624–i1 | TGGGTGGAGATTCATGTTCG | GAACTAATTGAGGGGGAAGACC | 573 |
| PHYB CL11161–i1 | TCTCATCCACGCTTTAAGAGG | AAACAGGCTTGTTGTCATGC | 552 |
| PIF4–i1 | TGATCTCCCTTCAACCAACC | GCCCATTTCTCCATAACAGC | 191 |
| PRR5–i2 | ATTCTGGGGCTGCAATGG | CCGCTCTAACGCCTCAGC | 232 |
| PRR5–i4 | CTCTCAGCCTCAAAATCTGG | TGCTCAGAGGAATAAGGAAACC | 194 |
| PRR7–i3 | GGGTGCTGTTGACTTCTTGG | TCGCTTCCAGTTCCACTACC | 393 |
| PRR7–i4 | GGAAGTGACAACGGAAGTGG | GCCATGAAAACACCTGTCG | 620 |
| SOC1–i1 | TCTCTATCCGCCTCATCTGG | ACTTGCTGGGTTCATTTTCC | 190 |
| SOC1–i2 | TGAGATCAGGTAGAAGATGACG | TTGTGTGTCTCTGCTCTTATGC | 620 |
| VIP3–i1 | GTTCCATGTGCTGGTGTCC | TCGGAAGTCTGGCAGTTACG | 420 |
| VRN1–i1 | CACAATCTATCTGTGACCTGAAGC | GCCATATGGGGAAGTCAGC | 222 |

# *4 Results*

## 4.1 Determining the phenotype and genotype of $F_2$ populations

During the 12 week observation period, none of the $F_1$ plants showed any signs of flowering. The plants that received a SD treatment flowered on the average 81 days after the end of the treatment (six weeks after sowing). Unexpectedly, almost all the LD–treated $F_1$ plants and *F. vesc*a seedlings started flowering only slightly later than the SD–treated plants.

In the $F_2$ populations, the first open flowers were observed in mid July, 51 days from sowing. Newly opened flowers on supposedly perpetually flowering plants were recorded until the end of August, 90 days from sowing. After this no new open flowers were observed until the unexpected flowering of all the plants (including *F. vesca* seedlings) at the end of experiment in early November.

The phenotypic segregation ratios for perpetual : seasonal flowering phenotypes were mostly in accordance with the 1:3 segregation ratios expected for a single recessive Mendelian gene. In populations FV × PI and FV × BS, the observed segregation ratios deviated from the expected at significance levels of 0.05 and 0.10, respectively (Table 6). In FV × PI, there was an excess of perpetually flowering phenotypes, whereas in FV × BS the number of perpetually flowering individuals was slightly lower than expected.

**Table 6. $\chi^2$ goodness–of–fit test for phenotypical segregation in five $F_2$ populations.**

| Population (n) | Expected phenotypes non–flowering : flowering | Observed phenotypes non–flowering : flowering |
|---|---|---|
| *F. vesca* × Alexandria (94) | 68 : 23 | 72 : 18 |
| *F. vesca* × Baron Solemacher (110) | 82 : 27 | 90 : 19 * |
| *F. vesca* × Hawaii–4 (110) | 82 : 27 | 81 : 27 |
| *F. vesca* × Piikkiö (110) | 82,5 : 27,5 | 73 : 37 ** |
| Yellow Wonder × *F. vesca* (110) | 82,5 : 27,5 | 88 : 22 |

*The phenotypes marked with * or ** deviate significantly from the expected Mendelian ratios at α = 0,1 and α = 0,05, respectively.*

Genotyping the three SCAR markers showed that while all markers produced fragments of the expected size, only SCAR2 detected a locus that was polymorphic between the parents. Therefore, only SCAR2, which amplified a codominant polymorphic product in all populations, was mapped. In some populations there was a high instance of missing values for the marker. However, SCAR2 and *Sfl* mapped to the same linkage group in all populations (Table 7).

**Table 7. The distance of marker SCAR2 from the *Sfl*.**

| Population (n) | Locus | Missing | Distance between the two loci (cM) |
|---|---|---|---|
| FV × AL (46) | SCAR2 | 8 | 13.3 |
| FV × BS (46) | SCAR2 | 7 | 10.1 |
| FV × H4 (46) | SCAR2 | 17 | 7.8 |
| FV × PI (48) | SCAR2 | 3 | 7.6 |
| YW × FV (46) | SCAR2 | 4 | 3.9 |
| All populations (232) | | 44 | 8.6 |

## 4.2 Linkage mapping and BAC library screening

BAC library screening was initiated by determining the closest flanking markers for the *Sfl*. 15 markers on the *Fragaria* linkage group six (FLG6) on the reference map of Sargent *et al.* (2007) were tested in the parents and $F_1$s of four populations. Out of the initial 15 tested markers, three markers were monomorphic, whereas 12 amplified codominant products that were polymorphic in the mapping populations. Six of these were mapped first in the FV × H4 and seven in the YW × FV population. Later on the six markers were mapped in the FV × BS and FV × AL populations (Table 8).

**Table 8. FLG6 markers tested for polymorphisms in the parents or F1 hybrids of the mapping populations. The markers in** bold **were chosen for mapping.**

| Name | Fluorescent dye | Polymorphic in parents | Product size range | Genotyped in populations | Originally published by |
|---|---|---|---|---|---|
| ARSFL–007 | HEX | N | – | | Lewers *et al.* (2005) |
| **BFACT–047** | HEX | Y | 158–161 | all | Denoyes–Rothan (unpublished) |
| CFVCT010 | 6–FAM | Y | 119–131 | | Monfort *et al.* (2006) |
| CFVCT030 | 6–FAM | N | – | | Monfort *et al.* (2006) |
| EMFn017 | 6–FAM | N | – | | Sargent *et al.* (2004) |
| EMFn117 | 6–FAM | Y | 151–153 | | Sargent *et al.* (2006) |
| EMFn123 | 6–FAM | Y | 202–210 | | Sargent *et al.* (2006) |
| **EMFn153** | HEX | Y | 220–222 | all | Sargent *et al.* (2006) |
| **EMFn185** | 6–FAM | Y | 211–213 | all | Sargent *et al.* (2006) |
| EMFv104 | 6–FAM | Y | 108–115 | | |
| **EMFvi025** | HEX | Y | 246–261 | YW × FV | Sargent *et al.* (2003) |
| **EMFvi133** | HEX | Y | 81–134 | all | Sargent *et al.* (2003) |
| FAC–004d | 6–FAM | Y | 322–333 | | Lewers *et al.* (2005) |
| **Fvi020** | 6–FAM | Y | 173–201 | all | Ashley *et al.* (2003) |
| **Fvi6b** | 6–FAM | Y | 274–276 | all | Ashley *et al.* (2003) |

All markers mapped to LG6 as expected, except for Fvi6b in FV × H4. However, mapping the selected markers in the FV × H4 and YW × FV populations produced linkage maps with different orders of markers on the map (Figure 3). The linkage map generated for population YW × FV was chosen to be representative of the actual locations of markers, as the number of missing marker scores was smaller in this population. The YW × FV map

covered a genetic distance of 35,2 cM with an average distance of about 4 cM between loci. Notably, SCAR2 mapped 11 cM away from the *Sfl*. After the chromosome walk had been initiated, populations FV × AL and FV × BS were genotyped with the six markers and the data from the four populations was combined to create a more precise genetic map. Marker EMFvi025 was excluded from the combined map, because it was scored only in population YW × FV. On the combined map, only the marker EMFn185 was reasonably close to the *Sfl* (Figure 3).

