QTL MAPPING AND MARKER-ASSISTED SELECTION IN BRASSICA AND ARABIDOPSIS

By

SAMODIMO NGWAKO

A thesis submitted to
The University of Birmingham
for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences
The University of Birmingham
August 2003

Synopsis

The study was aimed at applying molecular marker techniques to locate QTL and determine the efficiency of the marker-assisted selection. The research was done using *Brassica oleracea* and *Arabidopsis thaliana*. The Brassica DH lines represented a population of homozygous individuals while the F_2 and F_3 generations of Arabidopsis represented a segregating population. Marker-assisted selection was applied after the detection of QTL which allowed the identification of markers linked to the QTL and hence the selection for such markers.

In Brassica, 40 QTL were detected using the marker regression method. Between 1 and 6 QTL were located per trait, which individually explained 2-49% of the additive genetic variance. In Arabidopsis the marker regression method detected 23 QTL in the F_2, whereas 40 QTL were detected by the interval mapping method in the F_3 generation. 17 QTL mapping to similar positions and showing similar modes of action were detected by both methods. Alleles for various QTL were dispersed between parents in both crosses.

The efficiency of MAS was determined using various approaches, based on the number of top ranks, number of lines in a group, phenotypic value and as the ratio between response based on MAS and response obtained in the F_3 by applying phenotypic selection to the F_2 generation. The MAS gave generally better response compared to phenotypic selection, particularly when heritability was low. MAS for single QTL was always more effective while multiple QTL and QTL showing linkage posed some practical problems in MAS applications. Overall, MAS has to be applied in conjunction with phenotypic selection to get best results as QTL of minor effect cannot be tackled through marker/QTL associations.

Dedication

To members of my family for their support and encouragement during my studies

Acknowledgements

I would like to thank my supervisor Dr. H. S. Pooni, for his support and guidance on practical and theoretical issues throughout this PhD project. I also thank Professor M. J. Kearsey for his advice on many occasions on statistical, quantitative and laboratory issues.

I am very grateful to Sue Bradshaw and all staff at the glasshouses for their help during the practical work associated with growing Brassica and Arabidopsis plants. I am also very indebted to Dr. Tim Wilkes for his help, guidance and advice during the molecular work associated with Arabidopsis in the laboratory. I greatly appreciate Dr. Z. Luo for his help in the analysis of *Arabidopsis thaliana* F_3 families using the interval mapping method. I would also like to thank my wife Monde and my daughter Tshiamo for helping and encouraging me throughout my PhD studies.

Finally, I express my sincere thanks to Botswana College of Agriculture (an associate institution of the University of Botswana) for funding the PhD project and making my stay in Birmingham (United Kingdom) a success.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPENDICES

ABBREVIATIONS

σ^2_B	Between-families component of variance
σ^2_w	Within-families component of variance
a	additive effect
AFLPs	Amplified Fragment Length Polymorphisms
CI	Confidence Interval
cM	CentiMorgans
Col	Columbia, ecotype of Arabidopsis thaliana
d	dominance effect
df	Degrees of freedom
DH	Doubled haploid
ems	Expected mean squares
F_t	F-statistic
h_b^2	Broad sense heritability
h_f^2	Heritability of the means
h_n^2	Narrow sense heritability
Ler	Landsberg erecta, ecotype of Arabidopsis thaliana
LG	Linkage group
MAS	Marker-assisted selection
MS	Mean square
MS_B	Mean square between families
MS_W	Mean square within families
PCR	Polymerase Chain Reaction

QTL	Quantitative trait loci
r	Correlation
RAPDs	Random Amplified Length Polymorphisms
RILs	Recombinant Inbred lines
RFLPs	Restriction Fragment Length Polymorphisms
T_t	Student t-test
SSRs	Simple sequence repeats or microsatellites
V_A	Additive genetic variance
V_E	Environmental component of within family variance
$\chi^2_{(df)}$	Chi-squared

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 General

The application of biotechnology holds an enormous potential to improve the crop performance. Through molecular procedures, the plant breeder can manipulate alleles so as to obtain recombination of genes for useful characters. This is accomplished largely through hybridization, utilising normal plant reproductive pathways and plant breeding procedures. Plant breeding procedures were developed from Mendelian genetic principles. In contrast to earlier genetic research, which focussed on an extension of Mendelian genetics, current genetic research is concentrated on the biochemical and molecular aspects of the genetic process. Molecular genetics provides explanation on the molecular level and a more comprehensive understanding of the genetic process (Poehlman and Sleper, 1995). It is essential that plant breeders understand the potentials and limitations of the new technology, and as it unfolds, that they employ it appropriately for enhancement of existing breeding procedures.

A conventional breeding program involved crossing two plants with contrasting genotypes followed by selection of the superior recombinants from among the several segregation products. Such a procedure is laborious and time consuming, involving several crosses, several generations, and careful phenotypic selection. With the advent of DNA marker technology, several types of DNA markers, and molecular breeding strategies are now available to plant breeders and geneticists, helping them to overcome many problems faced by conventional breeding (Kumar, 1999).

Lande and Thompson (1990) point out that molecular genetic technique can never replace traditional methods of agricultural improvement, but instead they should be integrated to obtain maximum improvement in the economic value of domesticated population. The reason given by the authors are that most characters of economic importance are quantitative (influenced by numerous loci throughout the genome that often have small effects). As such, genes with small effects are difficult to map precisely and there may be practical problems of engineering polygenic traits once the genes are identified at the molecular level. The other reason given by Lande and Thompson (1990) is that the high mutability of polygenic characters guarantees genetic variation will arise within populations that can be selected to improve on whatever previous gain have been made.

The difficulty of manipulating quantitative traits is related to their genetic complexity, principally the number of genes involved in their expression and the interactions between genes (epistasis). Because several genes are involved in the expression of polygenic traits, they generally have smaller individual effects on the phenotype. This implies that several regions (QTL) must be manipulated at the same time in order to have a significant impact, and that the effect of individual regions can be easily identified. For this reason, repetitions of field tests are required to characterize accurately the effects of QTL and to evaluate their stability across environments. Although significant QTL effects should be detected across environments, the evaluation of the QTL by environmental interactions (GxE) remains a major constraint on the efficiency of marker-assisted selection (Ribaut and Hoisington, 1998).

Since molecular markers are almost unlimited in number and have a genome-wide distribution, this has renewed interest in the use of molecular markers to facilitate the identification of and selection for individual quantitative trait loci (QTLs) that control economically important traits. The steps in marker-assisted selection (MAS) consist of identifying associations between marker alleles and QTL, or ideally of estimating the contribution of marker loci to the phenotypic value of the trait by the association with the markers (MQTL effects). Marker effects are then combined with phenotypic information to rank individuals through an index and develop desired lines or populations (Xie and Xu, 1998; Lande and Thompson, 1990; Dudley, 1993; Gimelfarb and Lande, 1994a).

In a nutshell, the success of a molecular breeding programme depends on two factors: detection of QTL controlling a trait and their selection/manipulations using molecular markers. The present study, therefore, deals with these two aspects. In the study, the QTLs are determined using segregating populations of Arabidopsis and Brassica and later the same marker data are used to determine the efficiency of marker-assisted selection that can be applied to these QTLs.

1.2 The model species

1.2.1 *Arabidopsis thaliana*

Arabidopsis thaliana (L.) Heynh, is a small herbaceous annual weed of the Brassicaceae (Cruciferae) family with a broad natural distribution throughout Europe, Asia and North America (Meyerowitz, 1994). The main features that make the plant suited for molecular genetic analysis are its short generation (5 to 6 weeks) and its small size, which provides a limited space requirement. Arabidopsis is predominantly an inbreeding species, and it is relatively easy to cross by hand and inbreed artificially (Bowman, 1994). This, together with the ease of obtaining hundreds of seeds by manual crossing, makes handling the plant for genetic processes not very labour intensive (Koornneef, 1994). Another genetic advantage of Arabidopsis is the low chromosome number (2n=10), which allows a more efficient linkage analysis than in plants with more chromosomes. However, the small genome makes the examination of the chromosomes, their activity, segregation, pairing and aberrations in number or structure - that is cytogenetics difficult (Heslop-Harrison and Maluszynska, 1994).

Arabidopsis thaliana has a small genome (114.5 Mb/125 Mb total) that has been sequenced in the year 2000 (Arabidopsis Genome Initiative, 2000). The genome size is unusually small for a flowering plant and has remarkably little dispersed DNA. These properties facilitate a series of different types of experiments in molecular genetics and have allowed facile cloning of Arabidopsis genes by methods that would be difficult or impossible if the genome were larger or more typical in its content of repetitive sequences. Despite the unusual size and structure of the Arabidopsis genome, the structure of individual genes, the structure of

chromosomes, the genetic properties, and the overall compliment of genes in the genome are typical of other flowering plants (Meyerowitz, 1994).

The study of natural variation has proved useful for analyzing the genetic basis of some developmental processes in the model species *Arabidopsis thaliana*. Important contributions to their genetic dissection have been made by analyzing the progeny of inter-crosses involving ecotypes that differ in specific traits. The genetic control of flowering time has been studied generally through the isolation and characterization of both monogenic mutants and by the analysis of natural variation expressed as differences between ecotypes (Koorneef et al. 1994). These flowering-time genes include among others, *FLOWERING LOCUS C* (*FLC*; Michaels and Amasino, 1999; Poduska et al. 2003), *FLOWERING ALTERED* (*FLA*; Lee et al. 1993), also named *FRIGIDA* (*FRI*; Clarke and Dean, 1994; Johanson et al. 2000), *CONSTANS* (*CO*; Puterill et al. 1995; Suarez-lopez et al. 2001), *CHRYPTOCHROME2* (*CRY2*; Guo et al. 1999), *LUMIDEPENDENS* (*LD*; Lee et al. 1994a).

In addition, quantitative trait loci (QTL) analysis has been shown to be useful in identification of novel genes involved in some developmental processes such as those of *EARLY DAY-LENGTH INSENSITIVE* (*EDI*), *FLOWERING F* (*RLF*), *FLG*, and *FLH* (Alonso-Blanco et al. 1998) which affect flowering time; the *ROSETTE LEAF NUMBER* (*RLN1-RLN5*) loci (Clarke et al. 1995), which affect vernalization responsiveness; at least 11 QTL associated with several floral traits (Juenger et al. 2000); 110 QTL for inflorescence developmental traits (Ungerer et al. 2002); and 12 QTL affecting light and hormone responses (Borevitz et al. 2002) among others.

Critical regulatory genes that control meristem identity have been identified in Arabidopsis by mutations that disrupt normal inflorescence or flower development. Genes known to be important for meristem identity during the reproductive phase are *LEAFY* (*LFY*), *APETALA 1* (*AP1*), *APETALA 2* (*AP2*), *CAULIFLOWER* (*CAL*), *TERMINAL FLOWER 1* (*TFL1*) and *TFL2* (Irish and Sussex, 1990; Bowman et al. 1993; Shannon and Meeks-Wagner, 1993; Bradley et al. 1997; Larsson et al. 1998).

There are a number of late flowering mutants of Arabidopsis, most of which have been isolated in the early flowering ecotype Landsberg *erecta*. The mutations have been characterized on the basis of their vernalization responsiveness and their response to photoperiod. The groups includes one that responds to vernalization (*fca*, *fpa*, *fy* and *fve*), one group which is daylength responsive but not vernalization responsive (*fd*, *fe*, *fwa* and ft) and another group which is completely daylength neutral (*fg*, *fha* and *fb*) (Koorneef et al. 1991; Burn et al. 1993).

Quantitative genetic analysis permits investigations to examine the functions of previously identified developmental regulatory genes in new phenotypic contexts. The identification of naturally occurring variation allows the understanding of the developmental genetics and provides an opportunity to investigate the underlying genetic mechanisms that regulate developmental programs in plant architecture (Juenger et al. 2000; Mackay, 2001).

1.2.2 *Brassica oleracea*

Brassica oleracea belongs to the family Cruciferae and includes agronomically important species of vegetables such as cabbage, broccoli and cauliflower. *Brassica oleracea* is a diploid species (2n = 18, C genome), and it is closely related to other diploid species *B. nigra* (2n = 16, B genome) black mustard and *B. rapa* (syn. Campestris 2n = 20, A genome) turnip, rape-seed and oriental vegetables (Quiros, 2001).

The genome relationships among cultivated diploid and derived amphidiploid species were elucidated by U (1935). The three basic diploid species (*B. oleracea*, *B. nigra* and *B. rapa*) have produced the three amphidiploids, *B. carinata* (2n = 34, genomes BC), *B. juncea* (2n = 36, genomes AB) and *B. napus* (2n = 38, genomes AC) (Quiros, 2001). In general, the three diploids are considered to be mutually and partially homologous and presumably derived from a common ancestral genome. Results from molecular analyses not only confirmed the origin of the amphidiploids but also suggested that A and C genomes were closely related to each other forming a single lineage (McGrath and Quiros, 1991). The B genome is genetically distant from both A and C genomes forming a separate lineage (Song et al. 1990; Warwick and Black, 1991). The genomes of the diploids reveal striking conservation of content, although chromosome duplication and translocation have occurred during divergence (Lagercrantz & Lydiate, 1996).

Substantial advances have been made in understanding the genome organization of *B. oleracea*, and of other Brassica species, during the past few years by employing DNA-based markers. Several linkage maps have been produced using different sets of probes, as well as

different segregating populations (Slocum et al. 1990, Bohuon et al. 1996, Ramsay et al. 1996; Camargo et al. 1997; Sebastian et al. 2000; Howell et al. 2002). These developments have continued to facilitate the understanding of the relationships between various Brassica species.

1.2.3 Relationship between Arabidopsis and Brassica

The genus Arabidopsis shares membership of the Cruciferae family with the Brassica genus that includes several crop species. Investigations of many biologically important processes in Arabidopsis have identified genes fundamental to plant development, and this discovery has run hand in hand with the genetic studies in several Brassica species, where relatively comprehensive RFLP maps exist. Because the two plant species are related, it is hoped that any results for the genetic control of basic biological processes found in Arabidopsis can be related to Brassica. The comprehensive RFLP maps within the two species will facilitate the movement of cloned genes from Arabidopsis to Brassica commercial varieties (Teutonico and Osborn, 1994). Eleven regions of conserved homology have been found between the nuclear genome of Arabidopsis and Brassica. Because of the Arabidopsis smaller genome size and lower level of repetitive DNA, map-based cloning of the orthologous genes between the two species will be much easier in Arabidopsis compared to Brassica (Kowalski et al., 1994a).

Genetic studies have revealed that QTL controlling morphological traits (e.g. flowering time, lamina, stem length and disease resistance) in Brassica species have been mapped to the homologous regions of *Arabidopsis thaliana* (Osborn et al. 1997, Bohuon et al. 1998, Lan et al. 2000, Lan & Paterson, 2001, Kole et al. 2001). Some of the genes that have been cloned

in Arabidopsis may have homologues that regulate flowering in Brassica species. The two regions influencing flowering in *B. nigra* (on LG2 and LG8) have been found to be homologous to the CONSTANS (*CO*) region in Arabidopsis (Lagercrantz et al. 1996). These regions also carry quantitative trait loci (QTL) in the A genome of *B. napus*, and show large scale collinearity between regions of chromosomes O2, O3 and O9 of *B. oleracea* (Rae, et al. 1999). *VFR2* is a quantitative trait locus (QTL) controlling vernalization-responsive flowering time in *B. rapa* (2n = 2x = 20) (Teutonico & Osborne, 1994) and is found to be homologous to a region in *B. napus* (the hybrid of *B. rapa* and *B. oleracea*, 2n = 4x = 38) and a region at the top of chromosome 5 in Arabidopsis (Osborne *et al.* 1997). *FLC* is a gene from the top of chromosome 5 of Arabidopsis that also controls vernalization-responsive flowering time (Koornneef *et al.* 1994) and was found to have homology with *VFR2* (Kole *et al.* 2001). *FLC* was subsequently used to identify five similar sequences in *B. napus* that were each shown to delay flowering when individually expressed in Arabidopsis and it was found to delay flowering when expressed in *B. napus* (Tadege *et al.* 2001).

The collinearity between the genomes of the model species *A. thaliana* and Brassica genomes can increase the possibility of combining knowledge gained in the two genera. The identification of collinear regions will enable genetic information and molecular resources to be shuttled between the two genera (Lukens et al. 2003). Studies of genome conservation are also extended to other members of the dicots. Analogous studies revealed the existence of a network of micro-synteny between tomato and Arabidopsis that enabled the mapping of orthologous genes in Arabidopsis and tomato (Ku et al. 2000, Rossberg et al. 2001). Syntenous relationship has also been found between the linkage groups AZ, J and L of soybean and Arabidopsis chromosomes (Grant et al. 2000).

1.3 Genetic markers

Genetic markers can be regarded as specific locations on a chromosome, which serve as landmarks for genome analysis (Staub and Serquen, 1996; Young, 1994). There are basically two types of genetic markers: morphological and molecular markers. Morphological markers can be used as genetic markers provided their expression is reproducible over a wide range of environments. The inheritance of morphological markers can be monitored visually without specialized biochemical or molecular techniques. The main difficulties associated with the use of morphological markers in linkage mapping are:

a) Most of the morphological markers segregate as dominant-recessive alleles, thus only homozygous recessive genotypes can unambiguously be deduced from the phenotype.
b) Usually phenotypic markers are expressed at a specific growth stage.
c) Frequently morphological markers exhibit strong epistatic non-allelic interactions. This limits the number of segregating markers that can be unequivocally scored (Tanksley, 1993).
d) They are affected by environmental conditions.

Molecular markers are similar to morphological markers in that they are inherited in a Mendelian fashion. However, they have many advantages that have permitted the rapid advance of genetic mapping. Their main features are:

a) Not environmentally regulated and therefore unaffected by the conditions in which the plants are grown.

b) Detectable at all stages of plant growth and can extract DNA even at early stage.

c) Segregating loci are distributed across the entire genome. Although the level of polymorphism is different among species, it has been possible to identify enough genetic variation for DNA-based markers for most of the species in which the linkage maps are constructed.

d) Co-dominance. This enables all genotypes to be deduced from their phenotype in any generation.

e) Normally do not exhibit epistatic effects and alternate alleles do not produce changes in the phenotype (Tanksley, 1993)

Molecular markers are divided into biochemical and DNA markers. Biochemical markers are those that reveal polymorphism at the protein level. The most commonly used protein markers are isozymes, which are variant forms of the same enzyme (Vodenicharova, 1982). Protein markers actually reveal polymorphism in the gene sequence and are co-dominant markers. The proteins are produced as a result of gene expression, which can be separated by electrophoresis to identify the alleles.

DNA markers can be classified into hybridization-based markers, PCR-based markers and DNA chips and sequencing based markers. Hybridization-based polymorphism includes RFLPs (Sambrook et al. 1989) and VNTR loci (variable number of tandem repeats) (Rogstad, 1993; Weising, et al. 1992; Weising, et al. 1998). In this case probes such as random genomic clones, cDNA clones, and probes for microsatellite and minisatellite sequences are hybridized to filters containing DNA, which has been digested with restriction enzymes. The polymorphisms in VNTR loci are due to difference in the number of repeats,

while RFLPs are generated due to events such as point mutations, deletions or translocations that affect the distance between restriction sites. DNA marker analysis can be carried out at any stage of the life cycle of an organism and from almost any tissue including herbarium and mummified tissue.

PCR-based markers include random amplified polymorphism DNA (RAPD), amplified fragment length polymorphism (AFLPs), simple sequence repeats SSRs (or microsatellites), multiple arbitrary amplicon profiling (MAAP), arbitrary primed PCR (AP-PCR) and DNA amplification fingerprinting (DAF). PCR-based markers are cheaper and safe and there are more markers per unit of DNA (Westman and Kresovich, 1997). The steps involved in PCR reaction are template denaturation, primer annealing and enzymatic extension. PCR-based markers have opened new doors for genome manipulation since their use allows easier sampling because of the small amount of tissue required and faster DNA preparation done at minute quantity of template DNA. Also, they are more efficient in handling large sample sizes (Ribaut and Hoisington, 1998).

DNA chips are simply glass surfaces bearing arrays of DNA fragments at discrete sites, at which fragments are available for hybridization (Gerhold et al., 1999; Gibson, 2002). The DNA spots on the chip are hybridized to a complex sample of fluorescently labelled DNA or RNA. This technology is being used mainly for human genetics (Fan et al., 2000; Mei et al., 2000) and exploitation of its applicability for plant genetics (Rafalski, 2002) is initiated. Single nucleotide polymorphisms (SNPs) are DNA sequence variations that occur when a single base is altered so those different individuals may have different bases in the position

(Brookes, 1999). SNP markers could be used for all the usual applications in plant molecular biology, including construction of high resolution genetic maps (Cho et al., 1999).

The other advantages of molecular markers over traditional methods are that selection is not carried directly on the trait of interest but on markers linked to the trait or QTL and markers are unaffected by conditions in which the plants are grown. Also, the number of lines to be tested is reduced.

1.3.1 Isozymes

Isozymes were the first molecular markers used in genetics and breeding. The major advantages of isozymes are that they are easily analyzed by electrophoresis and they are expressed at an early stage of the plant life (Weber and Wricke, 1994). Isozyme marker loci are largely co-dominantly inherited, allowing complete classification of genotypes in segregating populations and they are less likely to have pleiotropic effects on the trait of interest (Tanksley, et al. 1992). Their major limitation is that the number of polymorphic loci is very limited within a gene pool. Isozymes are also phenotypic markers, in that they can be affected by the tissue and growth stage of the plant. Tissues need also to be fresh or properly treated before protein extraction, or erroneous results may be generated (Godwin et al. 2001).

1.3.2 Restriction Fragment Length Polymorphisms (RFLPs)

According to Poehlman and Sleper (1995), RFLPs can be defined as 'different fragment lengths of restriction endonuclease digested DNA detected by a defined probe between individuals'. In RFLPs, the restriction enzymes are used to cut DNA at specific sites and the

DNA is segmented into discrete parts. Restriction sites consist of specific nucleotide sequences. RFLPs posses the following properties: they are co-dominant, have a high allelic variation, most are phenotypically neutral, they generally show no epistasis, they can be tested in all tissues at any stage, they are not pleiotropic and they are unaffected by environmental conditions (Beckmann & Soller, 1983).

The number of steps for RFLP analysis from plant tissue include; isolation of genomic DNA, restriction enzyme digestion, gel electrophoresis, southern blotting, labelling of the probe, hybridization of the probe to the DNA with nylon membrane and auto-radiography (Poehlman and Sleper, 1995; Kochert, 1994). RFLPs are useful for cultivar identification, genetic mapping, germplasm evaluation and as indirect selection criteria. If the marker genes are used for indirect selection of other traits the initial population must be polymorphic for the markers and the economically important genes and there has to be strong linkage disequilibrium between markers and relevant QTL.

There are considerable disadvantages of the RFLP technology, not least of which is the low level of polymorphism seen within some species such as groundnut (Kochert et al., 1991). The generation of RFLP data is time-consuming, particularly with single copy probes, and the assay is one of the most costly to perform as many steps are involved and radioactivity is required. Large quantities of DNA are also required, generally 5-10 μg per digest and, as a result, whole plants would be needed for DNA extractions. Probes also need to be distributed to collaborating labs and, overall, the generation of RFLPs is moderately technically demanding (Godwin et al. 2001).

1.3.3 Random Amplified Polymorphic DNA (RAPD)

The random amplified polymorphic DNA (RAPD) was developed for rapid detection of polymorphism among individuals using a single primer of arbitrary sequence (usually 10 nucleotides) and the PCR (polymerase chain reaction) mediated amplification of random genome fragments (Poehlman and Sleper, 1995). RAPDs require no probe DNA and no advance information about the genome of the organism. The random primers are tried singly or in pairs in PCR reaction, and since the primers are so short, they often anneal to the template at multiple sites.

RAPDs depend on the recognition of single short oligonucleotide primers and similar sequences that are opposed to each other at distances close enough for the intervening sequence to be amplified by PCR. This makes RAPDs most valuable for finding new markers that are tightly linked with a specific locus. Because of the availability of random oligonucleotides and the relatively easy assay to look for linkage, it is a simple matter to screen many loci rather rapidly (Burr, 2001).

An important feature of RAPDs and PCR based methods is that the presence of a fragment is dominant to its absence. In other words, if one allele (+) supports amplification but the alternative allele (-) does not, then DNA from the genotype +/+ and +/- will support amplification equally well, whereas DNA from the genotype -/- will not support amplification. The allele + is therefore dominant to the - allele in regard to the corresponding RAPD fragment (Hartl, 2000).

Like isozymes and RFLP markers, RAPDs have been used in plants for constructing genetic maps in Arabidopsis and Brassica, estimating genetic relationships and tagging traits such as disease resistance (Cheung et al., 1997; Kuittinen et al., 1997; Burr, 2001). The advantages of RAPDs are that they are easy to generate, fast, multilocus and do not require radioactivity. Hence, they have many suitable qualities for use in a lab with little equipment except for a PCR thermal cycler, gel electrophoresis and photographic equipment.

However, there are some reliability problems and most of the markers generated are dominant. There is also a lack of cross-transferability, and it must be acknowledged that in some cases, fragments that are the same length may not necessarily be the same sequence. Some of these problems can be overcome by cloning and partially sequencing the fragments, to turn these into sequence characterised amplified regions (Paran and Michelmore, 1993), which are usually more robust than RAPDs. This does, however, lose the attraction of multilocus markers for diversity analysis, and does not always overcome the problem of dominance.

1.3.4 Microsatellites

Microsatellites are simple sequence repeats (SSRs) of 2-6 nucleotides. They are abundant, dispersed throughout the genome and show higher levels of polymorphism than other markers (Hardy et al. 2003). Microsatellites are PCR-based assays and it is sufficient to merely separate the amplification products by electrophoresis to observe the results. This consideration reduces the time required to obtain a result compared with methods that are based upon southern blotting.

Microsatellites are very powerful markers in that they are single locus, co-dominant and mutli-allelic. They do not require radioactivity for detection, although this is sometimes used on polyacrylamide gels to detect accurately alleles which differ by one repeat unit (as little as 2 pb) (Godwin et al. 2001). They are robust and easily exchanged between labs, and multiplex reactions can be run to speed up the assay, where the products have non-overlapping size ranges. SSRs are easier to use than RFLP's owing to the small amount of DNA required and the ability to automate assays. These features, coupled with the ease of detection, have made them useful molecular markers (Holton, 2001, Ribaut & Hoisington, 1998).

The greatest disadvantage of SSRs is the initial cost of finding and sequencing loci, because although they are ubiquitous, there needs to be considerable effort put into their isolation, hence they have a higher cost of establishment than other systems. However, this process is overcome by using probes from related species. SSRs also have limited use for phylogenetic analysis because of their high mutation rate (Godwin et al. 2001).

1.3.5 Amplified Fragment Length Polymorphic DNA (AFLP)

Amplified fragment length polymorhisms (AFLPs) are based on the amplification of restriction fragments generated by specific enzymes and oligo-nucleotide adapters of few nucleotide bases. The selective amplification is achieved by the use of primers that extend into the restriction fragments, amplifying only those fragments in which the primer extensions match the nucleotides flanking the restriction site. The technique involves three steps: (a) restriction of the DNA and ligation of oligo-nucleotide adapters, (b) selective

amplification of sets of restriction fragments, and (c) gel analysis of the amplified fragments (Vos, et al. 1995).

The AFLP method generates a large number of restriction bands facilitating the detection of polymorphisms. The number of DNA fragments that are amplified can be controlled by choosing the different band numbers and composition of nucleotide adapters. The high reproducibility, rapid generation and high frequency of identifiable AFLP polymorphisms make them an attractive technique for identifying polymorphism and for determining linkages by analysing individuals from segregating populations (Vos et al., 1995). The major limitations of AFLPs are that they are expensive to generate and bands are detected by silver staining, fluorescent dye or radioactivity (Mohan et al. 1997).

1.4 Linkage maps

The recent developments in molecular genetics, by which large numbers of markers are being generated, have caused a revival of the interest in classic genetic mapping. Scientists are constructing genetic linkage maps composed of DNA markers for a wide range of species. Detailed genetic maps have been constructed for several Brassica species using predominantly RFLPs, but also microsatellites, AFLPs and RAPDs (Cheung et al., 1997). All types of DNA markers detect sequence polymorphism and monitor the segregation of a DNA sequence among progeny of a genetic cross in order to construct a linkage map. The development of genetic maps based on DNA markers provide the essential tools in understanding molecular analysis of quantitative inheritance and comparative mapping.

In comparative mapping, the mapping information of one taxon is used to predict the linkage relationships in closely related or distant taxa by making use of a common set of DNA hybridization probes. Linkage maps have been used to compare the genomes of Brassica and Arabidopsis (Teutonico and Osborn, 1994; Lukens et al. 2003), tomato and potato (Bonierbale et al. 1988), maize and sorghum (Hulbert et al. 1990), and as well as rice, maize and wheat (Ahn et al. 1993). Comparative mapping is effective in identifying homologous loci and collinear chromosomal segments in species, provided common markers are used. For a genus such as Brassica, where the cultivated species have a close genomic relationship, the examination of the genome structure and organization is of great interest, not only in terms of locating QTLs, but also in classifying Brassica evolution, taxonomy and synteny with related species and genera, and not the least with Arabidopsis (Sebastian et al. 2000).

In order to construct a linkage map from DNA marker data statistical software packages capable of running chi-squared contingency table analysis are required (Young, 2001). This statistical test determines two-point linkage between markers, which can then form a basis for constructing linkage groups. A number of programs have been written that construct genetic linkage maps. Two of the most commonly used are MAPMAKER and JOINMAP. MAPMAKER is a program specifically written to implement maximum likelihood techniques with interval mapping, as proposed by Lander and Botstein (1989). JOINMAP (Stam, 1993; van Ooijen and Voorrips, 2001) is a computer program used to construct an integrated genetic map from different sets of mapping populations. The integration of genetic maps relies on the segregating markers common to the maps. Map integration can resolve variation between common markers on different maps and enable comparisons between maps to be undertaken more easily (Sebastian et al. 2000).

1.5 Marker-assisted selection (MAS)

Marker-assisted selection (MAS) can be regarded as the use of artificial selection and molecular markers combined together. Marker-based selection is an indirect method of selection for specific DNA sequences, which are part of the genes coding for the economically important traits. Although the cost effectiveness of MAS is widely debated, the usefulness of QTL mapping for finding new favourable alleles is hard to dispute. Tanksley et al. (1989) stressed how marker-assisted backcrossing could be used to minimize linkage drag and greatly speed up the development of near-isogenic lines.

What QTL mapping adds to marker-assisted selection is knowledge about the distribution of favourable alleles between parent and donor inbred lines and their progeny, estimates of gene effects without restrictive assumptions about their genetics and marker loci linked to the genes to be selected (Knapp, 2001). Once the alleles are found they must be introgressed from donor to elite inbred lines. MAS can be used to introgress the new favourable alleles through backcross breeding (Tanksley et al. 1989) or through pedigree or other variants of inbred line-breeding methods (Lande and Thompson, 1990; Lande, 1992).

Stuber and Edwards (1986) showed that marker-facilitated genotypic selection was effective for manipulating quantitatively inherited traits in two corn populations they studied. They noted that one generation of genotypic selection for yield, ear height, and ear number based on marker loci representing no more than 40% of heritable variation was as effective as one generation of mass selection. They concluded that increasing the number of markers to more

thoroughly cover the genome should increase the effectiveness of marker-facilitated selection. However, this must be accompanied by the detection of more QTL controlling the traits of interest.

Lande and Thompson (1990) observed that there are three practical considerations that limit the potential utility of MAS in applied breeding programs and these are; (i) the number of molecular marker loci necessary for the existence of significant associations (linkage disequilibrium) with the QTLs, (ii) sample sizes needed to detect QTLs for traits with low heritability, and (iii) sampling errors in the estimation of relative weights in the selection index combining molecular and phenotypic information. They noted that the potential efficiency of marker-assisted selection on a single trait utilizing a combination of molecular and phenotypic information, relative to standard methods of phenotypic selection, depends on the heritability of the character, the proportion of the additive variance associated with the marker loci, and the selection scheme.

Lande and Thompson (1990) proposed a method of marker-assisted selection that employs multiple regression of the phenotype on markers to identify a set of markers associated with QTLs as well as to estimate marker effects. The main conclusion from their deterministic analysis was that MAS based on an index incorporating marker effects together with phenotype yields better responses than selection based strictly on phenotype, provided there are sufficient markers and the population size is very large. They also noted that the use of phenotypic information from relatives reduces the relative efficiency of MAS, but the amount of reduction depends on family size. Unless family sizes are very large there is still opportunity for substantial increase in the efficiency of selection through the use of

molecular markers. Even with large families, the relative efficiency of MAS may be greatest if there are common family environmental effects, e.g. strong maternal effects on full sibs. Knapp (2001) also observed that the efficiency of MAS increases as heritability decreases, while the probability of finding a given QTL decreases as heritability decreases for a given sample size.

Moreau et al. (1998) evaluated the efficiency of marker-assisted selection (MAS) based on an index incorporating both phenotypic and molecular information with an analytical approach that takes into account the size of the experiment. They considered a population derived from a cross between homozygous lines, and studied the relative efficiency of MAS compared with selection based only on phenotype in the first cycle of selection. They used population sizes of 100, 300 and 500 individuals and they observed that population size and heritability were the key parameters of MAS efficiency. They found that for a given population size, the relative efficiency of MAS was high at low heritabilities.

Whittaker et al. (1995) developed a method for using information on the location of markers to improve the efficiency of marker-assisted selection in a population produced by a cross between two inbred lines. They observed that the method was closer to mapping QTL than the selection index approaches to MAS described by other authors. The authors also used computer simulation to compare map-based marker-assisted selection (MBMAS) with phenotypic selection and the approach to MAS detailed by Gimelfarb and Lande (1994a). The method used was based on the idea of interval mapping, first introduced by Lander and Botstein (1989) for a cross between two inbred lines. Whittaker et al. (1995) confirmed that the advantage of MAS declines as heritability increases, being high with low heritabilities.

Also, they found out that using a marker map gives at best a slight improvement on the simple regression on markers approach. They worked with simulated populations of 100 and 400 individuals.

In most evaluations of marker-assisted selection population size and heritability appear to be the key parameters that determine the efficiency of selection. Compared with analytical approach, simulation models have been used and have proven useful in marker-assisted evaluation, because they can be designed to be closer to real conditions of selection, and they allow one to evaluate the efficiency of the method for many successive generations of selection. Edwards and Page (1994) used computer simulation to make comparisons of gains expected from marker-assisted selection and phenotypic recurrent selection (PRS). Their objective was to examine the effects of the number of loci that are involved in the inheritance, distance between the markers and QTLs, and the use of single versus flanking markers in determining the usefulness of MAS for plant breeders in maize (*Zea mays* L).

Edwards and Page (1994) found that population size did not affect the rate of gain from selection. The average responses of the various population sizes were constant, due to the use of constant selection intensity of 10%. The population sizes used were 50, 100, 300 and 500 individuals. Smaller population exhibited a greater variance in the rate of responses across replicates of the selection process. They also observed that differing gene numbers did not affect the rate of response to selection and that distance between markers and QTLs was the most important factor affecting the gain achieved using MAS. They concluded that tight marker-QTL linkages were needed in order to achieve appreciable response from MAS with single markers. Relative to phenotypic recurrent selection (PRS), they further observed that

MAS produced rapid responses early in the selection process, but the rate of these responses diminished greatly within three to five cycles.

According to Gimelfarb and Lande (1994a), any method of selection that makes use of genetic markers requires markers that are associated with QTLs and genotypic values of the trait that are associated with each marker (marker effect). Gimelfarb and Lande (1994a) used simulations to investigate the efficiency of MAS as affected by several factors including total number of markers in the genome, number of markers contribution to the index, population size and heritability of the character. They observed that selection based on genetic markers can effectively utilize the linkage disequilibrium between genetic markers and QTL created by crossing inbred lines. Also, they observed that selection was more efficient if markers contributing to the index are re-evaluated each generation than if they are evaluated only once. Increasing the number of markers in the genome as well as the number of markers contributing to the index was found not to give higher efficiency of selection. Moreover, too many markers may result in a weaker response to selection. Population size was shown to be the most important factor affecting the efficiency of MAS.

1.6 QTL location and analysis

1.6.1 Importance of QTL

A QTL is a region of any genome that is responsible for variation in the quantitative trait of interest. Quantitative traits show continuous variation because phenotypes in the segregating generations do not fall into discrete classes as happens for monogenic-qualitative traits and in many cases the distribution approximates to a bell-shaped curve of the Normal distribution. The continuous nature of the variation for quantitative traits is due to there being only small differences between genotypes and an additional varying effect from non-genetic, or environmental sources which causes the phenotypic classes to merge with each other. Many important characteristics in the fields of medicine, agriculture, evolution and sociology show quantitative variation so there is considerable interest and activity in the study of quantitative genetics (Kearsey and Pooni, 1996; Mackay, 2001).

Quantitative traits do not afford the use of classic Mendelian rules to study and predict their genetics, as the phenotype is not particularly informative of the genotype. So geneticists, having been unable to benefit from the classical Mendelian ratios, resorted to the use of statistical tools in order to explain and quantify the components of quantitative variation. It is important to study QTL because of the fundamental knowledge gained from understanding how many loci govern a given character, what effect individual genes have, how these genes interact, how heritable they are, and what impact the environment has on the trait. The creation of a comprehensive genetic map for a plant species, gained from these investigations, will allow the plant breeder to select for particular genes based on the gene's linkage to specific markers.

1.6.2 QTL detection and methods for location

QTL detection can be undertaken by looking for associations between the quantitative trait and the marker alleles segregating in the population. A number of statistical approaches can be used to identify associations between the trait and particular markers, the technique used depending on the type of population. A strong association between the genotype at a marker locus and difference in the trait score indicates that there is a QTL in the vicinity of the marker. The statistical power of the approach will depend on the heritability of the trait and the size of the individual QTL effects, but it is now accepted that there is generally a very large confidence interval associated with the location of individual QTL.

Some of the major statistical methods used to map QTLs are; single marker analysis (Soller & Brody, 1976; Edwards et al. 1987), marker regression (Kearsey and Hyne, 1994), multiple regression (Haley and Knott, 1992), interval mapping (Lander and Botstein, 1989) and composite interval mapping (Jansen, 1996; Zeng, 1994), and they are briefly discussed below.

1.6.2.1 *Single-marker analysis*

The single marker method (Edwards et al. 1987, Beckmann & Soller 1988) is a traditional approach to mapping QTL and involves looking at a single marker at a time, and at all individual associations between the marker and the phenotype. This can be accomplished using any statistical analysis software package such as t-test, ANOVA and simple linear regression (Doerge, 2002).

In the single marker analysis, association of a polymorphic marker with the expression of a quantitative trait is detected as a significant difference in mean quantitative value between offspring groups, characterised by the marker alleles. A statistically significant difference in the mean phenotype of the marker groups indicates either that the marker gene has a pleiotropic effect on the quantitative character or that a QTL is linked to the marker. If a QTL is linked to the marker this indicates that the parents have different alleles at that locus for the QTL affecting the trait in question (Hyne, 1995).

The drawbacks of the single marker tests are that the phenotypic effects of QTLs are systematically underestimated and the genetic locations of QTLs are not well resolved because distant linkages cannot be distinguished from small phenotypic effects. Also, the number of progeny required for detecting QTLs is larger than necessary and the method neglects the problem that testing many genetic markers increases the risk that false positive will occur (Lander and Botstein, 1989). The other methods were developed to overcome the problems faced by single marker method.

1.6.2.2 *Interval mapping method*

The interval mapping method (Lander and Botstein, 1989) uses an estimated genetic map as the framework for the location of the QTL. The intervals that are defined by ordered pairs of markers are searched (for example, 2 cM), and statistical methods are used to test whether a QTL is likely to be present at the location within the interval or not. The results of the tests are expressed as LOD (logarithm of the odds) scores, which compare the evaluation of the likelihood function under the null hypothesis (no QTL) with the alternative hypothesis (QTL at the testing position) for the purpose of locating probable QTL (Doerge, 2002). The

maximum LOD profile has the potential to indicate multiple or ghost QTL incorrectly. Determining which of the many peaks indicates a single QTL leads to issues of determining statistically significant result (Lander & Botstein, 1989). Significance levels have to be adjusted to avoid false positives resulting from multiple tests, while confidence intervals are set at the map interval corresponding to a 1 LOD decline either side of the peak. Nowadays, the computer package, MAPMAKER (Lander et al., 1987) is freely available and does most of the QTL analysis.

The advantages of the approach are that the QTL likelihood map represents clearly the strength of the evidence for QTLs at various points along the entire genome, and in contrast to the traditional approach, the inferred phenotypic effects are asymptotically unbiased (Lander & Botstein, 1989). The important drawback of the method is that it may produce significant peaks (or 'ghost' effects) when actually there are no QTL (Haley and Knott, 1992; Martinez and Curnow, 1992). Existing QTL may produce peaks that exceed the threshold significantly and reveal QTL in neighbouring intervals even if there is no QTL in the interval under investigation. This is due to the fact that interval mapping does not provide an interval test i.e. a test of significance of the detected QTL within the particular interval (Doerge et al., 1997)

1.6.2.3 *Interval mapping by regression*

The interval mapping by regression approach developed by Haley and Knott, 1992, produces very similar results to interval mapping both in terms of accuracy and precision, but has the advantages of speed and simplicity of programming. The relative simplicity and computational rapidity makes it easier to fit models for two or more linked and interacting

QTL, and it can also give good estimates of QTL effects. The method has also been developed to handle complex pedigrees and to include a wide range of fixed effects in the model such as sex differences and environment. The use of regression not only eases the analysis of experimental data but also allows thorough study of the power of flanking marker methods both through simulation and theoretically (Haley and Knott, 1992).

1.6.2.4 *Marker regression approach*

Marker regression approach fits a model to all the marker means on a given chromosome simultaneously, and obtains significance tests by simulation (Kearsey and Hyne, 1994). It has the advantage of speed and of integrating all the marker information in a single test. The method tries to locate the QTL with respect to all markers simultaneously by regression onto marker means. It also estimates the additive or dominance effects, tests their significance and tests for more than one QTL. It is as reliable as the interval mapping and multiple regression approaches, but has wider application and is capable of hypothesis testing. However, because you do not know which markers flank the QTL or that there is just one QTL per chromosome, the multiple marker approach does provide an overall test of the model, no matter how the QTLs are organised on the chromosome.

1.6.2.5 *Composite Interval mapping (CIM)*

The method was developed as a combination of the standard method of interval mapping with multiple regression that involves additional markers as co-factors in the analysis (Jansen, 1996; Zeng, 1994). In general, the increase in power is accomplished because the residual variation for any position tested due to linked and unlinked QTL is reduced by the

introduction of the appropriate linked and unlinked markers in the analysis. The composite interval mapping leads to a dramatic increase in the power and accuracy of analysis.

1.7 Precision and bias of QTL mapping

The precision, with which the estimate of the QTL position is known, is dependent on a number of factors. These include the heritability of the trait, type and size of the mapping population, magnitude of the gene effect, the total number of QTL involved, the number of marker loci and their distribution (van Ooijen, 1992; Darvasi et al., 1993). One way of improving the power and accuracy of QTL detection is to increase the size of the mapping population and thus, reduce the sampling variation of each marker class (residual variance). Darvasi et al., (1993) showed that the precision of the estimate of QTL map position was biased, dependent on the distance between QTL and marker, giving a smaller confidence interval when the QTL was near a marker. Darvasi and Soller (1994) showed that under general conditions the spacing of the markers that gives the highest probability of detecting a QTL is 20 to 30 cM.

Simulation has shown that four or five well-spaced markers along the chromosome provide similar power of detecting QTL and more markers do not guarantee any substantial increase in power and precision. Recent surveys of experimental results with plants (Kearsey and Farqhuar, 1998), Arabidopsis in particular (Alonso-Blanco and Koornneef, 2000) and *Drosophila* (Mackay, 2001) conclude that confidence intervals for QTL are rarely less than 10 cM for the loci with largest effects and can be considerably greater for those with lower heritabilities.

The precision of QTL detection also depends on the method used. In general, single point analysis and interval mapping are shown to provide about the same level of power and accuracy when marker spacing is less than 30 cM. However, when marker density is larger than 30 cM, the interval mapping method is more powerful (Tanksley, 1993). Among all the methods, composite interval mapping gives the smallest confidence interval because it excludes from the error all the effects of the detected QTL. In general, however no method possesses the power of reducing the confidence interval below 10 cM (Kearsey, 1998).

1.8 Objectives of the study

The first objective of this study was to identify QTLs affecting morphological traits in the model dicot species *Arabidopsis thaliana* using the Columbia x Landsberg erecta cross and the model crop species *Brassica oleracea* using the Chinese kale (A12) and a Calabrese (GDD33) cross. Arabidopsis is a model plant for many molecular biology experiments due to its small size, short generation (5 to 6 weeks) and the ease of obtaining hundreds of seeds by manual crossing. The Brassica family is closely related to Arabidopsis and belongs to the same genus (Cruciferae) and it is hoped that any results for the genetic control of basic biological processes found in Arabidopsis can be related to Brassica.

The study culminated in investigating the efficiency of marker-assisted selection in Brassica and Arabidopsis by selecting for one, two or more QTL. As little is known about the efficiency of using markers in practical breeding programs, the study attempted to find out the efficiency of MAS using different approaches. It also looked at the number of markers that could be used in the selection process and markers that gave maximum efficiency.

CHAPTER 2

QTL mapping in *Brassica oleracea* doubled haploid lines

Abstract

A total of 40 QTL were detected in *B. oleracea* doubled haploid (DH) lines derived from a cross between a rapid-cycling Chinese kale *B. oleracea var alboglabra* (A12DHd) and a calabrese, *B. oleracea var italica* (GDDH33). Fifteen traits ranging from height, leaf and flowering were analyzed for QTL using the marker regression method. On average, 4.4 QTL per linkage group and 2.7 QTL per trait were detected in 89 DH lines. Between 1 and 6 QTL were located per trait, which individually explained 2-49% of the additive genetic variance. QTL controlling different traits were correlated and often localized to the same genomic regions.

2.1 Introduction

The genus Brassica includes many diverse types of plants, grown as vegetables, fodder or sources of oils and condiments. Six different Brassica species are widely cultivated throughout the world. Three of these species are diploids (*B. oleracea* (CC), *B. rapa* (AA) and *B. nigra* (BB)) and the remaining are their amphidiploid derivatives (*B. juncea*, AABB; *B. napus*, AACC; *B. carinata*, BBCC). The vegetable Brassicas (*B. oleracea*) are the most diverse morphological group and includes cabbage, cauliflower, broccoli, kale, brussels sprouts and kohlrabi. Although there have been active breeding programs in *B. oleracea*, very limited information is available on the inheritance of many morphological traits in this species, due to the complex inheritance of some traits and the difficulty of over-coming incompatibility. The fact that many members of *B. oleracea* are inter-crossable and able to generate fertile progenies provides an expedient route to investigate the genetic basis of quantitative traits.