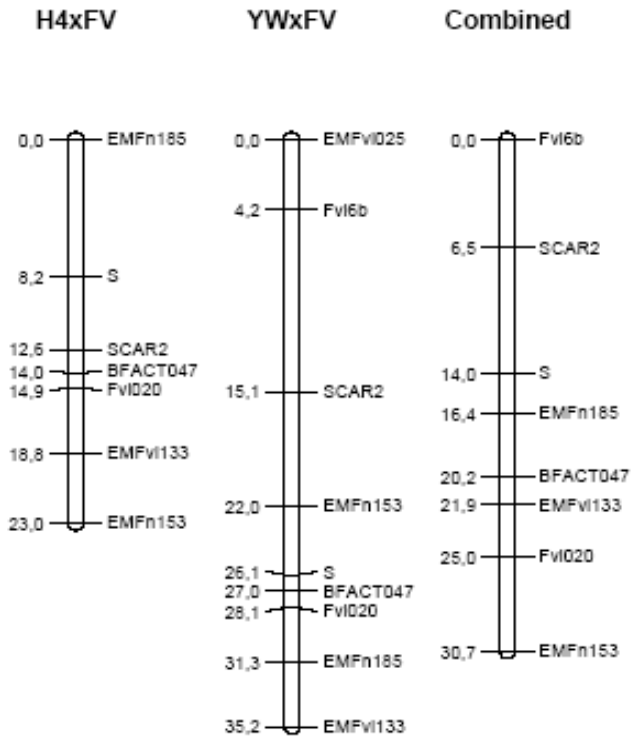



*Figure 3. The marker order on the FLG6 in two F2 populations and the map based on combined marker segregation data from four populations.'S' denotes the Seasonal flowering locus.*

On the YW x FV linkage map, the markers BFACT–047, Fvi020 and EMFn153 mapped the closest to the *Sfl*, and were used for initiating chromosome walk. During the first round of BAC library screening, BFACT–047, Fvi020 and EMFn153 identified eight, five and six positive clones, respectively. Further rounds of BAC library screening identified four, three and two positive clones with primer pairs derived from positive clones identified by BFACT–047, Fvi020 and EMFn153, respectively

(Table 9).

**Table 9. Positive clones identified with the initial flanking markers and the progress of chromosome walk thereafter.**

| Flanking marker | Clones identified | Primer pairs used for the subsequent screening | Clones identified | Primer pairs used for the subsequent screening | Clones identified |
|---|---|---|---|---|---|
| BFACT–047 | 15L02, 20C20, 29C22, 39F06, 39A15, 39I16, 45F03, 45I04 | 15L02–P1 | 05K19, 17K21, 29F06, 33P04 | 05K19–P1, 05K19–P2, 29F06–P1, 29F06–P2 | –<br>–<br>–<br>– |
| Fvi020 | 15J07, 19G01, 22P03, 28K02, 36P03 | 19G01–P1 | 31B24 | 31B24–P1, 31B24–P2 | 06D11, 40I12<br>– |
| EMFn153 | 03C05, 03D05, 03D06, 09I09, 09J09, 09J10 | 03D05–P1<br>03D05–P2 | –<br>30G06, 33C19 | –<br>no primers designed | |

End–sequencing the BAC clones produced sequences ranging from 900 to 1600 bp, which was enough for designing new primers. Each end–sequence was used for designing from one to four new primer pairs (see Table 4).

Screening the library with new primer pairs identified the same clones as the previous screening round, but also some new positive clones. Altogether, two whole rounds of BAC library screening were completed, and a third round was initiated. There was no indication of covering the entire genomic area between any two flanking markers.

## 4.3 Designing and mapping gene–specific primer pairs

The 91 EST sequences that showed homology to 30 flowering–related genes had initially an average length of about 600 bp. After assembly into larger contigs, the number of contiguous sequences was 58 with an average length of 761 bp. All consensus sequences of the 30 flowering–related genes found hits in the *Fragaria vesca* genome sequence assembly version 6 (Vladimir Shulaev, unpublished) and were aligned with their corresponding genomic

sequences to reveal intron–exon junctions. 16 consensus EST sequences were too short and could detect only one exon, or of too bad quality for inferring intron–exon junctions. Among the remaining alignments, 100 introns were detected, some of which were, however, either too short or too long for designing functioning primers. 54 introns were of suitable length, and were used for designing flanking primers. With the criteria for primer picking, 45 gene–specific primer pairs for 21 flowering–related genes were designed (see Table 5).

Nearly all primer pairs amplified products of the expected size in the $F_1$ hybrids of the mapping populations FV × H4 and FV × FN (see Table 5). However, some primer pairs amplified several loci and therefore had to be excluded from further mapping efforts. Fortunately, there was more than one primer pair designed for most genes, and only three genes had to be left out of further experiments because of non–functional primer pairs (Table 10).

Only GI–i2 amplified a locus which showed size polymorphism when visualised on gel and could be simply mapped by agarose electrophoresis. Markers PHYB CL8836–i1 and PRR7–i3 appeared to amplify loci visibly polymorphic on agarose gel, but fluorescently labelled forward primers were ordered for these markers for performing more reliable genotyping through capillary electrophoresis. PHYB CL8836–i1 amplified monomorphic fragments, but PRR7–i3 produced polymorphic fragments in the parents of the mapping populations and was mapped both in the bin set and the FV × H4 population.

For the primer pairs that did not appear polymorphic on gel, a SNP mapping strategy was employed. Altogether 16 primer pairs were used for amplifying DNA from the $F_1$ hybrid of the reference FV × FN population. Ten primer pairs produced fragments with SNPs (Appendix 3) and nine of these were mapped in the bin set. Five primer pairs amplified products with mixed sequences. Fragments amplified by two primer pairs, namely COP1–i5 and VRN1–i1 had identical sequences and could not be mapped in the FV × FN population (Table 10).

**Table 10. Mapping strategies for 21 flowering–related markers and the location of the mapped markers on the reference linkage map.**

| Origin of the ESTs | Gene | Mapping strategy | Mapping outcome |
|---|---|---|---|
| 454 project | AGL14 | SNP | Mixed sequence |
| GDR | AGL22 / SVP | no amplification | – |
| GDR | AP1 | no amplification / several bands | – |
| GDR | COP1 | SNP | no SNPs in FV×FN |
| 454 project | CRY1 | SNP | Mapped to FLG5 (5:5) |
| 454 project | ELF7 | SNP | mixed sequence |
| GDR | ELF8 | SNP | polymorphic in FV×FN, to be mapped |
| GDR | EMF2 | SNP | mapped to FLG3 (3:3) |
| GDR | FKF1 | SNP | mixed sequence |
| GDR | FVE MSI 4 | SNP | mapped to FLG1 (1:2) |
| GDR | GA2OX | SNP | mapped to FLG3 (3:3) |
| Hytönen *et al.* (2008) | GA3OX | no amplification | – |
| 454 project | GI | bin set scored on gel | mapped to FLG2 (2:4) |
| GDR | LHY | SNP | mapped to FLG7 (7:2) |
| 454 project | PHY B | labelled forward primer | not polymorphic |
| 454 project | PIF4 | SNP | mixed sequence |
| GDR | PRR5 | SNP | mapped to FLG7 (7:2) |
| GDR | PRR7 | labelled forward primer | mapped to FLG6 (6:3) |
| GDR | SOC1 | SNP | mixed sequence |
| GDR | VIP3 | SNP | mapped to FLG1 (1:7) |
| 454 project | VRN1 | SNP | no SNPS in FV×FN |

*The EST sequences mostly derived from a 454–sequencing project of the berry research group of the Dept. of Applied Biology (University of Helsinki), or were retrieved from Genome Database for Rosaceae, a public database.*

Nine mapped gene–specific markers could be localised to six of the seven linkage groups of the diploid *F. vesca*. The marker PRR7–i3 proved especially interesting, as it mapped close to the *Sfl* in the bin set. Mapping the gene, a *Fragaria* homologue of *PRR7* (*PSEUDO - RESPONSE REGULATOR 7*) in the FV × H4 population showed that PRR7–i3 indeed locates 0.8 cM away from the *Sfl* (Figure 4).

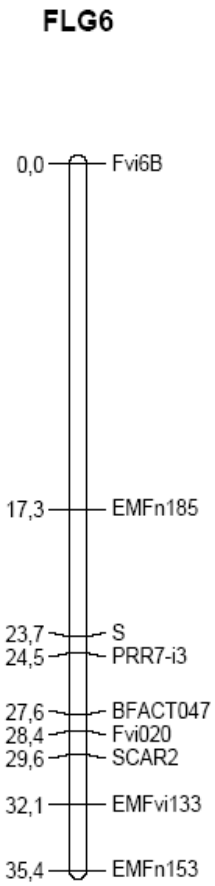



*Figure 4. Linkage group six in the FV x H4 population with the gene–specific marker PRR7–*i3. *'S' denotes the Seasonal flowering locus.*

# *5 Discussion*

## 5.1 Single–gene control of flowering in *F. vesca*

Similar patterns of co–segregation between the *Sfl* and SCAR2 in all five $F_2$ populations included in the current experiment indicates that perpetual flowering in divergent *F. vesca* f. *semperflorens* cultivars is controlled by the same recessive gene. Although it was stated already by Brown & Wareing (1965), that perpetual flowering habit in *F. vesca* f. *semperflorens* cultivar 'Baron Solemacher' is controlled by a single recessive Mendelian gene, no studies have previously been conducted to find out how universal this gene is in other perpetually flowering *F. vesca* cultivars.

Segregation of perpetual flowering in $F_2$ progenies of the crosses FV × AL, FV × BS, FV × H4 and YW × FV did not differ significantly ($\alpha = 0.05$) from the expected 3:1 ratio. These results are in accordance with those obtained by Brown & Wareing (1965) and Sargent *et al.* (2004), who found that the '*semperflorens*' gene segregates 3:1 in an $F_2$ progeny. Only the $F_2$ population FV × PI deviated significantly from the expected 3:1 ratio. Deviation can probably be explained by segregation distortion, which has been previously recorded to occur at morphological marker loci (Sargent *et al.* 2004) as well as at SSR and gene–specific marker loci in inter– and intraspecific $F_2$ mapping populations (Davis &Yu 1997, Sargent *et al.* 2006; Sargent *et al.* 2007). Reasons for segregation distortion are not entirely clear; partial incompatibility between crossing parents of wide crosses or self–incompatibility of either parent have been suggested as possible causes (Sargent *et al.* 2007). Co–segregation of *Sfl* and SCAR2 were, however, similar in FV × PI to the other populations. It appears safe to assume that the deviation observed in FV × PI is caused by segregation distortion rather than a different genetic model for inheritance of perpetual flowering.

The unexpected flowering of $F_1$ seedlings in early spring may be explained either by the effect of low temperature or by the age of plants at the beginning of flowering (more than 10 months??). Flowering of all $F_2$ progenies and *F. vesca* seedlings in late fall can be explained by the effect of the reported low temperature on flower induction. In an experiment

conducted in controlled environment, Heide & Sonsteby (2007) showed that flowering in *F. vesca* is induced independently of day length at temperatures below 15 ºC. During the current experiment, night temperatures below 13 ºC were recorded from September to November. Despite flowering of all $F_2$ progenies in November, the flowering data can be considered reliable, as the seeds were sown in May in very warm conditions and received LD treatment since the day of sowing. Plants that were scored as perpetually flowering flowered within three months from the day of sowing, whereas flowering of all plants occurred 5.5 months after germination. The effect of the reported cooler temperature during autumn should not have an effect on the phenotypic scores.

Genetic distance between the *Sfl* and SCAR2 in the five $F_2$ populations ranged from 3.9 to 13.3 cM, being 8.6 cM in the combined data of all populations. The differences can probably be attributed to the small number of individuals scored per each population, as well as to the relatively high number of missing scores in some populations. The genetic distances reported here are not in concordance with the results of Albani *et al.* (2004), who could not detect any recombination between SCAR2 and the *Sfl*. The population of 1049 seedlings scored by Albani *et al.* (2004) was grown in an unheated greenhouse from January to December. During this period, low temperatures are likely to have occurred, and day length–independent flowering may have been induced, rendering the determination of phenotype inaccurate. Moreover, SCAR2 has not been scored in the FV × FN reference population, and therefore its location on the map in relation to other markers and the *Sfl* has not been clear. Results gained from the current experiment suggest that SCAR2 is not close enough to the *Sfl* to make it especially suitable for further molecular studies, e.g. positional cloning.