The development of molecular markers (RFLP's, RAPD's, SSR's, isozymes, etc) has provided a sound framework for locating and measuring the effects of genes controlling quantitative traits (QTLs). Such markers have been used successfully used to locate QTL in many species including tomato (Paterson et al., 1988, 1991; Weller et al., 1988), rice (Wang et al., 1993), maize (Edwards et al., 1987; Stuber et al. 1992) and soybean (Keim et al., 1990). In Brassica, early inheritance studies focussed on morphological traits and complex inheritance was often observed, suggesting many genes control the traits. Kennard et al. (1994) attempted to resolve complex inheritance patterns for a variety of morphological traits in *B. oleracea* using single locus QTL in an F_2 population of a cabbage x broccoli

cross. The highly polymorphic nature and diversity of *B. oleracea* cultivars facilitated the construction of the first RFLP linkage map using a segregating F_2 population from a cross between cabbage (cultivar group capitata) and broccoli (cultivar italica) (Slocum et al., 1990). Bohuon et al., 1996 and Sebastian et al., 2000 also developed *B. oleracea* linkage maps using different populations of DH lines. The linkage map provides an opportunity to detect and measure the effects of genes controlling quantitative traits (Bohuon et al., 1998; Rae et al., 1999 and Sebastian et al., 2002). Once QTL for a particular trait have been identified then plant breeding through marker-assisted selection (MAS) can be carried out to produce commercial varieties with the desired characteristics.

QTL mapping uses different mapping populations such as the basic generations (Backcross, F_2) and advanced generations (recombinant inbred lines (RILs), or artificial populations such as doubled haploid (DH) lines. DH populations have an advantage over heterozygous populations as the homozygous lines can be maintained and multiplied by selfing so that similar experiments can be carried out in different environments. The homozygosity of DH lines also avoids complications such as dominance/recessive relationships and simplifies variance analysis.

In this study, the marker regression approach is used to detect and locate QTL controlling morphological traits in DH lines of a cross between *B. oleracea var. alboglabra* (A12) and *B. oleracea var. italica* (GD), prior to subjecting the same material to marker-assisted selection. The marker regression method by Kearsey and Hyne (1994) estimates QTL position and the QTL effects. This essentially involves regressing the additive difference

between marker genotype means at a locus against a function of the recombination frequency between that locus and a putative QTL. Considering the doubled haploid lines as in this case, with two pure breeding parental lines of P_1 and P_2. Suppose *R* represents the recombination frequency between the marker, M, and the QTL, Q. $\bar{x}_{11}$ can be defined as the mean value, of all the progeny whose marker genotype is M1M1, for the trait concerned. Via standard theory, we gain an expression relating $\bar{x}_{11}$ to the mid-parent (m), additive effect (*a*) of the QTL, and the recombination frequency between the QTL and the i^{th} marker locus (*Ri*):

$$\bar{x}_{11} = m + (1\text{-}2Ri)a$$

Via similar logic, we can define $\bar{x}_{22}$ as:

$$\bar{x}_{22} = m - (1\text{-}2Ri)a$$

Now, δ_i is defined as the difference between the mean trait values for the two marker genotypes:

$$\delta_i = \tfrac{1}{2}(\bar{x}_{11} - \bar{x}_{22})$$

δ_i is half the difference between the means at the i^{th} marker, Hence:

$$\delta_i = (1\text{-}2Ri)a$$

Thus, we have a clear expression relating δ_i, half the difference between the phenotypic effects of the two marker genotypes, and *a* and *Ri*, the additive genetic effect of the QTL, and the recombination frequency between a marker and the QTL, respectively. This relationship can be expressed as the equation of a straight line

$$\delta_i = (1\text{-}2Ri)a + 0$$

$$y = x.m + c$$

δ_i is represented on the y-axis, (1-2*Ri*) is plotted on the x-axis, the additive genetic effect of the QTL, *a*, is calculated from the gradient of m, when the intercept of the y-axis, c, is zero.

The positions of each marker are known, so the recombination frequency between each marker and the putative QTL position can be calculated, and the results represented on a graph. Because the intercept of the y-axis is zero, we use the uncorrected part of the sum of squares to calculate the regression items. This is a special case of regression analysis, and alters the values of the items in the regression analysis of variance, as the correction term is effectively zero (Burns, 1997; Kearsey and Hyne, 1994).

At the correct position of the QTL, there is a simple linear regression of δ_i onto (1-2*Ri*) with gradient a, which passes through the origin of the x and y-axis. The regression sum of squares item confirms that the additive effect (*a*) is not zero. This indicates that a significant difference exists between the mean trait values for the marker genotype classes at the locus concerned. The residual sum of squares item shows the model is adequate to explain the observed results: in this case, a one QTL per chromosome model. The most likely position of the QTL is where the residual sum of squares is minimal. The marker regression method is equally applicable to other generations derived from the F_1 e.g. backcrosses or single-seed descent lines.

The marker regression method produces estimates of QTL location and effects that are comparable to existing methods (Kearsey and Hyne, 1994). The unique features of the method are that the residual mean square can be used to test the adequacy of the simple one-QTL model and it incorporates all the marker information on that chromosome on a single test. The method also provides a simple test for whether the QTL, located on a given chromosome in different populations, are the same and this is achieved through joint regression analysis.

2.2 Materials and methods

2.2.1 Plant material

The mapping population consisted of 89 doubled haploid lines, derived from the F_1 plants of a cross between two doubled haploid parents, a rapid-cycling Chinese kale line, *B. oleracea var alboglabra* (A12DHd) and a calabrese, *B. oleracea var italica* (GDDH33), through microspore culture of the F_1 (Bohuon et al. 1996). A single plant of a microspore-derived double-haploid line of a calabrese was used to pollinate a plant from a DH line of Chinese kale. The double-haploid plants were propagated and multiplied via cuttings. This population was a subset of the 169 DH lines previously investigated by Bohuon et al. (1996). The parents were selected for standing ability and morphological uniformity. The abbreviations A12 and GD will be used to represent the original parents of DH lines in this study.

2.2.2 Trial design

The 89 DH lines were sown on the 17th May 2001 in the glasshouse. The pots were randomized and placed into two blocks and each block consisted of 3 replicate pots of each DH line. The glasshouse was unheated and unlit so that the temperature and day length depended on natural conditions.

When the seedlings attained the fourth or fifth leaf stage, they were moved to an open-ended polythene tunnel to harden off and then transplanted into the field 5 days later. The young plants were transplanted in their randomized order in a netted shade to protect them from bird damage. Empty positions with missing seedlings were transplanted with guard plants. The trial was surrounded by a row of guard plants to minimize edge effects. The plants were

spaced 76 cm between rows and 30 cm between plants in rows. Spraying was carried out to control the pest population whenever necessary.

2.2.3 Traits scored

Fifteen traits were scored in the trial (Table 2.1). The table below gives a brief description of the traits.

Table 2.1: Description of traits scored in the trial

Trait designation	Trait description
PH (Plant height)	Measured as height from the ground to the apex of the plant at 40 (PH1) and 67 days (PH2) in millimetres
NL (Number of leaves)	Total number of leaves at 40 days after sowing
LL (Leaf length)	Total length of leaf and petiole of the largest leaf at 40 days (LL1) and the same leaf at 67 days (LL2) in millimetres
LW (leaf width)	Length across the widest portion of the same largest leaf used for LL at 40 (LW1) and 67 days (LW2) in millimetres
PL (Petiole length)	Length from main stem to the start of the same leaf used for LL at 40 (PL1) and 67 days (PL2) in millimetres
FH (Flowering height)	Height from the ground to the apex of the plant at flowering time
FT (Flowering time)	Days from sowing to the first flower opening on each plant
MH (Maximum height)	Height from the ground to the tallest part of the plant at harvest (116 days after sowing)
AH (Apical height)	Height from the ground to the apex of the plant at harvest
SW (Stem width)	Length across the base of the stem at harvest
FW (Fresh weight)	Fresh weight of the harvested plant in grams

2.2.4 Data analysis

2.2.4.1 Phenotypic data

Analysis of variance (ANOVA) using the general linear model (GLM) was carried out on 89 DH lines to test the line, block, and line x block interaction effects for each trait. The lines and blocks were designated random effects. Each trait was tested for normal distribution using the Anderton-Darlington normality test (Minitab Release 13.31 for windows). Correlations between traits were calculated on family means using Minitab program and the trait distribution were obtained from the QTL café program.

2.2.4.2 Heritability

The narrow sense heritability represents the proportion of the variation that is due to the additive effects. The heritability was estimated by dividing the additive variance by the total variation for the trait. The proportion of the observed variation that is due to the effect of the genes can also be expressed as heritability of the means, when family means are used to estimate the genetic effects (Fehr, 1987). This was done for each trait by dividing the genetic variance component by the variance of the family means. The genetic variance component is equal to σ^2_L and the variance of the line means can be calculated from the mean squares for the lines as MS_L / br (Table 2.2). MS_L is equal to $\sigma^2_W+r\sigma^2_{LB}+br\sigma^2_L$ so the variance of the family means is equal to $[(\sigma^2_W + r\sigma^2_{LB}) / br] + \sigma^2_L$. This makes heritability of the means the genetic variance component divided by the genetic variance plus the environmental variance components reduced by the family size.

2.2.4.3 QTL analysis

QTL mapping was carried out using 89 DH lines which are a subset of 169 DH lines used previously to construct a detailed linkage map of *B. oleracea* (Bohoun et al, 1996). The

genetic map contained 310 loci arranged into nine linkage groups covering a total length of 875cM (Kosambi). QTL analysis was carried out using a subset of 90 loci evenly spaced at approximately 10cM intervals (Table 2.3 & Appendix I).

The marker regression analysis method of Kearsey and Hyne (1994) was used for the detection and location of QTL based on the line means. The QTL analysis was performed using a computer program that is called QTL café and is available on the web (http:\\web.bham.ac.uk/g.g.seaton). Firstly, significant associations between the marker loci and the trait means were detected using single factor ANOVA. The QTL was assumed present when the p-value for the regression MS was ≤ 1%. Given the number of tests carried out for each trait, residuals were considered significant only when probability level was ≤0.01. The regression analysis was based on 1000 simulations.

The additive genetic variance (V_A) associated with the QTL was calculated as the square of the additive effect (a), estimated by the marker regression. The percentage variance explained by the QTL was obtained by comparing the additive genetic variance of the QTL with the total variance for the character $\Sigma a^2 (=2V_A)$, estimated from the ANOVA. This is because the DH lines are homozygous, and the variance components between lines (σ^2_L) are equated to twice the additive genetic variance, V_A. For traits with two QTL the total additive variance explained by the QTL was calculated using the equation, $2V_A = a_1^2 + a_2^2 + 2a_1a_2(1 - 2R)$, Kearsey and Pooni (1996). The equation takes into account the bias when putative QTL are linked.

2.3 Results

2.3.1 Phenotypic variation

The frequency distributions of the 15 traits measured on the doubled-haploid (DH) lines are shown in Fig. 2.2a-e. Normal distribution measured by the Anderson-Darlington normality test showed that the traits are normally distributed except LL1, PL1 and MH. Further, transgressive variation is evident, which makes the population promising for detecting QTL. Highly significant differences ($P<0.05$) were observed between DH lines in all the traits (Appendix I). The analysis of variance detected significant differences between blocks in plant height1, number of leaves, plant height2, leaf length2, leaf width2, petiole length2, flowering height, maximum height, apical height and fresh weight. This significance indicated that the overall mean of the traits differed significantly in the two blocks. Significant interaction was observed in plant height1, number of leaves, leaf length1, leaf width1, petiole length1, flowering height, maximum height, apical height and stem width. This indicated that the line response to the block effects was not the same across all the lines.

The highly significant differences ($P<0.05$) between DH lines allowed the estimation of the additive genetic variance ($V_A=1/2\Sigma a^2$) for each trait (Kearsey and Pooni, 1996). Significance of genetic variation as indicated by significant differences between the lines suggests that at least one or possibly more QTL were segregating among the DH lines for each trait.

The narrow sense heritability ranged between 8% (stem width and fresh weight) and 70% (Flowering time) with an average of 28% (Table 2.4). Heritability was low for early and last traits measured on the DH lines. These values represent the relative importance of the

additive genetic effects and low values indicate that genetic effects can be easily masked by environmental or other non-heritable variation. Consequently traits with low heritability have an unreliable relationship between the genotype and the phenotype of individuals. Heritability of the means values range between 35% and 93% and represent the improved relationship between the genotype and the mean phenotype of a family.

Further, variability in heritability values can be due to different environmental components or high vs. low genetic variation or both. Therefore, to investigate if environmental variation (σ^2_W) has caused these differences, $\sqrt{\sigma^2_W}$ was divided by overall mean (μ) to make comparisons between traits. These values, given in Table 2.4 show that the environment has in general been responsible for reduced heritability. There is a highly significant correlation between heritability and $\sqrt{\sigma^2_W}/\mu$ values ($r = -0.62$) indicating that environmental variation is generally low for those traits that have high heritability and vice versa. However, this correlation accounts for only 38.44% of variation in heritability values ($r^2 = 0.38$). The rest (61.56%) must be due to differences between genetic variation. In other words, different traits show different levels of the genetic variation.

2.3.2 QTL analysis

QTL detection and location were carried out using the marker regression approach (Kearsey and Hyne, 1994). The QTL was assumed present when the p-value associated with the regression was below 1%. If the residual was below 1% it was assumed that a single QTL did not adequately explain the data and a further QTL was added to the model. Forty QTL were detected in the 15 traits measured (Table 2.5 & Figure 2.3a-d). No QTL was detected in linkage group 4.

Significantly more QTL were detected in linkage groups 1, 3, 8 and 9 (9, 6, 6 and 8 respectively) compared to others ($\chi^2_{(1)} = 8.10^{**}$). Further, the number of QTL detected in linkage groups 1, 3, 8 and 9 was statistically the same ($\chi^2_{(3)} = 0.94^{ns}$) despite a large variance in the length of these linkage groups.

QTL number varied from 1-6 per trait, and were more evenly distributed around a mean of 3 (Fig. 2.1). Plant heights seem to have more QTL (mean of 4) but the difference between these and other traits is not significant (all 5 heights vs. the rest: $\chi^2_{(1)} = 0.10$ns). The correlation between heritability of the mean and number of QTL detected is (r = 0.35) positive but non-significant and the same is true for QTL number and chromosome length, as measured by the linkage group (r = 0.07). This suggests clearly that the detection of QTL does not depend on heritability of the trait or the length of the chromosome.

2.3.2.1 Flowering traits

Three QTL were detected for flowering time on linkage groups 2 and 3 (Table 2.5 & Fig. 2.3a-b). A single QTL mapped to linkage group 2 at 26 cM with a decreasing effect on the A12 by 2.57 days. This is consistent with the results, since the A12 parent flowered earlier than the GD parent (Fig 2.2d). Two QTL mapped to linkage group 3 at positions 2 and 138 cM. The 2 QTL were observed to be in repulsion, thus their effect cancelled each other. The full model containing the QTL explained 49% of the phenotypic variance, with individual QTL models explaining 19-30% of the variance. The QTL for flowering time mapped to similar regions as the QTL for flowering height in linkage groups 2 and 3.

2.3.2.2 Height traits

The QTL for height were detected across all linkage groups except linkage group 4. The variation explained by the individual QTL varied from 2 to 48%. Six QTL were detected for height traits on linkage group 1 and between 2 and 3 QTL were detected for the other linkage groups (Fig. 2.3a-d). A lot of variation was observed in the location of the QTL within a linkage group, with some height traits mapping to at least two locations.

In linkage group 1, QTL for height traits were detected for all the developmental or measurement stages. The QTL were detected for plant height1, flowering height, maximum height and apical height at positions 0, 2, 40 and 42 cM, respectively. The first two and the last 2 positions are close, considering the confidence intervals associated with each QTL. QTL mappings to similar regions were observed in linkage group 6 for plant height1 and plant height2, and linkage group 9 for flowering height and maximum height. The QTL mapping to linkage group 9 for height traits showed increasing effects on the A12 parent, which is not consistent with the observed results.

2.3.2.3 Leaf traits

A total of 13 QTL were detected for leaf traits in linkage groups 1, 3, 7, 8 and 9 (Fig 2.3a-d). The QTL individually explained between 3 to 49% of the additive genetic variance. QTL controlling different leaf traits were often localized to the same genomic regions, and in some cases showed similar gene action. Five QTL were detected for leaf traits on linkage group 8, and four of them showed positive signs, indicating that the A12 parent has the increasing allele for the traits and they mapped to the same position, around 18 cM (Table 2.5 & Fig. 2.3d). The traits that map to this region are number of leaves, leaf length1, leaf

width1 and petiole length1. Leaf width2 mapped to 36 cM, which is very close to the other four traits considering the confidence interval associated with the QTL. This may suggest that the same genes on the same linkage group control the traits. QTL mapping to the same regions were also observed in linkage group 9 for leaf length1 and leaf width1, and linkage group 1 for leaf length2 and petiole length2. The single QTL mapping to linkage groups 1 and 9, showed positive signs, thus an increasing effect by the A12 parent for leaf length1, leaf length2 and leaf width1, consistent with the results (Fig. 2.2a-c).

2.3.2.4 Stem width

Four QTL were detected for stem width, two each on linkage groups 6 and 9 (Fig.2.3c-d). The 2 QTL on linkage group 6 both showed an increasing effect for the A12 parent and they explained 32% of the variation. The QTL on linkage group 9 are in repulsion and they explained 38% of the variation due to stem width. The QTL on linkage group 6 mapped to similar regions to QTL for plant height1 and plant height2. Traits plant height2, leaf length1, leaf width1, flowering height, maximum height and fresh weight map to the same regions as stem width in linkage group 9.

2.3.2.5 Fresh Weight

A single QTL mapping to linkage group 9 at 48 cM was detected for fresh weight (Fig. 2.3d). This QTL explained 20% of the variation for fresh weight and showed an increasing effect in the A12 parent, consistent with the results. This QTL for fresh weight mapped to a similar position to those for stem width, maximum height and flowering height.

2.3.3 Correlation between traits

Significant correlations between family means were detected between many traits measured on the DH lines (Table 2.6). Out of a total of 105 pairs, 81 (77%) showed significant correlations. However, many of these correlations were rather weak because they took values of $r < 0.3$. Pairs with $0.7 > r > 0.3$ were considered moderately correlated while those with $r > 0.7$ were assumed strongly correlated. Five pairs of traits were strongly correlated and 26 pairs were moderately correlated. Strong correlations were observed between leaf length1 with leaf width1, leaf length1 with petiole length1, leaf width1 and petiole length1, leaf length2 with petiole length2, and maximum height with apical height. Repeat measurements of the same trait recorded at different stages of development were found to be moderately correlated, such as plant height1 with flowering height, plant height2 with maximum height and apical height, flowering height with maximum height and apical height.

Correlation between those traits that were measured at the same time were also observed significant, such as number of leaves with leaf length1, leaf width1 and petiole length1, leaf length1 with leaf width1 and petiole length1, and maximum height with apical height. Pair-wise correlations were also observed between height and leaf measurements (e.g. number of leaves with plant height2, petiole length1 with plant height2, leaf length1 with plant height2). As expected, flowering time was positively correlated with flowering height.

Trait-wise, all traits seem to be equally correlated with others. Further, very few correlations (17) took negative and significant values, and they involved flowering time, leaf length2, leaf width2, plant height2 and petiole length2. Flowering time correlated negatively with the early measurements (number of leaves, leaf length1, petiole length1, leaf width1 and plant

height2) indicating that fast growing plants mostly flower early. However, none of these 17 negative correlations took a high value except between flowering time and plant height2 (r = -0.69).

QTL for the correlated traits were observed to map to the same linkage groups. For example, plant height1 and plant height2 mapped to linkage group 6; number of leaves, leaf length1, leaf width1, petiole length1 and leaf width2 mapped to linkage group 8; plant height1, flowering height, maximum height and apical height map to linkage group 1; leaf width1, plant height2 and leaf length1 map to linkage group 9; flowering height and flowering time map to linkage groups 2 and 3.

2.4 Discussion

A total of 40 QTL were detected for 15 traits measured on the DH lines. The QTL were detected in all linkage groups except linkage group 4. A strict significance of 1% was applied to the DH lines due to many traits measured and a large number of tests performed. A total of 90 marker loci were used for QTL mapping and the markers were selected such that they were spaced approximately 10cM apart to cover the whole genome. On average 4.4 QTL per linkage group and 2.7 QTL per trait were detected. In their review, Kearsey and Farqhuar (1998) observed that in most examples no more than 4 QTL per trait could be detected. In this study, the number of QTL per trait varied from 1 to 6. According to Hyne and Kearsey (1995), the number of QTL detected per trait can never be more than 12. This is attributed mainly to the low heritabilities associated with each additional QTL, making their detection rare at suitable significance levels.

The individual QTL explained between 2 and 49% of the additive genetic variation. Kearsey & Farquhar (1998) observed that individual QTL might explain 1-50% of the additive variation, which is in agreement with the present study. The 95% confidence intervals for the mapped QTL ranged between 14 and 60 cM, indicating a low accuracy of the estimates. Usually a confidence interval of up to 30 cM has been observed for the segregating populations. These confidence limits are very large and the reliability can be increased by increasing family size and the number of DH lines (van Ooijen, 1992; Darvasi et al., 1993; Kearsey & Farquhar, 1998). In this study, only 89 lines were used which is a small family size. van Ooijen (1992) observed that a minimum population of 200 backcross and F_2 individuals was necessary for detecting QTLs that explained at least 5% of the total variance for a trait.

The number of QTL detected per linkage group varied from 2 to 9. More than five QTL per linkage group were detected for linkage groups 1, 3, 8 and 9. In soybean, Mansur, et al. (1993) found that QTL for many traits were clustered on just three linkage groups, whereas Sebastian et al. (2002) observed six or more QTL in three linkages groups (01, 07 and 08) in another *B. oleracea* mapping population. In this study, QTL controlling different traits were also frequently localized to the same genomic regions. For example, plant height1, flowering height, maximum height, apical height and petiole length2 were associated with common marker loci on linkage group 1. Therefore, it could be concluded that the same genes could be involved in the control of two or more traits. In other words, the QTL are showing pleiotropic effects, particularly for traits that are measured at different stages of plants' life cycle (e.g. plant heights and leaf traits).

Associations of the same marker loci (and QTL) with several traits, e.g. number of leaves, leaf length1, leaf width1, leaf length2 and petiole length1, which map to linkage group 8, is also supported by Kennard et al. (1994) who found similar marker-trait associations for lamina width, lamina length and petiole length. Sebastian et al. 2002 has also mapped QTL affecting lamina width to linkage group 8. In linkage groups 8 the A12 parent showed the increasing effect and this is true as high values for A12 parent were observed. In linkage group 9, QTL seem to be located at around 54 and 92 cM. Traits that mapped to 54 cM are flowering height, maximum height, fresh weight and stem width, whereas those mapping to 92 cM are leaf width1, plant height2, leaf length1 and stem width. QTL mapping to similar regions were also present on linkage group 3 (flowering height and flowering time), linkage group 6 (plant height1 and plant height2) and linkage group 2 (flowering time and flowering

height). This clearly shows that the 40 QTL detected in this study may not all be separate loci, thus the actual number of QTL (detected) for which A12 and GD differ would well be as low as 20 or less.

Some functionally related traits such as plant height1 and flowering height, and number of leaves, leaf length1, leaf width1, petiole length1 and leaf width2, were correlated and the QTL controlling these traits were found to be located on the same linkage groups. For example, QTL controlling flowering height on linkage groups 2 and 3 were also involved in controlling flowering time. This is expected because many genes that affect flowering height also affect time to flower. The single QTL for flowering height and flowering time showed negative value, suggesting that the A12 parent had the decreasing allele. This is expected as the A12 was selected for short generation time, hence early flowering.

The results of QTL mapping in the present study can also be compared to those of Bohoun et al. (1998) and Rae et al (1999), because both studies are based on the same set of DH lines (Figure 3.2a-d). Bohuon et al. (1998) observed the QTL for flowering time on linkage groups O2, O3, O5 and O9, whereas Rae et al. (1999) observed the flowering QTL in substitution lines on linkage groups O1, O2, O3, O5 and O9. Rae (2000) also observed flowering time QTL in DH lines in linkage groups 2 and 3. In this study, flowering time QTL were only detected on linkage groups 2 and 3. The QTL detected in linkage group 2 was detected by Bohuon et al., 1998 and Rae et al., 1999, whereas that detected in linkage group 3 was also detected by Bohuon et al. (1998); Rae et al. (1999) and Rae (2000). Kennard et al. (1994) and Lan & Paterson (2000) also reported linked regions controlling both growth habit and flowering time in *B. oleracea*. Camargo and Osborn (1996) found

three linkage groups with significant QTL for flowering time using F_3 families in a cross between cabbage and broccoli. Experiments involving *B. napus* have shown that regions on linkage groups in the A genome which show homology to regions in O2, O3 and O9 linkage groups also carry QTL for flowering time (Salinas-Garcia, 1996; Osborn et al., 1997). The former regions are also homologous to regions on linkage groups 2 and 8 of *B. nigra* to which a flowering time QTL has been mapped (Lagercrantz et al., 1996) and to the region around the CONSTANS gene and several other flowering candidate genes on chromosome 5 of *A. thaliana.*

Comparative data help to highlight how tools from Arabidopsis might be used to quickly explore for genes that may directly account for Brassica QTL. In these studies, we have identified 4 flowering height and 3 flowering time QTL that associate for flowering in *B. oleracea.* The QTL detected fall in regions involving the top of linkage group 1 (flowering height) that correspond to a segment of Arabidopsis chromosome 5, that contains seven flowering mutations (*tfl1*, *flc*, *tfl2*, *co*, *fy*, *art1*, *emf1*); a homologous region in the lower half of linkage group 9 (flowering height), which corresponds to a region of Arabidopsis chromosome 1 containing the mutation *efs*; a region of *Brassica* linkage group 3 (flowering height and flowering time), which corresponds to a region of Arabidopsis chromosome 1 containing the mutation *fha* (Lan & Paterson, 2000). As the Arabidopsis genome sequence unfolds, the QTL detected and other candidates will provide a foundation for the sequencing methods to explore levels and patterns of allelic variation in *Brassica* that may help to implicate some of these in the genetic control of complex traits in well-defined *Brassica* gene pools (Wang et al., 1999).

Comparison of QTL data across studies, in terms of position, action and effect, can give greater credence to the results particularly when confidence intervals associated with QTL positions are large. The additional confidence gained from these comparisons may allow inferences about QTL position to be made between different populations within a species and between species, especially within a genus such as Brassica where there are close relationships between its member species (U, 1935). This in turn may help in the development of efficient breeding programs. However, the ability to compare the results of QTL analysis with other studies relies firstly, on being able to identify homologous or homeologous markers, between linkage maps to which putative QTL are linked, and secondly, to have trial data for the same or similar traits.

Table 2.2: Outline of a two-way ANOVA (number of DH lines, n = 89, number of blocks, b = 2 and number of replications, r = 3)

Source	df	MS	Ems
Lines	n-1	MS_L	$\sigma^2_W+r\sigma^2_{Lb}+br\sigma^2_L$
Block	b-1	MS_b	$\sigma^2_W+r\sigma^2_b$
LxB	(n-1)(b-1)	MS_{Lb}	$\sigma^2_W+r\sigma^2_{Lb}$
Error	nb(r-1)	MS_W	σ^2_W
Total	nbr-1		

Table 2.3: Number of QTL observed, number of markers used, average distance between the markers and length of the linkage group

Linkage group	Length of the linkage group (cM)	Number of QTL	Number of markers	Average distance between markers (cM)
1	83.8	9	8	10.48
2	106.6	3	9	11.84
3	139.0	6	15	9.27
4	99.6	0	10	9.96
5	97.7	2	11	8.88
6	72.6	4	8	9.08
7	80.8	2	9	8.98
8	86.8	6	10	8.68
9	103.0	8	10	10.30
Total	869.9	40	90	9.72

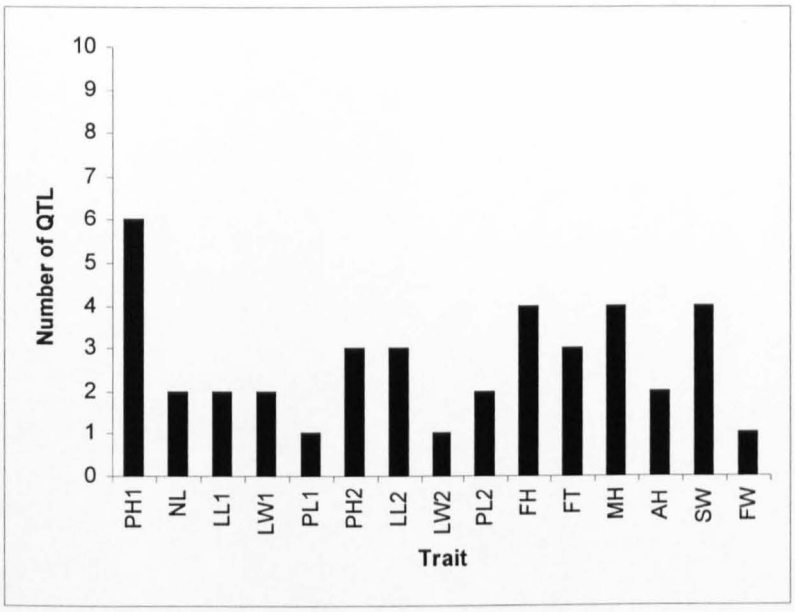


Figure 2.1: The number of QTL detected per trait in DH lines

Table 2.4: Population mean (μ), environmental variance (σ^2_W), interaction variance (σ^2_{LB}), additive variance (σ^2_A), total variance (σ^2_T), narrow sense heritability (h_n^2) and heritability of the mean (h_f^2) for the traits

Trait	Mean (μ)	σ^2_W ($\sqrt{\sigma^2_W}/\mu$)	σ^2_{LB}	σ^2_A	σ^2_T	h_n^2(%)	h_f^2 (%)
PH1	58.98	73.80 (0.15)	8.10	51.24	133.12	38.49	86.24
NL	6.58	0.41 (0.10)	0.08	0.05	0.54	9.26	50.38
LL1	88.66	422.80 (0.23)	47.17	87.08	557.05	15.63	64.93
LW1	54.19	150.80 (0.23)	22.30	16.73	189.83	8.81	47.98
PL1	25.53	74.51 (0.34)	10.47	9.96	94.94	10.49	53.02
PH2	291.36	3355.00 (0.20)	-	3088.25	6443.25	47.93	84.67
LL2	277.10	1594.00 (0.14)	-	589.08	2183.08	26.98	68.92
LW2	157.37	537.50 (0.15)	-	1099.50	1637.00	67.17	92.47
PL2	113.77	325.00 (0.16)	-	214.92	539.92	39.81	79.87
FH	387.81	2278.00 (0.12)	453.67	1778.08	4509.75	39.43	85.43
FT	74.17	7.57 (0.04)	-	17.47	25.04	69.77	93.27
MH	838.28	28454.00(0.20)	6161.33	7537.41	42152.74	17.88	65.84
AH	649.49	29342.00(0.26)	6826.67	7718.67	43887.34	17.59	65.02
SW	36.12	73.22 (0.24)	14.27	8.07	95.56	8.44	45.51
FW	486.64	46705.00(0.44)	-	4155.92	50860.92	8.17	34.81

NB: - shows no interaction

Table 2.5: Putative QTL position, additive effect, additive variance (V_A), percentage variance explained, confidence interval (CI), regression and residual mean squares for all the traits

Trait	Linkage group	QTL position (cM)	Additive effect	V_A	% V_A	CI	Regression P	Residual P
PH1	1	2.0	-5.84	34.11	23.13	23	0.004	0.129
	3	48.0	-7.07	49.98	33.89	17	0.009	0.125
	5	6.0 42.0	-5.84 4.25	25.15	17.05		0.811	0.008
	6	36.0	-6.95	48.30	32.75	14	0.002	0.066
	7	80.0	-4.57	20.88	14.16	25	0.009	0.957
PH2	6	56.0	-29.66	879.72	14.24	29	0.002	0.01
	8	48.0	28.53	813.96	13.18	27	0.002	0.058
	9	94.0	34.02	1157.36	18.74	32	0.006	0.079
FH	1	0.0	-10.35	107.12	2.08	60	0.426	0.01
	2	12.0	-23.22	539.17	10.47	58	0.068	0.007
	3	0.0	-33.12	1096.93	21.30	43	0.004	0.059
	9	54.0	24.13	582.26	11.31	32	0.008	0.111
MH	1	0.0 40.0	-90.64 95.39	9543.90	48.16		0.033	0.002
	2	60.0	29.08	845.65	4.27	55	0.159	0.008
	9	60.0	53.27	2837.69	14.32	32	0.007	0.064
AH	1	0.0 42.0	-90.64 85.44	8827.64	43.43		0.123	0.001
NL	7	0.0	-0.11	0.01	2.55	56	0.358	0.002
	8	18.0	0.44	0.19	49.30	14	0.003	0.325
LL1	8	20.0	13.15	172.92	39.97	15	0.006	0.277
	9	92.0	10.86	117.94	27.26	26	0.0	0.567
LL2	1	40.0	10.71	114.70	9.74	33	0.008	0.011
	3	24.0 80.0	19.52 -20.31	534.88	45.40		0.01	0.002
LW1	8	18.0	6.95	48.30	36.82	18	0.006	0.154
	9	94.0	5.81	33.76	25.73	29	0.003	0.543
LW2	8	36.0	-7.78	60.53	16.51	24	0.001	0.346
PL1	8	18.0	5.14	26.42	39.04	17	0.001	0.284
PL2	1	0.0 36.0	-13.11 13.18	177.36	41.26		0.037	0.002
FT	2	26.0	-2.57	6.60	18.90	34	0.002	0.428
	3	2.0 138.0	-2.86 1.73	10.52	30.10		0.0	0.0
SW	6	10.0 50.0	0.99 1.90	2.90	32.15		0.0	0.001
	9	36.0 90.0	1.44 -1.76	3.45	38.25		0.568	0.008
FW	9	48.0	40.94	1676.08	20.16	32	0.006	0.261

Table 2.6: Pearson correlation coefficients between the traits (only significant values are given)

	PH1	NL	LL1	LW1	PL1	PH2	LL2	LW2	PL2	FH	FT	MH	AH	SW
NL	0.19													
LL1	0.18	**0.66**												
LW1	0.22	**0.64**	**0.91**											
PL1	0.23	**0.61**	**0.90**	**0.80**										
PH2	0.24	**0.38**	**0.33**	0.25	**0.30**									
LL2		-0.15		-0.12	-0.11	-0.12								
LW2	0.16	-0.25	-0.26	-0.14	**-0.33**		**0.60**							
PL2	0.09	-0.16		-0.13		0.11	**0.70**	**0.41**						
FH	**0.37**					0.26	0.17	0.17	0.16					
FT	0.11	**-0.39**	-0.29	-0.20	-0.25	**-0.69**		0.19		**0.31**				
MH	0.11		0.11		0.10	**0.34**	**0.32**	0.22	**0.32**	**0.44**	-0.09			
AH	0.13		0.13		0.15	**0.31**	0.17		0.25	**0.52**		**0.74**		
SW				0.14		-0.16	0.20	**0.33**	0.17	0.16	0.25	0.20		
FW	0.21		0.09	0.11		0.14	**0.31**	**0.33**	0.25	**0.40**	0.14	**0.52**	**0.36**	**0.43**

Bold = correlation >0.30

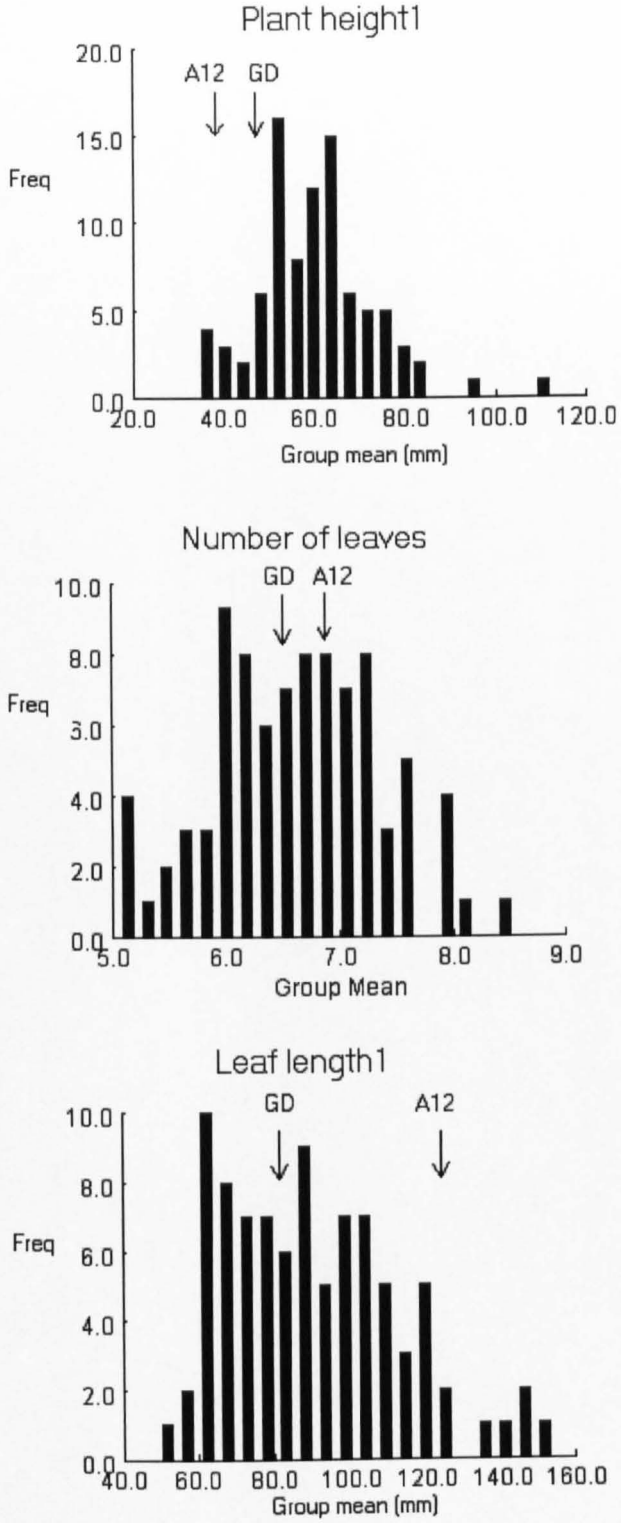


Figure 2.2a: Trait distribution for plant height1, number of leaves and leaf length1 in DH lines with arrows indicating the mean position of the parents

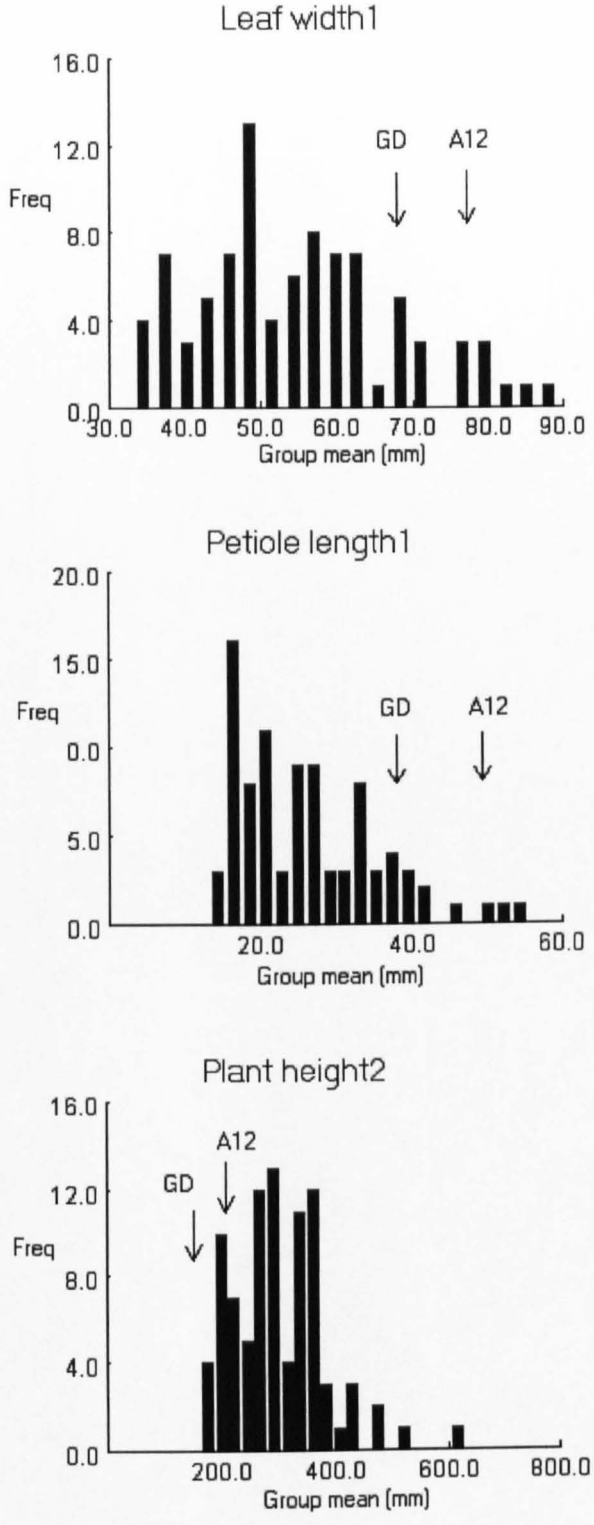
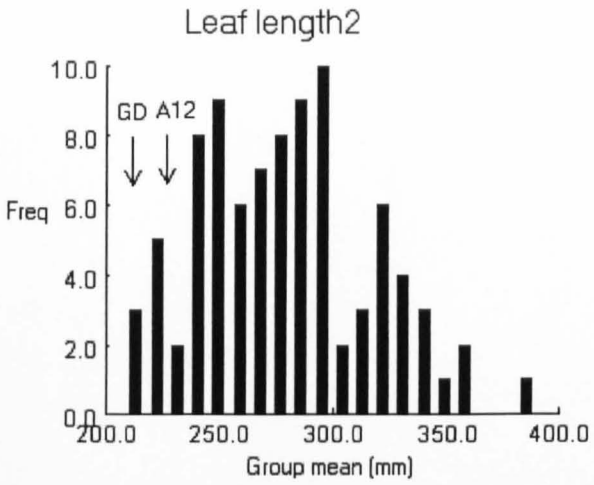
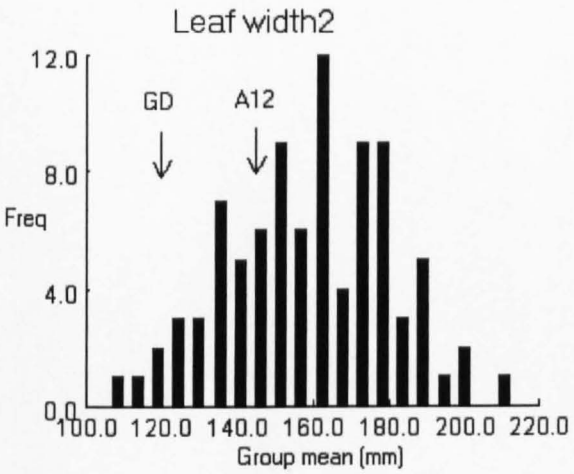


Figure 2.2b: Trait distribution for leaf width1, petiole length1 and plant height2 in DH lines with arrows indicating the mean position of the parents





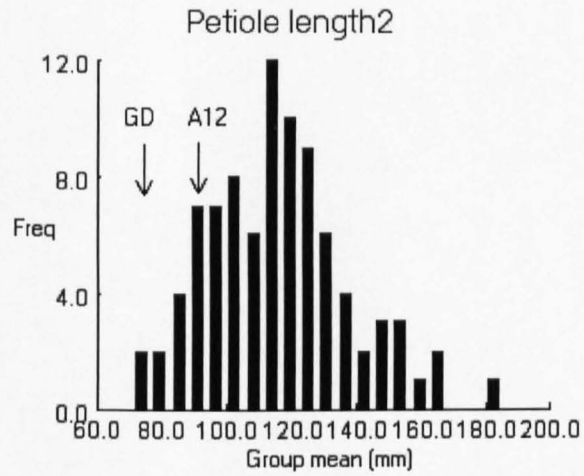


Figure 2.2c: Trait distribution for leaf length2, leaf width2 and petiole length2 in DH lines with arrows indicating the mean position of the parents

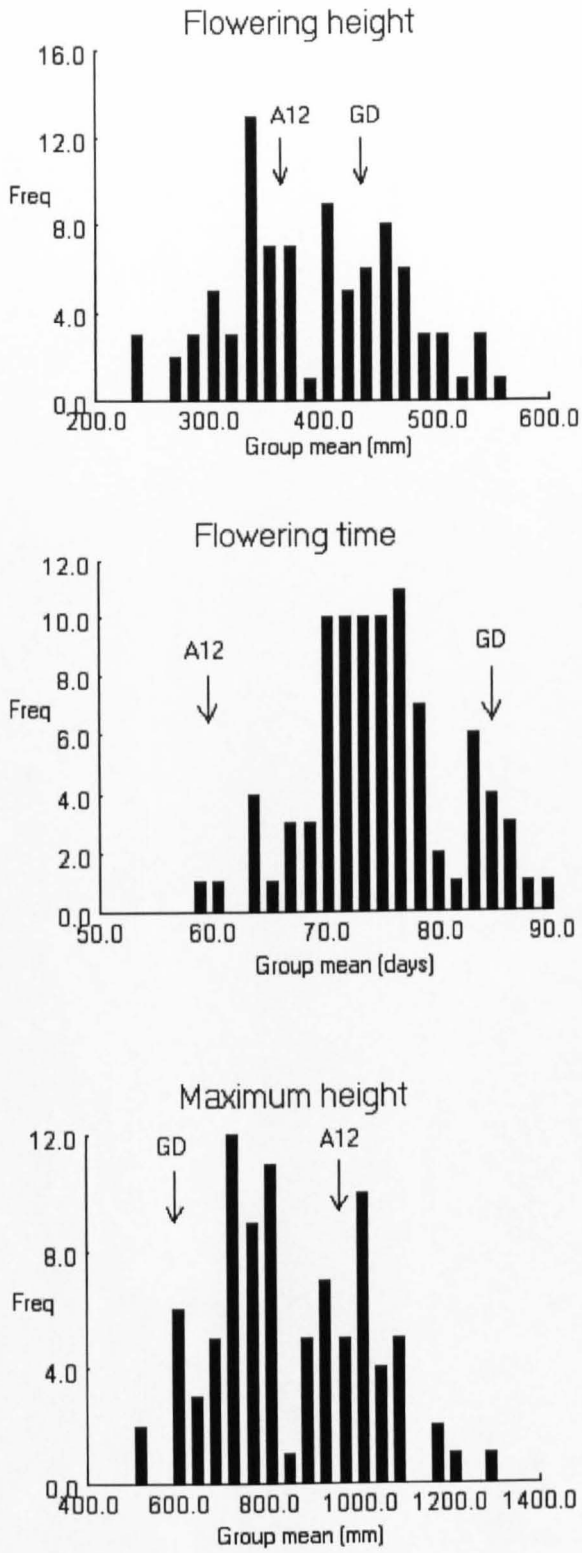
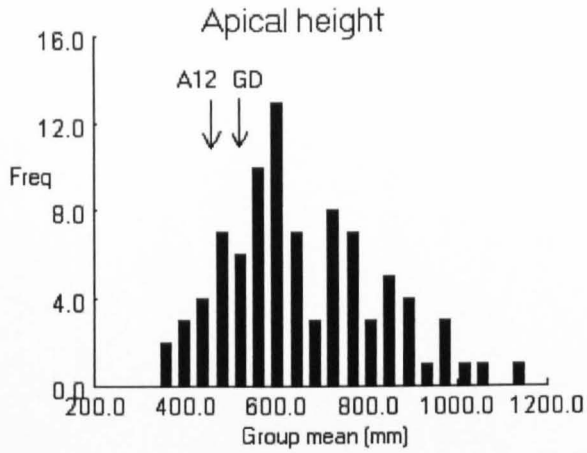
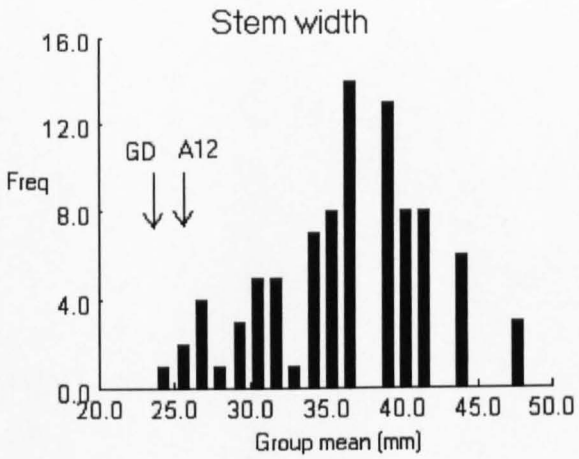