This study shows that perpetual flowering in all five *F. vesca* f. *semperflorens* accessions is controlled by a gene located at the *Sfl* approximately 8.6 cM away from the marker SCAR2. The gene residing at the *Sfl* appears central for regulation of flowering time in diploid *Fragaria*. Given the fact that perpetually flowering *F. vesca* and *F.* x *ananassa* cultivars react very similarly to temperature and day length (Sonsteby & Heide 2008), it is reasonable to assume that the same gene plays a role in

control of flowering time in both species.

## 5.2 Chromosome walking

The positional cloning attempt was initiated using marker order data obtained from mapping seven loci in a population of 45 seedlings (Figure 3). The markers BFACT–047, Fvi020 and EMFn185 and the primer pairs derived from BAC end–sequences were able to identify positive clones. In total, 28 positive clones were identified. The BAC library used in this study proved to be suitable for positional cloning. However, the identified clones were not fingerprinted nor mapped, making it impossible to deduce the size or the orientation of the contiguous fragments.

During the course of the experiment, the importance of selecting the closest initial flanking markers correctly became clear. After the chromosome walk had been initiated, six markers were used for genotyping the populations FV × AL and FV × BS, and the segregation data of the four populations (FV × AL, FV × BS, FV × H4 and YW ×FV) were combined. The resulting map (Figure 3) showed marker segregation patterns quite different from those obtained for the YW × FV population. Marker order and genetic distances between markers on this 'combined' map were mostly in accordance with the marker order on the FV × FN reference map (Sargent *et al.* 2007) and therefore the combined map can be considered more reliable than the YW × FV or FV × H4 maps alone. In other words, the initial flanking markers were chosen incorrectly, based on marker data from too few seedlings.

Obviously, the three linkage maps presented in Figure 3 are not similar. The differences between marker order and distance are likely to be due to the small number of genotyped individuals. In a small mapping population (and especially when there are plenty of missing values), the genotypic score of one individual has a large effect on the composition of the linkage map. Therefore incorrect genotyping has a relatively large effect on the marker order. Incorrect genotyping may arise from human errors while genotyping the data, sample contamination or, as reported by Tabernet *et al.* (1996), very low quantity of sample DNA. The data were scored three

times so as to minimize the possibility of erroneous genotyping. Although care was taken when handling the samples, contamination cannot be entirely ruled out. However, the most likely cause to possible incorrect genotyping is the low amount of DNA used in the PCR reactions. According to Tabernet *et al.* (1996), a very low amount of sample DNA may lead to scoring a heterozygous individual as a homozygote. Genotyping samples with small quantities of DNA incorrectly may have contributed to the differences in marker order on the linkage maps.

There has previously been one positional cloning attempt aimed at cloning the *Sfl* (Peak 2005). The experimenter used SCAR2 as starting point for chromosome walk, and completed three rounds of BAC library screens. Five positive clones were identified, covering a genetic distance of 200 kb. The greater number (29) of identified positive clones in the current experiment is probably due to the better quality of the BAC library. The library of Peak (2005) consisted of 17,000 clones with an average insert size of 60 kb and 19 % occurrence of false positive clones. The BAC library used in this experiment contains 18,432 clones with an average insert size of 98 kb and a low occurrence of false positive clones (Bonet *et al.* unpublished). Peak (2005) did not succeed in positionally cloning the *Sfl* for the same reason as the current experiment failed; the flanking markers were not close enough to the target locus.

Successful positional cloning attempts have usually been preceded by fine–mapping the genetic region of interest in mapping populations as large as 3095 plants (Yan *et al.* 2003). Fine–mapping typically results in markers flanking the locus of interest at distance of less than 1 cM, and sometimes markers co–segregating with the trait are found (Yang *et al.* 2001, Kaufmann *et al.* 2003). In future positional cloning efforts of the *Sfl*, care should be taken that the flanking markers are located at distance no greater than ca. 1 cM from the *Sfl*. Furthermore, the number of individuals in mapping population should be large enough to make genotyping reliable and reproducible.

## 5.3 Candidate gene approach

This study succeeded in bin mapping nine flowering time–related genes on the FV × FN reference map (Table 11). One of the genes, a *Fragaria* homologue of *PRR7* (*PSEUDO–RESPONSE REGULATOR7*) was mapped in the FV × H4 population and was proven to map closer to the *Sfl* than any other marker included in the experiment. Considering the small amount of individuals scored in the $F_2$, it seems possible that *FraPRR7* could be gene whose mutation confers perpetual flowering in *F. vesca* f. *semperflorens*.

**Table 11. Gene–specific markers mapped in the experiment, their map positions and the functions of their putative homologs in *Arabidopsis*.**

| Gene mapped in *Fragaria* | *Fragaria* map position | Function in *Arabidopsis* |
|---|---|---|
| *CRY1*–i2 | LG5 | Photoperiodic pathway: blue/UV–A light receptor |
| *EMF2–i4* | LG3 | Photoperiodic pathway: repression on reproductive development (Yoshida *et al.* 2001) |
| *FVE MSI4–i10* | LG1 | Autonomous pathway: positive regulator of flowering (Ausín *et al.* 2004) |
| *GA2OX–i1* | LG3 | Gibberellic pathway: inactivates bioactive gibberellins (Taiz & Zeiger 2001) |
| *GI–i2* | LG2 | Photoperiodic pathway: circadian clock component (Fowler *et al.* 1999) |
| *LHY–i1* | LG7 | Photoperiodic pathway: circadian clock component (Wang *et al.* 1997) |
| *PRR5–i4* | LG7 | Photoperiodic pathway: circadian clock component (Matsushika *et al.* 2000) |
| *PRR7–i3* | LG6 | Photoperiodic pathway: circadian clock component (Matsushika *et al.* 2000) |
| *VIP3–i1* | LG1 | Vernalisation pathway: activates *FLC* (Zhang *et al.* 2003) |