Figure 2.2d: Trait distribution for flowering height, flowering time and maximum height in DH lines with arrows indicating the mean position of the parents





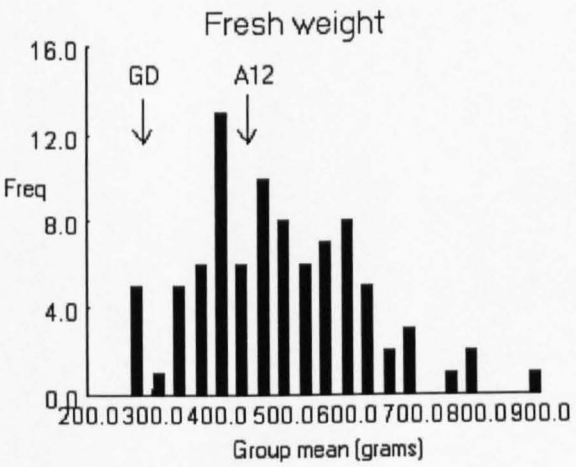


Figure 2.2e: Trait distribution for apical height, stem width and fresh weight in DH lines with arrows indicating the mean position of the parents

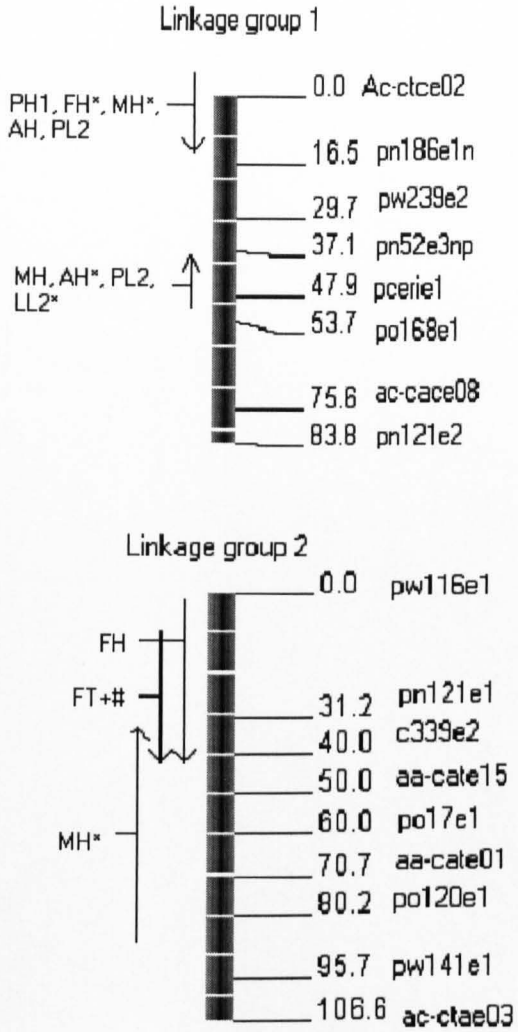
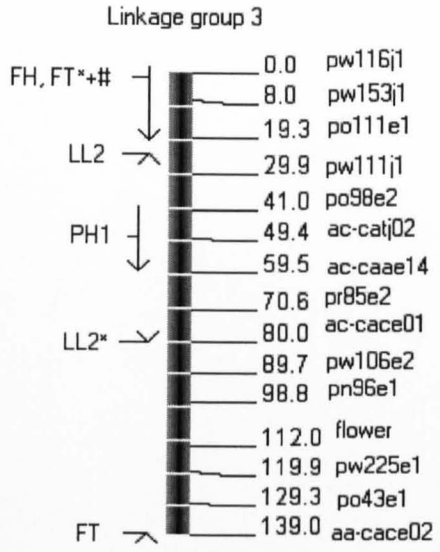


Figure 2.3a: The location of QTL on *Brassica oleracea* linkage groups 1 and 2, the confidence intervals of the QTL and the additive effect of the QTL indicated by the direction of the arrow (Symbols *, + and # involves QTL detected by Rae (2000), Bohuon et. al., 1998 and Rae, et. al, 1999, respectively)



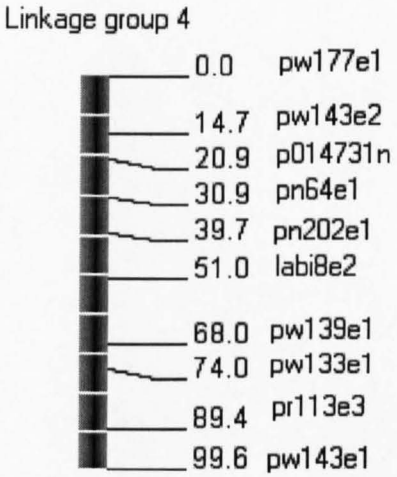


Figure 2.3b: The location of QTL on *Brassica oleracea* linkage groups 3 and 4, the confidence intervals of the QTL and the additive effect of the QTL indicated by the direction of the arrow (Symbols *, + and # involves QTL detected by Rae (2000), Bohuon et. al., 1998 and Rae, et. al, 1999, respectively)

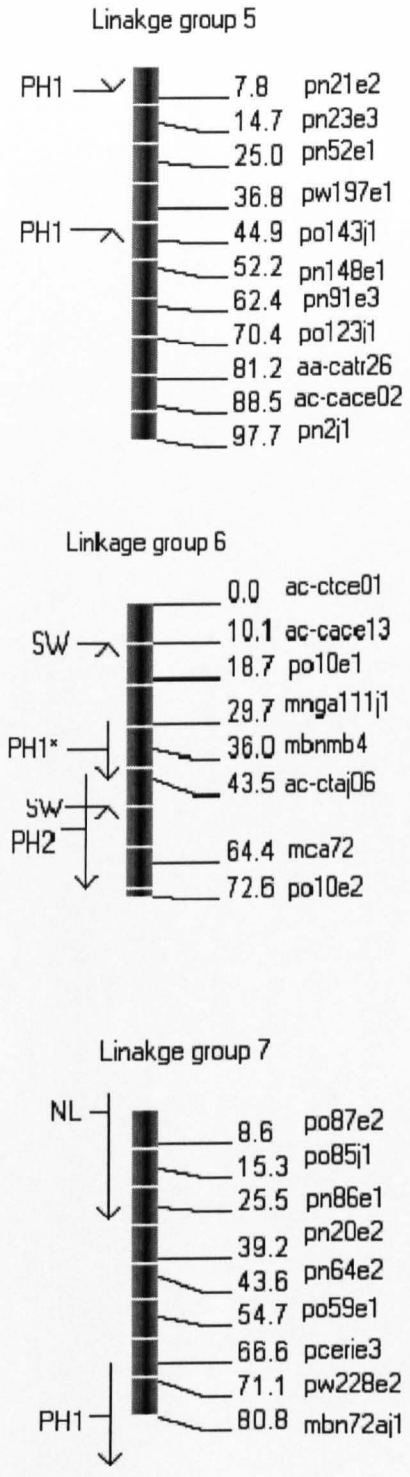


Figure 2.3c: The location of QTL on *Brassica oleracea* linkage groups 5, 6 and 7, the confidence intervals of the QTL and the additive effect of the QTL indicated by the direction of the arrow (Symbols *, + and # involves QTL detected by Rae (2000), Bohuon et. al., 1998 and Rae, et. al, 1999, respectively)

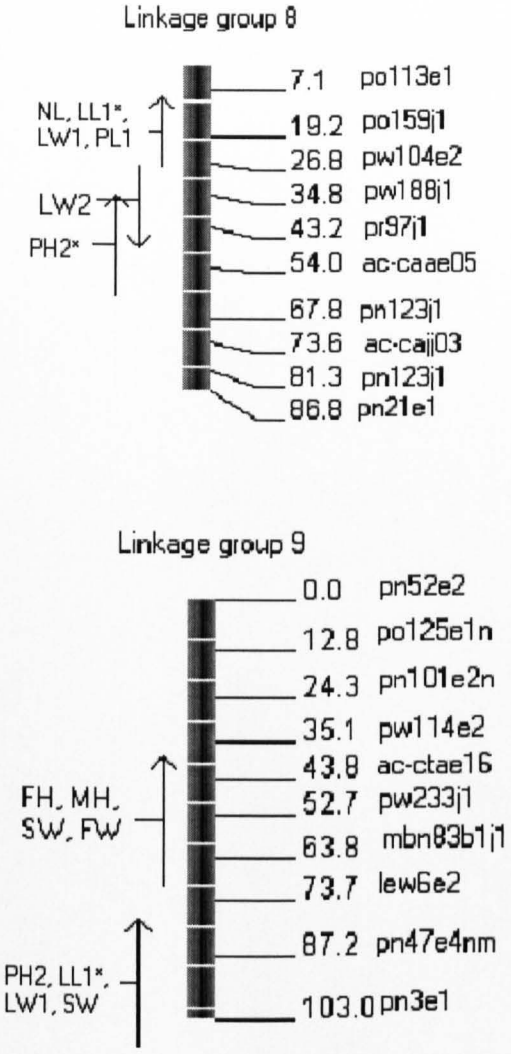


Figure 2.3d: The location of QTL on *Brassica oleracea* linkage groups 8 and 9, the confidence intervals of the QTL and the additive effect of the QTL indicated by the direction of the arrow (Symbols *, + and # involves QTL detected by Rae (2000), Bohuon et. al., 1998 and Rae, et. al, 1999, respectively). The RFLP markers are denoted by prefix 'p' and are named after the probe used for hybridization. Multiple loci resulting from hybridization with a single probe are given identical names with the addition of a number suffix. Lew and Lab are also RLFP markers. The microsatellites have the prefix 'm' and are similarly named after their primer pair with a number suffix added if more than one locus was identified. Nomenclature of the AFLP markers is based on their primer extensions with band number, denoted by a suffix, corresponding to the polymorphic bands numbered sequentially in order of increasing mobility.

CHAPTER 3

QTL mapping in *Arabidopsis thaliana* F_2 population

Abstract

A segregating F_2 population of *Arabidopsis thaliana* derived from a cross between early flowering ecotypes Columbia (Col) and Landsberg *eracta* (Ler) was analyzed for morphological variation. Twenty-three QTL were detected in traits including flowering, height and leaf measurements. The QTL affecting similar traits were localised to the same genomic regions. Time to flower and time to budding mapped to the same regions in chromosomes 1, 2, 3 and 5. Height at flowering and rosette leaves at flowering mapped to regions close to the QTL for flowering time in chromosome 2. The traits mapping to the same regions were correlated and showed similar mode of gene action, suggesting pleiotropic effects of the same genes. The detected QTL mapped close to previously identified genes such as *QLN-2*, *RLN-1*, *RLN-2*, *RLN-4*, *QLN-7*, *QLN-7*, *QLN-12* and mutations *FRI*, *flc*, *co*, *fy*, *erecta* and *fpa*.

3.1 Introduction

The development of genetic maps based on DNA polymorphisms is beginning to provide the geneticist and plant breeder with powerful tools for the study and manipulation of quantitative genetic variation. The use of molecular markers to detect and locate individual loci responsible for quantitative variation provides an even greater power than the segregation analysis without further information (Haley and Knott, 1992).

Marker genes have been used to identify chromosomal regions containing QTLs since 1923 (Sax, 1923), and the availability of an infinite number of DNA (or molecular) markers as well as new statistical methods (Lander and Botstein, 1989; Kearsey and Hyne, 1994; Haley and Knott, 1992) have made marker-based mapping efficient. These methods can give detailed information on the number of loci, effects of alleles and epistatic interactions between these alleles. The resolution of marker studies is better than that of traditional biometrical studies, which assume complete additivity, equal effects and independent segregation of the genes (Kuittinen, et al. 1997).

Flowering time is one of the quantitative traits that have widely been studied in most *Arabidopsis* ecotypes. The timing of germination and flowering in annual plants synchronises the growth and reproduction of plants within favourable periods. Flowering time can also have an effect on the seed set of a plant, leading to possible trade-offs between timing and magnitude of reproduction (Dorn and Mitchell-Olds, 1991). The extensive natural variation for flowering time and the ease with which mutations may be produced has

meant that *Arabidopsis* has been used as a model plant for the study of floral induction for over many years.

There are many possible candidate genes in *Arabidopsis thaliana* that account for the different flowering-time phenotypes. The candidate genes have been identified either through classical crossing experiments (Clarke and Dean, 1994, Clarke et al. 1995, Jansen et al. 1995) or mutation and physiological studies (Coupland, 1995) and most loci had been placed on the linkage map of *Arabidopsis thaliana* (Kuittinen, et al. 1997). For example, two independently segregating loci that affect 'days to first flowering' have been mapped to chromosome 5. These QTLs appear also to influence node number, leaf length at flowering and leaf length at 35 days (Kowalski et al. 1994b).

QTL affecting late-flowering have been identified in *Arabidopsis thaliana*, some of which confer phenotypes that are responsive to vernalization (Koornneef et al. 1991). The latest flowering ecotypes of Arabidopsis have late-flowering alleles at the *FRI* locus on the top of chromosome 4 (Lee et al. 1993; Clarke and Dean 1994) and at *FLC* locus on the top of chromosome 5 (Koornneef et al. 1994; Lee et al. 1994a). These two loci account for the late-flowering phenotype in progeny of crosses to Landsberg *erecta*, and plants with the late-flowering alleles are responsive to vernalization. The tops of chromosome 4 and 5 also contain *LD* loci on chromosome 4 (Lee et al. 1994b) and *FY* and *CO* on chromosome 5 (Koornneef et al. 1991, Putterill et al. 1995) for which recessive (*LD* and *FY*) or semi-dominant (*CO*) mutations confer late-flowering phenotypes. Despite flowering time, other studies focussed on QTL involving rosette leaf number (Clarke et al., 1995; Stratton, 1998), germination rate (van Schaar et al., 1997), seed dormancy (Alonso-Blanco et al, 2003), floral

characters (Juenger et al., 2000), inflorescence architecture (Weinig et al., 2002; Ungerer et al., 2002), leaf morphology (Robles and Micol, 2001; Pérez-Pérez et al., 2002). Some of the genes that have been cloned in Arabidopsis may have homologues that regulate flowering in Brassica species. The comparison of map positions of QTLs with those of induced mutations can also help integrate all the information on loci controlling morphological traits in *Arabidopsis*.

In the present study, attempts are made to find out how many QTLs are responsible for flowering time and other morphological traits using the F_2 plants of a cross between the early flowering *Arabidopsis* ecotypes Columbia (Col) and Landsberg *erecta* (Ler) as a mapping population. The ultimate objective of this study was to genotype each F_2 plant for a selected set of markers, identify markers correlated with various quantitative traits, identify the QTLs and measure their effects, apply MAS and determine its efficiency against phenotypic selection using different criteria.

3.2 Materials and methods

3.2.1 Plant material

The experimental material was produced by crossing the *Arabidopsis* ecotypes Columbia (Col) with Landsberg *erecta* (Ler). The ecotypes were obtained from Nottingham *Arabidopsis* Seed Collection (Nottingham, UK). Hybridization of these ecotypes was carried out by hand crossing using Col as the female parent. The reciprocal cross was also made and the F_1 was verified by microsatellite (SSR) analysis. The verified F_1 plants were selfed to generate F_2 seeds and 400 F_2 individuals were evaluated for QTL analysis together with the parents and the F_1.

3.2.2 Growth conditions and experimental design

Four hundred *Arabidopsis thaliana* F_2 plants were sown in 7.5 cm pots containing soil mix of 2 parts John Innes No. 1 compost, 2 parts peat based compost and 1 part silvaperl. The plants were sown during the summer of 2001 on the 15th June 2001 in the growth room. Three seeds were sown per pot, and pots were placed on benches with perforated matting for underneath watering. Guard plants to minimize edge effects surrounded the experiment. The F_2 plants were exposed to 16-hour photoperiod and 24°C temperature in the growth room. After two weeks, the seedlings were thinned to one per pot, and empty (experimental) pots were planted with guard plants.

3.2.3 Traits measured

Each experimental plant was measured for 11 traits involving leaves, height and flowering. Table 3.1 provides a description of each trait.

Table 3.1: Traits measured in the F_2 population

Trait	Symbol	Description
Time to germination	TTG	Number of days from sowing to the time the plant emerges from the soil
Time to true leaves	TTL	Number of days from sowing to the time the first true leaves appears
Cauline leaves at 20 days	CL20	Number of cauline leaves at 20 days after sowing
Cauline leaves at flowering	CLF	Number of cauline leaves at flowering time
Rosette leaves at 20 days	RL20	Number of rosette leaves at 20 days after sowing
Rosette leaves at flowering	RLF	Number of rosette leaves at flowering time
Height at 20 days	HT20	Plant height from the soil level to the apex of the plant at 20 days after sowing in millimeters
Height at flowering	HTF	Plant height at flowering time in millimeters
Height at 34 days	HT34	Plant height at 34 days after sowing in millimeters
Time to bud	TTB	Number of days from sowing to the appearance of the first flower bud
Flowering time	TTF	Number of days from sowing to the appearance of the first flower

3.2.4 Genetic mapping

Leaf tissue for DNA extraction was taken from large, green leaves, on three weeks old plants. The leaf samples were put into 1.5ml test tubes, temporarily stored in liquid nitrogen, and then transferred to –70°C for long term storage. DNA was extracted from frozen leaf samples using a GenElute^{TN} plant genomic kit (SIGMA-Aldrich) in accordance with the manufactures instructions (Appendix IV).

PCR was carried out in a 25 µl reaction volume consisting of 12.5 µl PCR master mix (0.625 units Taq DNA polymerase, 75 mM Tris-HCl, pH 8.8 at 25°C, 20 mM $(NH_4)_2SO_4$, 2.5 mM $MgCl_2$, 0.01% (v/v) Tween 20, 0.2 mM each of dATP, dCTP, dGTP and dTTP, precipitant

dye for electrophoresis), 10 ng template DNA, 1 µl (10 pmol/µl) each of the forward and reverse primer, and 5.5 µl sterile distilled water. The reaction contained in a 0.5ml tube was overlaid with 20 µl of mineral oil to prevent evaporation. The PCR reaction had one denaturation step of 4 minutes at 94°C, followed by 30 cycles of denaturation for 30 seconds, annealing for 30 seconds and extension for 30 seconds. The annealing temperature was dependent on the primer (Appendix II), whereas the denaturation temperature of 94°C, and extension temperature of 72°C were used. This was followed by an extension step of 72°C for 4 minutes, to ensure full extension of product molecules. The band sizes and primer sequences for the SSRs markers are given in Appendix II.

The 30 microsatellites (SSRs) used in this study had previously been mapped in a similar cross, so polymorphism did not need to be established (Koumproglou, 2002). The primer sequences and PCR conditions for all the PCR markers were obtained from the Arabidopsis information resource (TAIR) website. The markers were chosen such that they cover the genome at intervals of approximately 20cM in each chromosome (Table 3.2, Appendix II).

3.2.5 Agarose gel preparation and electrophoresis

PCR products (3–5µl) were resolved on a 3% (w/v) agarose gel (Bioline) stained with 0.5 µg/ml Ethidium Bromide solution after electrophoresis at 3.7V cm^{-1} for 2 hrs. When a greater resolution of allele differences was required the products were separated on pre-cast 'Spreadex EL400' gels in a 'SEA2000' gel electrophoresis system (Elchrom Scientific) at 10 V cm^{-1} for 2 hrs and subsequently stained with 5µl 'SYBR-green I' staining dye (Sigma) in 50ml of 10mM TAE, then de-stained in water for 30 minutes. The PCR products were

visualized using Flowgen IS500 imaging system. The F_2 individuals were scored as follows: Columbia = 1, Heterozygote = 2, Landsberg = 3 and missing or unscorable = 4.

Figures 3.1 and 3.2 show the segregation of markers nga139 and nga249 separated using the Spreadex gel for greater resolution of the bands.

3.2.6 High through-put fragment analysis

The extreme markers on each chromosome were genotyped by automated fragment analysis. PCR reactions were performed in 25 µl reaction volumes (10 ng template DNA, 1U Taq DNA polymerase (Bioline), 10mM Tris-HCl pH 8.9 (25°C), 1.5 mM $MgCl_2$, 100 mM KCl, 50 µg/mL bovine serum albumin (BSA), 0.05%(v/v) Tween® 20, 200 µM dNTP (each), 0.4 µM forward and reverse primer) using locus specific primer pairs. To allow target identification, the forward primer was 5' labeled by the manufacturers (TAG) with the fluorophore 5-FAM (5-carboxyflourescene). Thermal cycling was performed in a Robocyler® thermal cycler (Stratagene) for 30 cycles of 94°C for 40 seconds, 56°C for 40 sec and 72°C for 1 min. PCR product was then diluted 14:1 in 10µl of High Dye Formamide® (ABI) containing 0.5µl of ROX 500 size standard. The samples were then denatured for 3 min at 94°C and placed on ice before being loaded onto an ABI 3700 automated 96 capillary sequencer. Samples were run using the sequencers Genescan® 500 fragment analysis parameters and data collected as a series of fragment run times.

The fragment analysis run files were downloaded from the Birmingham genomics laboratory server by FTP (FTP//:www.genomics.bham.ac.uk) and examined using the Genescan® analysis package. This package enables an initial analysis of the quality of data to be carried

out (background noise, intensity) and fragment run times to be converted to allele sizes by comparison with the ROX 500 internal standard. Following initial analysis and manipulation, Genescan® files were exported to the application Genotyper® (version 3.7) for detailed genotype analysis. The individuals were scored as above (Columbia = 1, Landsberg = 3, Heterozygous = 2 and missing or unscorable = 4).

The ADI 3700-sequence analyzer reduces the amount of labour in PCR reactions because the loading, electrophoresis and imaging are done electronically. The machine also imports sample information and can export the analyzed data.

3.2.7 Data analysis

3.2.7.1 Phenotypic data

The quantitative data were initially subjected to ANOVA to test for differences between the parents, F_1 and F_2 generations. Further analysis involved the comparison of the F_2 mean with the mid-parent and F_1, estimation of the additive/dominance effects, scaling tests and tests for genetic variation. The inter-trait relationships were studied by calculating Pearson's correlation coefficient for each pair of traits, and an Anderson-Darlington normality test (Minitab Release 13.31) was carried out to test for significant deviations from normal distribution for each trait.

The mid parent (m), showing the mean between the parents was estimated by the formula:

$$m = \tfrac{1}{2}\,\bar{P}_1 + \tfrac{1}{2}\,\bar{P}_2$$

where $\bar{P}_1$ and $\bar{P}_2$ stand for the mean of Col and Ler parents respectively.

The additive [a] and dominance [d] effects, showing the direction of the additive and dominance effects were calculated by the formulas:

$$[a] = \tfrac{1}{2}\ \bar{P}_1 - \tfrac{1}{2}\ \bar{P}_2 \text{ and } [d] = \bar{F}_1 - \tfrac{1}{2}\ \bar{P}_1 - \tfrac{1}{2}\ \bar{P}_2.$$

The scaling test (C) was used to determine the presence of other factors such as differential viability, maternal effects or interactions between genes (often referred to as epistasis or non-allelic interaction). This was calculated using the student test (t_{df})) as follows:

$$t_{(389df)} = C/\sqrt{s^2_c}$$

where, $s^2_c = 16s^2_{\bar{F}2} + 2\ s^2_{\bar{F}1} + 2\ s^2_{\bar{R}F1} + s^2_{\bar{P}1} + s^2_{\bar{P}2}$

$$C = 4\ \bar{F}_2 - \bar{F}_1\ - R\ \bar{F}_1\ -\ \bar{P}_1 -\ \bar{P}_2$$

and $s^2_{\bar{F}2}$, $s^2_{\bar{F}1}$ represents the variance of the mean for each generation (Kearsey and Pooni, 1996).

The F-statistic (F) tested the presence of the genetic variation in the F_2 for each trait and was calculated using the formula:

$$F_{(397,92df)} = s^2_{F2}/V_E$$

where V_E is the environmental variation calculated as the average of the variance of the non-segregating populations, and s^2_{F2} is the variance of the F_2 population.

Finally, the broad sense heritability (h_b^2) of the F_2 population was calculated as:

$h_b^2 = (s^2_{F2} - V_E)/\ s^2_{F2}$ (Kearsey and Pooni, 1996).

3.2.7.2 Marker data

The mapping data were obtained by visual scoring of the spreadix or agarose gels. Only clear, unambiguous bands were scored. The goodness of fit of observed-to-expected allelic ratios was analyzed using the chi-square test (χ^2). Markers were defined as polymorphic

fragments that did not significantly depart from Mendelian ratios at the p = 0.05. The correlations between markers were calculated using the Minitab program.

3.2.7.3 QTL mapping

Thirty microsatellite markers, chosen to cover the *Arabidopsis* genome at intervals of approximately 20cM (Table 3.2) served as the basis for the QTL analysis. The 'marker regression' approach (Kearsey and Hyne, 1994) was used to determine the QTL positions and estimate their effects. A QTL was assumed present when the p-value associated with the regression was below 5%. If the residual was significant (p<1%) it was assumed that one QTL did not adequately explain the variation and a second QTL was added to the model.

3.3 Results

3.3.1 Phenotypic variation

Significant differences were observed between the parents in TTL, RL20, CL20, HT34, RLF and HTF (Appendix II). This significance between the parents suggest that there are differences between the parents, which brings about differences in the F_1 and F_2 population due to recombination of alleles. The Landsberg *erecta* parent grew fast (RL20, HT20), produced flower buds early (TTB) and hence flowered earlier (TTF) than Columbia parent (Fig. 3.4a-d). On the other hand, the Columbia parent produced more leaves (RL20, RLF, CLF) and grew taller than Landsberg *erecta* (HTF, HT34).

In all the traits there was no significant difference between the F_1 and its reciprocal (RF_1) (Appendix II). This may suggest that there is no difference whichever parent is chosen as the male or female in the Columbia and Landsberg *erecta* cross. Significant differences between the mid-parent and the F_2 mean were observed for TTG, TTL, RL20, TTB and HTF. The F_1 and F_2 means were intermediate to the parental means for most of the traits, except for HT20, HT34 and HTF where higher values than the parents were observed. Higher values for height traits suggest dominance or partial dominance. HT20 values were higher than both parents but closer to the Ler parent. The F_2 population included individuals with a wide range of phenotypes and transgressive segregates were observed for many traits (Fig. 3.4a-d). All the traits showed continuous variation in the F_2 population, exhibiting nearly normal distribution.

The additive effect [a] for the traits is very small except for HT34 and HTF (Table 3.3). The small value of the additive effect could imply that the parents slightly differ for the traits or

that there is gene dispersion. The high value of the dominance effect [d] for HT20, HT34 and HTF suggests that there is strong directional dominance for the height traits. The scaling test shows that the additive-dominance model is adequate, so only the additive and dominance effects are acting on the traits and there are no maternal or epistatic effects.

Significant genetic variation (measured by F-statistic) was observed in TTL, RL20, TTB, HT34, RLF, CLF, HTF and TTF, indicating that at least one or more QTL were segregating for these traits (Table 3.3). The broad sense heritability varied from 0.11 to 0.50 for TTG and RL20 respectively. The traits measured at flowering time (TTF, RLF, CLF and HTF) showed very close heritability values, between 0.34-0.39. These values represent the relative strength of the genotypic effects and low values indicate that the genotype is masked by environmental or other sources of variation. Consequently traits with low heritability have an unreliable relationship between the genotype and phenotype of individuals.

3.3.2 Marker data

The segregation analysis was carried out between homozygous and heterozygous alleles. Of the 30 markers segregating 1:1 only 20% and 27% showed distorted segregation between the homozygous, and between homozygous and heterozygous, respectively (Table 3.4). Among the distorted ratios between the homozygous, 5 out of 6 (83%) were due to excess of Columbia (P_1) genotype. The distortion between the homozygous and heterozygous was due to the excess of the heterozygous genotype. About 30% of the markers showed distortion from the 1:2:1 Mendelian ratio. Most of the distortions (about 56%) were observed in chromosome 5. The heterozygous genotype was in excess in all except marker nga249. This segregation distortion indicates that random sampling was not completely effective in

preventing some sort of natural selection (Lister and Dean, 1993; Alanso-Blanco et al. 1998a).

Significant correlations were observed between markers in a chromosome (Tables 3.7-3.11 and Appendix II). In most chromosomes highly significant correlations were observed between markers close to each other than markers located further from each other. For example, in chromosome 1 high correlation was observed between markers nga59 and F20D23 than between markers nga59 and nga111. Significant correlation showed the linkage relationship between the markers, with high correlation showing high linkage relationship. Significant correlation were not observed between markers in a chromosome only between markers nga8 and T18A10 and nga8 and T5J17 in chromosome 4, and MED24 and MMJB21A and MED24 and MM19 in chromosome 5. In this case, there was segregation distortion in nga8 (chromosome 4) and MED24 and MM19 (Chromosome 5).

3.3.3 QTL analysis

The correlation between markers and traits were determined using the single marker ANOVA in the QTL café (web.bham.ac.uk/g.g.seaton). A total of 94 correlations were detected between the markers and the traits (Appendix II). The marker regression method (Kearsey and Hyne, 1994) was then used to identify QTL affecting morphological traits in *Arabidopsis thaliana* F_2 plants. Of the 94 significant correlations, the marker regression method revealed 23 QTL affecting six morphological traits. Nineteen QTL were detected for flowering characters (TTB, TTF, RLF & HTF), 1 for HT34 and 3 for RL20 (Table 3.5, Fig. 3.3a-b). The QTL were not evenly distributed and showed some clustering in regions on chromosomes 2 and 5. More than 3 QTL were located close to markers nga1126 and nga139

in chromosomes 2 and 5, respectively. Where two QTL were detected for a trait in a chromosome, the QTL had opposite signs of their additive effects and this supports the biometric evidence for gene dispersion in the parents. Fig. 3.3 shows the location of the QTL on Arabidopsis chromosomes, with the direction of the additive effect indicated by the arrow and the length of the arrow indicating the confidence interval for the QTL positions. In positions where more than 1 QTL mapped, the length of the arrow shows the QTL with the largest confidence interval. The proportion of the genetic variance explained by individual QTL ranged from 5% to 63% of the additive variance and the trait totals ranged from 32% to 74% (Table 3.5).

3.3.3.1 QTL for flowering time

In the segregating Col x Ler F_2 population, 6 QTL affecting TTF were detected (Table 3.5, Fig. 3.3a-b). The QTL mapped to chromosomes 1, 2, 3 and 5. The QTL on chromosomes 2 and 5 mapped to 50 and 42 cM respectively, whereas in chromosomes 1 and 3, two linked QTL were detected. The QTL in chromosomes 1 and 3 are in repulsion, so their effects cancel each other. In chromosome 1 the QTL mapped to 34 and 90 cM, whereas in chromosome 3 they mapped to 28 and 60 cM. A full model containing these QTL explained 37% of the additive variance, with individual QTL explaining 5-11% of the variance. The allele effects were consistent with the difference between parents in chromosomes 2 and 5, as the Col alleles increased the time to flower (Fig. 3.4d).

3.3.3.2 QTL for budding time

Seven QTL were detected for TTB in all the 5 Arabidopsis chromosomes (Table 3.5 & Fig. 3.3a-b). The QTL mapped to the same regions as the QTL for flowering time in

chromosomes 1, 2, 3 and 5. The 2-linked QTL mapping to chromosomes 1 and 3 are also in repulsion and located to the same positions as those for TTF. The QTL detected on chromosome 4 for TTB was not detected for TTF. The additive effect of the same QTL detected for TTB and TTF are similar, suggesting the same genes may be involved. A full model containing these QTL explained 55% of the phenotypic variance, with individual models explaining 6-21% of the variance. The allele effects were consistent with the difference between parents in chromosomes 2, 4 and 5, as the Col alleles increased the time to produce flower buds (Fig. 3.4b).

3.3.3.3 QTL for rosette leaves

The QTL detected for RL20 in chromosomes 2, 4 and 5 at around 56, 34, and 38 cM respectively, were also detected for RLF (Table 3.5). The allele effects were consistent with the difference between parents in chromosomes 2, 4 and 5, as the Col alleles increased the number of rosette leaves in both traits. A major QTL explaining the most additive effect (about 33%) mapped to chromosome 2 for both RL20 and RLF. The QTL for RL20 explained 57% of the additive variance while those for RLF explained 72% of the variation.

3.3.3.4 QTL for height

Four QTL were detected for height traits (HTF and HT34) in chromosomes 1, 2 and 5 (Table 3.5). A major QTL was detected in chromosome 2 for HTF and HT34, explaining 63% and 33% of the additive variance, respectively. The Col allele in chromosome 2 increased height. The QTL for HT34 and HTF contributed 33% and 74% of the phenotypic variation for each trait, respectively.

3.3.4 Correlation between the traits

Highly significant correlation was found between traits in the F_2 population (Table 3.6). TTF was highly correlated with TTB (r=0.94). TTB and TTF also correlated positively with TTG, TTL, RLF and CLF. Negative correlation with TTB and TTF was observed in HT20, RL20, and CL20, suggesting that fast growing plants mostly flower early. Repeat measurements of the same traits were positively correlated such as RL20 with RFL (r=0.75), HT20 with HT34 (r=0.24) and HTF (r=0.30). Correlation was also observed between traits measured at the same time such as HT20 with CL20 (r=0.60). QTL for the correlated traits mapped to the same genomic regions. For example, QTL for RL20 and RLF mapped to chromosome 5; HT34 and HTF to chromosome 2; TTB and TTF to chromosomes 1, 2, 3 and 5. The correlated traits also showed similar gene action, suggesting that they maybe under the control of the same genes.

3.4 Discussion

In this study, the Columbia and Landsberg *erecta* were used as the parental lines. There were significant differences between the parents in TTL, RL20, CL20, HT34, RLF and HTF. No significance was observed between the F_1 and its reciprocal, suggesting that there is no difference whichever parent is used as the male or female. Significant differences between the mid-parent and the F_2 plants were observed between TTG, TTL, RL20, TTB and HTF. The frequency distributions of the F_2 population indicated the presence of transgressive variation (Fig. 3.4a-d). The wider spread of the F_2 scores can be attributed to the random association of increasing and decreasing alleles in the recombinants that occur dispersed in the parental lines (Kearsey and Pooni, 1996). The scaling test showed that the additive-dominance model is adequate, so only the additive and dominance effects are controlling the traits and there are no maternal or epistatic effects (Kearsey and Pooni, 1996).

The χ^2 test and correlations between markers pointed to segregation distortion by some markers. While χ^2 test indicated that segregation ratios differed from 1:2:1 for 9 markers, the major disturbances were observed on chromosome 5. It is possible that these distortions are independent of each other for various markers. But it is also possible that distortion at a marker nearer to the centromere is observed among markers that lie beyond this marker towards the terminal end.

Correlations between marker scores have an expectation of (1-2RF) because:

$$r = cov/\sqrt{(var_1 \times var_2)}$$

where, $cov = \frac{1}{2}(1-2RF)$, $var_1 = \frac{1}{2}$ and $var_2 = \frac{1}{2}$

This formula gives the correlations between various pairs of markers as follows:

Marker distance, cM	RF	Correlation
15	0.129	0.74
20	0.165	0.67
25	0.197	0.61
30	0.225	0.55
35	0.251	0.50
40	0.275	0.45
50	0.316	0.36

As hardly any pair of markers are less than 15cM away from each other, no correlation between them is expected to be >0.74. Also, as there were more heterozygous in general for many markers, this correlation will be even lower. Consequently, the correlation between markers that are next to each other is not likely to be much different from 0.50 and this is what is observed in most cases. While correlations between most markers on each chromosome show expected trend, indicating that there is little segregation distortion per se and its effects are minute, but the same is not true for chromosome 4. Segregation distortion is a serious problem in this chromosome. It makes correlations for markers T18A10 and nga8 non-significant when the expected value of r between these markers would be in the range 0.55-0.61. Large distortions observed for markers on chromosome 5 (Table 3.4), however, do not translate into very visible distortions in the correlation values in Table 3.11, largely because these distortions are consistent across markers.

A population of 400 F_2 plants was used to identify and map QTL that control heritable variation during the vegetative and reproductive phases. The type and size of the population used was essentially ideal and should contribute positively to the efficient detection of QTL for various traits (Hyne and Kearsey, 1995). Significant genetic variation in traits TTL,

RL20, TTB, HT34, RLF, CLF, HTF and TTF suggested that one or more QTL per trait were segregating in the F_2 population. The study revealed 23 QTL affecting six morphological traits. The number of QTL varied from 1 to 7 per trait and between 3 to 6 per chromosome. QTL controlling different traits were correlated and often localized to the same genomic regions.

Most QTL associated with TTB were also associated with TTF. This is expected, since many genes that affect time to bud formation would also affect time to flowering. The trait 'days from bud to flower' was a measure of the interim period between bud formation and floral opening. In previous studies a significant correlation has been observed between leaf number at flowering and the time of flowering (Clarke et al. 1995; Kuittinen et al. 1997). In this study a positive but low correlation was observed between number of rosette leaves at flowering and flowering time. The QTL for rosette leaves were detected at the same positions as those for flowering time in chromosome 2 and 5. The other trait measured at flowering that also has a QTL detected at the same position as the QTL for flowering time was HTF (in chromosome 2 and 5). The alleles in Columbia parent in chromosomes 2 and 5 increased TTF, TTB and RLF. These results are consistent with the difference between the parents. The Columbia parent had a decreasing allele for height in chromosome 5, and this is not consistent with the difference between the parents. Lynch and Walsh (1998) also observed that the direction of allele effects was not consistent with the direction of the difference between parental lines.

QTL for repeat measurements such as height and number of rosette leaves mapped to similar regions of the genome. For example, QTL for height at flowering and height at 34 days

mapped to similar regions in chromosome 2 and those for rosette leaves at 20 days and at flowering mapped to similar regions in chromosomes 2, 4 and 5. In both cases the QTL showed similar mode of action. QTL for consecutive measurements that map to overlapping regions would be strongly suspected as being the effects of a single gene that maintains influence over a trait for an extended period. Therefore, it could be concluded that the same genes could be involved in the control of these traits. In other words, the QTL are showing pleiotropic effects, particularly for traits that are measured at different stages of the plant's life cycle.

The precision of QTL mapping is still a problem because it depends on several factors such as the heritability of the character, number of QTL involved, the distribution of QTL over the genome, number of marker loci and their distribution and the size of the population (Knapp et al., 1990). But the precision of QTL position depends more on the population size than the number of markers and no notable increase in accuracy is obtained with more than five well spread markers for each chromosome (Kearsey & Pooni, 1996). However, it is worthwhile to use a mapping population of relatively large size and QTL of high heritability for reliable estimation of QTL effect.

The 95% confidence interval for the mapped QTL ranged from 0 to 42 cM with an average of 20cM, indicating a low accuracy of the estimates. However, this interval is much smaller than 30cM that is often observed for segregating populations (Kearsey and Farquhar, 1998). It is difficult to narrow down the confidence interval associated with the precision of QTL position with present mapping software while using formal mapping populations of RILs, F_2,

double haploids and backcrosses. It is necessary to use near-isogenic lines (NILs) and substitution lines to achieve finer mapping (Koumproglou et al., 2002).

Possible candidate loci

The availability of mapping data, coupled with the whole genome sequence map available for *Arabidopis thaliana* (Arabidopsis Genome Initiative, 2000), provides an opportunity to identify candidate genes for flowering and other morphological traits. Several QTL detected in this study span regions containing candidate genes with known functions, and the study will attempt to compare what appears to be similar 'significant' QTL from other studies. However, uncertainty may arise through these comparisons across studies as significance levels for the identification of QTL may vary, raising questions as to whether QTL studies identified in the same chromosomal region between studies are exactly the same QTL (Burns, et al., 2003).

The QTL for HTF, TTB and TTF were detected towards the middle of chromosome 1 and mapped to the same region as *QLN-2* detected in Col x Ler RIL population (Jansen et al. 1995) and *RLN-1* (Clarke et al. 1995). Clarke and co-workers studied the number of rosette leaves at flowering in Ler x H51 F_2 population. The TTB and TTF QTL detected at 90 cM corresponds with the *RLN-2* (Clarke et al. 1995). In their studies Clarke and co-workers observed that the *RLN-1* and *RLN-2* had allelic effects that reduced leaf number in H51. Mutations affecting flowering time have also been identified in the middle and towards the end of chromosome 1. The middle of chromosome 1 maps *FHA* and *GI* mutations, whereas *FT* and *FE* maps towards the end of the chromosome (Koornneef et al. 1991, Weinig et al., 2002). *CHRYPTOCHROME 2* (*CRY2*) is also an attractive candidate for the QTL at the top

of chromosome 1. This blue-light photoreceptor plays a central role in perception of long days (Guo et al., 1998) and has been identified as the QTL originally referred to as *EDI* in RILs from the Ler x Cvi cross (El-Assal et al., 2001).

Six QTL involving RL20, RLF, HTF, HT34, TTB and TTF mapped to around 50 cM (48 to 56) in chromosome 2. QTL with large effect were observed for HTF (62.52%), RL20 (33.92%), HT34 (32.74%) and RLF (32.40%). The candidate genes may be the *fpa* and *erecta* mutations. The *fpa* mutation has been found to be responsive to vernalization under long days and short day conditions in the Ler *ecotype* (Koornneef et al. 1991). The *erecta* mutation affects inflorescence architecture in *Arabidopsis* (Ungerer, et al. 2002). The traits mapping to chromosome 2 all showed increasing effect for the Col allele and are consistent with the results. The QTL for TTF and TTB were located at 30cM in chromosome 3 and the candidate gene maybe the *QLN-7* (Jansen et al. 1995) detected at the top of chromosome 3.

In chromosome 4, the QTL for TTB was located close to the *RFI* gene thought to delay flowering in Arabidopsis ecotype Stockholm (Koornneef et al. 1994). The *FRI* locus is thought to be allelic with the *FLA* locus mapped by Lee et al., 1993. The Col allele increased the time to produce flower buds in this study. The QTL for RL20 and RLF were located at around 30 cM on chromosome 4. The candidate gene may be the *QLN-9* QTL (Jansen et al. 1995). The Ler allele for *QLN-9* was reported to increase the number of leaves and early flowering (Jansen et al. 1995). In this study, the Col allele increased the number of rosette leaves.

The QTL affecting RL20, TTB, RLF and TTF mapped close to the *RLN4* gene detected towards the top of chromosome 5 (Clarke et al. 1995) and flowering mutations *CO*, *FY* and *FLC*. The Ler allele at *FLC* has been shown to suppress many of the late genotypes, like the *FRI* and *LD* alleles (Koornneef et al. 1994). The QTL for HTF was located at 80 cM in chromosome 5 and this is close to the flowering genes *QLN12* (Jansen et al. 1995) and *DFF-2* (Kowalski, et al. 1994b). The *DFF-2* QTL was identified in a cross between ecotypes Hannover/Münden and Wassilewskija (WS) F_2/F_3 population. An increasing effect for the Ler parent was observed for HTF, consistent with the results.

Table 3.2: Length of the chromosome covered, number of QTL, number of markers and average number of markers per chromosome

Chromosome	Length of chromosome (cM)	Number of QTL	Number of markers	Average distance between markers (cM)
1	115.55	5	6	19.25
2	73.77	8	5	14.75
3	87.88	5	7	12.55
4	119.00	4	6	19.83
5	116.90	7	6	19.48
Total	513.10	29	30	17.17

Table 3.3: Genetic and environmental components of variation for various traits

Trait	[a]	[d]	C	s_c^2	$t_{(92)}$	s^2_{F2}	V_E	$F_{(23,92)}$	h_b^2
TTG	0.31	0.31	-0.96	0.69	-1.16	2.08	1.85	1.12	0.11
TTL	0.55	0.73	-1.99	1.36	-1.71	3.20	4.64	1.38*	0.28
RL20	1.25	0.26	-0.51	0.42	-0.79	1.94	0.98	1.99**	0.50
CL20	0.80	0.81	-0.76	0.39	-1.22	2.00	1.75	1.14	0.13
HT20	1.25	33.44	-15.67	212.82	-1.07	463.54	389.72	1.19	0.16
TTB	1.13	1.90	-0.74	2.19	-0.5	7.05	4.91	1.44*	0.30
HT34	12.3	38.12	43.48	1352.77	1.18	4536.02	2911.93	1.56**	0.36
RLF	1.23	0.48	0.51	0.45	0.76	2.56	1.66	1.54**	0.35
CLF	0.09	0.29	0.36	0.23	0.75	1.18	0.78	1.51**	0.34
HTF	10.0	29.10	19.35	199.43	1.37	942.02	578.57	1.63**	0.39
TTF	0.90	1.37	0.29	1.78	0.22	6.14	3.85	1.59**	0.37

[a] = additive effect; [d] = dominance effect, t = Student T-test, F = F-statistic, s^2_{F2} = variance for F_2 population; V_E = environmental variance; h_b^2 = heritability in the broad sense; C = scaling test; s_c^2 = variance of the generation means

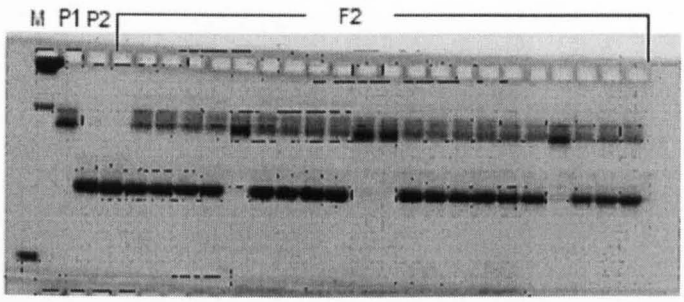


Figure 3.1: Microsatellite marker nga139 with M as the DNA ladder, P_1 is Columbia parent and P_2 Landsberg parent and F_2 plants with band sizes of 174 and 132 for Columbia and Landsberg, respectively on Spreadex gel

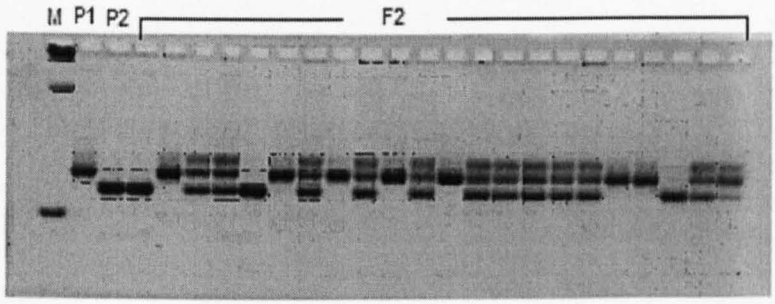


Figure 3.2: Microsatellite marker nga249 with M as the DNA ladder, P_1 is Columbia parent and P_2 Landsberg parent and F_2 plants with band sizes of 125 and 115 for Columbia and Landsberg, respectively on Spreadex gel

Table 3.4: χ^2 test for segregation distortion for the markers used to genotype F_2 plants

Marker	Marker number	Chr.	Number of individuals (P_1:Het:P_2)	X^2 test		
				1:1 (P_1 vs P_2)	1:1 (Homozygous vs heterozygous)	(1:2:1)
Nga59	38	1	79:212:97	1.84	3.34	5.18
F20O23	43	1	75:173:87	0.89	0.36	1.25
Nga392	42	1	102:190:78	3.20	0.27	3.47
T27k12sp	32	1	79:207:103	3.16	1.61	4.77
Nag280	14	1	93:213:83	0.57	3.52	4.09
Nga111	7	1	100:206:84	1.39	1.24	2.63
Nga145	15	2	56:255:59	0.08	52.97***	53.05***
MSF3A	73	2	88:179:99	0.65	0.17	0.82
Nga1126	16	2	97:212:74	3.09	4.39*	7.48*
Nga361	39	2	86:190:76	0.62	2.23	2.85
Nga168	10	2	96:183:81	1.27	0.10	1.37
Nga172	25	3	74:212:91	1.75	5.86*	7.61*
Nga162	24	3	87:212:88	0.01	3.54	3.55
Athgapar	31	3	76:168:89	1.11	0.03	1.14
MMJ24	76	3	89:203:95	0.20	0.93	1.13
TH620B	77	3	77:167:103	3.76	0.49	4.25
Nga707	66	3	85:204:97	0.79	1.25	2.04
Nga112	47	3	82:188:90	0.37	0.71	1.08
T18A10	86	4	105:168:74	5.37*	0.35	5.72
Nga8	59	4	105:216:69	7.45**	4.52*	11.97**
FCA9	79	4	73:167:79	0.24	0.71	0.95
F20O24	80	4	85:179:104	1.91	0.27	2.18
Nga1139	62	4	96:203:91	0.31	0.66	0.97
T5J17	81	4	86:188:82	0.10	1.12	1.22
MED24	82	5	93:228:59	7.61**	15.20***	21.81***
Nga249	13	5	117:194:75	9.19**	0.01	9.20**
Nga139	8	5	106:216:66	9.30**	4.99*	14.29**
Nga76	5	5	77:224:74	0.06	14.21***	14.27***
MJB21A	83	5	83:190:111	4.04*	0.04	4.08
MM19	84	5	84:216:77	0.30	8.02**	8.32**

Chr. = chromosome

Table 3.5: QTL position, nearest marker, additive effect (a), dominance effect (d), additive variance (V_A), percentage additive variance and the confidence interval (CI) of the QTL in F_2 plants using marker regression approach

Trait	Chr.	QTL position	Nearest marker	a	d	V_A	$\%V_A$	CI
RL20	2	56.0 ± 9.55	Nga1126	0.79	0.23	0.62	33.70	19
	4	34.0 ± 15.04	Nga8	0.40	-1.33	0.16	8.70	30
	5	38.0 ± 8.73	Nga139	0.52	-0.07	0.27	14.67	17
TTB	1	38.0	Nga392	-1.29	-0.11	0.81	11.50	
	1	90.0	Nga280	0.86	-0.14			
	2	48.0 ± 10.53	Nga1126	0.63	0.03	0.40	5.63	21
	3	30.0	Nga162	-1.41	0.19	0.77	10.97	
	3	60.0	Th620b	1.17	-0.19			
	4	8.0 ± 20.09	T18a10	0.65	0.60	0.42	5.99	40
	5	42.0 ± 7.46	Nga139	1.21	-0.27	1.46	20.77	15
HT34	2	52.0 ± 2.56	Nga1126	39.07	45.80	1526.46	32.71	5
RLF	2	54.0 ± 3.82	Nga1126	0.88	0.29	0.77	32.40	8
	4	26.0 ± 10.06	Nga8	0.49	-0.23	0.24	10.05	20
	5	40.0 ± 4.90	Nga139	0.84	-0.53	0.71	29.52	10
HTF	1	22.0 ± 19.29	F20D23	-6.68	-3.25	44.62	5.08	39
	2	52.0 ± 2.07	Nga1126	23.43	33.14	548.96	62.52	4
	5	74.0 ± 16.54	Mjb21a	-7.40	0.89	54.76	6.24	33
TTF	1	34.0	Nga392	-1.13	-0.25	0.62	10.05	
	1	90.0	Nga280	0.67	-0.07			
	2	50.0 ± 11.49	Nga1126	0.55	0.10	0.30	4.93	23
	3	28.0	Nga162	-1.23	-0.06	0.65	10.65	
	3	60.0	Th620b	1.11	-0.01			
	5	42.0 ± 10.04	Nga139	0.83	-0.34	0.69	11.24	20

NB: Chr. = chromosome

Table 3.6: Pearson's correlation coefficients between the traits (only significant values are given)

	TTG	TTL	RL20	CL20	HT20	TTB	HT34	RLF	CLF	HTF
TTL	0.845									
RL20	-0.340	-0.376								
CL20	-0.397	-0.413								
HT20	-0.331	-0.366		0.600						
TTB	0.573	0.611	-0.133	-0.713	-0.639					
HT34	-0.263	-0.328	0.361	0.129	0.236	-0.363				
RTF	-0.212	-0.235	0.746	-0.183	-0.174	0.188	0.135			
CLF			0.121	-0.197	-0.293	0.319	-0.121	0.274		
HTF			0.225		0.298		0.429	0.189		
TTF	0.599	0.635	-0.202	-0.668	-0.582	0.936	-0.408	0.154	0.352	

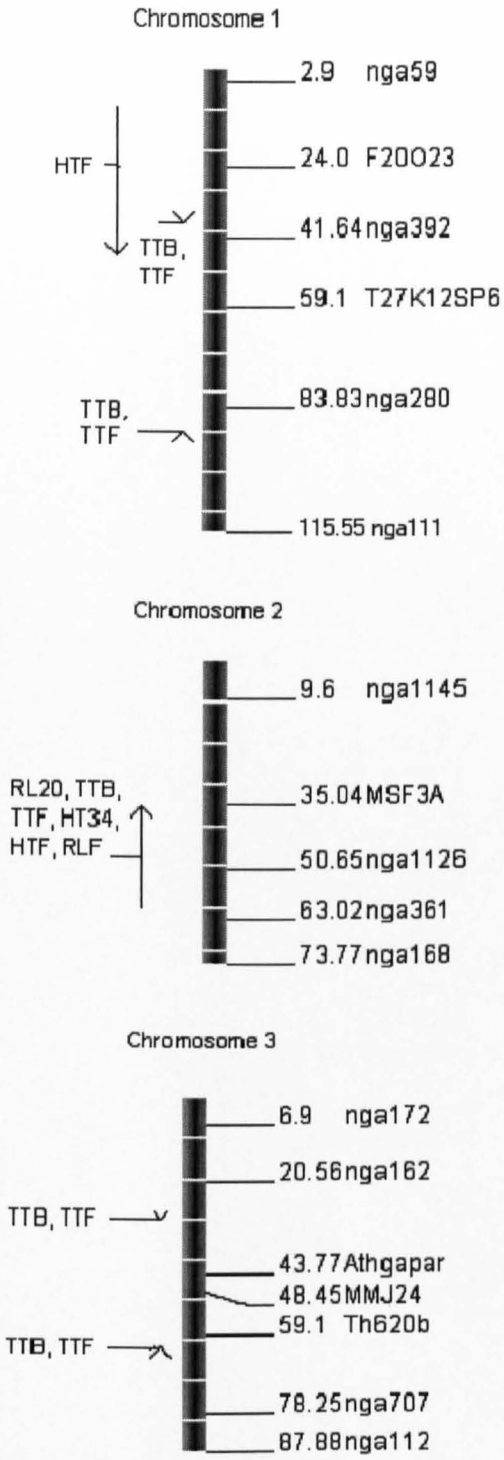


Figure 3.3a: The location of QTL on Arabidopsis chromosomes 1, 2 and 3 with the direction of additive effect indicated by the arrow

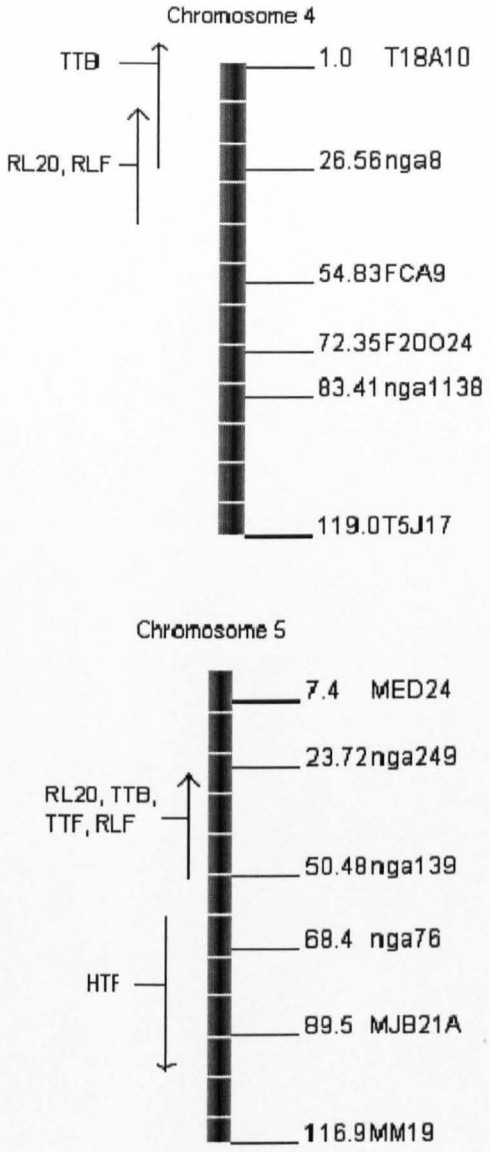


Figure 3.3b: The location of QTL on Arabidopsis chromosomes 4 and 5 with the direction of additive effect indicated by the arrow

Table 3.7: Correlation between markers in chromosome 1 and the position of the marker in brackets

	Nga59 (2.90)	F20D23 (24.00)	Nga392 (41.64)	T27K12SP (59.10)	Nga280 (83.88)
F20D23 (24.00)	0.41				
Nga392 (41.64)	0.67	0.37			
T27K12SP (59.10)	0.33	0.35	0.34		
Nga280 (83.88)	0.19	0.22	0.23	0.65	
Nga111 (115.55)	0.13	0.13	0.13	0.30	0.44

Table 3.8: correlation between markers in chromosome 2 and the position of the marker in brackets

	Nga1145 (9.60)	MSF3A (35.04)	Nga1126 (50.65)	Nga361 (63.02)
MSF3A (35.04)	0.35			
Naga1126 (50.65)	0.23	0.56		
Nga361 (63.02)	0.23	0.26	0.52	
Nga168 (73.77)	0.25	0.21	0.46	0.67

Table 3.9: Correlation between markers in chromosome 3 and the position of the marker in brackets

	Nga172 (6.90)	Nga162 (20.56)	Athgapar (43.77)	MMJ24 (48.45)	Th620b (59.10)	Nga707 (78.25)
Nga162 (20.56)	0.52					
Athgapar (43.77)	0.31	0.40				
MMJ24 (48.45)	0.33	0.57	0.59			
Th620b (59.10)	0.17	0.13	0.56	0.36		
Nga707 (78.25)	0.18	0.16	0.19	0.36	0.49	
Nga112 (87.88)	0.14	0.24	0.16	0.29	0.28	0.55

Table 3.10: Correlation between markers in chromosome 4 and the position of the marker in brackets

	T18A10 (1.00)	Nga8 (26.56)	FCA9 (54.83)	F20D24 (72.35)	Nga1138 (83.41)
Nga8 (26.56)	0.05ns				
FCA9 (54.83)	0.15	0.10			
F20D24 (72.35)	0.10	0.16	0.42		
Nga1138 (83.41)	0.13	0.17	0.30	0.67	
T5J17 (119.00)	0.19	0.04ns	0.27	0.41	0.63

Table 3.11: Correlation between markers in chromosome 5 and the position of the marker in brackets

	MED24 (7.40)	Nga249 (23.72)	Nga139 (50.48)	Nga76 (68.40)	MJB21A (89.50)
Nga249 (23.72)	0.52				
Nga139 (50.48)	0.35	0.67			
Nga76 (68.40)	0.21	0.37	0.51		
MJB21A (89.50)	0.09ns	0.15	0.28	0.41	
MM19 (116.90)	0.06ns	0.14	0.24	0.20	0.37

NB: ns = non-significant

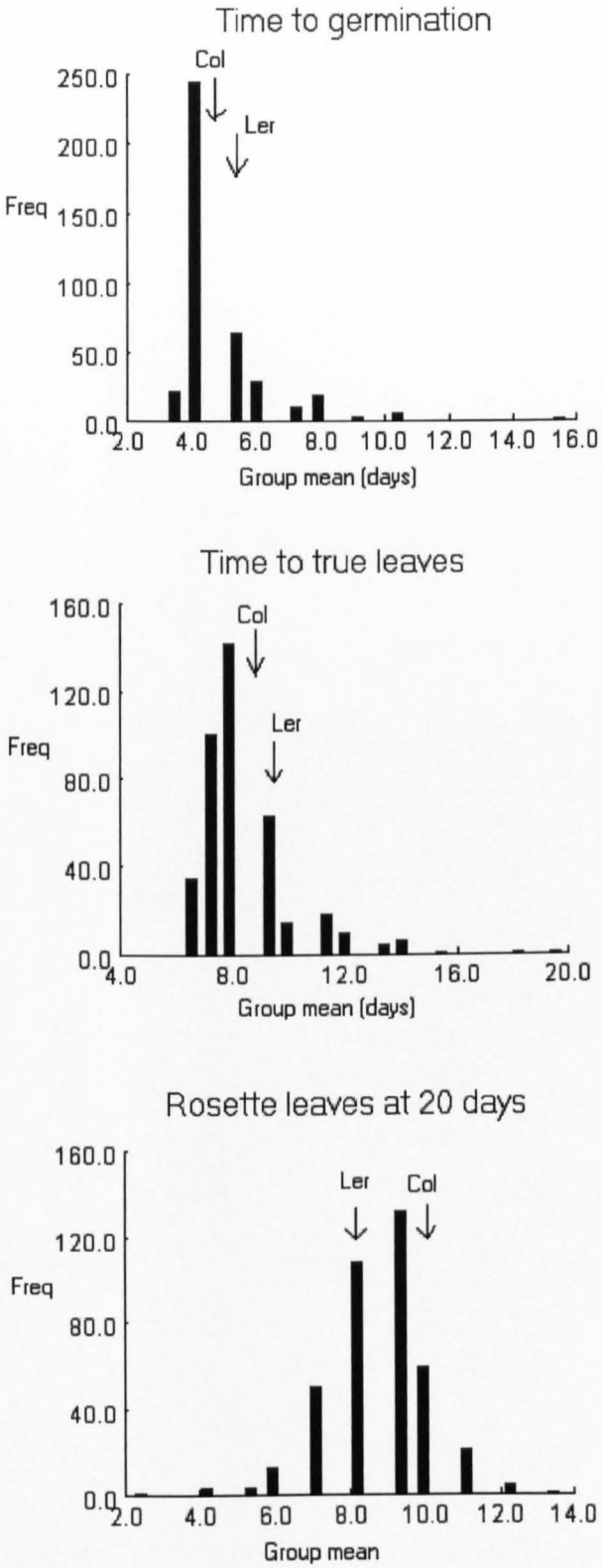


Figure 3.4a: Trait distribution for time to germination, time to true leaves and rosette leaves at 20 days in the F_2 population with arrows indicating the mean position of the parents

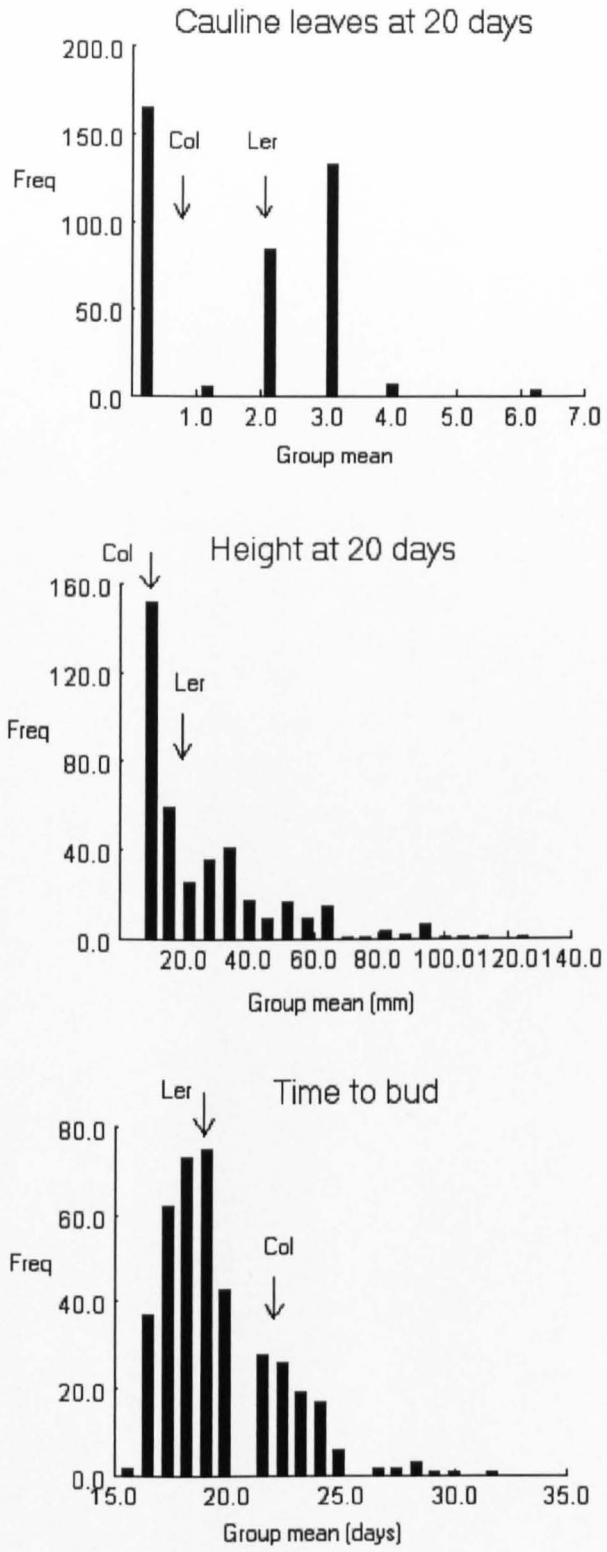


Figure 3.4b: Trait distribution for cauline leaves at 20 days, height at 20 days and time to bud in the F_2 population with arrows indicating the mean position of the parents

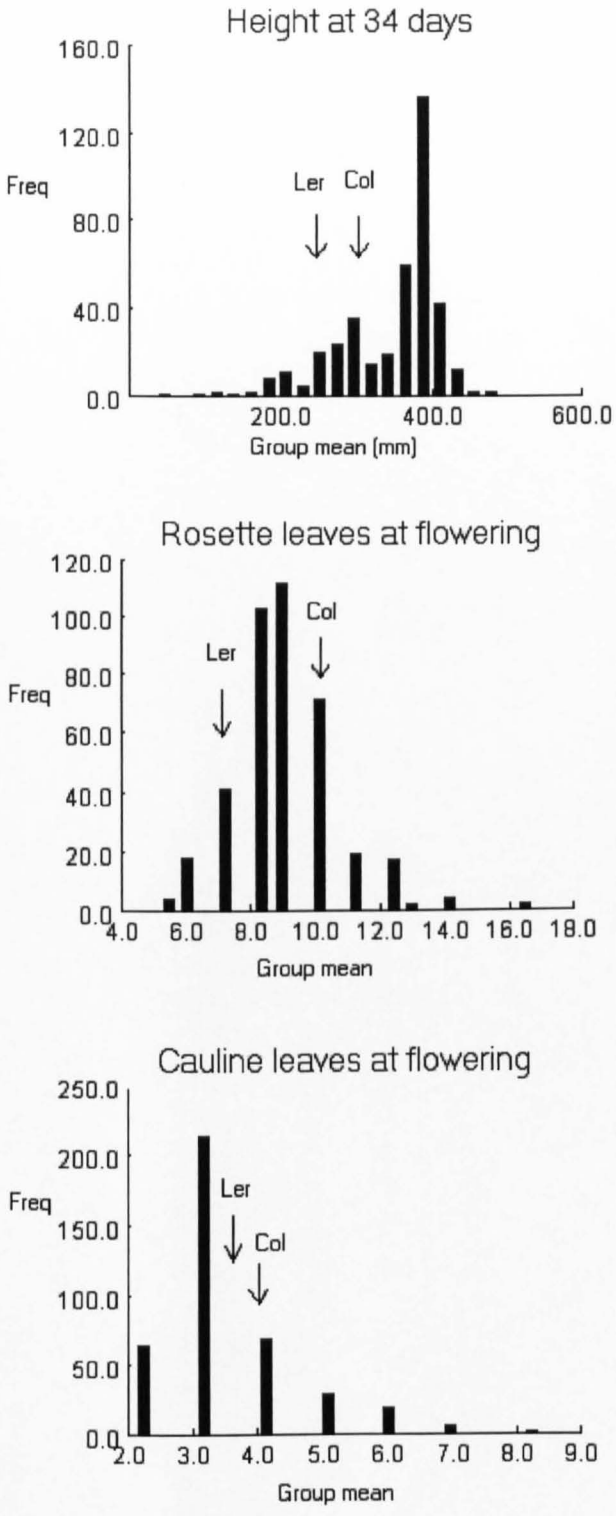


Figure 3.4c: Trait distribution for height at 34 days, rosette leaves at flowering and cauline leaves at flowering in the F_2 population with arrows indicating the mean position of the parents

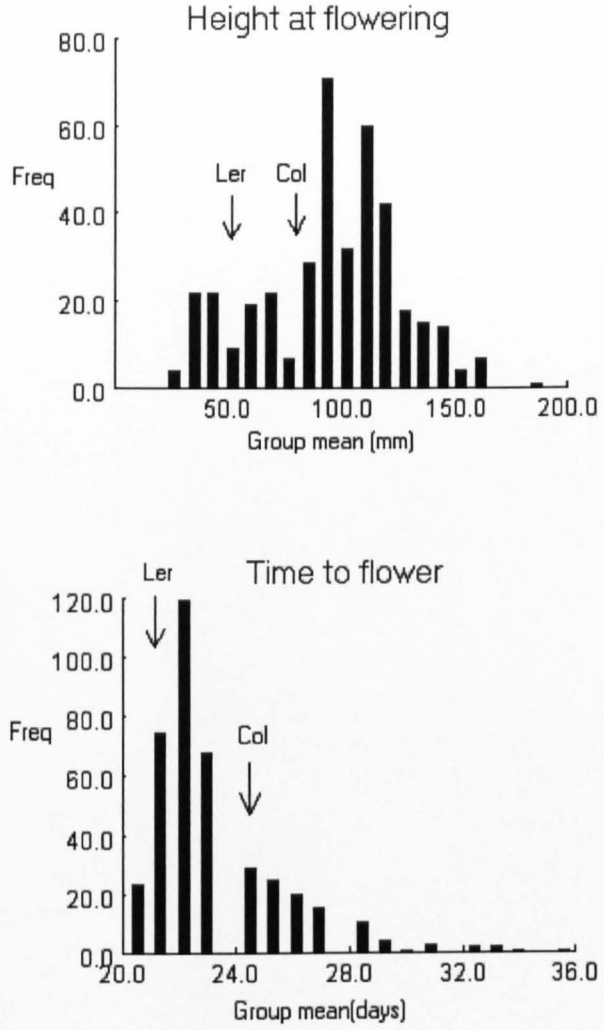


Figure 3.4d: Trait distribution for height at flowering and time to flower in the F_2 population with arrows indicating the mean position of the parents

CHAPTER 4

Mapping quantitative trait loci using the interval mapping method in *Arabidopsis thaliana* F_3 families

4.1 Introduction

Quantitative genetic studies, including the use of quantitative trait locus (QTL) mapping techniques, provide an opportunity to investigate the underlying genetic mechanisms that regulate developmental programs in plant architecture. QTL mapping approaches are some of the first steps in identifying naturally occurring allelic variation and can serve as a gene discovery tool by facilitating the identification of new developmental genes. The identification of naturally occurring allelic variation in genes not only contributes to a greater understanding of the developmental genetics, but also allows to explore the evolutionary and ecological implications of variation in the development patterns.

The present study is a continuation of the work described in the previous chapter where QTL were detected in the F_2 population. In this study, the interval mapping method was used for detecting QTL in *Arabidopsis thaliana* F_3 families. The interval mapping method (Lander and Botstein, 1989) uses an estimated genetic map as the framework for the location of the QTL. The intervals that are defined by ordered pairs of markers are searched (for example, 2 cM) and statistical methods are used to test whether a QTL is likely to be present within the interval or not. The results of the tests are expressed as LOD (logarithm of the odds) scores, which compare the evaluation of the likelihood function under the null hypothesis (no QTL) with the alternative hypothesis (QTL at the testing position) for the purpose of locating

probable QTL (Churchill and Doerge, 1994; Doerge, 2002). The maximum LOD profile has the potential to indicate multiple or ghost QTL incorrectly. Determining which of the many peaks indicates a single QTL leads to issues of determining statistically significant results (Lande & Botstein, 1989). Significance levels have to be adjusted to avoid false positives resulting from multiple tests, while confidence intervals are set at the map interval corresponding to a 1 LOD decline either side of the peak. The approach of interval mapping considers one QTL at a time and this can bias identification and estimation of QTL when multiple QTL are located in the same linkage group (Haley and Knott, 1992, Zeng, 1994).

The main purpose of this study was to confirm if the QTL detected in the F_2 were also expressed in the F_3 and if using the F_3 families will reduce the confidence interval to some degree. The other purpose of the F_3 evaluation was to test the efficiency of MAS when applied to the F_2. The QTL detection in the F_3 will also go a long way to explaining the realised response to selection in the F_2. Finally, a comparison of the QTL detection by marker regression method of Kearsey and Hyne (1994) and interval mapping of Lande and Thompson (1989) will also be possible.

4.2 Materials and methods

4.2.1 Plant material

The 400 F_2 plants evaluated in Chapter 3 were self-pollinated by hand generating F_3 families. A subset of 200 F_3 families was self-pollinated and evaluated for morphological traits together with the parents and F_1.

4.2.2 Evaluation of morphological traits

Seeds were sown in the glasshouse on 7th July 2002. The experiment consisted of 5 plants from each 200 F_3 families and 25 each of the parents and F_1 generation. Three seeds were sown per pot and seedlings were thinned to one per pot after germination. All plants were randomised within the experiment and surrounded by guard plants. Each plant was scored for the same 11 morphological traits that were measured in the F_2 population. The traits were: time to germinate (TTG), time to true leaves (TTL), height at 20 days (HT20), cauline leaves at 20 days (CL20), rosette leaves at 20 days (RL20), time to bud (TTB), time to flower (TTF), height at flowering (HTF), rosette leaves at flowering (RLF), cauline leaves at flowering (CLF) and height at 34 days (HT34).

4.2.3 Phenotypic data analysis

The analysis of variance (ANOVA) using the general linear model (GLM) was applied to the F_3 families to detect differences between the families, Appendix III. The skeletal ANOVA for the analysis of F_3 families is shown in Table 4.2. The normality of the phenotypic distributions was tested by the Anderson-Darlington normality test (Minitab release 13.1) prior to QTL analysis, and the correlations between traits were calculated using the Minitab

program. The percentage variation explained by the QTL was calculated as the ratio between the variance explained by the QTL (V_A) and the additive variance for the trait (V_A*).

4.2.4 Heritability

The total phenotypic variance was partitioned into sources attributable to additive (V_A^*) and environment (V_E). The components of variance were used to estimate narrow sense heritability (h_n^2). The narrow sense heritability measures the proportion of total phenotypic variation that is due to additive effects of genes.

The proportion of the observed variation that is due to the effects of genes can also be expressed as heritability of the means, when family means are used to estimate the genetic effects. This was done for each trait by dividing the genetic variance component by the variance of the family means. The genetic variance component is equal to σ^2_B and the variance of the line means can be calculated as the between families MS divided by the family size. MS_B is equal to $\sigma^2_W+r\sigma^2_B$ so the variance of the family means is equal to $(\sigma^2_W / r) + \sigma^2_B$ (Table 4.2).

4.2.5 QTL analysis

The marker information developed for the F_2 population was used for mapping QTLs in the F_3 families. This information consisted of 30 microsatellite (SSR) markers spaced at approximately 20cM in each Arabidopsis chromosome and covering 513.10cM of the genome. More details are available from the previous chapter. Firstly, the genotype of the F_3 lines was verified by scoring a few lines with a few selected markers.

The QTL were mapped using the interval mapping method (Lander & Botstein, 1989), Fig. 4.2. The method uses an estimated genetic map as the framework for the location of the QTL. This procedure test sequentially along each chromosome whether intervals flanked by two molecular markers contain a QTL while statistically accounting for other QTL segregating outside the tested interval. The identity (and number) of markers for this genetic background control was determined independently for each trait by forward selection, backward elimination stepwise regression. In this study, a 1 and 5 cM scan window was used when the distance between the markers was less than 20 and 50 cM respectively. The results of the test are expressed as LOD (logarithm of the odds) scores, which compare the evaluation of the likelihood function under the null hypothesis (no QTL) with the alternative hypothesis (QTL at the testing position) for the purpose of locating probable QTL. The location of the maximum LOD profile was taken to indicate the location of the QTL. A LOD score greater than 2, was used to declare the presence of a QTL within a marker in this study (Lander & Botstein, 1989). The confidence intervals were set at the map interval corresponding to a 1 LOD decline either side of the peak (Lander and Botstein, 1989, Haley and Knott, 1992, Zeng, 1994).

Significant associations between specific markers and morphological traits detected by interval mapping were first confirmed by the single factor analysis of variance of the trait means for the marker genotype classes at the marker locus closest to the peak LOD using the QTL café.

4.2.6 Analysis of GxE interactions

The presence of the parental and F_1 families allowed the test of GxE interactions both at the means and variance level. At the variance or micro-environmental level, the within variances of the parents and F_1 generations are not expected to differ from each other both within and across environments. The variances were compared within each experiment using variance ratio. The larger variance was divided by the smaller variance and the probability of F thus obtained corrected by multiplying by n(n - 1) (= 6) before determining its significance. The pooled values of these variances were then compared across experiments to see if average V_E differed between the two experiments. The probability in this case, however was doubled before determining the significance.

The presence of genotype by macro-environmental interactions, on the other hand, was tested using a 2-way ANOVA. In this ANOVA, variation between the parental and F_1 means was partitioned into 3 components, experiments (1df), generation (2df) and generation x environment interactions (2df). The GxE interaction MS was tested against the pooled within MS corrected for the fact that this ANOVA was carried out on generation means. The main effects were also tested against the within MS assuming that both genotypes and experiments are the fixed effects.

4.3 Results

4.3.1 Phenotypic variation

For all the traits, differences between families were highly significant ($P<0.001$), indicating substantial genetic variation for the morphological traits in *Arabidopsis thaliana*. The F_3 family means were normally distributed ($P<0.001$) for all the traits. The Columbia parent (P_1) scored higher than Landsberg parent (P_2) for RL20, TTB, TTF, HTF, RLF, CLF and HT34, and this was also true in the previous experiment (Table 4.1). The F_3 mean was intermediate between the parents in most of the traits except for HT20, TTB, TTF and CL20. In HT20, the F_3 mean was higher than the parents but less than the F_1 (Table 4.1). For TTB and TTF, the F_3 mean was only slightly higher than both parents and close to the Columbia parent, whereas for CL20 it was slightly less than the parents. In this experiment the F_3 mean was not higher than the parents for HTF and HT34 as it was in the F_2 experiment.

The narrow sense heritability of the F_3 families ranged for various traits from 18% (TTL) to 65% (TTF) with a mean of 42% (Table 4.1). The heritability tended to be high for traits associated with flowering time (TTF) such as TTB (0.63), HTF (0.50) and CLF (0.59). In the F_2 experiment traits measured at flowering time also showed heritability values very close to that for flowering time. Repeat measurements such as height, cauline leaves and rosette leaves showed very close heritability values. The heritability of the means ranged from 0.50 (TTL) to 0.83 (TTB and TTF) and represented the relative strength of the additive effects. The heritability of the means values are higher than the heritability values and represent an improved relationship between the genotype and the mean of a family.

There were significant correlations among traits (Table 4.3), and they ranged from –0.693 to 0.937. About 87% of the correlations were significantly different from zero. Strong positive correlations (>0.6) were observed between TTG and TTL, TTF and TTB, HTF and HT34, and RLF and RL20. Time to flower and time to produce some flower buds has previously been identified as strongly correlated in *Arabidopsis thaliana* (Koorneef et al., 1991; Peeters and Koorneef, 1996). TTF correlated positively with RLF and CLF, and negatively with HT20, RL20, CL20 and HT34. The strong correlation between traits may suggest that the traits are likely influenced by the same or by tightly linked loci.

4.3.2 QTL analysis

A total of 40 QTL were detected in the F_3 families (Table 4.4 & Fig. 4.1a-b). For each trait between 1 and 6 QTL were detected, with a mean of 3.6 QTL per trait. The number of QTL varied from 3 to 11 per chromosome and they clustered in two or three regions in each chromosome (Fig. 4.1a-b). The amount of phenotypic variation explained by the QTL varied from 5.76-164.56%. The QTL explaining the most variance mapped to the same region on chromosome 2, for RL20, CL20, HTF, RLF and HT34.

4.3.2.1 Time to germination and true leaves (TTG & TTL)

A single QTL affecting TTG was detected towards the end of chromosome 5. This QTL showed a similar effect as the QTL for TTL mapping to the same position. The Col parent has plus allele for both the TTG and TTL on this QTL. QTL affecting TTL were also detected in chromosomes 1 and 4, and both QTL showed decreasing effect for the Col parent. The 3 QTL for TTL explained 24.08% of V_A*, whereas the single QTL for TTG explained 8.65% of the variation for the trait.

4.3.2.2 Flowering time traits

QTL controlling traits TTB and TTF were detected on all five chromosomes, except that no QTL was detected for TTF in chromosome 3. The QTL for TTB and TTF mapped to similar positions (chromosomes 1, 2, 4 and 5) and showed similar additive effect. A major QTL having a large effect on TTB (LOD score = 29.15) was found in chromosome 5, whereas a major factor for TTF (LOD = 29.85) was found in chromosome 1. The QTL detected in chromosome 2 for TTF and TTB, on the other hand, had a minor effect. The traits TTF and TTB were also positively correlated (r = 0.937). The Col alleles increased TTB and TTF in chromosomes 2, 4 and 5, which is consistent with the results. The QTL for TTB and TTF explained 149.54 and 150.41% of the variation for the trait, respectively. The QTL for TTF and TTB mapping towards the middle of chromosome 1 and those mapping to chromosomes 2 and 5 were detected in F_2 plants using the marker regression method. This also applies to QTL for TTB detected in chromosomes 3 and 4 (Table 4.5 & Fig. 4.1a-b).

The other traits measured at flowering time are RLF, CLF and HTF. RLF and CLF were positively correlated with flowering time and no correlation was observed with HTF. However, the QTL for HTF mapped close to those for flowering time in chromosomes 2 and 4, together with the QTL for RLF. The QTL for CLF and RLF also mapped close to flowering time QTL in chromosomes 1 and 5.

4.3.2.3 Height traits

QTL for height traits mapped to similar regions in chromosome 2 for HTF and HT34, and in chromosome 4 for HT20 and HTF. The QTL detected in chromosome 2 for HTF and HT34

showed a large effect with LOD scores of 117.11 and 128.75 respectively. The QTL showed an increasing effect for the Col parent, consistent with the results. The QTL mapping to chromosome 2 for HTF and HT34 was also detected at a similar position using the marker regression, and also the QTL mapping to chromosome 5 for HTF. The traits HTF and HT34 were positively correlated (r = 0.67).

4.3.2.4 Rosette leaves

The QTL for RL20 and RLF mapped to similar regions in chromosome 2, 4 and 5, and showed similar mode of action. The correlation between the traits was 0.602. The Col allele increased rosette leaves in the 3 QTL. The 2-linked QTL in chromosome 1 for RLF showed a decreasing effect for the Col parent. QTL for rosette leaves mapping to chromosomes 2, 4 and 5 were also detected using the marker regression approach.

4.3.2.5 Cauline leaves

The QTL detected in chromosome 1 for CL20 and CLF mapped to similar positions, but showed different modes of action. The QTL detected in chromosomes 2, 3 and 4 for CL20 were not detected for CLF. The correlation between the cauline leaf traits though positive and significant was very low (0.10).

4.3.3 Comparison between QTL detected using interval mapping and marker regression approach

A total of 17 QTL out of 23 QTL detected for six traits in the F_2 population using the marker regression approach were also detected by the interval mapping method in the F_3 families at approximately the same positions (Table 4.5). The interval mapping detected 23 more QTL

than the marker regression method (Fig. 4.1a-b). In the F_2 population QTL were not detected for traits TTG, TTL, HT20, CL20 and CLF, whereas in the F_3 families QTL were detected for all the 11 traits. The same QTL detected using the interval mapping and marker regression methods showed very close and similar additive mode of action except for the QTL detected in chromosome 3 for TTB. The slight differences between the F_2 and F_3 additive effects may be accounted by environmental, epistasis or genotype by environmental differences. However, the replicated F_3 families provided a better estimate of the genotypic effects.

Figure 4.2 shows the LOD scores for TTB, TTF and HTF for the QTL detected in chromosome 4 at approximately 30cM using the interval mapping method. Figure 4.3 shows the QTL detected by the interval mapping method and the marker regression approach at around 52cM. In the interval mapping method the significant QTL is determined by the peak LOD value above a threshold of 2, whereas in the marker regression approach the significant QTL is based on the additive or dominance effect. The similarity between the results for marker regression and interval mapping may suggest that even though the two methods use different procedures and significant tests more or less the same results are obtained.

4.3.4 Genotype x environmental interactions

The environmental variation between the non-segregating generations in the first experiment (F_2 generation) and second experiment (F_3 families) was estimated using the variance and means (Tables 4.6 to 4.9). In the first experiment, significant environmental variation was observed in 10 traits, whereas, in the second experiment there was variation in all the traits. This shows that the parents and the F_1 responded differently to the same micro-environment

in the two experiments and that the segregating generations are heterogenous, and there is interaction between the micro-environment and the genotype. The pooled variance between the two experiments showed significant environmental variation in 9 of the 11 traits, showing that there are differences between the segregating generations in the two experiments.

The analysis of genotype by macro-environmental variation was partitioned into genotype, environment and genotype x environment interaction estimated from the means of the segregating populations (Table 4.9 & Appendix III). The 2-way ANOVA showed highly significant differences between the genotypes in 9 traits and significant differences between the environments were observed in 7 traits. The presence of GxE interaction was observed only in one trait (HT20).

4.4 Discussion

Previous studies used single F_2 plants as the experimental unit. In the present study, F_3 families were used instead to provide a better estimate of the genotype effects. The continuous distribution of F_3 families for the morphological traits suggested that they are under polygenic control. Flowering time showed the highest heritability of 0.65. The heritability was close to other traits measured at flowering such as CLF (0.59), HTF (0.50) and TTB (0.63). In the F_2 population the broad sense heritability of traits measured at flowering ranged from 0.30-0.39. The heritability of flowering time in the F_3 families tend to be close to that observed for developmental traits (mean of 0.648) by Ungerer et al., 2002. The heritability of the means were higher than the heritability values and represent an improved relationship between the genotype and the phenotype of a family.

The number of QTL detected in the study varied from 3 to 11 per chromosome and they clustered in two or three regions in each chromosome. This clustering of the QTL reflects a similar underlying genetical basis of the morphological traits. This result is also confirmed by strong correlations between several flowering and morphological traits. It remains unclear, however, whether the correlations are due to pleiotropy or the presence of closely linked loci in the same genomic region.

The amount of phenotypic variation explained by the QTL varied from 5.76-164.56%. The QTL explaining the most variance mapped to the same region on chromosome 2 at around 50cM for traits RL20, CL20, HTF, RLF, HT34, TTB and TTF. This QTL was also detected in the F_2 population. This QTL is likely to be the *erecta* mutation, which affects

inflorescence architecture (Ungerer et al. 2002). The *erecta* mutation is not a naturally occurring mutation, but was generated in the laboratory through mutagenesis. Overall, the largest proportion of QTL detected were relatively small effect and large-effect QTL were relatively rare. This result is consistent with findings of other QTL studies documenting that most differences between lines are due to a small number of QTL of large effect accompanied by a large number of QTL of smaller effect (Tanksley, 1993, Lynch and Walsh, 1998). The apparent decline of QTL in the class of smallest effect should not be interpreted as evidence that small effect QTL are rare, but rather simply reflects the statistical difficulties of detecting these loci (Ungerer et al. 2002).

Several of the morphological traits studied in this paper such as number of rosette and cauline leaves, time to bud, time to flower and flowering time have already been investigated (Clarke et al. 1995; Lee et al. 1993; Kowalski et al. 1994b; Alonso-Blanco et al. 1998b). Results of marker interval analysis confirmed that flowering time is under polygenic control, with at least five putative loci involved. Two QTL mapped to chromosome 1 and one each on chromosomes 2, 4 and 5. The QTL positions for TTF mapped to similar positions and showed similar mode of action as the QTL for time to produce flower buds (TTB). This is expected as the time to produce buds marks the beginning of the flowering period. The other traits measured at flowering time mapping to the same regions as the QTL for time to flower are HTF, RLF and CLF. Also, the flowering time QTL were observed to be associated with other morphological traits (e.g. HT34 mapping to chromosome 2).

The Ler alleles at all the loci contributed towards early flowering time, except in chromosome 1. Thus, the direction of allele effects was not consistently in the direction of

the relative differences among parental lines. This was also observed for other morphological traits. This phenomenon is commonly observed in QTL studies (Tanksley 1993; Lynch and Walsh 1998) and the sorting of these alleles likely explains transgressive segregation in the progeny of genetically differentiated parental lines (Rieseberg et al. 1999).

The 95% confidence interval for QTL detected using the interval mapping ranged from 2-25cM with a mean of 9cM, whereas that associated with marker regression ranged from 4-40cM with a mean of 22cM. In the F_3 experiment, the experimental size was larger than the F_2 population. This agrees with van Ooijen (1992), Darvasi et al., (1993), Kearsey & Farquhar (1998) who observed that the confidence limits and the reliability of the QTL studies can be improved by increasing the family size and the number of families.

The QTL mapping methods have evolved from marker analysis (t-test, single or multiple regression) to one-QTL models (interval mapping and composite interval mapping), and further to the multiple-QTL models such as the multiple interval mapping (MIM). In practice, the detected QTL can be used for selecting parents with desired genotypes for producing progeny or gene transfer to achieve the ultimate goal of trait improvement in later generations. QTL need to be mapped as precisely as possible to ensure good quality of the follow-up operation on QTL. Therefore, precision and unbiasedness in estimating the parameters of QTL should be more important than the ease of computation and implementation in QTL mapping (Kao et al. 1999). The methods of interval mapping and marker regression approach follow the procedure of creating a QTL model for the observed data and then testing that model for its suitability. In both cases the models are relatively simple and consider only one or two QTL thus giving a limited number of possibilities to be

considered. However, the two methods yield the same results with regard to the traits showing significant genetic variation in the F_2 population and F_3 families, even though the methods employed differ in the statistical significance.

The interval mapping method and marker regression does not include the analysis of other parameters such as epistasis. When epistasis is included the range of possible models to be tested becomes much larger and the process of model selection and testing becomes extremely demanding (Doerge, 2002). Although the results of the present study are based on a single population in one environment, but we do know that the same QTL would have significant effects in other environments because they have been consistently detected in the F_2 and F_3 experiments. The advantages of the marker regression over the interval mapping method are its ability to test for the presence of more than one QTL and the method incorporates all marker information on a chromosome in a single test (Hyne & Kearsey, 1995).

The detection of uncontrolled variation in the segregating population within and between the two experiments showed that one can never hope to eliminate the micro- and macro-environmental variability nor the developmental errors entirely and experiments must be randomised to reduce the variance (Kearsey and Pooni, 1996). The presence of GxE creates problems in estimating components of variance, undermines the heritability of the means of experiments and consequently reduces the efficiency of the selection. Therefore, since the F_2 and F_3 generation will be subjected to MAS in this study, the efficiency of selection is likely to be low due to the presence of environmental variation.

Table 4.1: Means, variance components, narrow sense heritability of the F_3 families and hertability of the means of various morphological traits in the F_3 families

Trait	Generation means				Variance components			
	Col	Ler	F_1	F_3	V_A*	V_E	h_n^2	h_f^2
TTG	3.50	4.40	3.17	4.29	0.26	1.05	0.20	0.52
TTL	7.71	9.36	6.79	8.71	0.25	1.14	0.18	0.50
HT20	17.92	19.20	95.63	25.91	263.50	321.45	0.45	0.74
RL20	11.33	7.96	9.83	9.78	0.78	1.43	0.35	0.68
CL20	3.08	3.24	5.17	2.88	0.68	1.02	0.40	0.71
TTB	17.88	17.12	15.88	18.40	2.43	1.31	0.63	0.83
TTF	22.88	22.00	20.63	23.05	2.22	1.20	0.65	0.83
HTF	93.54	65.80	122.08	92.43	533.04	538.58	0.50	0.77
RLF	11.92	8.12	9.96	10.65	1.82	2.34	0.21	0.81
CLF	5.58	4.60	6.75	5.40	0.62	1.24	0.59	0.54
HT34	453.75	319.80	503.33	395.24	3140.60	4510.70	0.41	0.72

Table 4.2: Analysis of *Arabidopsis thaliana* F_3 families. (i) Skeleton ANOVA and (ii) parameter estimates

i.