Mapping the nine genes has not only narrowed down the number of flowering–related genes possibly responsible for perpetual flowering in *F. vesca*, but has also given one clear candidate for the *Sfl*. *PRR7*, the gene that mapped to the proximity of the *Sfl*, has been studied in the model plant *Arabidopsis*. *PRR7* belongs to a small family of structurally similar pseudo–response regulator genes, including also TOC1, a core component of the circadian clock (Matsushika *et al.* 2000). In the long day plant *Arabidopsis* in LD conditions, PRRs indirectly activate *CONSTANS* (*CO*), whose expression in turn activates *FLOWERING LOCUS T* (*FT*) (Nakamichi *et al.*

2007, Taiz & Zeiger 2006). FT acts in the apical meristem, stimulating flowering (Taiz &Zeiger 2006). According to Hayama *et al.* (2003), the most striking difference between the flowering pathways of the LD plant *Arabidopsis* and the SD plant rice is the reversed regulatory effect of CO on *FT*. In *Arabidopsis*, CO activates *FT* and promotes flowering, whereas in rice CO represses *FT* and therefore suppresses flowering. Assuming that the circadian clock components and their functions are similar in the SD plants rice and *Fragaria*, PRRs could suppress flowering in *F. vesca* in LD conditions by promoting *CO*. This is consistent with the hypothesis of Vince–Prue & Guttridge (1972), who suggested that flowering in *Fragaria* is controlled by an LD–induced inhibitor. A loss–of–function mutation of *PRR7* would therefore promote flowering in LD and cause the flowering phenotype of *F. vesca* f. *semperflorens*. Moreover, mutations of *PRR7* homologs have been shown to cause altered photoperiodic responses in cereals, and have been used in 'green revolution' cultivars of wheat (Turner *et al.* 2005, Beales *et al.* 2007). It has been demonstrated that *F. vesca* and *F. vesca* f. *semperflorens* have altered responses not only to photoperiod but also temperature (Sønsteby & Heide 2008). Interestingly, a role for *PRR7* in the temprature–sensitive circadian regulation has been indicated by Salomé *et al.* (2005). Taken together, these results suggest that *FraPRR7* could be the gene conferring LD flowering in *Fragaria*, not only because it is located very close to the *Sfl*, but also because mutations in the gene cause day neutral phenotypes in several species.

Bin mapping of seven gene–specific markers failed either because no SNPs could be detected within the amplified regions, or due to poor quality mixed sequences. High failure rate of SNP discovery have been previously reported by Han *et al.* (2009), who associated the problem with non–specific primer annealing. In the current investigation, the primers were designed to have a GC–clamp at 3'–end to minimise non–specific primer binding. Furthermore, most primer pairs amplified a single product of the expected size. Instead of non–specific primer binding, mixed sequences may be caused by primers binding to conserved exon regions of related genes. For instance, *Arabidopsis AGL14* shows significant similarity to six other *AGL*–family genes, and it is probable that also *Fragaria* possesses families of

closely related genes. Monomorphic sequences prevented the mapping of two gene–specific markers, even though the primers were designed to amplify supposedly variable intron fragments. It appears that the parents of the bin mapping population, i.e. *F. vesca* and *F. nubicola* share notable similarity even in less–conserved intron areas.

Selective bin mapping has been shown to be a reliable method for assigning markers an approximate chromosomal position on the FV × FN reference map (Sargent *et al.* 2008), and was further validated in this investigation. Primers designed in the current experiment amplified mostly the genomic regions they were supposed to amplify, as confirmed by fragment size estimation on agarose gel and sequencing. The exact position of *FraPRR7* on the LG6 remains to be examined by mapping the gene in a larger population. *FraPRR7* and the other gene–specific flowering–related markers developed in this investigation are potentially transferable to the octoploid *F.* × *ananassa*, and therefore offer a good starting point for flowering time studies in the cultivated species.

## *6 Conclusions*

This Master's thesis has taken advantage of the genetic tools available for the diploid *Fragaria*, employing BAC library screening, public sequence information and unpublished 454 sequence data for elucidating the gene controlling flowering in *F. vesca*. Although BAC library screening did not result in positional cloning of the *Sfl*, the experiment proved the library suitable for such a study. Designing and mapping nine gene–specific markers lead to identification of a potential candidate for the *Sfl*, *FraPRR7*. The marker designed for *FraPRR7* locates closer to the *Sfl* than any other marker studied here, and is therefore a suitable starting point for further BAC library screening.

The utility of *PRR7* as a marker for the everbearing character in the octoploid strawberry will be investigated further in *F.* × *ananassa* breeding populations. Also the other gene–specific markers developed in this thesis are potentially useful in studying the genetic mechanisms of flowering in the octoploid strawberry.

## *Acknowledgements*

The traineeship at the Department of Applied Biology was sponsored by the University of Helsinki. The special training period at East Malling Research was from the project funds of MTT Mansikkaremontti. Dr. Vladimir Shulaev and the *Fragaria* genome sequencing consortium allowed access to the genome sequence before it was published. I'd like to express my gratitude to both my supervisors, Timo Hytönen and Dan Sargent, without whom this thesis would never have come to be. Big thanks go also to Kati Mouhu and Marja Rantanen for practical advise and help in the lab and in the greenhouse.

## *References*

**Ahmadi H, Bringhurst RS & Voth V**. 1990. Modes of Inheritance of Photoperiodism in *Fragaria.* Journal of the American Society for Horticultural Science 115(1): 146–152.

**Albani MC, Battey NH & Wilkinson MJ**. 2004. The development of ISSR–derived SCAR markers around the *SEASONAL FLOWERING LOCUS* (*SFL*) in *Fragaria vesca*. Theoretical & Applied Genetics 109: 571–579.

**Ausín I, Alonso–Blanco C, Jarillo JA, Ruiz–Garcia L & Martínez–Zapater JM**. 2004. Regulation of flowering time by FVE, a retinoblastoma–associated protein. Nature Genetics 36(2): 162–166.

**Beales J, Turner A, Griffiths S, Snape JW & Laurie DA. 200**7. A *Pseudo–Response Regulator* is misexpressed in the photoperiod insensitive *Ppd–D1*a mutant of wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics 115(5): 721–733.

**Bonet J, Sargent DJ, Monfort A, Muños–Torres MC, Ribes J, Abbott AG, Arús P, Simpson DW & Davik J**. unpublished. Construction and PCR–based characterisation of a BAC library from the *Fragaria vesca* (Rosaceae) cultivar 'Ali–Baba' and genome–wide anchoring of BAC clones to the diploid *Fragaria* reference map.

**Bortiri E, Jackson D & Hake S**. 2006. Advances in maize genomics: the emergence of positional cloning. Current Opinion in Plant Biology 9: 164–171.

**Brown T & Wareing PF**. 1965. The genetical control of the everbearing habit and three other characters in varieties of *Fragaria vesca*. Euphytica 14: 97–112.