Source	df	MS	ems
Between families	n-1	MS_B	$\sigma^2_w + r\sigma^2_B$
Within families	n(r-1)	MS_W	σ^2_w

ii.

$\sigma^2_B = (MS_B - MS_w)/r$

$\sigma^2_w = MS_w$

$V_A^* = \sigma^2_B$ (assuming dominance = 0)

$V_E = \sigma^2_w - 1/2\ V_A^*$

$h_n^2 = V_A^*/(V_A^* + V_E)$

r = number of replicates; V_A^* = additive variance for the trait; V_E = environmental variance

Table 4.3: Pearson's correlations between traits based on the F3 family means (only significant values are given)

	TTG	TTL	HT20	RL20	CL20	TTB	TTF	HTF	RLF	CLF
TTL	**0.827**									
HT20	-0.334	-0.432								
RL20	-0.363	-0.416	-0.098							
CL20	-0.444	-0.465	0.504							
TTB	0.542	0.592	-0.661		-0.693					
TTF	0.593	0.590	-0.641	-0.094	-0.700	**0.937**				
HTF	-0.064	-0.131	0.311	0.217		-0.098				
RLF			-0.363	**0.602**	-0.421	0.564	0.560	0.140		
CLF	-0.074	-0.062	-0.063	0.162	0.102	0.186	0.272	0.291	0.292	
HT34	-0.249	-0.317	0.316	0.356	0.096	-0.278	-0.294	**0.670**	0.091	

NB: bold values are those greater than 0.60.

Table 4.4: QTL detected using the interval mapping in *Arabidopsis thaliana* F_3 families

Trait	Chr.	QTL position	LOD score	Additive effect (a)	Dominance effect (d)	$\%V_A$	2 LOD support limit (cM)
TTG	5	115	8.65	0.15	-0.25	8.65	112-115
TTL	1	0	6.40	-0.17	0.14	11.56	0-7
	4	55	3.54	-0.13	-0.04	6.76	47-63
	5	115	6.88	0.12	-0.22	5.76	112-115
HT20	1	0	36.67	11.15	-5.31	47.18	0-3
	2	0	4.86	-5.24	-7.74	10.42	0-3
	4	27	27.01	-10.59	-1.93	42.56	24-29
	5	44	32.64	-12.60	-1.51	60.25	39-49
RL20	2	54	37.66	0.71	0.05	64.63	49-58
	4	37	21.33	0.71	0.14	64.63	35-39
	5	34	17.35	0.55	-0.02	38.78	30-42
CL20	1	0	18.59	0.44	-0.06	28.47	0-3
	2	46	22.79	-0.53	-0.06	41.31	42-50
	3	0	9.77	0.23	-0.34	7.78	0-7
	4	27	9.25	-0.29	0.01	12.37	21-32
	5	10	11.09	-0.39	0.28	22.37	7-12
	5	110	11.44	-0.32	0.53	15.06	108-115
TTB	1	0	26.85	-0.91	0.26	34.08	0-5
	1	34	12.26	-0.70	0.36	20.16	30-40
	2	40	8.57	0.61	-0.18	15.31	25-47
	3	48	10.06	-0.44	0.50	7.97	47-49
	4	27	28.16	0.91	-0.20	34.08	25-29
	5	49	29.15	0.96	-0.20	37.93	41-57
TTF	1	0	29.85	-0.92	0.14	38.13	0-5
	1	34	16.52	-0.92	0.14	38.13	31-39
	2	38	7.35	0.53	-0.01	12.65	20-45
	4	27	21.72	0.78	-0.17	27.41	23-29
	5	44	23.87	0.86	-0.36	33.32	39-60
HTF	2	50	117.11	27.49	10.77	141.77	49-51
	4	37	11.61	-16.14	9.19	48.87	27-47
	5	88	11.33	-14.66	11.77	40.32	78-94
RLF	1	5	14.53	-0.69	0.18	26.16	0-11
	1	34	9.50	-0.57	0.15	17.85	28-45
	2	50	34.33	0.94	0.02	48.55	47-54
	4	27	47.55	1.03	-0.15	58.29	25-29
	5	39	47.15	1.24	-0.28	84.48	37-44
CLF	1	10	7.85	-0.37	0.01	22.08	0-20
	5	20	18.05	0.51	-0.09	41.95	14-25
HT34	2	50	128.75	71.89	20.06	164.56	49-51
	3	87	9.26	-15.68	65.99	7.83	86-87

$\%V_A = V_A$ due to the QTL/V_A* for the trait, and $V_A = a^2$

Table 4.5: Comparison between QTL detected using interval mapping in F_3 and marker regression approach in F_2 generation

Trait	Chr.	Interval mapping				Marker regression		
		QTL position	LOD score	a	CI	QTL position	a	CI
RL20	2	54	37.66	0.71	9	56	0.79	19
	4	37	21.33	0.71	4	34	0.40	30
	5	39	17.32	0.57	12	38	0.52	17
TTB	1	34	12.26	-0.70	10	38	-1.29	
	2	40	8.57	0.61	22	48	0.63	21
	3	48	10.06	-0.44	2	60	1.17	
	4	27	28.16	0.91	4	8	0.65	40
	5	49	29.15	0.96	16	42	1.21	15
TTF	1	34	16.52	-0.92	8	34	-1.13	
	2	38	7.35	0.53	15	50	0.55	23
	5	44	23.87	0.86	21	42	0.83	20
HTF	2	50	117.11	27.49	2	52	23.43	4
	5	88	11.33	-14.66	16	74	-7.40	33
RLF	2	50	34.32	0.94	7	54	0.88	8
	4	27	47.55	1.03	4	26	0.49	20
	5	39	47.15	1.24	7	40	0.84	10
HT34	2	50	128.75	71.89	2	52	39.07	5

Chr. = chromosome

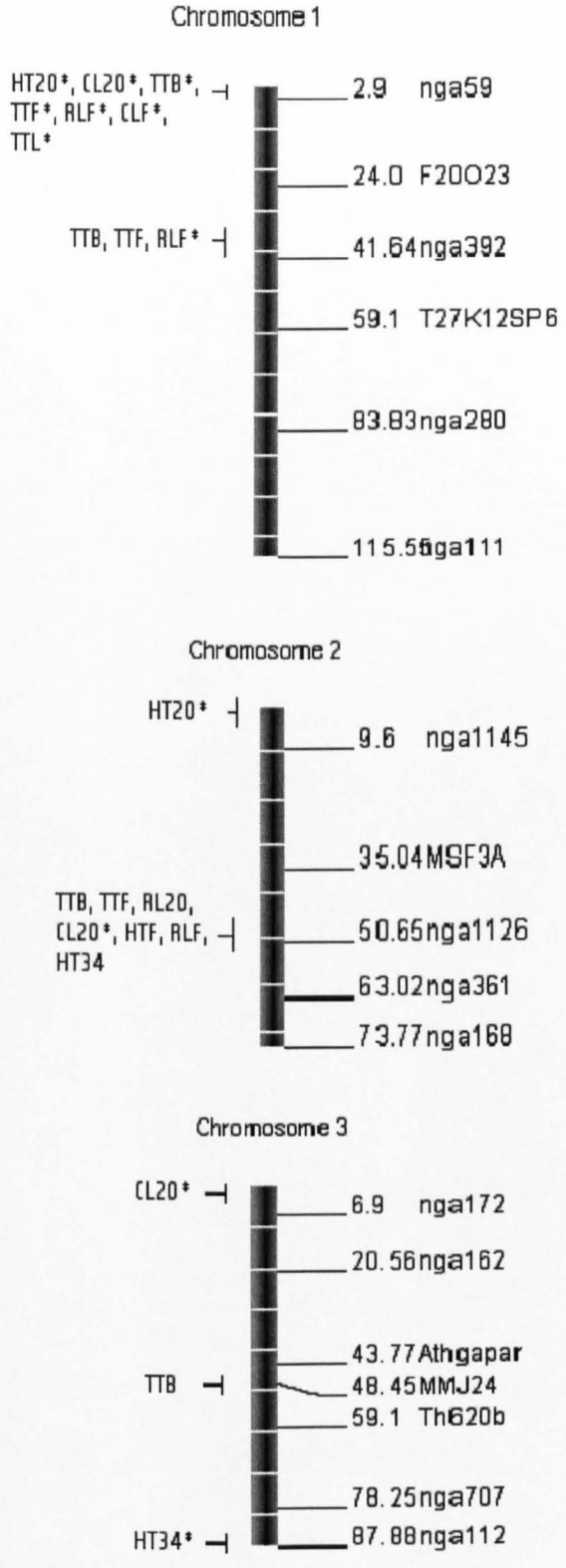


Figure. 4.1a: QTL location in Arabidopsis chromosomes 1, 2 and 3 (‡ indicates new QTL detected using the interval mapping method)

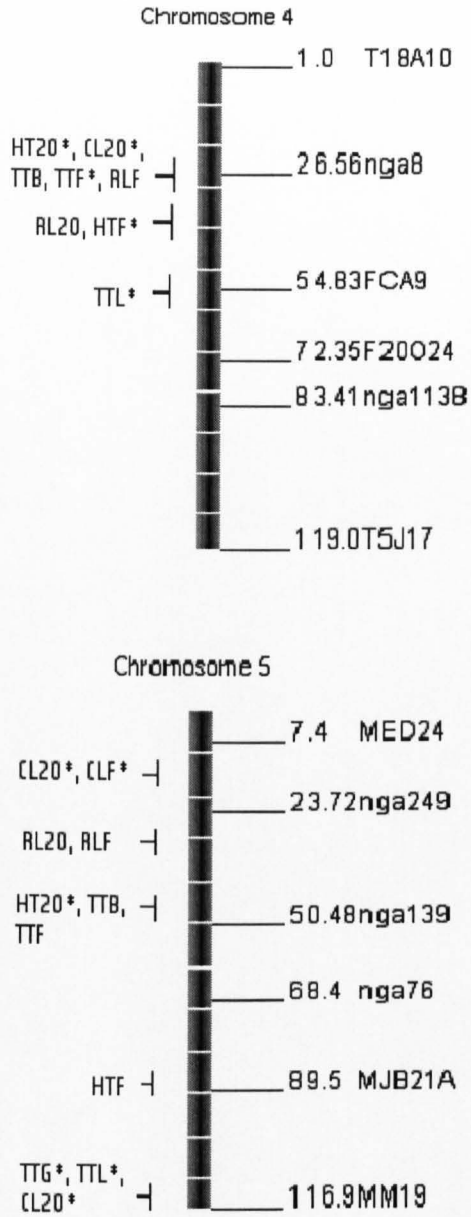


Figure. 4.1b: QTL location in Arabidopsis chromosomes 4 and 5 (‡ indicates new QTL detected using the interval mapping method)

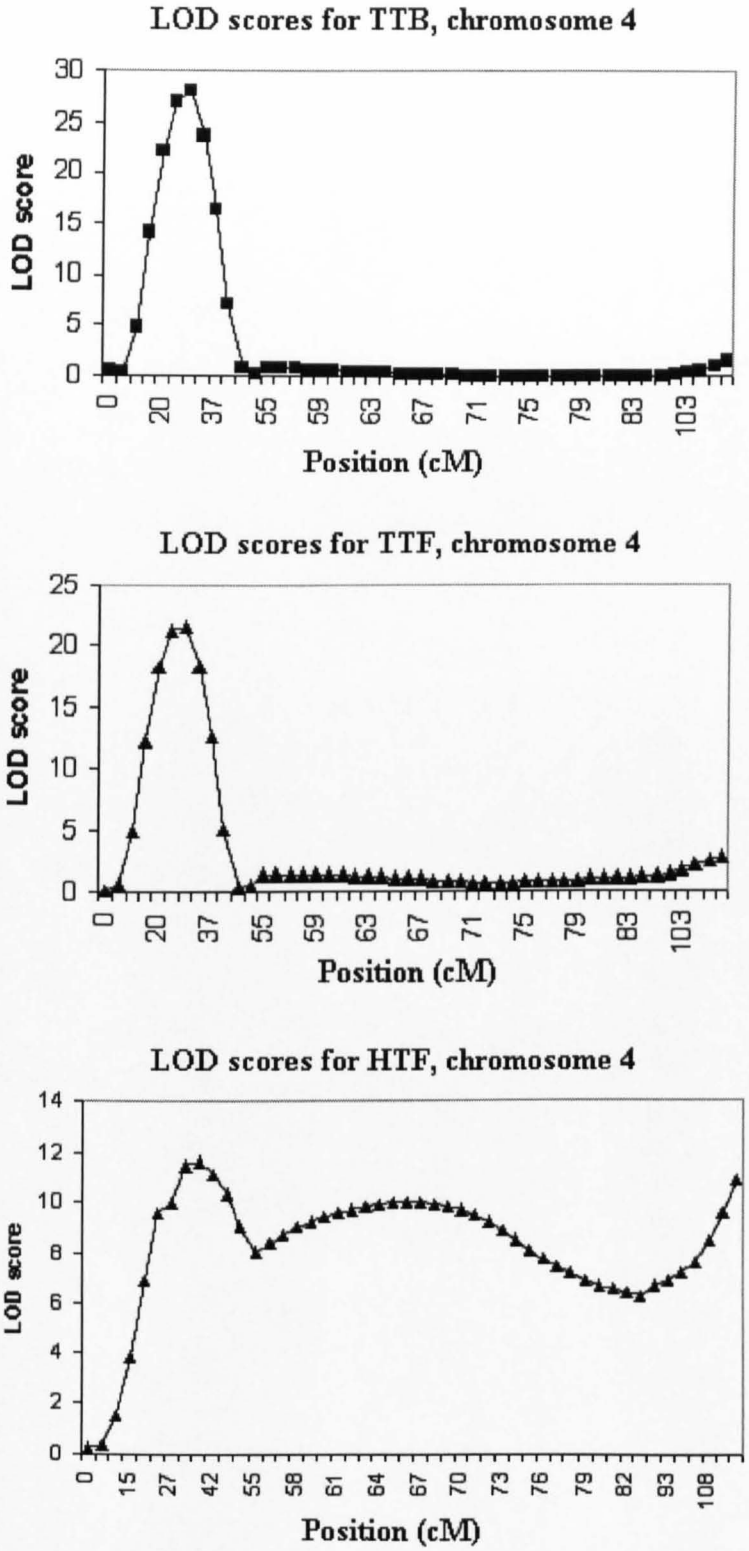
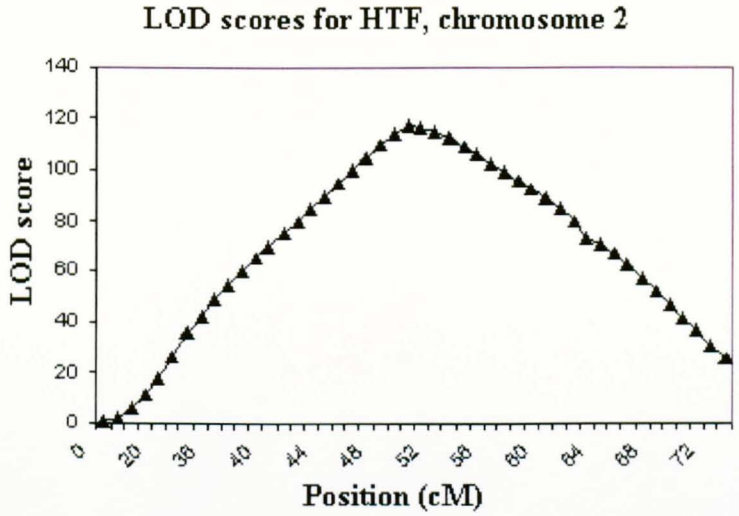


Fig 4.2: Peak LOD scores in chromosome 4 for QTL detected using interval mapping at approximately 30cM



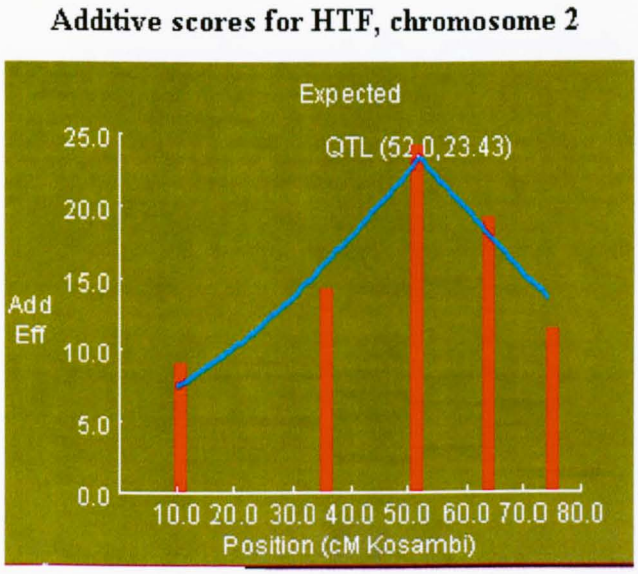


Figure 4.3: QTL for HTF detected at the same positions using interval mapping in F_3 families (top) and marker regression approach in F_2 plants at around 52cM

Table 4.6: Col, Ler, F_1 variances and the variance ratio in the first experiment (F_2 population)

Trait	Col (23df)	Ler (23df)	F_1 (47df)	Variance ratio, F
TTG	2.98	1.46	2.24	2.04*
TTL	2.30	2.32	5.39	2.34*
RL20	2.78	0.60	1.14	4.63***
CL20	1.30	1.77	1.92	1.48ns
HT20	36.97	32.60	1164.78	35.71***
TTB	8.85	1.72	8.87	5.16***
HT34	12144.04	1534.29	3524.80	7.92***
RLF	3.19	1.42	1.05	3.04***
CLF	0.35	1.45	0.66	4.14***
HTF	353.06	403.21	958.92	2.72***
TTF	10.08	1.83	6.51	5.51***

Table 4.7: Col, Ler, F_1 variances and the variance ratio in second experiment (F_3 family)

Trait	Col (23df)	Ler (24df)	F_1 (23df)	Variance ratio, F
TTG	1.39	0.42	0.14	9.93***
TTL	1.87	0.57	0.52	3.60***
RL20	2.06	0.37	0.93	5.57***
CL20	0.95	0.19	5.71	30.05***
HT20	106.30	34.69	1531.16	44.14***
TTB	2.98	0.53	0.81	5.62***
HT34	2937.64	118.16	1349.09	24.86***
RLF	1.73	0.44	0.82	3.11***
CLF	1.65	2.42	3.59	2.18*
HTF	194.60	882.09	464.83	4.53***
TTF	2.64	0.33	0.51	8.00***

Table 4.8: Pooled variance and the variance ratio between the two experiments (F_2 and F_3 generations)

Trait	Pooled variance (experiment 1, 93df)	Pooled variance (experiment 2, 70df)	Variance ratio, F
TTG	0.23	0.08	3.00***
TTL	0.31	0.13	2.38***
RL20	0.17	0.15	1.13ns
CL20	0.17	0.30	1.76*
HT20	27.54	72.64	2.64***
TTB	0.65	0.19	3.42***
HT34	669.71	191.30	3.50***
RLF	0.23	0.13	1.77*
CLF	0.09	0.42	4.67***
HTF	53.28	118.70	2.23***
TTF	0.66	0.81	1.23ns

Table 4.9: The variance ratio between genotypes, environments and genotype x environmental interaction (GxE)

Trait	Genotype	Environment	GxE
TTG	1.53ns	7.56***	0.14ns
TTL	4.44*	2.95ns	0.48ns
RL20	14.32***	1.27ns	0.70ns
CL20	3.64*	15.78***	1.05ns
HT20	20.34***	7.15***	3.17*
TTB	3.81*	15.12***	0.35ns
HT34	8.28***	18.75***	2.93ns
RLF	13.70***	4.62*	0.71ns
CLF	1.14ns	19.28***	1.70ns
HTF	9.57***	1.41ns	0.31ns
TTF	3.19*	3.58ns	0.17ns

CHAPTER 5

Marker-assisted selection in *Brassica oleracea* doubled haploid lines

Abstract

Selection based on the phenotype and using markers was carried out in *Brassica oleracea* DH lines of a cross between a rapid cycling Chinese kale (A12DHd) and a calabrese (GDDH33). In traits with single QTL, marker closest to the QTL did not seem to give better results than selection on flanking markers or the next marker nearer the QTL. However, the efficiency of selection was low when there were more missing data in a marker position. The efficiency of selection based on the phenotypic value (means) and top ranks provided a better method to measure the efficiency achieved by MAS, since the efficiency was high or maximum when most or all the top ranks were selected. The selection for individual QTL was better than selection for two or more QTL simultaneously on the same or different chromosomes in a trait. The efficiency of selection based on the marker loci varied with the additive variance and the heritability of the means.

5.1 Introduction

The development of highly polymorphic markers (e.g. SSRs, RFLPs, AFLPs, RADPs, etc) has opened a new era for geneticists and plant breeders. Molecular markers enable one to identify and map quantitative trait loci (QTL) that control variation in quantitative characters. Nowadays, markers are used to help in identifying particular genotypes and this phenomenon is called marker-assisted selection, or MAS. The steps in MAS consist of identifying association between marker alleles and QTL, or ideally of estimating the contribution of marker loci to the phenotypic value of the trait, and then combining these marker effects with the phenotypic information through an index and development of desired lines or populations (Xie and Xu, 1998; Lande and Thompson, 1990; Dudley, 1993).

The traditional approach to agricultural improvement was to simply select for desirable phenotypes either by eye or via field evaluation/assessment. However, this had the disadvantage of incorporating genotype by environmental interactions, and sometimes the effect of the environment would be so extreme, that it would mask the expression of the phenotype. Further, most characters of economic importance are quantitative traits and are influenced by numerous QTL that often have small effects individually. These characters are very difficult to improve because they are highly influenced by the environment.

Molecular markers are now being used to significantly reduce the amount of linkage drag, so that the breeder can select those lines that show recombination in the area of interest, without incorporating alleles from other genes that give undesirable effects (Tanksley et at. 1989; Kearsey and Pooni, 1996). The goals of quantitative trait mapping experiments are to find

sources of favorable alleles for developing superior cultivars and hybrids and to gain the knowledge necessary for maximizing gains through marker-assisted selection. What QTL mapping adds to the marker-assisted selection is the knowledge about the distribution of favorable alleles between the parental and donor inbred lines and their progeny, and estimation of gene effects without the restriction assumptions about their genetics and marker loci linked to the genes to be selected (Knapp, 2001). Once the alleles are found they can be introgressed from donor to elite inbred lines.

Another pertinent use of marker information would be to improve the efficiency of the traditional inbreeding procedures by integrating marker-assisted selection into such breeding programs. For example, markers may be used to increase the frequency of desirable alleles for a particular locus while conventional selection may be applied to select for genes with small effects. The gains obtained by implementing such strategy, however, will depend to a large extent on the number of markers used in the selection and on the linkage of such markers with the chromosome segments of interest. As little is known about the efficiency of such a strategy in practical breeding, the present study will ask some pertinent questions and strive to obtain answers for them. The major questions that we shall seek answers are listed below:

i) How many markers to use in the selection process?

ii) Which markers give maximum efficiency?

iii) Does heritability have an effect on the result?

iv) How do the results compare with phenotypic selection?

Although, many of these questions can be answered theoretically and using computer simulations, but we shall attempt to use real life data from the experiments already presented in this thesis. The first population to which we shall apply such selection is the *Brassica oleracea* DH lines of a cross between *Brassica oleracea var. italica* (A12DHd) and *Brassica oleracea var. alboglabra* (GDDH33) for which QTL have already been identified in the previous section. These lines also allow us to obtain comparisons between direct and indirect methods of selection because we shall be applying selection retrospectively, in this case.

5.2 Materials and methods

5.2.1 Brassica DH lines

The material consisted of 89 *Brassica oleracea* doubled haploid (DH) lines, derived from the F_1 plants of a cross between two doubled haploid parents, a rapid cycling Chinese kale, *Brassica oleracea var. italica* (A12DHd) and a calabrese, *Brassica oleracea var. alboglabra* (GDDH33), through microspore culture of the F_1 (Bohoun, et al. 1998). This population was used for QTL analysis in the previous section (Chapter 2), so the design and the data collection are outlined in that section.

The 89 lines were firstly subjected to the ANOVA in the QTL café (http:\\web.bham.ac.uk/g.g.seaton) to identify the markers significantly associated with the QTL. The marker regression method of Kearsey and Hyne (1994) revealed 40 QTL in 15 traits that involved leaf, height and flowering time measurements. The QTL served as the basis of marker-assisted selection applied in this section.

5.2.2 Application of selection

5.2.2.1 Phenotypic selection

Phenotypic selection is the most common method that is used to select desired phenotypes. It is also called truncated selection as the top scoring individuals are selected and propagated in the next generation and others are rejected. This type of selection is usually applied direct to the trait of interest, for example, selection for early maturity will involve the selection of early flowering individuals.

The most frequent use of direct selection is in the case of populations, such as cattle herds and sheep flocks. Consequently, the method and the predictions are also better elaborated for populations, particularly those that have a high degree of random mating. The selection can be applied to individuals or families or combined over individuals/families and the basic procedure takes the following form.

Suppose there is a population of individuals that is normally distributed for the trait under selection and that the trait has a mean μ and variance σ_p^2. It is also assumed that we are selecting top p ($0<p<1$) of the population and using the selected individuals to generate the next generation. The mean of the selected proportion and that of their descendants is predicted following Falconer and Mackay (1996).

In such cases, the mean of the selected individuals is $\mu + S$, where S stands for 'selection differential' and,

$$S = i\sigma_p.$$

Here i is the intensity of selection (= height/p) and σ_p is the phenotypic standard deviation of the trait.

When selection is applied to individuals, S is a phenotypic value and the genetic advance (R) is then calculated as:

$$R = h_n^2 S$$

Or $R = h_n^2 i\sigma_p$, where h_n^2 is the narrow sense heritability of the trait.

The procedure is better illustrated diagrammatically in Fig. 5.1.

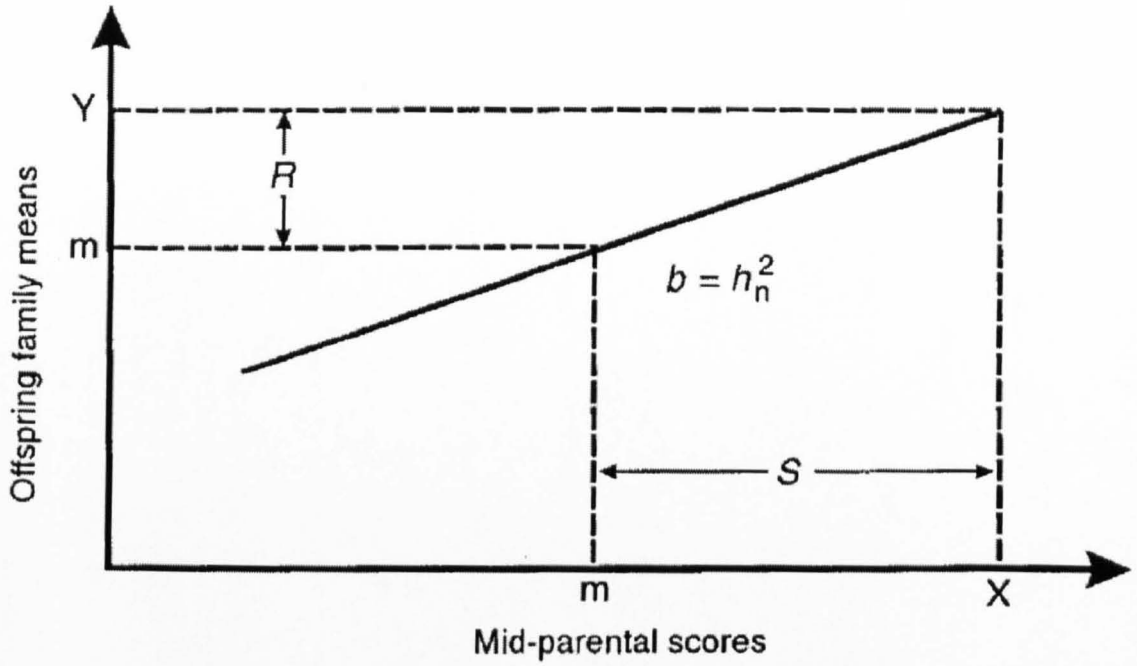


Figure 5.1: Predicting the response to selection (Illustrate the concept of selection differential (S) and the response to selection (R)), extracted from Kearsey and Pooni, 1996, p316.

Selection in our case will essentially follow the same procedure as above, except that it will be applied to DH means. Selection on means differs from that on the individuals only in its R-value because heritability of means is usually much higher than that of the individuals. In fact, h_n^2 is replaced by h_f^2 (heritability of the means) and now:

$R = h_f^2 S$

Where $h_f^2 = (\sigma_L^2/(\sigma_L^2 + \sigma_{LB}^2/r + \sigma_w^2/br)$, σ_L^2 = between DH family component, σ_w^2 = within DH family component, σ_{LB}^2 = interaction variance, b = number of blocks and r = family size.

While the above procedure will give the expected genetic advance if we applied direct selection to the 89 DH lines, but in practice we can select retrospectively say 5 highest and 5 lowest scoring DH lines for each trait as well. This type of selection will normally be followed in the present case because only random samples of DH lines can be extracted from a cross and therefore selection will be applied later. This situation is also akin to assuming that in a conventional breeding program selection has been delayed till inbreds are produced and the desirable line or lines were selected at the end.

The relevant statistics obtained from these selections, i.e., their means, variances and range of the selected lines have been presented for various traits in Tables 5.4 (highest scores) and 5.5 (lowest scores). Tables 5.1, 5.2 and 5.3 show the predicted values of R and S together with other relevant data, e.g., population mean, heritability of the mean, phenotypic standard deviation and the observed response. The value of p used for predictions was p =5/89 = 0.056.

5.2.2.2 Single marker selection

It is a matter of observation that usually one or two QTL are detected for many traits irrespective of the level of variation present in the population. Thus, marker-assisted selection will involve one (for one QTL), or at the most 2 or 3 marker loci that are located close to the QTL in question. As selection for one QTL is procedurally quite different from when it is applied to 2 or more QTL, we initially consider those traits where only one QTL was detected. These traits are leaf length1 (LL1), petiole length1 (PL1) and fresh weight (FW).

The selection in these traits will consider the following possibilities:

(1) Use a marker that lies close to the QTL and therefore is expected to show strong association with it.

(2) Use flanking markers that will ensure that the QTL is carried with them, unless there is a double cross over within the short segment flanked by the markers.

(3) It is also possible that data are either not complete or available for the closest marker, but instead are available for the next marker nearer to the QTL, or (4) second marker that lies on the other side of the QTL.

Similarly, it would be of academic interest to see what sort of results are obtained for the situations where:

(5) selection is applied to a marker that is located close to the centromere

(6) a marker located far from the QTL (i.e. at the farthest terminal end).

In practice, marker selection is applied to put individuals or lines into groups that would either be retained for breeding or rejected. Phenotypic selection is then applied to the selected group as all individuals or families in this group will not show the same performance, unless the trait is controlled by just one QTL with large effect (that is, it is a qualitative trait). Further, to compare the efficiency of MAS with direct selection, we apply marker-assisted selection to all 89 DH lines and then identify the best 5 lines present among the selected families. Tables 5.7-5.9 present the results of such a selection for traits with single QTL and for the markers listed above under (1) to (6). In the selection process, the selection for the highest and lowest scores was dependent on the additive effect of the QTL, with a positive additive effect showing an increasing effect for the A12, and a negative effect showing an increasing effect for the GD parent.

5.2.2.3 Observed efficiency of MAS

The efficiency of MAS was measured using the following criteria:

(a) The number of the most extreme lines among the 5 selected by markers. In this case the score will vary from 0 and 5, the last when all top-ranking lines based on phenotype are selected.

(b) The proportion of the ranked lines among the marker group. This method is a little complex. Initially we determine how many DH lines possess a particular marker. That is, we count the number of lines in each marker group. Suppose there are 20 such lines in a group. Then we take the equivalent number of the top ranking lines (=20) for the trait and count how many of these lines are present in our marker group. We again suppose there are 16 such lines, and we calculate the efficiency as 16/20 = 0.8.

(c) The proportion of phenotypic advance (S) achieved by using MAS. In this case:

Efficiency = {(Mean of 5 top scoring lines selected under MAS – μ)/OS}

Where μ = population mean

OS = observed response

5.2.2.4 Selection for two QTLs

Out of the 15 traits considered in DH lines, the following had two or more QTL:

Trait	QTL1		QTL2		QTL3		QTL4	
	Chr.	Pos.	Chr.	Pos.	Chr.	Pos.	Chr.	Pos.
PL2	1	0	1	36				
AH	1	0	1	42				
NL	7	0	8	18				
LL1	8	20	9	92				
LW1	8	18	9	94				
FT	2	26	3	2	3	138		
PH2	6	56	8	48	9	94		
LL2	1	40	3	24	3	80		
FH	1	0	2	12	3	0	9	54
MH	1	0	1	40	2	60	9	60
SW	6	10	6	50	9	36	9	90
PH1	6 QTL (on chromosomes 1, 3, 5, 6(2) and 7)							

Selection for two or more QTL presents several practical problems, particularly when it is applied retrospectively. The major problem is the limited sample size. With only 89 DH lines available and marker data being incomplete for some markers, it is possible that simultaneous selection for two or more markers may identify only one or a few lines with the desired genotype. This situation will not allow an efficient comparison between the methods. Second, two QTL can be located on different chromosomes and this will make the above problem less severe. But, when the QTL are on the same chromosome and linked, the

desired marker genotypes may be even more rare than otherwise expected. In this situation only homozygous loci are selected at the markers with selection for the highest and lowest scores dependent on the additive effect of the QTL.

It is also possible that we may have data on only one marker and thus we are forced to select for one QTL while we know there are actually two QTL controlling the trait. MAS for two QTL is therefore applied as follows, using one marker per QTL that is closest to the QTL:

(1) Select for QTL1 only (single marker selection)

(2) Select for QTL2 only (single marker selection)

(3) Select for both QTL.

Based on the lines selected under these conditions, 5 best lines were selected and the results are summarized in Tables 5.10-5.14 for various traits.

5.2.2.5 Selection for more than two QTL

The selection for traits with more than two QTL was based on one marker per QTL (marker location closest to the QTL), then selection for all the QTL simultaneously. QTL were arranged according to their magnitude, such that QTL1 had the largest effect and QTL3 the smallest when the trait had 3 QTL. The results are presented in Tables 5.15-5.17.

5.2.2.6 Selection index, relative efficiency and the efficiency of selection based on the marker loci

Lande and Thompson (1990) have shown that statistics like selection index (I), relative efficiency (RI) and the efficiency of selection (ES) based on the marker loci can be easily determined using the following formulas.

$I = (1/ h_n^2 - 1)(1 - A)$

$RI = (A/ h_n^2 + (A - 1)^2)/\sqrt{(1 - h_n^2 A)}$

$ES = \sqrt{(A/ h_n^2)}$

where, A = the percentage additive variance of the QTL

These calculations are presented for individual QTL detected in DH lines in Tables 5.18. The heritability of the mean (h_f^2) was used in the formulas because the genetic effects were estimated using family means.

5.3 Results

5.3.1 Phenotypic selection

Tables 5.1 and 5.2 show the population mean, the phenotypic standard deviation, the heritability of the mean, selection differential, mean of the 5 selected lines, the genetic advance (R) and the observed response (OS) for the highest and lowest scores, respectively. These results are indeed very interesting. For instance, the heritability of the means varies from 0.35 to 0.93 indicating that various traits differ in their inheritance patterns and the level of genetic variation displayed by the DH families. The traits with the lowest heritability of the means are LW1, SW and FW, and the highest is observed for FT and LW2. Plant heights (PH1, PH2 and FH) also have a high heritability value (0.85-0.86) while LL, MH, PL and NL etc. show a moderate level of heritability.

The phenotypic variation further shows some considerable differences between traits. As a proportion of μ, its value differs from 8% for FT to 36% for PL1. However, there is no visible relationship between the standardized values of phenotypic variation and the heritability values and the correlation between them is not significant (r= -0.31ns). The observed values of OS also seem to differ between the high and low selections. OS takes a larger value for high selections for 10 traits than low selections and the opposite is true for the remaining 5 traits. A chi-squared test, however, reveals that this difference is not significant ($\chi^2_{(1)}$= 1.67ns). Comparison of calculated S with the OS, on the other hand, shows that the observed S is larger than the calculated S on 6 and 12 occasions for the high and low selections respectively. Collectively, the large: small (18:12) ratio does not differ significantly from 1:1 ($\chi^2_{(1)}$=1.2ns) while separately the same (12:3) ratio shows a significant

difference between calculated S and observed S for low selections ($\chi^2_{(1)}$= 5.4*). Calculated S vs. observed S for high selections (6:9), on the other hand, shows a non-significant chi-squared ($\chi^2_{(1)}$ = 0.60ns).

The calculated S values themselves differ considerably between traits, largely due to differences in the scale of measurement. For example, S has the highest and lowest values of 342.31 and 1.44 for AH and NL respectively. When presented as a proportion of the overall mean μ, the ranking changes considerably, now the lowest advance is observed for FT (16%) and the highest for PL1 (73%). Clearly, response to selection differs with traits as well as with its likely direction. The realized response R, however, will be very much different from S as there are large differences between the heritability of the mean values. The highest value of R (% of μ) is observed for PH2 (48%) and the lowest for NL (11%).

Table 5.3 gives the mean observed selection response OS and difference in its value between the high and the low selections for the various traits. It is interesting to note that the observed response varies considerably between traits. As a proportion of μ, its magnitude is as large as 67% (PL1), which is phenomenal. However, most of the values are more realistic, e.g. in the region of 20-40% while the minimum gain that can be obtained is 17% for FT. Differences between the high and low selections are also sometimes large. For example, this difference is 44%, 27% and 22% of μ for PL1, PH2 and LL1 respectively. In fact, there is a strong association between the averaged S values and these differences and their correlation is positive and significant (r=0.87**). It is also apparent that whenever the difference was negative its magnitude was rather small (1-3%).

The mean, variance and range for the highest and lowest selections based on the phenotypic score are given in Tables 5.4 and 5.5. These results are rather complex to comprehend. Apparently it seems as if different lines are picked up for various traits and there is little commonality between traits. But a closer look reveals a different picture. Altogether 39 lines are chosen as high selections and the number of low selections is 38. Further, each line appears nearly twice in a group and there are 14 lines that appear in both high and low selections (common to both tables). Therefore, virtually 71% (63 out of 89) of the lines possess some good trait or traits. An elaborated summary of these results is given below.

(i) High selections: single traits

Line	11	23	25	28	31	35	36	44	45	51
Trait	FT	FT	FW	FH	PH1	LW2	NL	FH	FT	FH

Line	54	56	58	63	65	75	81	90	91	92
Trait	LW1	PL2	SW	LW1	PL2	LW2	FW	SW	PH1	LL2

(ii) High selections: two traits

Line	7	9	13	20	33	49	53	76	77	95
Trait-1	NL	FH	LL2	PH2	NL	LL2	PL2	PH1	LW2	LW2
Trait-2	PL1	FT	AH	MH	PH2	PL2	SW	PH2	FT	FW

(iii) High selections: 3 or more traits

Line	Traits
12	PH1, NL, LW1
17	LL1, LW1, PL1
18	LL1, LW1, SW
57	PH2, MH, AH
93	LL2, LW2, MH
4	LL1, PL1, PH2, AH
42	PH1, NL, LL1, PL1
38	LL2, PL2, MH, AH, SW, FW
47	LL1, PL1, FH, MH, AH, FW

(iv) Low selections: one and two traits

Line	2	5	12	23	24	40	68	69	74	76	82	87	88	89	94
Trait	FW	FT	LL2	NL	PH2	PH2	FH	PL2	MH	FT	AH	SW	LL1	NL	AH

Line	1	4	6	7	9	11	26	31	37	49	66	75	77	86	95
Trait-1	PH1	FH	LW2	LW2	LL1	LL1	PL2	LW1	MH	NL	NL	PH2	NL	LL2	LL1
Trait-2	PL2	FT	SW	MH	LW1	PL1	AH	PL1	FW	FH	PL1	MH	PH2	LW2	LW1

(v) Low selections: three or more traits

Line	Traits
16	LL2, LW2, SW
34	LL1, LW1, PL1
45	LL2, PL2, FW
57	PH1, LW1, SW
67	PL1, LL2, PL2
29	PH1, LW2, FT, SW
85	PH1, FH, FT, AH, FW
21	PH1, PH2, FH, MH, AH, FW

(vi) Common lines in the high and low selections

Line	Highest scores	lowest scores
4	LL1, PL1, PH2, AH	FH, FT
7	NL, PL1	LW2, MH
9	FH, FT	LL1, LW1
11	FT	LL1, PL1
12	PH1, NL, LW1	LL2
23	FT	NL
31	PH1	LW1, PL1
45	FT	LL2, PL2, FW
49	LL2, PL2	NL, FH
57	PH2, MH, AH	PH1, LW1, SW
75	LW2	PH2, MH
76	PH1, PH2	FT
77	LW2, FT	NL, PH2
95	LW2, FW	LL1, LW1

These results show that some lines are superior for one (28) trait and more possess desirable genes for 2 or more characters (10, 13, 8, 1 and 3 for 2, 3, 4, 5 and 6 traits respectively). But there are only a handful of lines that show promising performance for several traits simultaneously, for example, numbers 38 and 47 for high score, 85 and 21 for low score and 4 and 57 for the combination of the two types. All of these lines show the best performance for 5 or more traits.

These results allow a test of independent segregation of various traits in the form of a Chi-squared where lines can be assumed to have a Poisson distribution with a mean of 1.52. The expected numbers can then be calculated by the formula $e^{-\mu}(\mu^r/r!)$ where $\mu=1.52$, r is the number of traits for which the line has a higher or lower score (r varies from 0 to 15), e = 2.71828 and ! = factorial. Application of the above formula gives the following results:

Traits	0	1	2	3	≥4	Total
Observed	26	28	10	13	12	89
Expected	19.46	29.59	22.49	11.39	6.07	89
$(O-E)^2/E$	2.20	0.09	6.94	0.23	5.79	15.25

As an example, the expected value for r=3 was calculated as:

$89 \times [\{1.52^3 \div (3x2x1)\} \div (2.71828)^{1.52}] = 11.39$

This gives a $\chi^2_{(4)} = 15.25$ **, suggesting that at least one pair of traits is not independent of each other. Further, the observed values are less than the expected for r = 2 while the opposite is true for r = 0, 3 and 4 respectively.

Another feature of the results in Tables 5.4 and 5.5 is that the variance and spread of the high selections are larger than those of low selections for eleven traits. There is also a complete association between range and variance (this is expected), that is, whenever one is bigger the other is also bigger and vice versa. The variance ratio shows that the variances differ significantly for PH1, PL1, PH2, LL2 and FW. However, the ratio of 11 larger to 4 smaller variances does not deviate significantly from 1:1 ($\chi^2_{(1)} = 3.27^{ns}$). The large difference between the variance of the low and high selections maybe due to the skewness observed in traits PH1, LL1, LW1, PL1, PH2, PL2, AH and FW (Table 5.6).

5.3.2 Marker-assisted selection

The results of MAS for traits controlled by a single QTL are shown in Tables 5.7-5.9. In each case selection was based on: (1) the marker nearest to the QTL, (2) flanking markers, (3) next marker close to the QTL, (4) second marker that lies on the other side of the QTL, (5) marker close to the centromere and (6) marker farthest from the QTL on the same chromosome, at the opposite terminal end.

One consistent feature of these results in all the tables is that the number of lines in various marker groups is highly variable. In theory, this number is expected to be 44.5 (=89/2) because the lines are homozygous doubled haploid and the two bands for each marker have an equal chance of being present or absent. While in practice the observed number can be higher or lower than 44.5 due to sampling error but the numbers in the 2 categories (high and low) must add up to 89 for each marker. The total number considered in most of cases, on the other hand, was always <89. This was mainly due to missing data as some lines could not be scored at various marker positions.

The selection efficiency also seems to vary between the markers. Considering top ranks, selection on the marker closest to the QTL gives better results frequently. For example, at least 4 top scoring lines could be selected using this marker for LW2 and PL1 (Tables 5.7 and 5.8). Selection on flanking markers seems to be rather ineffective as only one or two of the best lines are selected with this procedure. Further, there are more missing marker data for the best lines in FW and PL1 compared to LW2. The total number of best lines (27, 32 and 26 for LW2, PL1 and FW) identified by the various marker selection procedures are also very comparable between traits.

Efficiencies of the other two methods (last 2 columns in each table) also follow closely that of the marker based selection. The number of best lines in a marker group is highly correlated with the number of lines in each group. For example, the highest (0.74) and the lowest (0.23) values of efficiency are observed for LW2 where the numbers of double haploid considered are 62 and 13 respectively (Table 5.7). The corresponding values for PL1 and FW are 0.79:0.36::62:11 (Table 5.8) and 0.73:0.31::56:22 (Table 5.9) respectively.

The highest values are observed for the efficiency on the means. These values are ≥0.90 on 14 out of 36 cases and these numbers are equally distributed between traits (LW2:PL1:FW, 4:6:4) as well as between high and low selections (High:Low, 8:6). The lowest value of the efficiency on means is observed for the flanking markers (0.58) for low selections in LW2 and only two other values are below 0.70 (0.64 and 0.68 for high selections in PL1, Table 5.8). Two of these 3 values are also associated with the small number of lines that form the marker group (13 for LW2 and 11 for PL1).

Perusal of the best lines that had marker data missing or did not get selected by various marker selection schemes (Tables 5.7-5.9) does not provide a clear indication as to what may be the cause of loss of efficiency in many cases. And this can only be pinpointed by a closer look at the marker profiles of the 5 high and 5 low selections given in Tables 5.4 and 5.5. These profiles are shown below for LW2, PL1 and FW where scores of 1 and 3 represent the marker bands for A12 and GD parents respectively.

Marker data from chromosome 8 for LW2 (high selections)

Line	M1	M2	M3(4)	M4(1)	M5(3/5)	M6	M7	M8	M9	M10(6)
93	3	3	-	3	3	-	3	-	3	3
75	-	-	1	1	3	3	-	3	3	3
77	1	3	3	3	3	3	3	3	3	3
35	1	3	-	3	3	-	3	-	3	3
95	1	3	-	3	3	3	3	3	3	3

NB: The upper case numbers in brackets represent the marker positions used in selection, with flanking markers involving markers 3 and 4. Lines in their descending order of score for high selections

Marker data from chromosome 8 for LW2 (low selections)

Line	M1	M2	M3(4)	M4(1)	M5(3/5)	M6	M7	M8	M9	M10(6)
7	-	1	1	-	3	-	3	-	3	3
86	-	3	-	3	3	3	-	3	-	-
6	-	1	1	-	-	3	1	1	1	1
29	-	-	1	1	1	1	1	1	1	1
16	3	1	-	1	1	-	-	-	1	1

NB: Lines in their ascending order of score for low selections

Marker data from chromosome 8 for PL1 (high selections)

Line	M1(4)	M2(1)	M3(3)	M4	M5(5)	M6	M7	M8	M9	M10(6)
4	1	1	-	3	3	-	3	-	3	3
42	1	1	1	1	1	1	3	3	3	3
47	-	1	1	1	1	1	1	1	1	1
17	1	1	3	3	3	-	3	-	3	3
7	-	1	1	-	3	-	3	-	3	3

Marker data from chromosome 8 for PL1 (low selections)

Line	M1(4)	M2(1)	M3(3)	M4	M5(5)	M6	M7	M8	M9	M10(6)
34	3	3	-	3	3	3	3	3	3	3
67	1	1	-	1	1	1	-	1	1	1
31	3	3	-	1	1	-	1	-	1	1
11	3	3	3	3	3	3	3	3	3	3
66	3	3	3	1	1	-	3	-	3	3

Marker data from chromosome 9 for FW (high selections)

Line	M1	M2	M3	M4(4)	M5(1)	M6(3/5)	M7	M8	M9	M10(6)
81	3	1	1	1	1	1	-	1	1	1
47	1	-	1	1	1	1	1	1	1	1
25	1	1	1	1	-	1	-	3	3	3
38	1	1	1	1	3	3	-	3	3	3
95	1	1	3	-	3	3	-	3	-	1

Marker data from chromosome 9 for FW (low selections)

Line	M1	M2	M3	M4(4)	M5(1)	M6(3/5)	M7	M8	M9	M10(6)
2	1	-	3	3	-	3	-	1	-	-
37	1	1	1	3	3	3	3	3	3	3
21	3	3	3	-	-	3	-	1	1	1
85	1	1	1	1	-	3	-	1	1	1
45	1	1	1	1	-	1	-	3	3	3

The above information allows the scrutiny of the marker profiles of various lines in more detail and helps to make an informed opinion about their possible genotypes. For example, the missing marker data (on markers M3, M6 and M8) on line 35 are likely to bear a score of 3 because that is the score of all the adjacent markers (markers M2 to M10). This in turn means that line 35 would be selected if data were available and selection was carried out on any of the markers in this region. Using these arguments we can safely assume that the following results will be obtained if the marker data were complete for the above lines in markers positions 1 to 6 for single QTL.

Marker position		1	2	3	4	5	6
LW2	High selections	4	4	5	4	5	5
	Low selections	3	3	3	4	3	3
PL1	High selections	5	3	3	5	2	1
	Low selections	4	3	3	4	2	3
FW	High selections	3	3	3	4	3	3
	Low selections	3	3	4	3	4	2
Totals		22	19	21	24	19	17
Range		3-5	3-4	3-5	3-5	2-5	1-5

The above results show that there is little difference between various positions and selection for the markers gives similar totals and range, except for the marker at the other end of the chromosome (6).

Another method of assessing efficiency would be to determine the occasions on which the lines definitely do not possess the desired marker genotype. In other words count the number of lines that would be discarded were we to make selections based on the molecular marker data that are available. The following results are obtained on this criterion.

Marker position	1	2	3	4	5	6	Totals
	Lines with wrong marker profiles						
LW2	2	3	2	1	2	1	11
PL1	1	2	1	1	6	6	17
FW	2	4	3	2	3	4	18

The above results show that lines are miss-classified regularly by every marker position. But marker positions 5 and 6 give more miss-classifications. There is also a visual trend between traits. Few lines are miss-classified for LW2 compared to PL1 and FW.

Selection for 2 QTL was carried out for AH, PL2, NL, LW1 and LL1 traits (Tables 5.10-5.14). These results divide into two groups, the first 2 traits were controlled by two QTL on the same linkage group while the remaining traits had one QTL on different linkage groups. Initially, we looked at the marker profiles of the 5 high and 5 low selections to establish if there was any correlation between the phenotypic performance and marker genotype of these lines. The data are shown below for AH and PL2:

Apical Height			Petiole Length-2		
		High selections			
Line	m1 (0.0)	m2(47.9)	Line	m1(0.0)	m2(37.1)
QTL effect	*-90.64*	*95.39*		*-13.11*	*13.18*
57	3/1	1	38	3	1
38	3	1	49	1/3	1
47	3	1	53	3	1
4	3/1	3	65	3/1	1
13	3/1	1	56	1/3	1
Desirable	(3	1)		(3	1)
		Low selections			
21	3/1	3	67	1	3
85	1	1	1	1	1
94	3	1	26	1	1
82	1	1	69	1	3
26	1	1	45	3/1	3
Desirable	(1	3)		(1	3)

The above results show that appropriate marker bands appear more frequently at the expected places. The alternative scores given for some of the markers, point to the suggested genotypes of these lines as the data were missing. For example, 3/1 means that a score of 3 is more likely than 1, based on the scores of the adjacent markers. It is also evident that such scores are more frequent for marker locus m1 because it is a terminal marker and it is not possible to speculate on its score with confidence.

Given below is the relevant information for NL, LW1 and LL1, the traits that are controlled by 2 QTL located on different linkage groups.

Number of leaves			Leaf Width-1			Leaf Length-1		
Line	LG7	LG8	Line	LG8	LG9	Line	LG8	LG9
Position	0.0	18.0		18.0	92.0		18.0	92.0
QTL effect	*-0.11*	*0.44*		*6.95*	*5.81*		*13.15*	*10.86*
			High selections					
33	1/3	1	18	1	1	18	1	1
12	3	1	17	1	1	17	1	1
7	3	1	63	1	1	47	1	1
36	1/3	1	12	1	3	4	1	1
42	3/1	1	54	1	1	42	1	1
Desirable	(3	1)		(1	1)		(1	1)
			Low selections					
23	1	3	9	3	3	34	3	3
89	3	3	34	3	3	9	3	3
66	1	3	95	3	3/1	88	3	1
77	1	3	11	3	3	11	3	3
49	1	3	31	3	3/1	95	3	3/1
Desirable	(1	3)		(3	3)		(3	3)

These results are much clearer than those described earlier because the desired marker alleles are present at the predicted sites. Apart from missing data, there are hardly any discrepancies. Further, such discrepancies occur more frequently for the marker on linkage group 7 that has a QTL with a very small effect and is a terminal marker. Marker data of linkage groups 8 and 9 seems to be fit better with the predictions, perhaps due to fewer missing values, better consistency of the marker results and markers being located some distance away from the terminal ends of the chromosome.

Results in Tables 5.10-5.14 further confirm that the efficiency of selection is highly reduced when 2 or more QTL are located on the same linkage group and when QTL have small effect and markers involved are located close to the chromosome ends. The quality of data, that is a few or no missing values, also improves the efficiency, Table 5.14. Another result worth reporting from Tables 5.13-5.14 is that the efficiency of selection is really high, particularly

on means, even when selection is carried out on both markers. The same however does not happen for NL where QTL have very unequal effects and one of the QTL is located close to the top of linkage group 7.

The efficiency is usually low for selecting 2 QTL simultaneously because the number of lines considered in such cases is usually small. These results are further influenced by missing data. The problem of missing data becomes even more acute when there are 3 or more QTL and the selection has to be applied to 3 or more markers simultaneously. Table 5.15 shows that sometimes all 5 top scoring lines have some missing marker data. Results in Tables 5.15, 5.16 and 5.17 further confirm that sample size becomes very critical when selection is applied to 3 or more markers/QTL simultaneously. As this number becomes very small the choice becomes limited and the efficiency on top ranks, lines and means is reduced. Selection based on a single marker clearly seems to give better efficiency, particularly for efficiency on means.

The relative efficiency and efficiency of selection based on the marker loci are presented for various traits in Table 5.18. It can be seen from the table that various QTL account for between 2% (A=0.02 for FH) to 49% (A=0.49 for NL) of the additive variance of various traits. Large differences are also observed for the selection index, I, whose values vary from 0.06 (FT) to 1.49 (FW), and efficiency of selection ES (0.02-0.98). The relative efficiency RI (0.90-1.43) and heritability of the means (0.35-0.93), on the other hand display less variation in their range compared to I, A and ES.

The selection index (I) is high for traits with low heritability (1.49, 0.80 & 0.97 for FW, LW1 & NL respectively) and low for traits with high heritability (0.06 & 0.07 for FT & LW2 respectively). The relative efficiency (RI) follows the same pattern as the selection index and increase as the heritability is reduced. The efficiency of selection based on the marker loci (ES) was observed to be low for most of the QTL detected. The ES is high for traits with high additive effects and it is more than 50% for LL1 (LG8), PL1 (LG8), NL (LG8), LW1 (LG8 and 9) and FW (LG9). The QTL detected in linkage group 8 for LL1, PL1, NL, LW1 and LW2 also show high efficiency of selection based on number of top ranks, lines and means.

Finally, the comparative costs of the methods are compared in the following paragraphs. First of all, the costs of direct selection will involve the following:

(1) costs of producing the F_1 hybrid seed (manually)

(2) costs of raising the F_1 plants

(3) costs of anther/ovule culture to produce haploids

(4) costs of producing doubled haploids

(5) costs of selfing these DHs.

(6) Costs of evaluating the DHs in the field and laboratory analyses of quality traits.

The comparative costs of the MAS procedure will also involve several of the above and some additional lab costs such as:

(1) costs of producing the F_1 hybrid seed (manually)

(2) costs of raising the F_1 plants

(3) costs of anther/ovule culture to produce haploids

(4) costs of producing doubled haploids

(5) costs of evaluating DH plants for marker and quantitative traits for locating QTL and identifying markers for selection.

(6) costs of selfing selected DHs.

(7) Costs of evaluating the DHs in the field and laboratory analyses of quality traits.

The above listings show that MAS will cost more in terms of money and manpower. In the present case, marker profiling of 89 DH lines for 90 (10 per chromosome and 9 chromosomes) microsatellite marker loci will cost approximately £8,000 and take about 10 months of lab work (Appendix III, calculations on Arabidopsis). The data would have to be recorded for more than 89 DH (say 150 lines) were we to locate QTL with reasonable precision.

5.4 Discussion and conclusions

Quantitative traits such as seed yield and plant weight are highly complex and controlled by many genes whose effects are also influenced by random and non-random environmental factors. These traits clearly show partial inheritance and it is not possible to trace each and every gene that is segregating for them. Further, genes also have unequal effects and those with large effect are termed as major genes while others are called modifiers (Kearsey and Pooni, 1996). The phenotypic selection of such traits is often less effective, particularly when applied to individuals. As many breeding programs have faltered due to this inefficiency of selection, scientists have suggested the use of marker-assisted selection to improve the chances of identifying and isolating desired genotypes, at least for those quantitative traits that have low heritability. The argument is based on the fact that molecular/DNA markers have 100% heritability and therefore are more amenable to selection than say a QTL that may be located close by (Tanksley et al, 1989, Kearsey and Pooni, 1996).

In a traditional breeding program, the marker-assisted selection is applied to say the F_2 individuals and the selected genotypes are propagated in the subsequent generations. However, breeders are also trying to reduce the total span of the breeding program either by rapidly advancing generations using methods such as single seed descent or double haploids (Poehlman & Sleper 1995). The present study concentrates on evaluating the efficiency of marker-assisted selection when applied to doubled haploids derived from a cross of *Brassica oleracea* lines. The efficiency of MAS is measured using three different criteria; based on

the number of top ranks, the proportion of the ranked lines in a marker group and proportion of the phenotypic advance based on means.

5.4.1 Phenotypic selection

The impact of phenotypic or direct selection on the population of 89 doubled haploids is very remarkable indeed. It is highly apparent from Tables 5.1 and 5.2 that such a selection would be very effective in the present case. Although the selection pressure was moderately severe (<5%) but the phenotypic gains that are possible for both high and low scores of each trait are really impressive. For example, S can be as large as 73% of the overall mean of the doubled haploids and when converted to R the highest value of R can be as large as 48% of the μ. Such levels of advance are possible only in those cases where selection has not been applied before. Incidentally, the present cross falls in to this very category because it is a cross between Chinese kale and calabrese and has not been used in any breeding program before. The extent of genetic variability displayed by such a wide cross can be very large and that seems to be borne out by the results in Tables 5.1-5.3.

Another pertinent point worth discussion is the marked difference between the observed responses for the high and low scores of each trait. Response has been higher for high score in all but 5 traits (see Table 5.3) and this difference between high and low scores was as big as 44% of the μ. This shows clearly that either there is genotype x environment interaction or a scalar effect that is making the top end of the distribution more spread out. In other words, variances of the DH lines may be significantly larger when they take an above average score and the opposite applies to the lower scores. Such a situation often leads to skewness and that is what we have for PH1, LL1, LW1, PL1, PH2, PL2, AH and FW. There is no

discernable relationship between observed response (% of μ) and heritability indicating that genetic variation does not correlate with μ in any simple manner.

Tables 5.4 and 5.5 further confirm that variances and spread of lines differs for the high and low scores because the range of scores displayed by the 5 highest scoring lines is much wider than that displayed by the lines with low scores. A significant difference between the variances of the high and low selections for 5 traits and the presence of skewness for 4 of these traits clearly show that scalar effects are present in these materials (Table 5.6). The selection of several lines for more than one trait also indicates that some of the traits are critically associated with each other.

5.4.2 Marker-assisted selection

Many scientists have argued that MAS will give better results compared to direct or phenotypic selection, particularly when heritability is low (Whittaker et al., 1995; Moreau et al., 1998; Knapp, 2001; Lange and Whittaker, 2001). The present results confirm this assertion conclusively as in all the cases at least some of the best lines could be identified using their marker profiles. Perusal of the 5 top scoring lines for each trait revealed that most of these lines possessed the marker genotype that was expected on the basis of the QTL analysis, even when the situation was complicated due to missing data. Positive identification of those lines that possessed a wrong marker (for single QTL traits) revealed clearly that there is some correlation between misclassification and low heritability. In other words, better results will be expected when selection is made on the basis of markers where heritability is low. However, it is also apparent that the various QTL considered for each of

the single QTL traits have unequal effects and they account for 17-39% of the total additive genetic variation, Table 5.18. Furthermore, the efficiency of MAS is similar for PL1 and LW2 because the small effects of the QTL controlling LW2 is balanced by its high heritability while the opposite holds for PL1. MAS efficiency as measured by the top scoring lines is low in the case of FW because its QTL has a small effect and it has a low heritability value of only 0.35.

As for the markers concerned, marker closest to the QTL does not seem to give any better results than say selection based on flanking markers or the next nearer marker (Tables 5.7-5.9). If the numbers mean anything, then the worst results are obtained using flanking markers. Missing data for the markers will have the biggest impact in this case because now there are twice as many chances of information being incomplete than otherwise. Selection based on a single marker that is located up to 25cM away from the QTL seems to give as good results as the one that is very close to the QTL. So, any marker that has complete set of data and lies within this distance can be used for selection without loosing much accuracy.

Another important outcome of this study is that marker selection is quite effective even when the marker used for selection is located far away from the QTL, at the terminal end of the opposite arm. This result has very significant implications for practical breeding because it shows that one can obtain at least some of the top 5% recombinants by using practically any marker on the relevant chromosome. On the face of it, this seems to be a rather unexpected result, but a closer scrutiny shows that we can expect as many as 20% intact chromosomes among the F_1 gametes (Koumproglou et al., 2002). And it is these

chromosomes and the ones with double recombination (not involving the QTL region) that are being picked up by the selection for this terminal marker.

Based on the above interpretation/discussion, it is also appropriate to suggest that one does not need more than 3 or 4 markers per chromosome to implement MAS effectively (Bouchez et al., 2002). Any more markers are not likely to improve its efficiency by any significant margin, unless the chromosomes are exceptionally long and there are recombination hot spots nearer to the QTL under selection.

The 3 methods employed to measure the efficiency of marker selection also give different outcomes (see Tables 5.7-5.9). Top ranks among the best 5 lines give perhaps the most confusing results due to the missing data. In any case, we can not expect all 5 lines to possess the expected marker profiles because each line has the best phenotype but it may not be the best genotype. Although heritability values are generally high for most traits but still there is a substantial level of non-heritable variation that can change the ranks of the lines that are genotypically close to the best but are not the best genotypes. Also, there is some probability that recombination may have occurred in the region between the QTL and the marker.

The second method (lines in the marker group) also seems to be rather unreliable and visibly more influenced by the sample size. However, it should provide a better discrimination between marker positions 1-6 when marker data are complete and there are no missing values for traits with single QTL. Perhaps the most suitable method to measure the efficiency of MAS is the 'method of means'. This method provides a measure of the S that

will be achieved by employing MAS. The results in Tables 5.7-5.9 show that the efficiency of MAS is high for LW2 and it also gives better results (efficiency of 0.9 or above) for FW where heritability is low (see Table 5.9).

The efficiency measures of MAS for 2 QTL provide a mixed picture. When these QTL are located on the same chromosome, the marker profiles of the top scoring lines do not match with the expected on 5 (out of 40) occasions (see results section for AH and PL2). However, there is no line that has both markers at variance with the expected genotype. So, it seems that either one of the markers has a poor association with the QTL or the effects of the QTL are small. Low efficiency estimates in Tables 5.10 and 5.11 confirm the above explanation and further indicate that selection for a single marker can give marked improvements in this case. As can be seen from these tables, selection for either of the two markers ranges between 2 to 4 of the best lines.

Selection for 2 QTL on different chromosomes seems to be much more efficient than those located on the same chromosome. About 5% (3 out of 60) of the top scoring lines possess the unexpected marker allele, with one such case identified for NL, LW1 and LL1. The high consistency of the above results is also due to the fact that the detected QTL account for all the significant variation among the DH lines for NL, LW1 and LL1 traits (see Table 2.5 of the QTL location chapter).

For more than 2 QTL (Tables 5.15-5.17), the results confirm the same trends as those observed for 2 QTL. The basic problem of multiple marker selection is to have a large sample of lines after the selection and this is possible only if very large populations are

screened (Lande and Thompson, 1990; Edwards and Page, 1994; Gimelfarb and Lande, 1994a). As this cannot be done retrospectively, the results that we obtain are rather constrained for interpretation. The comparison of multiple QTL selection with that based on one or two markers, therefore is not valid in the present case because we will be comparing samples of 3-4 lines with those of 20+ and even of 62 lines (Table 5.15). However, one conclusion that is possible despite the above mentioned problems is that selection for one QTL can be more effective than all the QTL that contribute significantly to the variation of a trait.

The estimates of relative efficiency (RI) in Table 5.18 simply indicate that MAS will not be more effective when heritability is high but it will clearly be so when there is more environmental variation. As the selection can be applied to several traits simultaneously, using the same marker data and they may show different levels of heritability, MAS therefore will always have some advantage over phenotypic selection under most circumstances. If nothing else, MAS will allow the breeder to cut down the experimental size by a large factor and that must accrue some savings in materials and monetary terms. RI has little bearing on the relative size of the QTL effect and therefore would apply to any QTL that is detected to have a significant effect in the population. Efficiency of selection (ES), on the other hand, depends both on the heritability and the QTL effect. In fact ES = A/h_f^2 (see values in Table 5.18). It simply shows that MAS will be more efficient when heritability is low (Lande and Thompson, 1990; Luo et al. 1997).

Finally, it is not easy to determine precisely the exact costs of the two breeding programs. In the present case, both programs would require the production of the haploid plants from the

F_1 gametes. For direct selection, these haploids will be converted into diploids using either tissue culture or colchicine and the resulting (single plant) families would be evaluated in the field for a choice of traits. Selection will then be applied to these families based on their phenotypic performance.

The MAS program will also start from the same point and end up with a selection or selections. It will however differ from the direct selection program in the intervening steps. For example, if the marker and QTL information is already available (as in the present case), then MAS can be applied to the haploids, thus reducing their number. How big this reduction will be, depends on the objectives. Were we to improve only one or two traits controlled by single QTL, then the reduction can be large, up to half the sample. This will lead to very substantial savings in the cost of further breeding and experimentation. If, however, we were to improve complex traits like yield and quality simultaneously and there are not any QTL with large effect, then a large sample will need to be retained so that good recombinants can be found among the selected DH which already have some QTL fixed in them.

Both programs will, however, take the same length of time to complete and MAS will not delay the breeding work in any significant manner. The cost of marker profiling would also be rather low because, based on the present study, no more than 3-4 markers will be needed per chromosome to implement MAS. In addition, genotyping of the haploids can be staggered to reduce the lab costs even further, i.e. select using one marker first and then select for the second only among the selected group of individuals.

Table 5.1: The population mean (μ), phenotypic standard deviation (σ_p), heritability of the means (h_f^2), selection differential (S), response to selection (R) and observed response (OS) for the 5 highest lines

Trait	μ	σ_p (% of μ)	h_f^2	S (% of μ)	Mean of selected lines	R (% of μ)	Observed response, OS
PH1	58.98	12.80 (22)	0.86	25.89 (44)	88.83	22.27 (38)	29.85
NL	6.58	0.71 (11)	0.50	1.44 (21)	7.97	0.72 (11)	1.39
LL1	88.66	22.95 (26)	0.65	46.43 (52)	141.00	30.18 (34)	52.34
LW1	54.19	12.94 (24)	0.48	26.18 (48)	82.33	12.57 (23)	28.14
PL1	25.53	9.24 (36)	0.53	18.69 (73)	48.17	9.91 (39)	22.64
PH2	291.36	82.07 (28)	0.85	166.03 (56)	496.17	141.13(48)	204.81
LL2	277.10	38.00 (14)	0.69	76.87 (28)	356.83	53.04 (19)	79.73
LW2	157.37	21.36 (14)	0.92	43.21 (27)	196.83	39.75 (25)	39.46
PL2	113.77	22.00 (19)	0.80	44.51 (39)	163.00	35.61 (31)	49.23
FH	387.81	75.60 (19)	0.85	152.94 (39)	531.17	130.00(33)	143.36
FT	74.17	6.21 (8)	0.93	12.56 (16)	86.03	11.68 (15)	11.86
MH	838.28	166.25 (20)	0.66	336.32 (40)	1171.67	221.97(26)	333.39
AH	649.49	169.21 (26)	0.65	342.31 (52)	1020.33	222.50(34)	370.84
SW	36.12	5.32 (15)	0.46	10.76 (30)	45.67	4.95 (14)	9.55
FW	486.64	126.87 (26)	0.35	256.66 (52)	779.00	89.83 (18)	292.36

Table 5.2: The population mean (μ), phenotypic standard deviation (σ_p), heritability of the means (h_f^2), selection differential (S), response to selection (R) and observed response (OS) for the 5 lowest lines

Trait	μ	σ_p	h_f^2	S	Mean of lines selected	R	Observed response, OS
PH1	58.98	12.80	0.86	-25.89	34.83	-22.27	-24.15
NL	6.58	0.71	0.50	-1.44	5.13	-0.72	-1.45
LL1	88.66	22.95	0.65	-46.43	55.67	-30.18	-32.99
LW1	54.19	12.94	0.48	-26.18	34.17	-12.57	-20.02
PL1	25.53	9.24	0.53	-18.69	14.00	-9.91	-11.53
PH2	291.36	82.07	0.85	-166.03	165.67	-141.13	-125.69
LL2	277.10	38.00	0.69	-76.87	212.33	-53.04	-64.77
LW2	157.37	21.36	0.92	-43.21	115.33	-39.75	-42.04
PL2	113.77	22.00	0.80	-44.51	75.83	-35.61	-37.94
FH	387.81	75.60	0.85	-152.94	240.33	-130.00	-147.48
FT	74.17	6.21	0.93	-12.56	61.10	-11.68	-13.07
MH	838.28	166.25	0.66	-336.32	555.67	-221.97	-282.61
AH	649.49	169.21	0.65	-342.31	372.33	-222.50	-277.16
SW	36.12	5.32	0.46	-10.76	25.33	-4.95	-10.79
FW	486.64	126.87	0.35	-256.66	262.33	-89.83	-224.31

Table 5.3: The heritability of the means (h_f^2) and the average and difference of observed responses (OS) for the highest and the lowest scores of each trait

Trait	h_f^2	Average observed response OS (% of μ)	Difference in observed response (% of μ)
PH1	0.86	27.00 (46)	5.70 (10)
NL	0.50	1.42 (22)	-0.06 (1)
LL1	0.65	42.67 (48)	19.35 (22)
LW1	0.48	24.08 (44)	8.12 (15)
PL1	0.53	17.09 (67)	11.11 (44)
PH2	0.85	165.25 (57)	82.12 (27)
LL2	0.69	72.25 (26)	14.96 (5)
LW2	0.92	40.75 (26)	-2.58 (2)
PL2	0.80	43.59 (38)	11.29 (10)
FH	0.85	145.42 (38)	-4.12 (1)
FT	0.93	12.47 (17)	-1.21 (2)
MH	0.66	308.00 (37)	50.78 (6)
AH	0.65	324.00 (50)	93.78 (14)
SW	0.46	10.17 (28)	-1.24 (3)
FW	0.35	258.34 (53)	68.05 (14)

Table 5.4: The mean, variance and range of the 5 highest scoring DH lines

Trait	Mean	Variance	Range	Magnitude of range	Selected lines (line numbers)
PH1	88.83	143.61	107.5-79.17	28.33	12, 42, 91, 76, 31
NL	7.97	0.05	8.33-7.83	0.50	33, 12, 7, 36, 42
LL1	141.00	33.47	147.50-132.50	15.00	18, 17, 47, 4, 42
LW1	82.33	7.78	85.83-79.17	6.66	18, 17, 63, 12, 54
PL1	48.17	21.67	52.50-41.67	10.83	4, 42, 47, 17, 7
PH2	496.17	3995.69	595.00-431.67	163.33	4, 76, 57, 20, 33
LL2	356.83	211.25	378.33-340.00	38.33	38, 92, 49, 13, 93
LW2	196.83	39.03	206.67-190.00	16.67	93, 75, 77, 35, 95
PL2	163.00	75.21	177.50-154.17	23.33	38, 49, 53, 65, 56
FH	531.17	110.97	543.33-516.67	26.66	44, 51, 28, 9, 47
FT	86.03	2.02	88.17-85.00	3.17	11, 77, 23, 9, 45
MH	1171.67	4181.94	1260.00-1095.00	165.00	47, 20, 57, 38, 93
AH	1020.33	3508.89	1105.00-961.67	143.33	57, 38, 47, 4, 13
SW	45.67	1.88	46.67-44.17	2.50	38, 53, 90, 18, 58
FW	779.00	4307.78	866.67-693.33	173.33	81, 47, 25, 38, 95

Table 5.5: The mean, variance and range for the 5 lowest scoring DH lines and the F ratio testing their variance against those DH with the highest scores in Table 5.4

Trait	Mean	Variance	Variance ratio, $F_{(4,4df)}$	Range	Magni-tude of range	Selected lines (line numbers)
PH1	34.83	2.57	55.8***	33.33-37.50	4.17	29, 21, 85, 1, 57
NL	5.13	0.02	2.5ns	5.00-5.33	0.33	23, 89, 66, 77, 49
LL1	55.67	24.44	1.4ns	47.50-60.00	12.50	34, 9, 88, 11, 95
LW1	34.17	1.74	4.5ns	32.50-35.83	3.33	9, 34, 95, 57, 31
PL1	14.00	1.18	18.4**	12.50-15.00	2.50	34, 67, 31, 11, 66
PH2	165.67	88.33	45.2**	156.67-180.00	23.33	24, 77, 40, 75, 21
LL2	212.33	22.01	9.6*	206.67-218.33	11.66	16, 12, 86, 67, 45
LW2	115.33	47.43	1.2ns	104.17-121.67	17.50	7, 86, 6, 29, 16
PL2	75.83	24.65	3.1ns	69.17-80.83	11.66	67, 1, 26, 69, 45
FH	240.33	443.61	4.0ns	223.33-265.00	41.67	85, 21, 49, 68, 4
FT	61.10	5.65	2.8ns	57.50-63.00	5.50	4, 29, 85, 76, 5
MH	555.67	2134.17	2.0ns	488.33-593.33	105.00	7, 37, 74, 75, 21
AH	372.33	1109.17	3.2ns	325.00-405.00	80.00	21, 85, 94, 82, 26
SW	25.33	1.94	1.0ns	23.33-26.67	3.33	57, 6, 87, 16, 29
FW	262.33	92.50	46.6**	253.33-278.33	25.00	2, 37, 21, 85, 45

Variance ratio is of larger variance/smaller variance

Table 5.6: The mean, variance, normality test, skewness and kurtosis for the 89 DH lines

Trait	Mean	Variance	Normality (P-value)	Skewness	Kurtosis
PH1	58.98	161.98	0.19	0.64*	4.62*
NL	6.58	0.50	0.50	-0.03	2.64
LL1	88.66	520.81	0.02*	0.54*	2.64
LW1	54.19	165.57	0.07	0.47*	2.60
PL1	25.53	84.34	0.00***	0.88*	3.29
PH2	291.36	6660.02	0.08	0.83*	4.32
LL2	277.10	1427.65	029	0.32	2.59
LW2	157.37	450.97	0.79	-0.12	2.49
PL2	113.77	478.59	0.36	0.43*	3.03
FH	387.81	5650.72	0.28	-0.01	2.35
FT	74.17	38.14	0.07	0.00	3.02
MH	838.28	27329.62	0.04*	0.24	2.38
AH	649.49	28309.66	0.06	0.47*	2.70
SW	36.12	28.03	0.05	-0.30	2.59
FW	486.64	15915.18	0.35	0.49*	3.16

Table 5.7: The mean, variance and range of selected lines, total number of lines in the marker group and efficiency of selection (EOS) based on top ranks, lines and means (phenotypic advance) for Leaf width2 when QTL is located in linkage group 8 at 36cM

Marker (position)	Mean	Var.	Range	Lines in group	EOS top rank	EOS lines	EOS mean
High selections							
1 (34.8)	195.00	52.78	206.67-188.33	54	4	0.69	0.95
2 (26.8&43.2)	188.17	25.83	196.67-183.33	36	1	0.47	0.78
3 (26.8)	196.83	39.03	206.67-190.00	62	5	0.74	1.00
4 (43.2)#	188.17	25.83	196.67-183.33	39	1	0.51	0.78
5 (43.2)#	188.17	25.83	196.67-183.33	39	1	0.51	0.78
6 (86.8)	196.83	39.03	206.67-190.00	56	5	0.64	1.00
Low selections							
1 (34.8)	124.83	16.81	120.00-130.00	26	2	0.46	0.77
2 (26.8&43.2)	129.00	43.89	120.00-136.67	13	1	0.23	0.58
3 (26.8)	126.00	31.39	120.00-133.33	22	2	0.41	0.75
4 (43.2)#	119.17	95.83	104.17-130.00	21	3	0.43	0.91
5 (43.2)#	119.17	95.83	104.17-130.00	21	3	0.43	0.91
6 (86.8)	121.67	12.50	116.67-125.00	32	3	0.38	0.85

NB: # - same marker at the same position (so same information for both markers)

Table 5.8: The mean, variance and range of selected lines, total number of lines in the marker group and efficiency of selection (EOS) based on top ranks, lines and means (phenotypic advance) for Petiole Length 1 when QTL is located in linkage group 8 at 18cM

Marker (position)	Mean	Var.	Range	Lines in marker Group	EOS top ranks	EOS lines	EOS means
High selections							
1 (19.2)	48.17	21.67	52.50-41.67	34	5	0.74	1.00
2(26.8&7.1)	41.00	41.46	51.67-35.00	11	1	0.36	0.68
3 (26.8)	44.83	31.39	51.67-39.17	21	3	0.52	0.85
4 (7.1)	46.00	35.21	52.50-39.17	33	3	0.55	0.90
5 (43.2)	43.50	50.83	51.67-35.00	22	2	0.50	0.79
6 (86.8)	40.00	32.99	50.00-35.83	32	1	0.28	0.64
Low selections							
1 (19.2)	14.33	1.18	12.50-15.00	51	4	0.76	0.97
2(26.8&7.1)	15.67	0.49	15.00-16.67	22	2	0.32	0.86
3 (26.8)	15.50	0.21	15.00-15.83	39	2	0.56	0.87
4 (7.1)	14.33	1.18	12.50-15.00	37	4	0.59	0.97
5 (43.2)	14.67	1.60	12.50-15.83	62	2	0.79	0.94
6 (86.8)	14.50	1.25	12.50-15.00	56	3	0.61	0.96

Table 5.9: The mean, variance and range of selected lines, total number of lines in the marker group and efficiency of selection (EOS) based on top ranks, lines and means (phenotypic advance) for fresh weight when QTL is located on linkage group 9 at 48cM.

Marker (position)	Mean	Variance	Range	Lines in group	EOS top ranks	EOS lines	EOS means
High selections							
1 (43.8)	717.33	12249.43	866.67-605.00	22	2	0.31	0.79
2 (52.7&35.1)	760.00	7715.59	866.67-660.00	26	3	0.50	0.94
3 (52.7)#	763.00	7011.11	866.67-675.00	33	3	0.55	0.95
4 (35.1)	776.00	4995.59	866.67-678.33	39	4	0.59	0.99
3 (52.7)#	763.00	7011.11	866.67-675.00	33	3	0.55	0.95
6 (55.0)	732.00	9343.64	866.67-641.67	40	3	0.45	0.84
Low selections							
1 (43.8)	325.67	1782.64	256.67-255.00	27	1	0.44	0.72
2 (52.7&35.1)	309.67	2539.38	253.33-353.33	27	2	0.56	0.79
3 (52.7)#	269.33	617.41	253.33-313.33	56	4	0.73	0.97
4 (35.1)	309.67	2539.38	253.33-353.33	30	2	0.47	0.79
3 (52.7)#	269.33	617.41	253.33-313.33	56	4	0.73	0.97
6 (55.0)	314.67	1960.18	256.67-353.33	41	2	0.49	0.77

NB: # - same marker at the same position (so same information for both markers)

Table 5.10: The mean, variance and range of selected lines, total number of lines in the marker group and efficiency of selection (EOS) based on top ranks, lines and means (phenotypic advance) for apical height on linkage group 1 at 0 and 42cM

QTL (Position)	Mean	Variance	Range	Lines in group	EOS top ranks	EOS lines	EOS means
High selections							
1 (42.0)	1015.67	4185.63	1105.00-948.33	51	4	0.65	0.99
2 (0.0)	929.33	9708.16	1051.67-825.00	21	2	0.33	0.75
Both	637.33	5022.92	763.33-498.33	6	0	0.00	-0.03
Low selections							
1 (42.0)	447.67	4895.16	325.00-595.00	26	1	0.23	0.73
2 (0.0)	431.67	426.49	403.33-451.67	27	1	0.56	0.79
Both	470.00	1831.89	421.67-521.67	18	0	0.22	0.65

Table 5.11: The mean, variance and range of selected lines, total number of lines in the marker group and efficiency of selection (EOS) based on top ranks, lines and means (phenotypic advance) for petiole length2 in linkage group 1 at 0 and 36cM

QTL (Position)	Mean	Variance	Range	Lines in group	EOS top ranks	EOS lines	EOS means
High selections							
1 (36.0)	160.84	114.90	177.50-149.17	51	4	0.69	0.96
2 (0.0)	151.67	303.82	177.50-137.50	21	2	0.44	0.76
Both	131.50	117.83	148.33-120.83	18	0	0.11	0.36
Low selections							
1 (36.0)	81.33	52.29	69.17-86.67	26	2	0.46	0.86
2 (0.0)	77.50	42.01	69.17-85.00	27	3	0.29	0.96
Both	102.00	165.48	86.67-117.50	6	0	0.00	0.31

Table 5.12: The mean, variance and range of selected lines, total number of lines in the marker group and efficiency of selection (EOS) based on top ranks, lines and means (phenotypic advance) for number of leaves in linkage group 7 and 8 at 0 and 18cM, respectively

Linkage group (position)	Mean	Variance	Range	Lines in group	EOS top ranks	EOS lines	EOS means
High selections							
LG8 (18.0)	7.97	0.05	8.33-7.83	34	5	0.78	1.00
LG7 (0.0)	7.60	0.09	8.00-7.33	28	2	0.43	0.73
Both	7.17	0.26	7.83-7.67	7	0	0.14	0.42
Low selections							
LG8 (18.0)	5.13	0.02	5.00-5.33	50	5	0.71	1.00
LG7 (0.0)	5.23	0.04	5.00-5.50	26	4	0.42	0.93
Both	5.87	0.24	5.00-6.17	13	1	0.08	0.49

Table 5.13: The mean, variance and range of selected lines, total number of lines in the marker group and efficiency of selection (EOS) based on top ranks, lines and means (phenotypic advance) for leaf width1 in linkage groups 8 and 9 at 18 and 94cM, respectively

Linkage group (position)	Mean	Variance	Range	Lines in group	EOS top ranks	EOS lines	EOS means
High selections							
LG8 (18.0)	82.33	7.78	85.83-79.17	34	5	0.71	1.00
LG9 (92.0)	82.00	10.28	85.83-78.33	41	4	0.61	0.99
Both	82.00	10.28	85.83-78.33	19	4	0.68	0.99
Low selections							
LG8 (18.0)	34.33	2.22	32.50-35.83	50	4	0.80	0.99
LG9 (92.0)	34.50	2.29	32.50-35.83	41	3	0.51	0.98
Both	34.67	264	32.50-35.83	26	2	0.58	0.98

Table 5.14: The mean, variance and range of selected lines, total number of lines in the marker group and efficiency of selection (EOS) based on top ranks, lines and means (phenotypic advance) for leaf length1 in linkage groups 8 and 9 at 18 and 92cM, respectively

Linkage group (position)	Mean	Variance	Range	Lines in group	EOS top ranks	EOS lines	EOS means
High selections							
LG8 (18.0)	141.00	33.47	147.50-132.50	34	5	0.68	1.00
LG9 (92.0)	141.00	33.47	147.50-132.50	41	5	0.61	1.00
Both	141.00	33.47	147.50-132.50	19	5	0.74	1.00
Low selections							
LG8 (18.0)	55.67	24.44	47.50-60.00	50	5	0.78	1.00
LG9 (92.0)	57.00	33.19	47.50-61.67	41	3	0.54	0.96
Both	57.00	33.19	47.50-61.67	26	3	0.54	0.96

Table 5.15: The mean, variance and range of selected lines, total number of lines in the marker group and efficiency of selection (EOS) based on top ranks, lines and means (phenotypic advance) for plant height2 in linkage groups 6, 8 and 9 at 56, 48 and 94cM, respectively

Linkage group (position)	Mean	Variance	Range	Lines in group	EOS top ranks	EOS lines	EOS means
High selections							
LG9 (94.0)	487.17	4919.44	595.00-428.33	41	4	0.61	0.96
LG8 (48.0)	432.83	2777.64	517.50-373.33	22	2	0.50	0.69
LG6 (56.0)	363.00	926.94	413.33-331.67	11	0	0.09	0.35
Both	303.89	2337.04	351.89-225.00	3	0	0.00	0.06
Low selections							
LG9 (94.0)	172.33	149.44	160.00-186.67	41	4	0.65	0.95
LG8 (48.0)	165.67	88.33	156.67-180.00	62	5	0.76	1.00
LG6 (56.0)	213.00	629.72	186.67-240.00	15	0	0.13	0.62
Both	262.78	2400.93	276.67-208.33	3	0	0.00	0.23

Table 5.16: The mean, variance and range of selected lines, total number of lines in the marker group and efficiency of selection (EOS) based on top ranks, lines and means (phenotypic advance) for flowering time in linkage group 2 and 3

Linkage group (position)	Mean	Variance	Range	Lines in group	EOS top ranks	EOS lines	EOS means
High selections							
LG2 (26)	86.03	2.02	88.17-84.67	44	3	0.66	1.00
LG3 (138)	85.34	2.85	88.17-83.67	31	3	0.52	0.94
LG3 (2)	85.27	4.73	88.17-83.00	20	2	0.40	0.94
Both	71.00	13.94	75.33-66.83	4	0	0.00	-0.27
Low selections							
LG2 (26.)	62.70	17.86	57.50-66.83	41	3	0.63	0.88
LG3 (138)	61.50	7.85	57.50-64.50	34	4	0.62	0.97
LG3 (2)	63.23	6.58	59.83-67.00	28	3	0.32	0.84
Both	78.44	32.27	75.00-85.00	3	0	0.00	-0.33

Table 5.17: The mean, variance and range of selected lines, total number of lines in the marker group and efficiency of selection (EOS) based on top ranks, lines and means (phenotypic advance) for flowering height in linkage group 1, 2, 3 and 9 at 0, 12, 0 and 54cM, respectively

Linkage group (position)	Mean	Variance	Range	Lines in group	EOS top ranks	EOS lines	EOS means
High selections							
LG9 (54.0)	506.33	210.34	526.67-495.00	33	2	0.58	0.83
LG1 (0.0)	502.67	350.83	526.67-480.00	21	0	0.38	0.80
LG2 (12.0)	521.17	640.14	543.33-480.00	34	4	0.50	0.93
LG3 (0.0)	513.00	817.22	543.33-480.00	31	3	0.52	0.87
Both	432.33	3587.86	480.00-328.33	5	0	0.00	0.31
Low selections							
LG9 (54.0)	240.33	443.65	223.33-265.00	56	5	0.75	1.00
LG1 (0.0)	293.83	190.83	278.33-308.33	27	0	0.41	0.64
LG2 (12.0)	251.33	672.78	223.33-281.67	41	4	0.56	0.93
LG3 (0.0)	251.33	581.11	225.00-278.33	34	4	0.65	0.93
Both	349.72	2820.88	294.17-400.00	3	0	0.00	0.26

Table 5.18: The additive effect (a), additive variance (A), heritability of the means (h_f^2), selection index (I), relative efficiency (RI) and efficiency of selection (ES) for QTL in DH lines

Trait	LG	QTL position (cM)	Marker close to QTL	a	A	h_f^2	I	RI	ES
FT	2	26.0	Pn121e1	-2.57	0.19	0.93	0.06	0.95	0.20
LW2	8	36.0	Pw188j1	-7.78	0.17	0.92	0.07	0.95	0.18
PH1	1	2.0	Ac-ctce02	-5.84	0.23	0.86	0.13	0.90	0.27
	3	48.0	Accatj02	-7.07	0.34	0.86	0.11	0.98	0.40
	6	36.0	Ac-ctaj06	-6.95	0.33	0.86	0.11	0.98	0.38
	7	80.0	Mbn72aj1	-4.57	0.14	0.86	0.14	0.96	0.16
PH2	6	56.0	Mca72	-29.66	0.14	0.85	0.15	0.96	0.16
	8	48.0	Pr97j1	28.53	0.13	0.85	0.15	0.96	0.15
	9	94.0	Pn47e4nm	34.02	0.18	0.85	0.14	0.96	0.21
FH	1	0.0	Ac-ctce02	-10.35	0.02	0.85	0.17	0.99	0.02
	2	12.0	Pw116e1	-23.22	0.10	0.85	0.16	0.97	0.12
	3	0.0	Pw116j1	-33.12	0.21	0.85	0.14	0.96	0.25
	9	54.0	Pw233j1	24.13	0.11	0.85	0.16	0.97	0.13
LL2	1	40.0	Pn52e3np	10.71	0.10	0.69	0.40	0.99	0.14
MH	2	60.0	P017e1	29.08	0.04	0.66	0.49	1.00	0.06
	9	60.0	Mbn83b1j1	53.27	0.14	0.66	0.44	1.00	0.21
LL1	8	18.0	P0159j1	13.15	0.40	0.65	0.32	1.13	0.62
	9	92.0	Pn47e4nm	10.86	0.27	0.65	0.39	1.04	0.42
PL1	8	18.0	P0159j1	5.14	0.39	0.53	0.54	1.24	0.74
NL	7	0.0	P087e2	-0.11	0.03	0.50	0.97	1.01	0.06
	8	18.0	P0159j1	0.44	0.49	0.50	0.51	1.43	0.98
LW1	8	18.0	P0159j1	6.95	0.37	0.48	0.68	1.29	0.77
	9	94.0	Pn47e4nm	5.81	0.26	0.48	0.80	1.16	0.54
FW	9	48.0	Ac-ctae16	40.94	0.20	0.35	1.49	1.26	0.57

CHAPTER 6

Marker-assisted selection in *Arabidopsis thaliana*

6.1 Introduction

Marker-assisted selection (MAS) is an indirect method of selection for specific DNA sequences, which are part of genes coding for economically important traits. Although the cost effectiveness of MAS is widely debated, the usefulness of QTL mapping for finding new favourable genes is hard to dispute (Tanksley et al. 1989). Marker-assisted selection helps in the redistribution of alleles between the parents and progeny, hence selecting for the appropriate genotype.

Marker-assisted selection has been widely studied using different approaches. Lande and Thompson (1990) proposed a method of marker-assisted selection that employs multiple regression of the phenotype on markers to identify a set of markers associated with QTLs as well as to estimate marker effects. The main conclusion from their deterministic analysis was that MAS based on an index incorporating marker effects together with phenotype yields better responses than selection based strictly on phenotype, provided there are sufficient markers and the population size is very large.

Moreau et al. (1998) evaluated the efficiency of marker-assisted selection (MAS) based on an index incorporating both phenotypic and molecular information with an analytical approach that takes into account the size of the experiment. They found that for a given population size, the relative efficiency of MAS was high at low heritabilities. Gimelfarb and

Lande (1994a) used simulations to investigate the efficiency of MAS as affected by several factors including total number of markers in the genome, number of marker contribution to the index, population size and heritability of the character. They observed that selection was more efficient if markers contributing to the index are re-evaluated each generation and increasing the number of markers in the genome as well as the number of markers contributing to the index gave higher efficiency of selection.

In this study, marker assisted selection is applied to the F_2 plants of a cross between two Arabidopsis ecotypes, Landberg *erecta* and Columbia. The efficiency of marker based selection is estimated based on number of top ranks and as the ratio between response based on MAS and response obtained in the F_3 by applying phenotypic selection to the F_2 generation. The efficiency of population selection based on the marker loci is determined using the formulae of Lande and Thompson (1990).

6.2 Materials and methods

6.2.1 Mapping population

The population consisted of 200 F_2 individuals derived from the cross between the Arabidopsis ecotypes, Landberg *erecta* and Columbia, and their descendent F_3 families. Firstly, 400 F_2 plants were grown in the glasshouse and scored for molecular markers and quantitative traits with the objective of QTL detection and location. Then, 200 of these individuals were selfed to produce 200 F_3 families that were raised during the next season in a single experiment with 5 replicates. The F_3 families were also analysed for the presence of QTL and 17 QTL detected in the F_2 were also detected in the F_3 generation. The QTL detected in both the F_2 and F_3 were used for marker-assisted selection in this section.

The following traits were scored on both generations TTG, TTL, HT20, RL20, CL20, TTB, TTF, HTF, RLF, CLF and HT34. The study will focus on the QTL located in the F_2/ F_3 lineages for the six traits RL20, TTB, TTF, HTF, RLF and HT34. The F_2 were grown during the summer of 2001, from 15th June 2001 and the F_3 were raised during the summer of 2002, from 7th July 2002. In both experiments the plants reached maturity within eight weeks of sowing.