**Böhlenius H, Huang T, Charbonnel–Campaa L, Brunner AM, Jansson S, Strauss SH & Nilsson O**. 2006. CO/FT Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. Science 312: 1040–1043.

**Cai L, Taylor JF, Wing RA, Gallagher DS, Woo S–S & Davis SK**. 1995. Construction and Characterization of a Bovine Bacterial Artificial Chromosome Library. Genomics 29: 413–425.

**Corbesier L & Coupland G**. 2006. The quest for florigen: a review of recent progress. Journal of Experimental Botany 57(13): 3395–3403.

**Davis TM & Yu** H. 1997. A Linkage Map of the Diploid Strawberry, *Fragaria vesca*. Journal of Heredity 88: 215–221.

**Doyle JJ & Doyle JL**. 1990. Extraction of plant DNA from fresh tissue. Focus 12: 13–15.

**Folta KM & Davis TM**. 2006. Strawberry Genes and Genomics. Critical Reviews in Plant Sciences 25: 399 – 415.

**Food and Agricultural Organisation of the United Nations**. 2009. Production Statistics. Available at www.faostat.fao.org/ [Accessed February 2009]

**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. The EMBO Journal 18(17): 4679–4688.

**Gardner MJ, Hubbard KE, Hotta CT, Dodd AN & Webb AAR**. 2006. How plants tell the time. Biochemical Journal 397: 15–24.

**Georgi LL, Wang Y, Reighard GL, Mao L, Wing RA & Abbott AG.** 2003. Comparison of peach and *Arabidopsis* genomic sequences: fragmentary conservation of gene neighbourhoods. Genome 46: 268–276.

**Han Y, Gasic K, Marron B, Beever JE & Korban S**. 2007. A BAC–based physical map of the apple genome. Genomics 89: 630–637.

**Hayama R & Coupland G**. 2003. Shedding light on the circadian clock and the photoperiodic control of flowering. Current Opinion in Plant Biology 6: 13–19.

**Hättasch C, Flachowsky H, Kapturska D & Hanke M–**V. 2008. Isolation of flowering genes and seasonal changes in their transcript levels related to flower induction and initiation in apple (*Malus domestica*). Tree Physiology 28: 1459–1466.

**Heide OM & Sønsteby A**. 2007. Interactions of temperature and photoperiod in the control of flowering of latitudinal and altitudinal populations of wild strawberry (*Fragaria vesca*). Physiologia Plantarum 130: 280–289.

**Izawa T, Takahashi Y & Yano M**. 2003. Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and *Arabidopsis*. Current Opinion in Plant Biology 6: 113–120.

**James CM, Wilson F, Hadonou AM & Tobutt KR**. 2003. Isolation and characterisation of polymorphic microsatellites in diploid strawberry (*F. vesca* L.) for mapping, diversity studies and clone identification. Molecular Ecology Notes 3: 171–173.

**Kaufmann H, Mattiesch L, Lörz H & Debener** T. 2003. Construction of a BAC library of *Rosa rugosa* Thunb. And assembly of a contig spanning *Rdr1*, a gene that confers resistance to blackspot. Molecular Genetics & Genomics 268: 666–674.

**Li W, Lin Z & Zhang X**. 2007. A Novel Segregation Distorion in Intraspecific Population of Asian Cotton (*Gossypium arboretum* L.) Detected by Molecular Markers. Journal of Genetics and Genomics 34(7): 634–640.

**Maa- ja metsätalousministeriön tietopalvelukeskus.** 2008. Puutarhayritysrekisteri 2007. Edita Oy. Helsinki.

**Matsushika A, Makino S, Kojima M & Mizuno T**. 2000. Circadian Waves of Expression of the *APRR1/TOC1* Family of Pseudo–Response Regulators in *Arabidopsis thaliana*: Insight into the Plant Circadian Clock. Plant Cell Physiology 41(9): 1002–1012.

**McClung CR.** 2001. Circadian Rhythms in Plants. Annual Review in Plant Physiology and Plant Molecular Biology 52: 139–162.

**Mockler T, Yang H, Yu XH, Parikh D, Cheng Y, Dolan S & Lin C**. 2003. Regulation of photoperiodic flowering by *Arabidopsis* photoreceptors. PNAS 100(4):2140–2145.

**Mouhu K, Hytönen T, Folta K, Rantanen M, Paulin L, Auvinen P & Elomaa P**. 2008.Characterization of flowering pathways in strawberry – a perennial SD plant. Unpublished.

**Murakami M, Tago Y, Yamashino T & Mizuno T**. 2006. Comparative Overviews of Clock–Associated Genes of *Arabidopsis thaliana* and *Oryza sativa*. Plant Cell Physiology 48(1): 110–121.

**Nakamichi N, Kita M, Niinuma K, Ito S, Yamashino T, Mizoguchi T & Mizuno T**. 2007. *Arabidopsis* Clock–Associated Pseudo–Response Regulators *PRR9*, *PRR7* and *PRR5* Coordinately and Positively Regulate Flowering Time Through the Canonical CONSTANS–Dependent Photoperiodic Pathway. Plant Cell Physiology 48(6): 822–832.

**Van Ooijen JW**. 2006. JoinMap®, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.

**Peak E**. 2005. Towards isolation of a gene controlling seasonal flowering in *Fragaria vesca* L. PhD thesis. University of Reading.

**Peters JL, Cnudde F & Gerats T**. 2003. Forward genetics and map–based cloning approaches. Trends in Plant Science 8(10): 484–491.

**Puutarhaliitto ry. 2009.** PUUTARHA&kauppa 13.

**Rozen S & Skaletsky HJ**. 2000 Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ. pp 365–386.

**Salomé PA & McClung CR**. 2005. *PSEUDO–RESPONSE REGULATOR* 7 ans *9* Are Partially Redundant Genes Essential for the Temperature Responsiveness of the *Arabidopsis* Circadian Clock. The Plant Cell 17: 791–803.

**Sargent DJ, Hadonou AM & Simpson DW**. 2003. Development and characterisation of polymorphic microsatellite markers from *Fragaria viridis*, a wild diploid strawberry. Molecular Ecology Notes

**Sargent DJ, Davis TM, Tobutt KR, Wilkinson MJ, Battey NH & Simpson DW**. 2004. A genetic linkage map of microsatellite, gene–specific and morphological markers in diploid *Fragaria*. Theoretical and Applied Genetics 109(7): 1385–1391.

**Sargent DJ, Rys A, Nier S, Simpson DW & Tobutt KR**. 2007. The development and mapping of functional markers in *Fragaria* and their transferability and potential for mapping in other genera. Theoretical & Applied Genetics 114: 373–384.

**Sargent DJ, Cipriani G, Vilanova S, Gil–Ariza D, Arús P, Simpson DW, Tobutt KR & Monfort A**. 2008. The development of a bin mapping population and the selective mapping of 103 markers in the diploid *Fragaria* reference map. Genome 51: 120–127.

**Sargent DJ, Marchese A, Simpson DW, Howad W, Fernández–Fernández F, Monfort A, Arús P, Evans KM & Tobutt KR**. 2009. Development of "universal" gene–specific markers from *Malus* spp. CDNA sequences, their mapping and used in synteny studies within Rosaceae. Tree Genetics & Genomes 5: 133–145.

**Shaw DV & Famula TR**. 2005. Complex segregation analysis of day–neutrality in domestic strawberry (*Fragaria* × *ananassa* Duch). Euphytica 145: 331–338.

**Silva C, Garcia–Mas J, Sánchez AM, Arús P & Oliveira MM**. 2005. Looking into flowering time in almond (*Prunus dulcis* (Mill) D. A. Webb): the candidate gene approach. Theoretical and Applied Genetics 110: 959–968.

**Simpson GG, Gendall AR & Dean C**. 1999. When to Switch to Flowering. Annual Review of Cell and Developmental Biology 99: 519–550.

**Sønsteby A & Heide OM**. 2008. Long–day rather than autonomous control of flowering in the diploid everbearing strawberry *Fragaria vesca* ssp. s*emperflorens*. Journal of Horticultural Science and Biotechnology. 83(3): 360–366. Trends in Genetics 11(2): 63–68.

**Tabernet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP & Bouvet J**. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Research 24(16): 3189–3194.

**Taiz L & Zeiger E**. 2006. Plant Physiology 4th edition. Sinauer Associates, Inc., MA, USA.

**Tanksley SD, Ganal MW & Martin GB**. 1995. Chromosome landing: a paradigm for map–based gene cloning in plants with large genomes. Trends in Genetics 11(2): 63-68.

**Thomas SG & Sun T**. 2004. Update on Gibberellin Signaling. A Tale of the Tall and the Short. Plant Physiology 135:668–676.

**Turner A, Beales J, Faure S, Dunford RP & Laurie DA**. 2005. The Pseudo–Response Regulator *Ppd–H1* Provides Adaptation to Photoperiod in Barley. Science 310(11): 1031–1034.

**Vilanova S, Sargent DJ, Arús P & Monfort A**. 2008. Synteny conservation between two distantly–related Rosaceae genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry). BMC Plant Biology 8: 67.

**Vince–Prue D & Guttridge CG**. 1973. Floral Initiation in Strawberry: Spectral Evidence for the Regulation of Flowering by Lond–day Inhibition. Planta 110: 165–172.

**Vision TJ, Brown DG, Shmoys SB, Durrett RT & Tanksley SD**. 2000. Selective Mapping: A Strategy for Optimizing the Construction of High–Density Linkage Maps. Genetics 155: 407–420.

**Voorrips RE**. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. The Journal of Heredity 93(1): 77–78.

**Wang Z–Y, Kenigsbuch D, Sun L, Harel E, Ong MS & Tobin EM**. 1997. A Myb–Related Transcription Factor Is Involved in the Phytochrome Regulation of an *Arabidopsis Lhcb* Gene. The Plant Cell 9: 491–507.

**Weebadde CK, Wang D, Finn CE, Lewers KS, Luby JJ, Bushakra J, Sjulin TM & Hancock JF**. 2007. Using a linkage mapping approach to identify QTL for day–neutrality in the octoploid strawberry. Plant Breeding 127:94–101.

**Xu JC, Yang DC, Dominngo J, Ni JJ & Huang N**. 1998. Screening for overlopping bacterial artificial chromosome clones by PCR analysis with an arbitrary primer. Proceedings of the National Academy of Science of the United States of America 95: 5661–5666.

**Yan L, Loukolanov A, Tranquilli G, Helguera M, Fahima T & Dubcovsky J**. 2003. Positional cloning of the wheat vernalization gene *VRN1*. Proceedings of the National Academy of Science of the United States of America 100(10): 6263–6268.

**Yang Z–N, Ye X–R, Choi S, Molina J, Moonan F, Wing RA, Roose ML & Mirkov TE**. 2001. Construction of a 1.2–Mb contig including the citrus tristeza virus resistance gene locus using a bacterial artificial chromosome library of *Poncirus trifoliata* (L.) Raf. Genome 44: 382–393.

**Yoshida N, Yanai Y, Chen L, Kato Y, Hiratsuka J, Miwa T, Sung ZR & Takahashi S**. 2001. EMBRYONIC FLOWER2, a Novel Polycomb Group Protein Homolog, Mediates Shoot Development and Flowering in *Arabidopsis*. The Plant Cell 13: 2471–2481.

**Zhang H, Ransom C, Ludwig P & van Nocker S**. 2003. Genetic Analysis of Early Flowering Mutants in *Arabidopsis* Defines a Class of Pleiotropic Developmental Regulator Required for Expression of the Flowering–time Switch *Flowering Locus C*. Genetics 164: 347–358.

**Zhou Y, Sun X–D & Ni M**. 2007. Timing of Photoperiodic Flowering: Light Perception and Circadian Clock. Journal of Integrative Plant Biology. 49(1): 28–34.

**Zhebentyayeva TN, Swire–Clark G, Georgi LL, Garay L, Jung S, Forrest S, Blenda AV, Blackmon B, Mook J, Horn R, Howas W, Arús P, Main D, Tomkins JP, Sosinski B, Baird WV, Reighard GL & Abbott AG**. 2008. A framework physical map for peach, a model Rosaceae species. Tree Genetics & Genomes 4: 745–756.