The QTL located from the F_2 and F_3 data are presented in the previous chapter, but are given in Table 6.9 for maintaining the continuity of the present section.

6.2.2 Molecular marker analysis

Each F_2 plant was genotyped for 30 molecular markers and these markers are listed in Appendix II. Microsatellites were chosen as the marker for the present study because they are known to be the most consistent and repeatable of all the PCR markers. This gave the average distance between markers as approximately 20cM.

6.2.3 Phenotypic selection in F_2 plants

Selection of 10 F_2 plants (5%) was applied exactly in the same manner as in the case of DH families. Initially, response to selection was predicted following Falconer and Mackay (1996). This procedure is described for the DH lines previously. In the present case:

$$R = h_b^2 S \text{ and } S = i\sigma_p.$$

6.2.4 Response to selection in F_3 families

The corresponding 10 F_3 families that have descended from each of the highest and the lowest scoring F_2 plants were then identified and their overall means were used to compare observed and predicted response for various traits.

With the availability of the whole set of F_3 data, it is also possible to apply direct selection in this generation. Therefore, 10 highest and 10 lowest scoring F_3 families were identified for each trait and their mean, variance, range etc were calculated for comparison. The identification of the F_3 families was also extended to those individuals that were selected by MAS followed by phenotypic selection. In this case, the number of F_3 lines identified using markers are given by the efficiency of selection (EOS) based on number of top ranks.

The observed response (OS) and genetic advance (OR) were calculated from the F_2 and F_3 data as follows:

Observed response (OS) = F_2 mean (5%) – F_2 population mean

Observed genetic advance (OR) = F_3 mean (5%) – F_3 population mean.

6.2.5 Marker-assisted selection in F_2

For the traits where single QTL were detected (HT34 only), the marker genotypes of the above 10 plants were investigated to see if the desired marker has been picked up by the phenotypic selection. The marker profiles of the 10 highest and 10 lowest scoring plants for HT34 were noted. The MAS was also applied to the F_2 data exactly in the same manner as to the DH lines earlier. In short, 6 different positions were used and they are:

1) marker closest to the QTL
2) flanking markers
3) next marker nearer to the QTL
4) second marker that lies on the other side of the QTL
5) marker located around the middle of the chromosome
6) marker located farthest away from the QTL, at the distal end of the chromosome.

The application of MAS to those traits that were controlled by two or more QTL was restricted to those markers that were located closest to the putative position of the QTL. The selection was also applied using single markers representing individual QTL, as in the case of the DH lines.

6.3 Results and Interpretation

6.3.1 Selection in the F_2 generation

The statistics presented in Table 6.1 show that the F_2 and F_3 generations have very similar performance. For example, the population means are very similar. If anything, the F_2 mean was slightly lower than the F_3 mean for 4 traits, indicating that F_3 was grown under marginally better conditions.

The broad sense heritability of the F_2 varies from 0.30 (TTB) to 0.50 (RL20), whereas the F_3 heritability differs from 0.68 (RL20) to 0.83 (TTB and TTF). Clearly, there is no overlap between the two sets of estimates. This is precisely because the F_3 values were estimated from the family means and provided a measure of the heritability of the means. As family means are measured more accurately and subjected to low non-heritable variation, the heritability of the means thus takes a larger value than heritability for each trait.

On the other hand, the F_2 heritability values were different between traits (range = 0.20) and the heritability of the means were similar (range = 0.15). In fact, the highest and the lowest values differ in the heritability ($\chi^2_{(1)} = 5.00^*$) and the heritability of the means do not differ ($\chi^2_{(1)} = 1.49$ns). Consequently, there is little agreement between the relative rankings of these parameters. Further, this lack of association between heritability and heritability of the means may also be due to a high non-additive variance in the F_2. Overall, the estimates of heritability can be assumed moderate to high for various traits.

The mean, variance and range for the highest and lowest selections based on the phenotypic scores for the F_2 are given in Tables 6.2 and 6.3. These results show that a total of 69 plants have been selected for their good characteristics. Therefore, approximately 34.5% of plants (69 out of 200) possess some good traits. More precisely, 41 plants were selected for high scores while 34 were chosen as low selections. Six of these plants were present in both sets because they were selected for high and low scores for different traits. However, 13 of these 69 plants possess good scores for at least 3 traits and therefore can be considered as desirable recombinants. Plants 2, 14 and 178 show promising performance for 4 or more traits.

An elaborated summary of the above results is given below.

(i) High selections: single traits

Plant	10	37	57	62	67	68	78	80	85
Trait	HTF	HT34	TTB	HT34	RL20	HTF	HTF	HT34	RL20
Plant	88	100	103	105	118	129	130	145	150
Trait	TTB	HT34	HTF	HTF	TTF	HTF	TTF	HT34	HTF
Plant	153	157	163	164	183	193	195	196	
Trait	TTF	HTF	HT34	RLF	HT34	HT34	RL20	RL20	

(ii) High selections: two traits

Plant	38	47	59	60	75	86	89	111	124	135	176
Trait1	TTB	TTB	TTB	RL20	RL20	RL20	TTB	RL20	RL20	TTB	HTF
Trait2	TTF	TTF	TTF	RLF	RLF	RLF	HTF	RLF	RLF	TTF	HT34

(iii) High selections: 3 or more traits

Plant	Traits
30	RL20, RLF, HT34
171	RLF, TTB, TTF
178	RLF, TTB, TTF
189	RLF, TTB, TTF

(iv) Low selections: single trait

Plant	4	10	53	55	59	72	82	93	96
Trait	RL20	HT34	TTF	HT34	HT34	HTF	TTB	HTF	HTF

Plant	99	101	110	133	141	164	178	190
Trait	RLF	HTF	RLF	RL20	RLF	HT34	HT34	HTF

(v) Low selections: two traits

Plant	28	37	38	39	40	52	56	71	76	106	165
Trait1	TTF	RL20	RL20	HTF	RL20	TTB	TTB	TTB	TTB	RL20	RL20
Trait2	HTF	RLF	RLF	HT34	RLF	TTF	TTF	TTF	TTF	RLF	HT34

(vi) Low selections: 3 or more traits

Plant	Traits
16	TTB, TTF, HTF
42	RL20, RLF, HT34
79	TTB, TTF, HTF
91	TTB, HTF, HT34
2	RLF, RL20, TTB, TTF
14	RL20, RLF, TTB, TTF, HT34

(vii) Common plants in the high and low selections

Plant	High selections	Low selections
10	HT34	HTF
37	RL20, RLF	HT34
38	RL20, RLF	TTB, TTF
59	HT34	TTB, TTF
164	HT34	RLF
178	HT34	TTB, TTF, RLF

Character-wise, the most common associations observed among the selections are between TTB and TTF (15 plants) and RL20 and RLF (13 plants) and HTF and HT34 (3 plants). These 3 pairs account of 31 of the 36 plants that have desirable scores for 2 or more traits.

Comparison between the variances of the high and low selections shows that these variances differ significantly on 4 occasions. The variance ratio for RL20 comes close to significance

as well, with a p value of around 0.10. However, these differences are not consistently unidirectional.

6.3.2 Response in the F_3 generation

The mean, variance and range of the F_3 families that have descended from the F_2 selections are listed in Table 6.4. Comparison of the means reveals that high selections have indeed resulted in raising the score compared to low selections and these means are different from each other for every trait. The variances of these selections also show a similar trend and the variance of the high selections is significantly larger than that of the low selections for RLF, TTB, TTF and HTF. This point clearly to the presence of scalar effects, i.e., larger means have larger variances.

Table 6.5a gives the observed and expected S values and other related statistics for the various traits. The observed values are calculated as mean of selections minus mean of F_2 or F_3 and expected values are obtained by following the procedures of Falconer and Mackay (1996). While the magnitude of observed S (OS) differs between the high and low scores there is no trend to these difference. In fact, the difference is positive on 3 occasions and negative on 3 occasions as well (Table 6.5a). As a proportion of the F_2 mean, these differences are substantial (10-23% of the mean), except for HTF (3%) and RL20 (6%).

While some differences can occur between these estimates due to sampling error, non-normality and dominance are also expected to make the differences larger. The positive and negative values may also indicate the direction of dominance for various traits but it cannot

be a very reliable indicator because factors like scalar effects and genotype environment interaction can produce similar effects as well.

The average values of observed response (OS), on the other hand, are very close to the expected response (S), suggesting that predictions provide a good indication of the S that would be obtained by the application of direct selection.

Table 6.5b compares the F_2 heritability with the realised heritability that has been calculated as observed R/Observed S (from Table 6.5a) for each trait. The realised heritability has two values, one for high selections and the other for low selections, and a comparison reveals that these values differ from each other for various traits. It is also apparent that in general the mean realised heritability has the same or similar value (0.38 and 0.42 for high and low scores) as the F_2 heritability (0.38). But, there are major differences between individual values. For instance, the range of realised heritability is 0.19-0.58 and 0.09-0.71 for high and low selections respectively while the range in the F_2 heritabilities is only 0.30-0.50. This shows clearly that genotype x environment interaction is playing some part in determining the phenotypic expression in *Arabidopsis thaliana.*

Tables 6.6 and 6.7 give the statistics related to the 10 highest and 10 lowest scoring F_3 families. Comparison of the overall means of these families with those in Table 6.4 reveal that the former are always more diverse than the latter. This indicates clearly the influences of the non-heritable variation on the selection process. This variation is high at the F_2 level and it seems that the highest or lowest scoring F_2 plants are not always the genotypes with the highest or the lowest potential scores. The F_3 means themselves will also be influenced

by the non-heritable variation to some extent, because the family means are based on only a small family size.

The number of the highest or the lowest scoring F_3 families that are present among the 10 high and 10 low F_2 selections are as follows.

Trait	High selections	Low selections
RL20	85	38
RLF	86, 164, 171, 178	38, 110
TTB	89, 171, 178	16, 76, 79
TTF	118, 171, 178	16, 76, 79
HTF	68, 103, 105, 176	72, 93, 96
HT34	37, 100	14, 164, 165

In summary, 23 lines listed among the high and low selections of F_2 were also present among the high and low selections of the F_3 families. This puts the relative efficiency of the F_2 selection at 33%, because 23 out of 69 selected F_2 plants have yielded the F_3 families with a desired score. Incidentally, 33% is also close to but slightly lower than the average heritability for the F_2 generation. This further reveals that the F_3 family means themselves are subjected to some non-heritable influences.

Results in Table 6.8 show that direct selection will be much more effective when applied to F_3 families compared to F_2 individuals. On average, the R-value is expected to double for the F_3 compared to F_2 selection. However, this will require an evaluation experiment that will incur some financial, manpower and time costs.

6.3.3 Marker-assisted selection: single QTL

In a population of individuals such as the F_2, MAS can be applied in two different forms. In the first case, only those individuals that possess the desired marker in homozygous form may be selected. For a single QTL controlled trait, this effectively reduces the sample by 75%. This form of selection, therefore, will need a large sample to start with, so that a reasonable number of plants are retained in the selected group.

The alternative to the above scenario is the selection for the desired allele, in the homozygous or the heterozygous form. This form of selection will retain 75% of the sample and therefore would not require a large sample at the start of the breeding programme. However, this will require more than one cycle of MAS and may give superior results under some situations, particularly when there are modifier genes involved in the control of a trait.

The results of MAS as applied to HT34, a trait controlled by a single QTL, are presented in Table 6.10. In this case, selection was done on the marker nearest to the QTL and then using flanking markers and various other criteria specified in the Materials and Methods section. The selection was also applied to homozygous individuals (marker profiles) only.

Table 6.10 shows that the number of F_2 individuals falling into each marker group is indeed very variable. This number varies from 9 to 52. Since the number of parental type genotypes among the homozygous F_2 plants should be equal (1:1) and because the heterozygous individuals are not selected, so, out of 200 individuals there should be 100 homozygotes. Further, half of these should be homozygous for one marker and the remainder for the other.

In many cases, the number of homozygotes is less than 100 due to missing data, as some individuals could not be scored at various marker positions. Also, in some cases the number of homozygous individuals is more than 50 due to sampling error, e.g. at a marker close to the QTL for HT34, the ratio is 52:34. However, none of these ratios differ significantly from 1:1 for any marker. The ratios of homozygotes vs. heterozygotes were also compared using chi-square test and deviation from 1:1 was observed in the two of the 5 markers in chromosome 2, with the excess of the heterozygotes (Appendix III).

Below is a brief summary of the marker profiles for HT34 with 1, 2, 3 and 4 representing the marker bands for Col, heterozygous, Ler and unscorable respectively.

	High selections					**Low selections**			
Line	**15$^{(6)}$**	**73$^{(4/5)}$**	**16$^{(1)}$**	**39$^{(3)}$**	Line	**15$^{(6)}$**	**73$^{(4/5)}$**	**16$^{(1)}$**	**39$^{(3)}$**
183	1	2	2	2	59	1	2	2	3
163	2	2	2	2	178	2	2	1	4
176	2	2	2	4	39	1	1	3	3
30	2	3	2	2	91	2	3	3	3
145	2	2	2	2	164	2	3	3	3
62	2	2	2	4	165	2	3	3	3
80	2	2	3	3	55	2	2	2	1
193	4	2	2	4	14	4	3	3	3
37	2	2	2	2	10	2	2	2	2
100	2	1	1	2	42	2	2	3	3

NB: The upper case numbers in brackets represent the marker positions used in selection, with flanking markers involving markers 3 and 4.

The above information allows the scrutiny of the marker profiles of various lines and markers in more detail. In summary, there are as many as 46 (58%) heterozygous among the selections, 28 (35%) homozygous, 4 (5%) unscorable lines and 8 (10%) lines were out of place in the markers used in the selections. The distribution of heterozygous among the high and low selections (30:16) indicates directional dominance has influenced the selection in the F_2 ($\chi^2_{(1)} = 4.26^*$). The efficiency based on the top ranks is much lower in the high

selections than low selections and this can be attributed to the high number of heterozygous that were not considered for selection. Also, the best F_2 plants may not be the best F_3 plants, since the F_2 may have a large environmental variance than the F_3 plants.

As expected, the selection efficiency seems to vary with the markers and sample size. Efficiency is clearly low when sample size is small. Figures in the last column in Table 6.10 are very interesting. These values are the ratio between the response based on MAS (in F_3 based on F_2 selections) and the response (R) obtained in the F_3 by applying phenotypic or direct selection at the F_2 level. This ratio is >1 in all the cases except when the marker is farthest away from the QTL. A value of >1 means MAS is giving better results than direct selection and therefore is more effective in identifying the desired genotypes. It is also apparent from the results that selection for any marker close to the QTL will give better results than direct selection, particularly when the trait is controlled by a single QTL. MAS is not expected to give better results when there is little correlation between the QTL and the marker, as it would be when the marker is located at the other arm of the chromosome.

6.3.4 Marker-assisted selection: more than one QTL

Selection for traits with more than one QTL was carried out for individual QTL first and then for all the QTL simultaneously. Two QTL were detected for HTF, three for RL20, RLF and TTF, whereas 5 QTL were detected for TTB. In all the cases, the QTL were detected on different chromosomes. The results of MAS for these traits are given in Tables 6.11-6.15 and the marker profiles of the F_2 selections are listed below.

(a) Selections for RL20 and RLF

	RL20				RLF		
	Chr2	Chr4	Chr5		Chr2	Chr4	Chr5
QTL effect	0.79	0.40	0.52		0.88	0.49	0.84
			High selections				
30	2	2	2	178	1	1	1
86	1	1	1	111	1	3	3
111	1	3	3	189	2	1	1
60	1	1	1	171	2	1	1
67	2	1	1	30	2	2	2
75	2	1	1	75	2	1	1
85	1	1	1	86	1	1	1
124	1	2	2	124	1	2	2
195	1	2	2	164	3	1	1
196	1	1	1	60	1	1	1
Desirable	1	1	1		1	1	1
			Low selections				
37	2	1	1	37	2	1	1
40	1	2	2	40	1	2	2
38	3	3	3	2	2	2	2
165	3	2	2	14	3	2	2
2	2	2	2	38	3	3	3
4	2	3	2	42	3	3	3
14	3	2	2	99	2	2	2
42	3	3	3	106	2	2	2
106	2	2	2	110	1	2	2
133	1	2	2	141	1	2	2
Desirable	3	3	3		3	3	3

(b) Selections for TTF and HTF

	TTF				HTF	
	Chr1	Chr2	Chr5	Line	Chr2	Chr5
QTL effect	-1.13	0.55	0.83		23.43	-7.40
			High selections			
59	4	2	2	89	1	1
135	4	2	1	78	1	2
178	3	1	1	103	2	3
189	2	2	1	150	1	2
47	1	2	1	157	2	2
171	3	2	1	68	2	3
38	2	3	3	105	2	3
118	4	2	1	129	2	3
130	2	2	1	176	2	3
153	3	2	2	10	2	2
Desirable	3	1	1		1	3
			Low selections			
2	3	2	2	28	3	2
14	1	3	2	91	3	1
16	1	3	2	16	3	1
28	2	3	1	39	3	2
52	3	3	2	72	3	2
53	3	2	1	93	3	2
56	2	3	2	101	3	2
71	1	3	3	79	3	2
76	2	2	2	96	3	2
79	2	3	2	190	3	2
Desirable	1	3	3		3	1

(c) Selections for TTB

	Chr1	Chr2	Chr3	Chr4	Chr5
QTL effect	-1.29	0.63	1.17	0.65	1.21
		High selections			
59	4	2	4	2	2
178	3	1	4	2	1
135	4	2	1	2	1
38	2	3	4	2	3
47	1	2	2	1	1
171	3	2	1	2	1
189	2	2	1	2	1
57	3	2	2	2	2
88	4	1	4	2	1
89	4	1	3	2	1
Desirable	3	1	1	1	1
		Low selections			
14	1	3	4	2	2
2	3	2	3	2	2
16	1	3	1	2	2
52	3	3	4	3	2
56	2	3	2	2	2
71	1	3	2	4	3
76	2	2	2	2	2
79	2	3	2	3	2
82	3	1	1	4	2
91	3	3	4	4	2
Desirable	1	3	3	3	3

The above results show that there are as many (142) heterozygotes among the selections as there are homozygotes (161). They are present in a 1:1 ratio and the chi-square is not significant ($\chi^2_{(1)}$=1.19ns). The distribution of heterozygotes among the high and low selections (62:80) also agree with 1:1 ratio ($\chi^2_{(1)}$=2.28ns) indicating that directional dominance has not influenced the selection of individuals in the F_2 by any significant margin. If there is any marginal effect, it does not seem to have gone with the dominance because there are comparatively more heterozygotes (80) in the low selections.

The number of undesirable homozygous markers (1 or 3) that are observed out of place among these selections is also low. There are 34 (=13+21) such marker scores, constituting approximately 11% of the total. This number falls within the margins of error as markers are spaced approximately 20cM apart and consequently sufficient recombination can occur between the markers and the QTL. In general, however, there is good association between the markers alleles and the phenotypic performance in the F_2 generation. For instance, excluding heterozygotes, correct markers are found among the selections on 127 out of 161 occasions.

In summary, therefore, out of a total of 320 marker points, 17 were not scored, 142 were heterozygous, 34 possessed wrong allele to the one expected and 127 had desirable marker alleles. Thus, assuming that heterozygotes are desirable, desired marker profiles were observed in 269 out of 303 cases, giving a marker/phenotypic association of 269/303 = 0.89, which is quite high considering that heritability in the F_2 is moderate, around 0.38 on average.

The results of MAS for traits controlled by more than one locus are more or less similar to those described for HT34, particularly when selection is carried out for individual QTL. In Table 6.11, the ratio R_{MAS}/R_{F3} always has a value >1, indicating that MAS is better than direct selection for each QTL. These results really show that MAS selection for any one of three QTL controlling RL20 will yield twice as much response as we could get by direct selection. These selections would also include many of the top 10 F_3 families (28 out of 60 cases), indicating that they would identify at least some of the best genotypes that can be extracted from the cross under study. MAS selection for all 3 QTL simultaneously also

provides better results than direct selection, but now the efficiency is reduced (for high score) because the sample size is reduced by a large margin. The efficiency based on top ranks was low in the low selections due to many heterozygous (17) observed compared to only 9 observed for the high selections (see marker profiles for RL20).

Results in Tables 6.12 and 6.13 are amenable to similar explanations as above, except that R_{MAS}/R_{F3} value is <1 on 6 out of 14 occasions. The MAS for low score in RLF, on the other hand, has a R_{MAS}/R_{F3} value >>1 on all 4 occasions and <1 on 2 occasions in the high selections. The efficiency based on number of top ranks in RLF is similar to those for RL20, since the same QTL were detected with similar additive effect. In HTF, all top ranks were detected in the low selections in LG2 and the QTL effect for the linkage group was more pronounced (Table 6.9). The results clearly show that MAS for multiple QTL, whether applied to individual or all the QTL simultaneously, is not consistently efficient.

Results for TTB and TTF, presented in Tables 6.14 and 6.15 respectively, differ from the rest in several aspects and the efficiency of MAS for these traits is rather low. For instance, the number of top families among the 10 selected by MAS is 3 or less on 16 out of 20 occasions and this number goes up to 7 only when selection is applied to markers on LG2 and LG5. Reference to Tables 6.9 and 6.16 shows that QTL effects are generally more pronounced for both TTB and TTF on LG5 but not on LG2. It is also apparent from these results that MAS for QTL on LG5 is more effective in identifying the phenotypically high scoring lines while the MAS for QTL on LG2 produces the 7 lowest scoring lines for TTB and TTF, respectively. The R_{MAS}/R_{F3} value is low in all the cases and it is greater than one in

7 out of 20 cases. Clearly, MAS is not very effective on those traits that are controlled by several genes and in selecting for all the QTL simultaneously in a trait.

The results for MAS selection based on the marker loci in the context of population improvement are presented in Table 6.16. It can be seen from the table that various QTL account for between 5% (TTF) to 63% (HTF) of the additive variance for the traits. Large differences are also observed for the selection index (I), whose values ranges from 0.58 to 2.19, relative efficiency RI (1.05-2.33) and efficiency of selection ES (0.14-1.62). The heritability of the F_2 plants on the other hand display less variation (0.30-0.50). The selection index is high when the heritability is low (e.g. TTB, LG2 and 4). The values of ES depend on both the heritability and the additive variance of the trait, and also I and RI. In TTB (LG2 and 4) the lower additive effect resulted in lower efficiency of selection (0.20). The efficiency of selection based on the marker loci was greater than 1 when the additive effect was more than the heritability (e.g. HTF, LG2). The efficiency based on top ranks also identified all the top 10 lines based on the phenotype in HTF in the low selections. Other traits showing high efficiency of selection based on the marker loci also showed high efficiency of MAS such as HT34, RL20 and RLF in LG2 and RLF in LG5.

6.4 Discussion and Conclusions

One of the most important applications of molecular markers in practical breeding is considered to be in the improvement of selection efficiency through MAS. Many studies have been conducted since the late 1980's in order to show how useful MAS would be and how it can transform the process of crop improvement (Lande and Thompson 1990; Gimelfarb and Lande, 1994; Xie and Xu, 1998; Lange and Whittaker, 2001). While most of these studies have used computer simulations (Zhang & Smith, 1992; Ollivier, 1998; Luo et al., 1997) and explained procedures of selection and developed methods of assessing comparative advantage of selecting for QTL of different effect, rarely is it shown how to implement MAS in practice and what problem one will face when carrying out such a task. The present study was therefore conducted to see if MAS can be applied effectively at the experimental and field levels and what level of advantage it accrues to the breeder in terms of improving the efficiency of selection viz. a phenotypic or direct selection.

To make MAS effective, the breeder must integrate it into a standard inbreeding programme so that the maximum savings in time/space/resources can be obtained. So, in an inbreeding programme MAS must be applied at the F_2/F_3 level because after F_3 the level of recombination is reduced and very large samples would have to be raised for the application of selection. Further, only one or two cycles of MAS can be envisaged in practice because intense selection in the F_2 and F_3 generations reduces genetic diversity very quickly, particularly if the procedure of pedigree inbreeding were to be followed (Poehlman and Sleper, 1995).

In the present study, phenotypic or direct selection was applied initially in order to provide a comparison for MAS. A sample of 200 F_2 individuals was used in this selection and the same individuals were used to obtain F_3 families and for molecular marker evaluation. While there can be arguments for increasing this sample to 300 or 400 individuals, particularly in the present case when a large amount of marker information is already available and it will cost much less to increase the sample, it was however felt that a sample of 200 will be sufficient for practical purpose, including QTL location and selection, even when starting with a new or unknown cross. The phenotypic selection has clearly resulted in significant improvements in shifting the F_3 mean in the desired direction and the rate of change was within the margins of error of the predicted levels. This exercise has also confirmed that the magnitude of realised response depends very much on the heritability and the level of phenotypic variation observed for the trait. Furthermore, this selection will be much more effective if applied to F_3 families, confirming once again that the selection of F_2 individuals on a phenotypic basis can not be very effective under any circumstances, except when there is only one gene and there is no non-heritable variation.

The application of MAS has yielded mixed results. For traits that are controlled by a single QTL, MAS has been rather effective, particularly when the QTL accounted for a larger proportion of heritable variation. For traits with 2 QTL, the situation is more or less similar. But the major problem arises when 3 or more QTL are controlling the trait because not only now each QTL then accounts for a smaller proportion of additive variance but selection has to be carried out for several markers simultaneously. Simultaneous selection for several QTL makes the selected sample rather small and consequently there is little opportunity to select for those QTL that remain undetected because their effects are small.

The present and previous studies (on Brassica) on MAS have also shown clearly that for maximum efficiency MAS had to be combined with phenotypic selection. Application of MAS on its own would not yield the best results because selection for any marker or combinations of markers will yield several genotypes that will differ in their genetic potential because all variation is rarely accounted for by the QTL detected. This suggests that genes with small effect cannot be ignored while applying selection because they may in the end determine the success or failure of a breeding programme.

It is also essential that marker located close to each QTL is used in MAS when applied to F_2 and F_3 generations. This is because each F_2 individual is a true recombinant possessing maternal and paternal chromosomes and therefore will be homozygous under very rare conditions. Therefore, markers located at a distance from the QTL would nearly always show recombination and heterozygosity and therefore would be less reliable for selection. MAS for doubled haploids, on the other hand, was found to be effective even for distant markers because there was always a 20% chance that the selected chromosome may in fact be an intact, non-recombinant chromosome and thus carry the desired QTL (Koumploglou et al., 2002).

Finally, when dealing with many QTL, MAS need only be carried out for those genes that have large effects and it can be applied in two cycles. Initially, heterozygotes can be included in the selection so that a large sample can be kept for the second cycle of selection. It would then be comparatively easy to choose the homozygotes in the next cycle because sample size is now small and more manageable in the laboratory and in the field.

Table 6.1: The overall mean, variance and heritability of the F_2 and F_3 generations

Trait	F_2			F_3		
	Pop mean	Variance	h_b^2	Pop mean	Variance	H_f^2
RL20	8.60	1.80	0.50	9.78	1.15	0.68
RLF	8.84	2.34	0.35	10.65	2.26	0.81
TTB	19.18	6.10	0.30	18.40	2.96	0.83
TTF	22.88	5.29	0.37	23.05	2.71	0.83
HTF	93.80	920.52	0.39	92.46	695.31	0.77
HT34	342.88	3863.87	0.36	395.36	4395.60	0.72

Table 6.2: The mean, variance and range of the 10 highest scoring F_2 plants

Trait	Mean	Variance	Range	Magnitude of range	Selected plants
RL20	11.30	0.23	12-11	1	30, 86, 111, 60, 67, 75, 85, 124, 195, 196
RLF	12.80	2.18	16-11	5	178, 111, 189, 171, 30, 75, 86, 124, 164, 60
TTB	25.50	2.94	29-24	5	59, 178, 135, 38, 47, 171, 189, 57, 88, 89
TTF	29.50	2.94	33-28	5	59, 135, 178, 189, 47, 171, 38, 118, 130, 153
HTF	154.00	104.44	180-140	40	89, 78, 103, 150, 157, 68, 105, 129, 176, 10
HT34	423.50	133.61	440-410	30	183, 163, 176, 30, 145, 62, 80, 193, 37, 100

Table 6.3: The mean, variance and range of the 10 lowest scoring F_2 plants

Trait	Mean	Variance	Range	Magnitude of range	$F_{(9,9df)}$	Selected plants
RL20	5.40	0.71	4-7	3	3.09ns	37, 40, 38, 165, 2, 4, 14, 42, 106, 133
RLF	5.80	0.18	5-6	1	12.11***	37, 40, 2, 14, 38, 42, 99, 106, 110, 141
TTB	15.90	0.10	15-16	1	294.0***	14, 2, 16, 52, 56, 71, 76, 79, 82, 91
TTF	20.00	0.00	20-20	0	0.00ns	2, 14, 16, 28, 52, 53, 56, 71, 76, 79
HTF	30.50	13.61	25-35	10	7.67***	28, 91, 16, 39, 72, 93, 101, 79, 96, 190
HT34	185.00	544.44	135-210	75	4.07**	59, 178, 39, 91, 164, 165, 55, 14, 10, 42

Variance ratio = highest/lowest variance

Table 6.4: The mean, variance and range among the F_3 families that have descended from the 10 highest and 10 lowest scoring F_2 selections

Trait	High selections			Low selections		
	Mean	Variance	Range (magnitude)	Mean	Variance	Range (magnitude)
RL20	10.30	0.80	11.8-8.6(3.2)	9.48	1.05	7.8-10.8(3)
RLF	12.96	2.70	16-11.2(4.8)	10.18	2.40	8-12.8(4.8)
TTB	20.69	5.16	24.6-16.8(7.8)	16.72	1.63	15-18.8(3.8)
TTF	25.63	5.36	29.8-21.6(8.2)	21.44	1.13	20.4-23.8(3.4)
HTF	121.50	546.28	146-74(72)	47.40	97.60	31-63(32)
HT34	418.40	2027.60	487-352(135)	319.30	8586.57	150-449(299)

Table 6.5a: The observed and expected response (S) obtained from F_2 data and the observed and expected genetic advance (R) among the F_3 families based on F_2 selections for various traits

Trait	Highest scores		Lowest scores		Expected $S(F_2)$	Expected $R(F_2)$	Difference between OS (% of μ)
	$OS(F_2)$	$OR(F_3)$	$OS(F_2)$	$OR(F_3)$			
RL20	2.70	0.52	-3.20	-0.30	2.76	1.38	-0.50 (6)
RFL	3.96	2.31	-3.04	-0.47	3.16	1.10	0.92 (10)
TTB	6.32	2.29	-3.28	-1.68	5.12	1.53	3.04 (16)
TTF	6.62	2.58	-2.88	-1.61	4.74	1.76	3.74 (16)
HTF	60.20	29.04	-63.30	-45.06	62.59	24.41	-3.10 (3)
HT34	80.62	23.04	-157.88	-76.06	128.24	46.16	-77.26 (23)

NB: Expected S and R calculated using formulas of Falconer and Mackay (1996)

Table 6.5b. Realised versus observed heritability in the F_2 generation

Trait	Realised heritability		Observed heritability
	High selections	Low selections	
RL20	0.19	0.09	0.50
RLF	0.58	0.15	0.35
TTB	0.36	0.51	0.30
TTF	0.39	0.56	0.37
HTF	0.48	0.71	0.39
HT34	0.29	0.48	0.36

NB: Realised heritability = Observed $R(F_3)$/Observed $S(F_2)$

Table 6.6. The mean, variance and range of the 10 highest scoring F_3 families

Trait	Mean	Variance	Range (magnitude)	The highest scoring families
RL20	11.86	0.04	12.2-11.6(0.6)	87, 6, 13, 185, 43, 64, 85, 94, 55, 66
RLF	14.20	0.83	16-13 (3)	171, 89,86, 178, 67,118, 164, 123, 10, 13
TTB	22.94	1.16	24.8-21.8 (3)	164, 171,118, 123,89,178,155,86,161, 67
TTF	27.71	1.42	29.8-26.3(3.5)	171, 164,118, 89,123,86,178,155,67, 153
HTF	143.30	34.46	158-137 (21)	64, 176, 74, 160, 82, 105, 9, 68, 103, 20
HT34	484.75	52.40	499-474 (25)	61, 45, 95, 100, 6, 166, 68, 37, 87, 81

Table 6.7. The mean, variance and range of the 10 lowest scoring F_3 families

Trait	Mean	Variance	Range (magnitude)	The lowest scoring families
RL20	7.72	0.05	7.4-8.0(0.6)	15, 103, 16, 79, 38, 96, 123, 154, 41, 131
RLF	7.94	0.12	7.4-8.4 (1)	15, 16, 103, 96, 41, 79, 110, 38, 25, 28
TTB	15.50	0.08	15-15.8 (0.8)	79, 15, 16, 103, 34, 76, 128, 25, 41, 96
TTF	20.55	0.03	20.4-20.8(0.4)	25, 34, 41, 79, 166, 28, 76, 103, 15, 16
HTF	38.00	33.78	28-45 (17)	84, 72, 91, 164, 14, 168, 17, 96, 194, 93
HT34	240.90	1149.12	150-263 (113)	164, 72, 194, 84, 17, 67, 14, 93, 165, 11

Table 6.8: The observed response (OS) and the genetic advance (R) that is possible when direct selection is applied to the F_3 families

Trait	High scores			Low scores		
	OS	Exp'ted R	Efficiency	OS	Exp'ted R	Efficiency
RL20	2.08	1.41	1.02	-2.06	-1.40	1.01
RLF	3.55	2.88	2.62	-2.71	-2.20	2.00
TTB	4.54	3.77	2.46	-2.90	-2.41	1.58
TTF	4.66	3.87	2.20	-2.50	-2,08	1.18
HTF	50.84	39.15	1.60	-54.46	-41.93	1.72
HT34	89.39	64.36	1.39	-154.46	-111.21	2.40

Efficiency = Expected R(F_3)/ Expected R(F_2 from table 6.5a)

Table 6.9: QTL detected in F_2 and F_3 using the marker regression and interval mapping procedures respectively

Trait	Chr.	Interval mapping				Marker regression		
		QTL position	LOD score	A	CI	QTL position	A	CI
RL20	2	54	37.66	0.71	9	56	0.79	19
	4	37	21.33	0.71	4	34	0.40	30
	5	39	17.32	0.57	12	38	0.52	17
TTB	1	34	12.26	-0.70	10	38	-1.29	
	2	40	8.57	0.61	22	48	0.63	21
	3	48	10.06	-0.44	2	60	1.17	
	4	27	28.16	0.91	4	8	0.65	40
	5	49	29.15	0.96	16	42	1.21	15
TTF	1	34	16.52	-0.92	8	34	-1.13	
	2	38	7.35	0.53	15	50	0.55	23
	5	44	23.87	0.86	21	42	0.83	20
HTF	2	50	117.11	27.49	2	52	23.43	4
	5	88	11.33	-14.66	16	74	-7.40	33
RLF	2	50	34.32	0.94	7	54	0.88	8
	4	27	47.55	1.03	4	26	0.49	20
	5	39	47.15	1.24	7	40	0.84	10
HT34	2	50	128.75	71.89	2	52	39.07	5

Table 6.10: The mean and variance of the 10 high and 10 low selections obtained using MAS, the total number of homozygotes in each marker group, the efficiency of selection based on top scoring individuals/families among the 10 selections and the ratio R_{MAS}/R_{F3} for the trait HT34. The QTL controlling this trait is located in linkage group 2 at 52cM

Marker (Position)	F_2 mean	F_2 variance	F_3 mean	F_3 variance	No. of lines	EOS top ranks	R_{MAS}/R_{F3}
High selections							
1(50.65)	398.50	44.72	457.30	953.79	52	1	2.67
2(35.04&63.02)	383.00	145.57	451.20	267.29	21	0	2.42
3(63.02)	392.00	45.56	444.55	775.91	38	0	2.13
4(35.04)	397.00	45.56	448.00	1267.11	43	1	2.28
6(9.60)	383.00	595.56	415.30	2866.01	18	1	0.87
Low selections							
1(50.65)	206.00	510.00	256.20	1926.40	34	6	1.83
2(35.04&63.02)	229.50	1524.72	266.80	2255.73	21	4	1.69
3(63.02)	197.00	968.06	277.70	5488.23	36	7	1.55
4(35.04)	224.50	1324.72	262.50	2343.17	37	4	1.75
6(9.60)	362.78	338.19	421.17	3307.00	9	0	0.34

NB: The next marker on the other side of the QTL (3) was detected close to the centromere (5), so centromere information the same as for the marker

Table 6.11: The mean and variance of the 10 high and 10 low selections obtained using MAS, the total number of homozygotes in each marker group, the efficiency of selection based on top scoring individuals/families among the 10 selections and the ratio R_{MAS}/R_{F3} for the trait RL20. The QTL controlling this trait are located in linkage groups 2, 4 and 5

Linkage group (Position)	F_2 mean	F_2 variance	F_3 mean	F_3 variance	No. of lines	EOS top ranks	R_{MAS}/R_{F3}
High selections							
LG2(56)	10.90	0.54	10.86	0.79	52	7	2.08
LG4(34)	10.70	0.46	10.78	0.71	59	6	1.92
LG5(38)	10.70	0.46	10.78	0.71	61	6	1.92
All	10.50	0.50	10.10	1.12	15	4	1.38
Low selections							
LG2(56)	6.40	0.71	8.38	0.44	34	4	4.67
LG4(34)	7.00	1.11	9.14	1.28	30	3	2.13
LG5(38)	7.20	1.07	8.86	1.03	27	2	3.07
All	6.60	1.30	8.36	0.21	5	1	4.73

Table 6.12: The mean and variance of the 10 high and 10 low selections obtained using MAS, the total number of homozygotes in each marker group, the efficiency of selection based on top scoring individuals/families among the 10 selections and the ratio R_{MAS}/R_{F3} for the trait RLF. The QTL controlling this trait are located in linkage groups 2, 4 and 5

Linkage group (Position)	F_2 mean	F_2 var.	F_3 mean	F_3 var.	No. of lines	EOS top ranks	R_{MAS}/R_{F3}
High selections							
LG2(54)	12.00	2.89	12.44	1.41	52	5	0.77
LG4(26)	12.30	2.68	13.22	2.52	59	7	1.11
LG5(40)	12.30	2.68	13.22	2.52	61	7	1.11
All	11.30	3.12	12.74	2.35	15	5	0.28
Low selections							
LG2(54)	6.60	0.27	8.82	0.96	34	3	3.89
LG4(26)	7.50	0.72	9.40	1.71	30	2	2.66
LG5(40)	7.50	0.72	9.30	1.36	27	2	2.87
All	7.40	2.80	8.56	0.11	5	1	4.45

Table 6.13: The mean and variance of the 10 high and 10 low selections obtained using MAS, the total number of homozygotes in each marker group, the efficiency of selection based on top scoring individuals/families among the 10 selections and the ratio R_{MAS}/R_{F3} for the trait HTF. The QTL controlling this trait are located in linkage groups 2 and 5

Linkage group (Position)	F_2 mean	F_2 var.	F_3 mean	F_3 var.	No. of lines	EOS top ranks	R_{MAS}/R_{F3}
High selections							
LG2(52)	140.50	324.72	111.20	273.07	52	3	0.65
LG5(74)	138.50	183.61	122.65	624.45	26	5	1.04
Both	95.00	1366.7	94.70	285.57	11	1	0.07
Low selections							
LG2(52)	30.50	13.61	47.40	97.60	34	10	1.00
LG5(74)	46.50	144.72	64.40	670.71	49	2	0.62
Both	50.00	50.00	53.20	140.20	5	0	0.87

Table 6.14: The mean and variance of the 10 high and 10 low selections obtained using MAS, the total number of homozygotes in each marker group, the efficiency of selection based on top scoring individuals/families among the 10 selections and the ratio R_{MAS}/R_{F3} for the trait TTB. The QTL controlling this trait are located in linkage groups 1, 2, 3, 4 and 5

Linkage group (Position)	F_2 mean	F_2 var.	F_3 mean	F_3 var.	No. of lines	EOS top ranks	R_{MAS}/R_{F3}
High selections							
LG1(38)	23.60	3.38	20.54	4.32	38	3	0.93
LG2(48)	23.40	3.16	19.86	4.15	52	3	0.64
LG3(60)	23.60	2.71	20.44	2.62	33	3	0.89
LG4(8)	22.30	1.79	18.97	1.85	50	1	0.25
LG5(42)	24.90	1.66	21.85	4.63	61	7	1.51
All	16.00	0.00	16.80	0.00	1	0	-
Low selections							
LG1(38)	16.30	0.40	16.92	2.52	57	3	0.88
LG2(48)	15.90	0.10	16.44	0.86	34	7	1.17
LG3(60)	16.80	0.18	17.62	0.99	47	1	0.46
LG4(8)	16.80	0.40	17.29	1.35	30	2	0.66
LG5(42)	17.00	0.44	16.50	0.94	27	1	1.13
All	-	-	-	-	-	-	

Table 6.15: The mean and variance of the 10 high and 10 low selections obtained using MAS, the total number of homozygotes in each marker group, the efficiency of selection based on top scoring individuals/families among the 10 selections and the ratio R_{MAS}/R_{F3} for the trait TTF. The QTL controlling this trait are located in linkage groups 1, 2 and 5

Linkage group (Position)	F_2 mean	F_2 var.	F_3 mean	F_3 var.	No. of lines	EOS top ranks	R_{MAS}/R_{F3}
High selections							
LG1(34)	26.70	4.32	25.12	5.24	38	3	0.80
LG2(50)	26.50	2.94	24.08	3.94	52	1	0.40
LG5(42)	28.70	2.68	26.43	5.35	61	7	1.31
Both	22.00	0.67	22.93	1.06	4	0	-0.04
Low selections							
LG1(34)	20.50	0.28	21.32	0.56	57	3	1.07
LG2(50)	20.10	0.10	21.24	0.41	34	7	1.12
LG5(42)	21.10	0.32	20.96	0.37	27	1	1.30
All	-	-	-	-	-	-	

Table 6.16: The QTL position, additive effect (a), percentage additive variance (A), heritability (h_b^2) of the trait, selection index (I), relative efficiency (RI) and the efficiency of selection (ES) based on the marker locus for QTL in Arabidopsis F_2 lines

Trait	LG	QTL position	Nearest marker	a	A	h_b^2	I	RI	ES
RL20	2	56.0	Nga1126	0.79	0.34	0.50	0.66	1.34	0.68
	4	34.0	Nga8	0.40	0.09	0.50	0.91	1.06	0.18
	5	38.0	Nga139	0.52	0.15	0.50	0.85	1.11	0.30
HTF	2	52.0	Nga1126	23.43	0.63	0.39	0.58	2.33	1.62
	5	74.0	Mjb21a	-7.40	0.06	0.39	1.47	1.06	0.15
TTF	1	34.0	Nga392	-1.13	0.21	0.37	1.96	1.05	0.57
	2	50.0	Nga1126	0.55	0.05	0.37	1.62	1.06	0.14
	5	42.0	Nga139	0.83	0.11	0.37	1.51	1.13	0.29
HT34	2	52.0	Nga1126	39.07	0.33	0.36	1.19	1.55	0.92
RLF	2	54.0	Nga1126	0.88	0.32	0.35	1.26	1.55	0.91
	4	26.0	Nga8	0.49	0.10	0.35	1.67	1.14	0.29
	5	40.0	Nga139	0.84	0.30	0.35	1.30	1.51	0.86
TTB	1	38.0	Nga392	-1.29	0.25	0.30	1.75	1.51	0.83
	2	48.0	Nga1126	0.63	0.06	0.30	2.19	1.10	0.20
	3	60.0	Th620b	1.17	0.21	0.30	1.84	1.41	0.70
	4	8.0	T18a10	0.65	0.06	0.30	2.19	1.10	0.20
	5	42.0	Nga139	1.21	0.21	0.30	1.84	1.41	0.70

CHAPTER 7

GENERAL DISCUSSION AND CONCLUSIONS

7.1 QTL mapping in *Brassica oleracea*

In *Brassica oleracea*, 89 DH lines were used to investigate QTL affecting vegetative and flowering traits. A total of 40 QTL were identified in 8 of the 9 linkage groups in *Brassica oleracea*. The individual QTL explained between 2 and 49% of the additive genetic variation. Kearsey & Farquhar (1998) observed that individual QTL might explain 1-50% of the phenotypic variation, which is in agreement with the present study. Overall, the largest number of QTL detected had relatively small effect and QTL with large effect were relatively few. This is consistent with findings of other studies documenting that most differences between lines are due to a small number of QTL of large effect accompanied by a large number of smaller effects (Tanksley, 1993; Lynch and Walsh, 1998).

QTL controlling different traits were frequently localized to the same genomic regions. For example, plant height1, flowering height, maximum height, apical height and petiole length2 were associated with common marker loci on linkage group 1. Associations of the same marker loci (and QTL) with several traits, e.g. number of leaves, leaf length1, leaf width1, leaf length2 and petiole length1, which map to linkage group 8, is also supported by Kennard et al. (1994) who found similar marker-trait associations for lamina width, lamina length and petiole length. QTL affecting lamina width have also been mapped to linkage group 8 by Sebastian et al., 2002. The QTL in DH lines showed pleiotropic effects, particularly for traits that were measured at different stages of plants' life cycle (e.g. plant heights and leaf traits). This clearly shows that the 40 QTL detected in this study may not all be separate loci, thus the actual number of QTL

(detected) for which A12 and GD differ would well be as low as 20 or less. The results of QTL mapping in the present study were compared to those of Bohoun et al. (1998) and Rae et al (1999), because both studies are based on the same set of DH lines. In this study, flowering time QTL were only detected on linkage groups 2 and 3. The QTL in LG2 was detected by Bohuon et al. (1998) and Rae et al. (1999), whereas that detected in LG3 was also detected by Bohuon et al (1998), Rae et al. (1999) and Rae (2000).

7.2 QTL mapping in *Arabidopsis thaliana*

The marker regression approach revealed 23 QTL affecting six morphological traits in the Col x Ler F_2 population. In the F_3 families, 17 of 23 QTL in the F_2 population were detected in the six morphological traits using the interval mapping method. The QTL were detected at similar positions and showed the same mode of action. The replicated F_3 descendants provided a better measure of the genotypic effects of the F_2 plants. Most QTL associated with TTB were also associated with TTF. This is expected, since many genes that affect time to bud formation also would affect time to flowering. The other traits measured at flowering mapping close to the flowering time genes are RLF and HTF in chromosomes 2 and 5. Previous studies have noted a significant positive correlation between leaf number at flowering and the time of flowering (Clarke et al. 1995). In this study, rosette leaves at flowering was correlated with time to flower. Repeat measurements such as height and number of rosette leaves were correlated and mapped to similar regions of the genome.