## *Appendices*

# Appendix 1. Markers used for genotyping and mapping the $F_2$ populations

| Name | Forward primer | Reverse primer | Published in |
|---|---|---|---|
| ARSFL–007 | GCGCGCATAAGGCAACAAAG | GCGAATGGAATGACATCTTCTCT | Lewers *et al.* (2005) |
| BFACT–047 | | | Denoyes–Rothan |
| CFVCT010 | TGACAGAGACAAATGCATCACA | CGACGTTTGCCCTCTTTCTA | Monfort *et al.* (2006) |
| CFVCT030 | TCGATCGATTATCACCATGAA | CGTCCTCAACATCTCCTTCC | Monfort *et al.* (2006) |
| EMFn017 | TTTTCAAATTGTTACCCCATCC | CTAAAAATCCCCCAAATTGTGA | Sargent *et al.* (2004) |
| EMFn117 | ATCGGATCAACAAGCAAAGC | ATGGATGAGGGGAGAAGAGG | Sargent *et al.* (2006) |
| EMFn123 | CATTTCGGGCACACTTCC | AGACGGCAAAGAGACTCACC | Sargent *et al.* (2006) |
| EMFn153 | CTCGAGCTCCCTTTCTATCG | TGGCCAAATGTTCTCACTAGC | Sargent *et al.* (2006) |
| EMFn185 | GTAACGACGGCTGCTTCTCC | CGCTCGCTCTTATAAACTTCC | Sargent *et al.* (2006) |
| EMFv104 | TGGAAACATTCTTACATAGCCAAA | CAGACGAGTCCTTCATGTGC | Sargent *et al.* (2004) |
| EMFvi025 | TTGTGATCTCGTAGAAGGAGCA | GGGTTCCGTGAAACTAAACTTG | Sargent *et al.* (2003) |
| EMFvi133 | ATGCAGCTCAGAGAAAGGGTAG | AGTTGGGAAGAGGGAAGAAAAG | Sargent *et al.* (2003) |
| FAC–004d | GCCAATGTTCGATGTTTCACTA | TCCTTGGGTCGATCACATAAAT | Lewers *et al.* (2005) |
| Fvi020 | GAGTTTGTCATCCTCAGACACC | AGTGACCCAGAAACCCAGAA | Ashley *et al.* (2003) |
| Fvi6b | TCCTGATTCAACCACAAGAT | GTAACACTCATTGCTTCAGGTA | Ashley *et al.* (2003) |
| SCAR1 | AGAGAGAGAGAGAGAGTCAC | GGGTGAAACTGATTTCTTACC | Albani *et al.* (2001) |
| SCAR2 | GAAAAGTGAGGCGGATTTCG | CTTGAATTGTCTCCATTCCC | Albani *et al.* (2001) |
| SCAR3 | GGAAGGTCCTCGATATTCG | GAGATCGATAACGGTACC | Albani *et al.* (2001) |

As SCAR1 and SCAR3 are dominant markers, they had to have control primers to distinguish a failed PCR from a true null – mutant. The control primers for SCAR1 and SCAR3 were CACGCTTAAATAGGAGTTCG and GAGGAACGAAGAGAAACC, respectively.

## Appendix 2. An example of how to identify a positive colony.

| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | 7 | | 8 | | 9 | | 10 | | 11 | | 12 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | A01 | A02 | A03 | A04 | A05 | A06 | A07 | A08 | A09 | A10 | A11 | A12 | A13 | A14 | A15 | A16 | A17 | A18 | A19 | A20 | A21 | A22 | A23 | A24 |
| | B01 | B02 | B03 | B04 | B05 | B06 | B07 | B08 | B09 | B10 | B11 | B12 | B13 | B14 | B15 | B16 | B17 | B18 | B19 | B20 | B21 | B22 | B23 | B24 |
| B | C01 | C02 | C03 | C04 | C05 | C06 | C07 | C08 | C09 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 |
| | D01 | D02 | D03 | D04 | D05 | D06 | D07 | D08 | D09 | D10 | D11 | D12 | D13 | D14 | D15 | D16 | D17 | D18 | D19 | D20 | D21 | D22 | D23 | D24 |
| C | E01 | E02 | E03 | E04 | E05 | E06 | E07 | E08 | E09 | E10 | E11 | E12 | E13 | E14 | E15 | E16 | E17 | E18 | E19 | E20 | E21 | E22 | E23 | E24 |
| | F01 | F02 | F03 | F04 | F05 | F06 | F07 | F08 | F09 | F10 | F11 | F12 | F13 | F14 | F15 | F16 | F17 | F18 | F19 | F20 | F21 | F22 | F23 | F24 |
| D | G01 | G02 | G03 | G04 | G05 | G06 | G07 | G08 | G09 | G10 | G11 | G12 | G13 | G14 | G15 | G16 | G17 | G18 | G19 | G20 | G21 | G22 | G23 | G24 |
| | H01 | H02 | H03 | H04 | H05 | H06 | H07 | H08 | H09 | H10 | H11 | H12 | H13 | H14 | H15 | H16 | H17 | H18 | H19 | H20 | H21 | H22 | H23 | H24 |
| E | I01 | I02 | I03 | I04 | I05 | I06 | I07 | I08 | I09 | I10 | **I11** | **I12** | I13 | I14 | I15 | I16 | I17 | I18 | I19 | I20 | I21 | I22 | I23 | I24 |
| | J01 | J02 | J03 | J04 | J05 | J06 | J07 | J08 | J09 | J10 | **J11** | **J12** | J13 | J14 | J15 | J16 | J17 | J18 | J19 | J20 | J21 | J22 | J23 | J24 |
| F | K01 | K02 | K03 | K04 | K05 | K06 | K07 | K08 | K09 | K10 | K11 | K12 | K13 | K14 | K15 | K16 | K17 | K18 | K19 | K20 | K21 | K22 | K23 | K24 |
| | L01 | L02 | L03 | L04 | L05 | L06 | L07 | L08 | L09 | L10 | L11 | L12 | L13 | L14 | L15 | L16 | L17 | L18 | L19 | L20 | L21 | L22 | L23 | L24 |
| G | M01 | M02 | M03 | M04 | M05 | M06 | M07 | M08 | M09 | M10 | M11 | M12 | M13 | M14 | M15 | M16 | M17 | M18 | M19 | M20 | M21 | M22 | M23 | M24 |
| | N01 | N02 | N03 | N04 | N05 | N06 | N07 | N08 | N09 | N10 | N11 | N12 | N13 | N14 | N15 | N16 | N17 | N18 | N19 | N20 | N21 | N22 | N23 | N24 |
| H | O01 | O02 | O03 | O04 | O05 | O06 | O07 | O08 | O09 | O10 | O11 | O12 | O13 | O14 | O15 | O16 | O17 | O18 | O19 | O20 | O21 | O22 | O23 | O24 |
| | P01 | P02 | P03 | P04 | P05 | P06 | P07 | P08 | P09 | P10 | P11 | P12 | P13 | P14 | P15 | P16 | P17 | P18 | P19 | P20 | P21 | P22 | P23 | P24 |

If, for example, colony – PCR on pooled colonies had identified row five and column six on the plate 32 as positive, the colonies to pick from the 384 – well plate no. 32 would be the ones shown underlined from the plate 32.

## Appendix 3. Bin mapping; the first six sequences represent the bin set seedlings, the 7th is the *F. vesca* parent and the last the heterozygous $F_1$.

Polymorphism can be seen at nucleotide position 141.