The QTL explaining the most variance mapped to the same region on chromosome 2 at around 50cM in traits RL20, CL20, HTF, RLF, HT34, TTB and TTF. This QTL was detected in both the F_2 and F_3 analysis. This QTL is likely to be the *erecta* mutation, which affects inflorescence

architecture (Ungerer et al. 2002). The *erecta* mutation is not a naturally occurring mutation, however, but was generated in the laboratory through mutagenesis. The results shown in this study may be useful to plant breeders and geneticists alike. The QTL identified in Arabidopsis may provide a basis for identifying these QTL in other members of the Cruciferae family, or even introducing these QTL into other plant species (Burns, 1997). The similarities between the genomes of Arabidopsis and Brassica species reveal that the estimated positions of QTL in Arabidopsis correspond well with QTL on the Brassica genome (Kole et al. 2001, Tadege et al. 2001). Cloning the QTL detected in Arabidopsis, and introducing them into commercial Brassica species, may increase agricultural productivity. Also, the QTL detected can be selected for by using molecular markers for further evaluation and improvement, as was done for the Brassica and Arabidopisis in this study.

7.3 Precision of QTL mapping

The 95% confidence intervals for the mapped QTL in Brassica DH lines ranged between 14-60cM, whereas in Arabidopsis it ranged between 2-42cM in F_2 plants and 4-40cM in F_3 families. Usually a confidence interval of up to 30cM has been observed for the segregating populations. These confidence limits are very large and the reliability may be increased by increasing family size and the number of lines in the study (van Ooijen 1992; Darvasi et al. 1993; Kearsey & Farquhar, 1998). In the DH lines, only 89 lines were used which is a small family size. van Ooijen (1992) observed that a minimum population of 200 backcross and F_2 individuals was necessary for detecting QTLs that explained at least 5% of the total variance for a trait.

Precision also means more QTL detected per chromosome or trait. The efficiency of QTL detection calculated as the ratio of the number of QTL detected and the product of the number

of chromosomes and number of traits in which QTL were detected was similar in the F_2 (0.77) and F_3 (0.73) generations in Arabidopsis. In the DH lines the efficiency was low (0.30). A total of 40 QTL were detected for the DH lines. The DH lines represented a wide cross so more QTL were expected. In Arabidopsis the ecotypes Columbia and Landsberg are very similar in many aspects and less variation is expected. However, 40 QTL were detected by the interval mapping method in 11 traits whereas the marker regression method detected 23 QTL in six traits. The efficiency of QTL detection was however, the same in Arabidopsis.

7.4 QTL mapping methods

The QTL mapping methods have evolved from using marker analysis such as t-test, simple or multiple regression, to one-QTL model (interval mapping and composite interval mapping), and further to the multiple-QTL model such as the multiple interval mapping (MIM). QTL have to be mapped as precisely as possible to ensure good quality of the follow-up operation on QTL. QTL mapping methods should look at problems such as effects of multiple QTL, bias, and the precision of the QTL position. These problems were particularly evident in earlier mapping methods, and the new methods are addressing the problems to increase the accuracy and reliability of QTL mapping.

Therefore, precision and unbiasedness in estimating the parameters of QTL should be more important than the ease of computation and implementation in QTL mapping (Kao et al. 1999). The methods of interval mapping and marker regression approach follow the procedure of creating a QTL model for the observed data and then testing that model for its suitability. In both cases the models are relatively simple and consider only one or two QTL thus giving a limited number of possibilities to be considered. The two methods yield similar results even

though different significance tests are applied. Investigations have shown that QTL mapping methods yield comparable estimates of the QTL position and their accuracy is generally poor unless very large populations are used and the heritability is high (Hyne and Kearsey, 1995).

The interval mapping and marker regression methods do not include the analysis of other parameters such as epistasis and genotype x environmental interaction. When epistasis and genotype by environmental interactions are included the range of possible models to be tested becomes much larger and the process of model selection and testing becomes extremely demanding. Methods for dealing with such complexity is an area of active research and methods need to be developed to allow the detection of epistatic QTL that might normally pass undetected (Doerge, 2002).

7.5 Marker-assisted selection (MAS)

Quantitative traits such as seed yield and plant weight are highly complex and controlled by many genes whose effects are also influenced by random and non-random environmental factors. These traits clearly show partial inheritance and it is not possible to trace each and every gene that is segregating for them. Further, genes also have unequal effects and those with large effect are termed as major genes while others are called modifiers (Kearsey and Pooni, 1996). The phenotypic selection of such traits is often less effective, particularly when applied to individuals. As many breeding programs have faltered due to this inefficiency of selection, scientists have suggested the use of MAS to improve the chances of identifying and isolating desired genotypes, at least for those quantitative traits that have low heritability or cannot be selected directly (e.g. milk yield in bulls and egg production in chickens). The argument is based on the fact that molecular/DNA markers have 100% heritability and therefore are more

amenable to selection than say a QTL that may be located close by. While most of these studies have used computer simulations (Zhang & Smith, 1992; Ollivier, 1998; Luo et al., 1997) and explained procedures of selection and developed methods of assessing comparative advantage of selecting for QTL of different effect, rarely it is shown how to implement MAS in practice and what problem one will face when carrying out such a task.

Selection based on the phenotype and using markers was carried out in *Brassica oleracea* DH lines and *Arabidosis thaliana* F_2 plants/F_3 families. The efficiency of MAS was measured using different approaches based on the top ranks, number of lines in a marker group and phenotypic value in Brassica. In Arabidopsis the efficiency was measured based on the top ranks and as the ratio between response based on MAS and response obtained in the F_3 by applying phenotypic or direct selection to the F_2 generation.

For traits with single QTL, marker closest to the QTL did not seem to give better results than the next marker nearer the QTL or flanking markers. So, any marker that has complete set of data can be used for selection without loosing much accuracy. However, if the numbers mean anything, then the worst results are obtained using flanking markers. Missing data for the flanking markers had the biggest impact in this case because now there were twice as many chances of information being incomplete than otherwise. It is also important to suggest that one does not need more than 3 or 4 markers to implement MAS effectively, provided the markers have sufficient information (Bouchez et al., 2002). Any more markers are not likely to improve its efficiency by any significant margin, unless the chromosomes are exceptionally long and there are recombination hot spots nearer to the QTL under selection.

The selection for individual QTL was better than selection for two or more QTL simultaneously on the same or different chromosomes in a trait. The basic problem of multiple marker selection is to have a large sample of lines after the selection and this is possible only if very large populations are screened (Lande and Thompson, 1990; Edwards and Page, 1994; Gimelfarb and Lande, 1994a). The efficiency of selection based on the phenotypic value (means) and top ranks provided a better method to measure the efficiency achieved by MAS, since the efficiency was high or maximum when most or all the top ranks were selected in Brassica. In Arabidopsis the efficiency of selection based on the ratio R_{MAS}/R_{F3} was better than top ranks, since the efficiency on top ranks was low when there were missing or more heterozygotes in the population, because only homozygous genotypes were selected. Scientists have argued that MAS will give better results compared to direct or phenotypic selection, particularly when heritability is low (Whittaker et al., 1995; Moreau et al., 1998; Knapp, 2001; Lange and Whittaker, 2001). The present results confirm this assertion conclusively as in all the cases at least some of the best lines could be identified using their marker profiles even when the heritability was low.

The study has shown clearly that for maximum efficiency MAS had to be combined with phenotypic selection. Application of MAS on its own would not yield the best results because selection for any marker or combinations of markers will yield several genotypes that will differ in their genetic potential because all variation is rarely accounted for by the QTL detected. This suggests that genes with small effect cannot be ignored while applying selection because they may in the end determine the success or failure of a breeding program.

The costs of MAS in breeding programs also need to be documented to assess if the benefit exceeds the costs of implementation. In this study, it cost around £6,002.05 to grow and genotype 200 *Arabidopsis thaliana* plants using 30 microsatellite markers, excluding labour and space charges (Appendix III). This shows that to run PCR experiments using selected markers is quite expensive and takes time. However, MAS may not reduce the time to get inbreds but will cut down the costs of phenotypic evaluation. How big this reduction in experimental cost will depends on the objectives. Were we to improve only one or two traits controlled by single QTL, then the reduction can be big. This will lead to very substantial savings in the cost of further breeding and experimentation. If, however, we were to improve complex traits like yield and quality simultaneously and there are not any QTL with large effect, then a large sample will need to be retained so that good recombinants can be found among the selected DH or F_3 which already have some QTL fixed in them.

The cost of marker profiling would also be rather low because, based on the present study, no more than 3-4 markers will be needed per chromosome to implement MAS. In addition, genotyping can be staggered to reduce the lab costs even further, i.e. select using one marker first and then select for the second and third marker, etc. Finally, one cycle of MAS need to be applied when breeding inbred lines. Any more cycles would help only when selection is being applied in stages, i.e. linkage group 1 and trait or linkage group 2. Multiple cycles of MAS however would be more effective in population improvement, e.g. maize or cattle or sheep populations.

7.6 Further implications in plant breeding

The application of molecular genetics in breeding is currently constrained by the precision of the allele effects associated with markers, and the efficiency cum cost effectiveness of MAS. Marker-based selection is definitely useful to manipulate chromosome regions (QTL) and rapidly design new genotypes combining favorable regions. The clear example is MAS in backcross breeding schemes for the introgression of one or a few target genes in a given genetic background. Molecular markers can be used to assess the presence of the introgressed region or accelerate the return to the recipient parent genotype at other loci. Also, marker information can be used to facilitate selection when developing inbred lines from the F_2 in either self-pollinated or cross-pollinated crops. Phenotypic selection in plant breeding is limited by the ability to estimate genetic parameters for the traits of interest, using statistical analysis of phenotypic data. The use of marker information relaxes some of the constraints of quantitative selection, and provides better estimates of breeding values.

Costs, both financial and manpower related are the other key determinants for the application of molecular genetics in breeding programs. Clearly, MAS is efficient and valuable for simple traits and/or traits for which increase of genetic gain per unit of time is of high economic return. However, the advent of MAS for the ordinary breeding of complex traits (e.g. yield) relies on a re-thinking of breeding strategies, and on the availability of both statistical and molecular techniques that would provide precise estimates of gene effects in selected populations at low cost. In general, however, MAS will give the best results when compared with phenotypic selection (as the present study has shown) because this doubles the surety of success.

APPENDIX I

A. ANOVA results for 15 traits measured in DH lines

1. Analysis of Variance for Plant height1(PH1)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	86497.94	86497.94	982.93	10.02	0.000
Block	1	322.52	322.52	322.52	4.37	0.010
Line*Block	88	8631.65	8631.65	98.09	1.33	0.038
Error	356	26266.67	26266.67	73.78		
Total	533	121718.77				

2. Analysis of Variance for Number of leaves(NL)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	264.3633	264.3633	3.0041	4.64	0.000
Block	1	4.1367	4.1367	4.1367	10.04	0.010
Line*Block	88	57.0300	57.0300	0.6481	1.57	0.002
Error	356	146.6667	146.6667	0.4120		
Total	533	472.1966				

3. Analysis of Variance for Leaf length(LL1)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	278113.5	278113.5	3160.4	5.60	0.000
Block	1	731.5	731.5	731.5	1.73	0.250
Line*Block	88	49656.0	49656.0	564.3	1.33	0.036
Error	356	150516.7	150516.7	422.8		
Total	533	479017.6				

4 Analysis of Variance for Leaf Width1(LW1)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	88416.5	88416.5	1004.7	4.61	0.000
Block	1	10.5	10.5	10.5	0.07	0.820
Line*Block	88	19160.3	19160.3	217.7	1.44	0.011
Error	356	53683.3	53683.3	150.8		
Total	533	161270.6				

5. Analysis of Variance for Petiole length1(PL1)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	45037.30	45037.30	511.79	4.83	0.000
Block	1	181.13	181.13	181.13	2.43	0.100
Line*Block	88	9320.71	9320.71	105.92	1.42	0.014
Error	356	26526.00	26526.00	74.51		
Total	533	81065.13				

6. Analysis of Variance for Plant height2(PH2)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	3556453	3556453	40414	12.04	0.000
Block	1	25291	25291	25291	7.54	0.006
Line*Block	88	283813	283813	3225	0.96	0.579
Error	356	1194383	1194383	3355		
Total	533	5059941				

7. Analysis of Variance for Leaf length2(LL2)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	762363	762363	8663	5.45	0.000
Block	1	32532	32532	32532	20.41	0.000
Line*Block	88	119812	119812	1361	0.85	0.812
Error	356	567289	567289	1594		
Total	533	1481997				

8. Analysis of Variance for Leaf Width2(LW2)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	240815.8	240815.8	2736.5	5.09	0.000
Block	1	26828.1	26828.1	26828.1	49.91	0.000
Line*Block	88	46484.4	46484.4	528.2	0.98	0.528
Error	356	191350.0	191350.0	537.5		
Total	533	505478.3				

9. Analysis of Variance for Petiole length2(PL2)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	255567.4	255567.4	2904.2	8.94	0.000
Block	1	16910.2	16910.2	16910.2	52.03	0.000
Line*Block	88	36627.3	36627.3	416.2	1.28	0.062
Error	356	115716.7	115716.7	325.0		
Total	533	424821.6				

10. Analysis of Variance for Flowering height(FH)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	3038779	3038779	34532	9.49	0.000
Block	1	92812	92812	92812	40.74	0.000
Line*Block	88	320255	320255	3639	1.60	0.002
Error	356	811067	811067	2278		
Total	533	4262913				

11. Analysis of Variance for Flwowering time(FT)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	19116.97	19116.97	217.24	28.70	0.000
Block	1	7.67	7.67	7.67	1.01	0.372
Line*Block	88	840.00	840.00	9.55	1.26	0.074
Error	356	2693.33	2693.33	7.57		
Total	533	22657.97				

12. Analysis of Variance for Maximum height(MH)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	14594015	14594015	165841	3.53	0.000
Block	1	878569	878569	878569	30.88	0.000
Line*Block	88	4130565	4130565	46938	1.65	0.001
Error	356	10129467	10129467	28454		
Total	533	29732615				

13. Analysis of Variance for Apical height(AH)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	15117359	15117359	171788	3.45	0.000
Block	1	942312	942312	942312	32.11	0.000
Line*Block	88	4384319	4384319	49822	1.70	0.000
Error	356	10445696	10445696	29342		
Total	533	30889685				

14. Analysis of Variance for Stem width(SW)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	14967.51	14967.51	170.09	1.47	0.037
Block	1	179.96	179.96	179.96	2.46	0.210
Line*Block	88	10211.70	10211.70	116.04	1.58	0.002
Error	356	26066.67	26066.67	73.22		
Total	533	51425.84				

15. Analysis of Variance for Fresh Weight(FW)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Line	88	8498704	8498704	96576	2.07	0.000
Block	1	1848132	1848132	1848132	39.57	0.000
Line*Block	88	3834106	3834106	43569	0.93	0.646
Error	356	16626950	16626950	46705		
Total	533	30807891				

B. List of markers used for *Brassica oleracea* QTL mapping and their position

Marker name	Linkage group	Marker position
1_AC-CTCE02	1	0
1_pN186E1N	1	16.5
1_pW239E2	1	29.7
1_pN52E3NP	1	37.1
1_pCeriE1	1	47.9
1_pO168E1	1	53.7
1_AC-CACE08	1	75.6
1_pN121E2	1	83.8
2_pW116E1	2	0
2_pN121E1	2	31.2
2_C339E2	2	40
2_AA-CATE15	2	50
2_pO17E1	2	60
2_AA-CATE01	2	70.7
2_pO120E1	2	80.2
2_pW141E1	2	95.7
2_AC-CTAE03	2	106.6
3_pW116J1	3	0
3_pW153J1	3	8
3_pO111E1	3	19.3
3_pW111J1	3	29.9
3_pO98E2	3	41
3_AC-CATJ02	3	49.4
3_AC-CAAE14	3	59.5
3_pR85E2	3	70.6
3_AC-CACE01	3	80
3_pW106E2	3	89.7
3_pN96E1	3	98.8
3_flower	3	112
3_pW225E1	3	119.9
3_pO43E1	3	129.3
3_AA-CATE02	3	139
4_pW177E1	4	0
4_pW143E2	4	14.7
4_pO147E1N	4	20.9
4_pN64E1	4	30.9
4_pN202E1	4	39.7
4_labi8E2	4	51
4_pW139E1	4	68
4_pW133E1	4	74
4_pR113E3	4	89.4
4_pW143E1	4	99.6
5_pN21E2	5	7.8
5_pN23E3	5	14.7
5_pN52E1	5	25

5_pW197E1	5	36.8
5_pO143J1	5	44.9
5_pN148E1	5	52.2
5_pN91E3	5	62.4
5_pO123J1	5	70.4
5_AA-CATR26	5	81.2
5_AC-CACE02	5	88.5
5_pN2J1	5	97.7
6_AC-CTCE01	6	0
6_AC-CACE13	6	10.1
6_pO10E1	6	18.7
6_mNGA111J1	6	29.7
6_mBNMB4	6	36
6_AC-CTAJ06	6	43.5
6_mCa72	6	64.4
6_pO10E2	6	72.6
7_pO87E2	7	8.6
7_pO85J1	7	15.3
7_pN86E1	7	25.5
7_pN20E2	7	39.2
7_pN64E2	7	43.6
7_pO59E1	7	54.7
7_pCeriE3	7	66.6
7_pW228E2	7	71.1
7_mBN72AJ1	7	80.8
8_pO113E1	8	7.1
8_pO159J1	8	19.2
8_pW104E2	8	26.8
8_pW188J1	8	34.8
8_pR97J1	8	43.2
8_AC-CAAE05	8	54
8_pN123J1	8	67.8
8_AA-CATJ03	8	73.6
8_pN23J1	8	81.3
8_pN21E1	8	86.8
9_pN52E2	9	0
9_pO125E1N	9	12.8
9_pN101E2N	9	24.3
9_pW114E2	9	35.1
9_AC-CTAE16	9	43.8
9_pW233J1	9	52.7
9_mBN83B1J1	9	63.8
9_LEW6E2	9	73.7
9_pN47E4NM	9	87.2
9_pN3E1	9	103

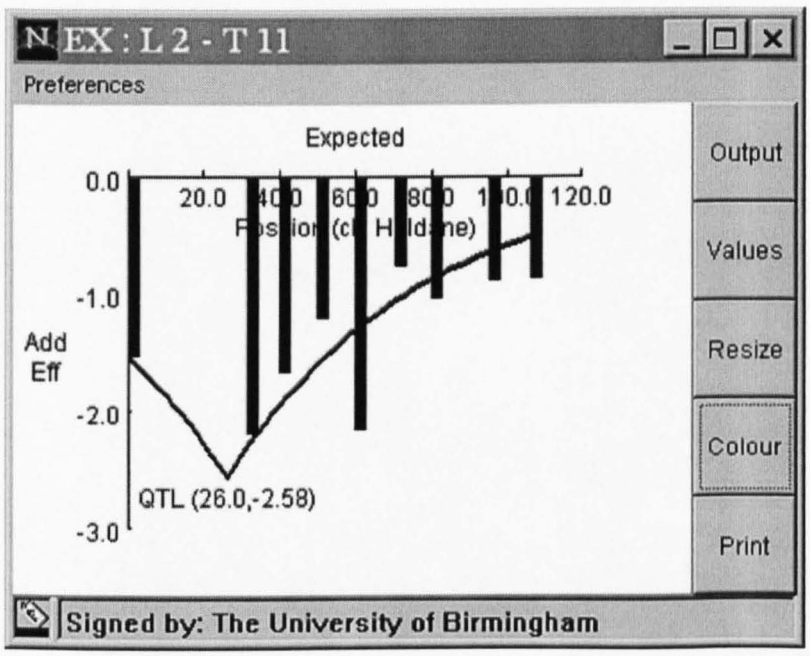
C. An example of QTL detected in DH lines using the marker regression method in flowering time

Linkage Group: 2
Trait: Flowering time
QTL located at 26.0 cM
Additive effect = -2.5771055

Source	df	MS	F	P	
Add Reg	1	1631.09		79.12	0.01
Residual	7	16.28	0.79	0.444	
Error	80	20.62			

Simulated QTL position is 29.96 +/- 17.261

Simulated Additive QTL effect is -2.464 +/- 0.749



APPENDIX II

A. One-way ANOVA between the parents, F_1, its reciprocal and F_2 plants

Analysis of variance for TTG

Source	DF	SS	MS	F	P
Plant	4	11.98	2.99	1.40	0.233
Error	489	1045.91	2.14		
Total	493	1057.89			

Analysis of variance for TTL

Source	DF	SS	MS	F	P
Plant	4	47.61	11.90	3.43	0.009
Error	489	1696.26	3.47		
Total	493	1743.86			

Analysis of variance for RL20

Source	DF	SS	MS	F	P
Plant	4	78.16	19.54	10.61	0.000
Error	489	900.68	1.84		
Total	493	978.84			

Analysis of variance for CL20

Source	DF	SS	MS	F	P
Plant	4	48.81	12.20	6.25	0.000
Error	489	954.77	1.95		
Total	493	1003.58			

Analysis of variance for HT20

Source	DF	SS	MS	F	P
Plant	4	30037	7509	15.40	0.000
Error	489	238469	488		
Total	493	268507			

Analysis of variance for TTB

Source	DF	SS	MS	F	P
Plant	4	153.04	38.26	5.43	0.000
Error	489	3446.17	7.05		
Total	493	3599.22			

Analysis of variance for HT34

Source	DF	SS	MS	F	P
Plant	4	51269	12817	2.75	0.028
Error	489	2281437	4666		
Total	493	2332706			

Analysis of variance for RLF

Source	DF	SS	MS	F	P
Plant	4	81.03	20.26	8.48	0.000
Error	489	1168.15	2.39		
Total	493	1249.18			

Analysis of variance for CLF

Source	DF	SS	MS	F	P
Plant	4	3.70	0.92	0.84	0.503
Error	489	541.17	1.11		
Total	493	544.87			

Analysis of variance for HTF

Source	DF	SS	MS	F	P
Plant	4	36097	9024	10.27	0.000
Error	489	429553	878		
Total	493	465651			

Analysis of variance for TTF

Source	DF	SS	MS	F	P
Plant	4	85.94	21.48	3.51	0.008
Error	489	2995.59	6.13		
Total	493	3081.52			

B. Data on generations in F_2 plants

Parameter	TTG	TTL	RL20	CL20	HT20
P_1 mean	4.75	8.58	10.08	0.54	10.00
P_2 mean	5.38	9.67	7.58	2.13	12.50
F_1 mean	4.63	8.21	8.63	2.21	51.04
RF_1 mean	4.88	8.58	8.50	2.08	38.33
F_2 mean	4.67	8.26	8.57	1.55	24.05
P_1 variance	2.98	2.30	2.78	1.30	36.97
P_2 variance	1.46	2.32	0.60	1.77	32.60
F_1 variance	2.24	5.48	1.46	1.83	1195.78
RF_1 variance	2.81	5.47	0.87	2.08	1099.59
F_2 variance	2.09	3.20	1.94	2.00	463.54
M	5.07	9.13	8.83	1.34	11.25
mF_1	4.76	8.40	8.57	2.15	44.69
P_1 vs. P_2	ns	<0.01	<0.01	<0.01	ns
F_1 vs. RF_1	ns	ns	ns	ns	ns
m vs. mF_1	ns	ns	<0.01	ns	<0.01
m vs. F_2	<0.01	<0.01	<0.01	ns	ns
mF_1 vs. F_2	ns	<0.01	ns	<0.01	<0.01

Parameter	TTB	HT34	RLF	CLF	HTF	TTF
P_1 mean	21.63	321.46	10.33	3.50	83.54	24.58
P_2 mean	19.38	296.87	7.88	3.33	63.54	22.79
F_1 mean	18.88	347.29	8.88	3.00	116.46	22.50
RF_1 mean	18.33	347.29	8.38	3.25	88.75	22.13
F_2 mean	19.37	339.10	8.87	3.36	92.91	23.07
P_1 variance	8.85	12150.65	3.19	0.35	353.06	10.08
P_2 variance	1.72	1534.29	1.42	1.45	403.21	1.83
F_1 variance	13.42	4656.70	0.64	0.17	535.92	9.57
RF_1 variance	4.15	2545.20	1.38	1.15	1022.08	2.90
F_2 variance	7.05	4536.02	2.56	1.18	948.02	6.14
M	20.51	309.17	9.11	3.42	73.54	23.69
mF_1	18.61	347.29	8.63	3.13	102.61	22.32
P_1 vs. P_2	ns	<0.01	<0.01	ns	<0.01	ns
F_1 vs. RF_1	ns	ns	ns	ns	ns	ns
m vs. mF_1	<0.01	ns	<0.01	ns	<0.01	<0.01
m vs. F_2	<0.01	ns	ns	ns	<0.01	ns
F_1 vs. F_2	ns	ns	ns	ns	ns	ns

m = mid-parent value; mF_1 = mid- F_1 value; P1 vs. P2, F1 vs. RF1 = values of the probabilities associated with the one-way AVOVA; ns = non-significant probabilities (>5%)

C. Markers, their band sizes in Col and Ler, the annealing temperature (T_m) and the Mg^{2+} concentration in PCR analysis and primer sequences

Chr.	Marker name(No.)	Pos.	Col (bp)	Ler (bp)	T_{ann} (°C)	[Mg^{2+}] (mM)	Forward Primer (5'-3')	Reverse Primer (5'-3')
1	Nga59 (38)	2.90	111	115	50	1.5	gcatctgtgttcactc gcc	ttaatacattaagcc cagacccg
1	F20O23 (43)	24.00	186	-	45	2.5	ccaacc cct tat ata tcg ttc a	cac atg aac gtt ggg ata aat a
1	Nga392 (42)	41.64	170	162	55	2.5	ttgaataatttgta gccatg	ggtgttaaatgcggt gttc
1	T27k12sp (32)	59.10	146	152	55	2.5	gga caa cgt tct aaa cgg tt	gga ggctat acg aatcttaca
1	Nag280 (14)	83.88	105	85	55	2.5	ctgatctcacgga caatagtgc	ggctccataaaaa gtgcacc
1	Nga111 (7)	115.55	128	162	55	2.5	ctccagttggaag ctaaaggg	tgttttttaggacaaa tggcg
2	Nga1145 (15)	9.60	213	217	55	2.5	ccttcacatccaa aacccac	gcacatacccaca accagaa
2	MSF3A (73)	35.04	173	-	50	2.5	cat tgg att tca ttt tat tcc c	gca tcg ttc cac aaa aat aaa a
2	Nga1126 (16)	50.65	191	199	50	2.5	cgctacgcttttcg gtaaag	gcacagtccaagtc acaacc
2	Nga361 (39)	63.02	114	120	55	3.5	aaagagatgag aatttggac	acatatcaatatatt aaagtagc
2	Nga168 (10)	73.77	151	135	55	2.5	tcgtctactgcact gccg	gaggacatgtatag gagcctcg
3	Nga172 (25)	6.90	162	136	55	2.5	agctgcttccttat agcgtcc	catccgaatgccatt gttc
3	Nga162 (24)	20.56	107	89	55	2.5	catgcaatttgca tctgagg	ctctgtcactcttttcc tctgg
3	Athgapab (31)	43.77	142	150	55	2.5	caccatggcttcg gttactt	tcctgagaattcagt gaaaccc
3	MMJ24 (76)	48.45	159	120	50	2.5	cct ctt att tct aac gga agc a	tgg agt agc aaa acc atc aat a
3	TH620B (77)	59.10	142	100	50	2.5	cag aaa tag acg tcg ata cga a	ggg cag aga gaa cta aaa aag c
3	Nga707 (66)	78.25	132	128	50	2.5	tgaatgcgtccat ggagaag	ctctctgcctctcgct gg
3	Nga112 (47)	87.88	197	189	55	2.5	taatcacgtgtat gcagctgc	ctctccacctcctcc agtacc
4	T18A10 (86)	1.00	155	-	50	2.5	taa gag gag gaa tct gat acg	aat gtg tgg tca gga att aac
4	Nga8 (59)	26.56	154	198	55	2.5	gagggcaaatct ttatttcgg	tggctttcgtttataa acatcc
4	FCA9 (79)	54.83	165	-	50	3.5	Tgaagaatgatg	ttgtgattaatttgga

							tctcctttg	catgg
4	F20D24 (80)	72.35	239	-	55	3.5	tca aaa act agc atg caa cag	aca cgg cta aac aaa taa tcg
4	Nga1139 (62)	83.41	114	118	50	2.5	tagccggatgag ttggtacc	tttttccttgtgttgcat tcc
4	T5J17 (81)	119.00	245	-	50	1.5	cag aga gag gga tat gga ttg	act cca cca ctt gtt cta agc
5	MED24 (82)	7.40	218	190	50	1.5	tgt aca tgt tgg att ctc ctc	tag ttg ggt tca tga tga atg
5	Nga249 (13)	23.72	125	115	55	2.5	taccgtcaatttct acgcc	ggatccctaactgt aaaatccc
5	Nga139 (8)	50.48	174	132	55	2.5	agagctaccaga tccgatgg	ggtttcgtttcactat ccagg
5	Nga76 (5)	68.40	231	>25 0	50	2.5	ggagaaaatgtc actctccacc	aggcatgggagac atttacg
5	MJB21A (83)	89.50	164	-	45	3.5	ttt taa ttg tca tgc aac acg	att tca aac gta cct ggt gaa
5	MM19 (84)	116.90	255	-	45	2.5	ctg act atc aag taa atc tac g	gtt gat gta atg aag tat gaa g

- unknown band sizes

D. Correlations between markers in F_2 population

	38	43	42	32	14	7	15	73	16	39	10	25	24	31	76	77	66	47
43	**0.41**																	
42	**0.67**	**0.37**																
32	**0.33**	**0.35**	**0.34**															
14	**0.19**	**0.22**	**0.23**	**0.65**														
7	**0.13**	**0.13**	**0.13**	**0.30**	**0.44**													
15	0.16	0.04	0.07	0.05	0.06	0.07												
73	0.21	0.17	0.12	0.17	0.16	0.08	**0.35**											
16	0.11	0.08	-0.01	0.15	0.14	0.08	**0.23**	**0.56**										
39	0.08	-0.04	0.12	0.08	0.09	0.06	**0.23**	**0.26**	**0.52**									
10	0.02	-0.01	0.04	0.04	0.05	0.01	**0.25**	**0.21**	**0.47**	**0.69**								
25	0.09	0.04	0.05	0.12	0.04	0.08	0.05	0.05	0.09	-0.01	-0.03							
24	0.12	0.04	0.04	0.12	0.07	0.08	0.07	0.03	0.04	0.02	0.05	**0.52**						
31	0.08	0.03	0.20	0.13	0.08	0.08	0.00	0.04	0.05	0.13	0.12	**0.31**	**0.40**					
76	0.13	0.07	0.04	0.10	0.10	0.06	0.06	0.06	0.08	0.02	0.07	**0.33**	**0.57**	**0.59**				
77	0.10	0.01	0.20	0.13	0.15	0.13	0.05	0.03	0.01	0.14	0.12	**0.17**	**0.13**	**0.56**	**0.36**			
66	0.14	0.15	0.02	0.16	0.17	0.07	0.17	0.20	0.12	0.09	0.06	**0.18**	**0.16**	**0.19**	**0.36**	**0.49**		
47	0.11	0.10	0.00	0.11	0.18	0.01	0.25	0.15	0.10	0.11	0.15	**0.14**	**0.24**	**0.16**	**0.29**	**0.28**	**0.55**	
86	0.04	0.05	0.14	0.07	0.11	0.11	0.11	0.07	0.12	0.22	0.29	0.08	0.08	0.19	0.08	0.20	0.06	0.18
59	0.13	0.12	0.05	0.17	0.13	0.10	0.10	0.12	0.18	0.07	0.06	0.15	0.16	0.12	0.21	0.01	0.15	0.07
79	0.04	0.17	0.07	0.05	0.04	0.04	0.12	0.14	0.13	0.05	0.13	0.12	0.09	0.06	0.16	0.09	0.15	0.19
80	0.15	0.19	0.10	0.06	0.09	0.10	0.12	0.17	0.16	0.11	0.10	0.10	0.12	0.05	0.11	0.09	0.15	0.12
62	0.17	0.06	0.15	0.13	0.15	0.13	0.12	0.10	0.16	0.20	0.11	0.12	0.14	0.13	0.14	0.14	0.12	0.13
81	0.11	0.04	0.13	0.01	0.06	0.10	0.20	0.06	0.13	0.33	0.26	0.09	0.13	0.17	0.12	0.22	0.13	0.17
82	0.08	0.06	0.09	0.12	0.14	0.13	0.16	0.16	0.19	0.11	0.13	0.12	0.04	0.03	0.15	0.07	0.11	0.07
13	0.11	0.07	0.09	0.19	0.18	0.11	0.09	0.14	0.13	0.11	0.08	0.17	0.10	0.09	0.10	0.03	0.12	0.08
8	0.12	0.12	0.04	0.15	0.12	0.09	0.10	0.11	0.17	0.05	0.05	0.12	0.15	0.10	0.19	0.01	0.16	0.08
5	0.13	0.06	0.01	0.05	0.05	0.12	0.25	0.13	0.12	0.22	0.21	0.16	0.19	0.09	0.23	0.09	0.18	0.21
83	0.18	0.15	0.04	0.10	0.09	0.18	0.14	0.22	0.17	0.05	0.07	0.11	0.10	0.06	0.17	0.05	0.08	0.04
84	0.14	0.05	0.08	0.11	0.09	0.14	0.04	0.08	0.18	0.05	0.08	0.08	0.07	0.04	0.12	0.02	0.05	0.01

Correlation between markers (continued)

	86	59	79	80	62	81	82	13	8	5	83
59	**0.05**										
79	**0.15**	**0.10**									
80	**0.10**	**0.16**	**0.42**								
62	**0.13**	**0.17**	**0.30**	**0.67**							
81	**0.19**	**0.04**	**0.27**	**0.41**	**0.63**						
82	0.10	0.36	0.08	0.07	0.08	0.03					
13	0.08	0.70	0.08	0.10	0.09	-0.02	**0.52**				
8	0.03	0.95	0.09	0.15	0.14	0.04	**0.35**	**0.67**			
5	0.16	0.54	0.12	0.14	0.13	0.17	**0.21**	**0.37**	**0.51**		
83	0.06	0.31	0.19	0.25	0.22	0.11	**0.09**	**0.15**	**0.28**	**0.41**	
84	0.03	0.24	0.05	0.16	0.21	0.10	**0.06**	**0.14**	**0.24**	**0.20**	**0.37**

NB: Bold numbers are correlation between markers in a chromosome.

E. Correlation between markers and traits in *Arabidopis thaliana* F_2 plants

Marker	Chr	TTG	TTL	RL20	CL20	HT20	TTB	HT34	RLF	CLF	HTF	TTF
Nga59	1	-	-	-	***	***	***	-	-	***	-	***
F20O23	1	-	-	-	-	-	*	-	-	*	*	*
Nga392	1	-	-	-	***	**	***	-	-	**	-	***
T27k12sp	1	-	-	-	-	-	-	-	-	-	-	-
Nag280	1	-	-	-	-	-	-	-	-	-	-	-
Nga111	1	-	-	-	***	-	*	-	-	*	-	**
Nga145	2	-	-	*	-	-	-	-	**	-	**	-
MSF3A	2	-	-	***	-	-	-	***	***	-	***	-
Nga1126	2	-	-	***	*	*	**	***	***	-	***	**
Nga361	2	-	-	***	-	-	-	***	***	-	***	-
Nga168	2	-	-	***	-	-	-	**	***	-	***	-
Nga172	3	-	-	-	*	-	**	-	-	-	-	*
Nga162	3	-	-	-	***	-	**	-	-	-	-	**
Athgapar	3	-	-	-	-	-	-	-	-	-	-	-
MMJ24	3	-	-	-	-	-	-	-	-	-	-	-
TH620B	3	-	-	*	-	*	*	*	-	*	-	**
Nga707	3	-	-	-	-	-	-	-	-	-	-	-
Nga112	3	-	-	-	-	*	-	-	-	-	-	-
T18A10	4	-	-	-	-	-	*	-	-	*	-	*
Nga8	4	-	-	**	*	***	***	-	***	-	*	*
FCA9	4	-	-	-	-	-	-	-	-	-	-	-
F20O24	4	-	-	-	-	-	-	-	-	-	-	-
Nga1139	4	-	-	-	-	-	-	-	-	-	-	-
T5J17	4	-	-	-	-	-	-	-	-	-	-	-
MED24	5	-	-	-	**	*	**	-	**	-	-	*
Nga249	5	-	-	**	*	-	***	-	***	-	-	*
Nga139	5	-	-	***	*	***	***	-	***	-	*	*
Nga76	5	-	-	-	-	*	*	-	*	-	*	-
MJB21A	5	-	-	-	*	***	**	-	-	-	**	**
MM19	5	-	-	-	*	*	*	-	*	-	-	-
Total		-	-	**9**	**12**	**11**	**16**	**5**	**11**	**6**	**10**	**14**

APPENDIX III

A. ANOVA using the General Linear Model (GLM) in F3 families

1. Analysis of Variance for Time to Germination (TTG)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
LINE	199	491.320	491.401	2.469	2.09	0.000
REP	4	16.040	16.040	4.010	3.39	0.009
Error	787	929.777	929.777	1.181		
Total	990	1437.136				

2. Analysis of Variance for Time to True leaves (TTL)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
LINE	199	499.095	499.616	2.511	1.99	0.000
REP	4	11.627	11.627	2.907	2.30	0.057
Error	787	993.889	993.889	1.263		
Total	990	1504.612				

3. Analysis of Variance for Height at 20 days (HT20)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
LINE	199	351763.6	352370.5	1770.7	3.91	0.000
REP	4	4944.0	4944.0	1236.0	2.73	0.028
Error	787	356690.7	356690.7	453.2		
Total	990	713398.3				

4. Analysis of Variance for Rosette leaves at 20 days (RL20)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
LINE	199	1140.460	1140.410	5.731	3.14	0.000
REP	4	4.641	4.641	1.160	0.64	0.637
Error	787	1436.059	1436.059	1.825		
Total	990	2581.160				

5. Analysis of Variance for cauline at 20 days (CL20)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
LINE	199	943.170	944.962	4.749	3.49	0.000
REP	4	10.049	10.049	2.512	1.84	0.118
Error	787	1071.968	1071.968	1.362		
Total	990	2025.187				

6. Analysis of Variance for Time to Bud (TTB)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
LINE	199	2915.104	2919.388	14.670	5.81	0.000
REP	4	33.611	33.611	8.403	3.33	0.010
Error	787	1986.639	1986.639	2.524		
Total	990	4935.354				

7. Analysis of Variance for Time to Flower (TTF)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
LINE	199	2669.511	2673.383	13.434	5.80	0.000
REP	4	29.510	29.510	7.377	3.19	0.013
Error	787	1821.557	1821.557	2.315		
Total	990	4520.577				

8. Analysis of Variance for Height at Flowering (HTF)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
LINE	199	690707.2	690595.1	3470.3	4.31	0.000
REP	4	1791.5	1791.5	447.9	0.56	0.694
Error	787	633639.7	633639.7	805.1		
Total	990	1326138.4				

9. Analysis of Variance for Rosette leaves at Flowering (RLF)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
LINE	199	2237.796	2237.814	11.245	5.23	0.000
REP	4	8.215	8.215	2.054	0.95	0.432
Error	787	1693.785	1693.785	2.152		
Total	990	3939.796				

10. Analysis of Variance for Cauline leaves at Flowering (CLF)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
LINE	199	1147.088	1147.535	5.767	2.18	0.000
REP	4	19.603	19.603	4.901	1.85	0.117
Error	787	2084.664	2084.664	2.649		
Total	990	3251.354				

11. Analysis of Variance for Height at 34 days (HT34)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
LINE	199	4344972	4334961	21784	3.58	0.000
REP	4	609627	609627	152407	25.06	0.000
Error	787	4785521	4785521	6081		
Total	990	9740119				

N.B: REP = replications

B. 2-way ANOVA for detecting macro-environmental variation

1. Analysis of Variance for TTF

Source	DF	SS	MS	F	P
Experiment	1	2.89	2.90	3.58	ns
Genotype	2	5.16	2.58	3.19	*
Experiment*Genotype	2	0.27	0.14	0.17	ns
Error	163	-	0.33		

2. Analysis of Variance for TTB

Source	DF	SS	MS	F	P
Experiment	1	12.70	12.70	15.12	***
Genotype	2	6.41	3.20	3.81	*
Experiment*Genotype	2	0.58	0.29	0.35	ns
Error	163	-	0.84		

3. Analysis of Variance for TTG

Source	DF	SS	MS	F	P
Experiment	1	2.42	2.42	7.56	***
Genotype	2	0.98	0.49	1.53	ns
Experiment*Genotype	2	0.09	0.06	0.14	ns
Error	163	-	0.32		

4. Analysis of Variance for TTL

Source	DF	SS	MS	F	P
Experime	1	1.30	1.30	2.95	ns
Genotype	2	3.91	1.96	4.44	*
Experime*Genotype	2	0.43	0.21	0.48	ns
Error	163	-	0.44		

5. Analysis of Variance for CL20

Source	DF	SS	MS	F	P
Experiment	1	7.41	7.41	15.78	***
Genotype	2	3.43	1.71	3.64	*
Experiment*Genotype	2	0.99	0.49	1.05	ns
Error	163	-	0.47		

6. Analysis of Variance for RL20

Source	DF	SS	MS	F	P
Experiment	1	0.41	0.41	1.27	ns
Genotype	2	9.1646	4.58	14.32	***
Experiment*Genotype	2	0.4503	0.23	0.70	ns
Error	163	-	0.32		

7. Analysis of Variance for HT20

Source	DF	SS	MS	F	P
Experiment	1	716.40	716.40	7.15	***
Genotype	2	4074.40	2037.20	20.34	***
Experiment*Genotype	2	634.90	317.40	3.17	*

Error	163	-	100.18		

8. Analysis of Variance for RLF

Source	DF	SS	MS	F	P
Experiment	1	1.66	1.66	4.62	*
Genotype	2	9.86	4.93	13.70	***
Experiment*Genotype	2	0.51	0.26	0.71	ns
Error	163	-	0.36		

9. Analysis of Variance for CLF

Source	DF	SS	MS	F	P
Experiment	1	8.10	8.10	19.28	***
Genotype	2	0.96	0.48	1.14	ns
Experiment*Genotype	2	1.43	0.71	1.70	ns
Error	163	-	0.42		

10. Analysis of Variance for HTF

Source	DF	SS	MS	F	P
Experiment	1	167.90	167.90	1.41	ns
Genotype	2	2272.43	1136.22	9.57	***
Experiment*Genotype	2	74.38	37.19	0.31	ns
Error	163	-	118.70		

11. Analysis of Variance for HT34

Source	DF	SS	MS	F	P
Experiment	1	16147.1	16147.10	18.75	***
Genotype	2	14259.0	7129.50	8.28	***
Experiment*Genotype	2	5040.3	2520.20	2.93	ns
Error	163	-	861.01		

C. χ^2 test for segregation distortion for the markers used in the F_3 families

Marker	Marker number	Chr.	Number of individuals (P1:Het:P2)	χ^2 test		
				1:1 (P1 vs P2)	1:1 (Homozygous vs heterozygous)	(1:2:1)
Nga59	38	1	40:107:52	1.57	1.13	2.57
F20O23	43	1	40:103:48	0.73	1.18	1.85
Nga392	42	1	57:84:38	3.80	0.68	4.71*
T27k12sp	32	1	44:105:50	0.38	0.61	0.96
Nag280	14	1	51:108:40	1.33	1.45	2.67
Nga111	7	1	52:101:46	0.37	0.05	0.40
Nga145	15	2	18:157:9	3.00	91.85***	92.72***
MSF3A	73	2	43:110:37	0.45	4.74*	5.12*
Nga1126	16	2	52:113:34	3.77	3.66	6.92**
Nga361	39	2	38:87:36	0.05	1.05	1.10
Nga168	10	2	47:88:34	0.95	0.05	2.34
Nga172	25	3	32:113:47	2.85	6.02*	8.36**
Nga162	24	3	43:113:41	0.05	4.27*	4.30*
Athgapar	31	3	34:66:43	1.05	0.84	1.98
MMJ24	76	3	45:102:50	0.26	0.25	0.50
TH620B	77	3	33:76:47	2.45	0.10	2.61
Nga707	66	3	44:109:44	0.00	2.24	2.24
Nga112	47	3	43:95:34	1.05	1.88	2.82
T18A10	86	4	50:80:30	5.00*	0.00	5.00*
Nga8	59	4	59:110:30	9.45**	2.22	10.67**
FCA9	79	4	39:91:48	0.93	0.09	1.01
F20O24	80	4	48:100:50	0.04	0.02	0.07
Nga1139	62	4	50:97:52	0.04	0.13	0.16
T5J17	81	4	44:86:37	0.60	0.15	0.73
MED24	82	5	46:131:19	11.22***	22.22***	29.66***
Nga249	13	5	68:93:34	11.33***	0.42	12.27***
Nga139	8	5	61:111:27	13.14***	2.66	14.27***
Nga76	5	5	49:111:26	7.05**	6.97**	12.65***
MJB21A	83	5	48:96:54	0.35	0.18	0.55
MM19	84	5	40:120:29	1.75	13.76***	15.04***

Chr. = chromosome

D. Costs of growing, DNA extraction and PCR reaction in *Arabidopsis thaliana* during the 2001/2002 academic year

i. Costs of growing the plants in the field and glasshouse

	£/m^2	Plants/m^2		£ per plant	£ per 200 plants
Growth room	15+10%	218	Arabidopsis	0.076	15.14
Field	0.1	6	Brassica	0.017	3.33
Poly tunnel	10	9	Brassica	1.11	222.22

ii. Costs of running a PCR

	Normal quantity bought	**Price**	**price/ML**	**price/uL**	**price/10 uL**	**price/ reaction**
TAQ	500U	76.14		0.76		0.061
Mg						0
Buffer						0
						0
DNTP's	500	45.3		0.1396		0.55
Primers	44	at 0.26/base	11.44			0
Av 22 bases X2						0.006
Pcr tubes	1000	75		0.075		0.075
White tips				0.01		0.01
Loading buffer					0.001897	0.001897
						0.703897
				for 200 samples		140.7794
FOR 6 markers on 5 chromosome i.e. 30 markers						**4223.382**

iii. Costs of DNA extraction

	Normal quantity bought	**Price**	**price/unit**	**price/uL**	**price/ 10uL**	**price/ reaction**
Ctab	100.00	24.80	0.248			4.96
Sodium chloride	500.00	2.94	0.005875			0.2403345
Edta	100.00	5.85	0.0585			0.217737
Tris	500.00	21.38	0.04276			0.2589118
Ethanol	2500.00	10.22	0.0041	0.000004	0.00004	
Iso-amyl alcohol	500.00	6.80	0.0136	0.000014	0.00014	
Chloroform	2500.00	6.90	0.00276	0.000003	0.00003	
Propan2-ol	2500.00	6.15	0.00246	0.000003	0.00003	
Ethanol	2500.00	10.22	0.004088	0.000004	0.00004	
Mercapto-ethanol	100.00	8.80	0.088	0.000088	0.00088	
Ball bearings	5000.00	75.00	0.015			
CTAB buffer 500ml		4.96				
(700 reactions)		0.24				
		0.22				
		0.26				
		5.68				0.0081
Iso-amyl alcohol 20ml	0.27	0.27				
Chloroform 480ml	1.32	1.32				
(810 reactions)		1.60				0.0020
Propan2-ol 500uL	0.00123					0.0012
Ethanol 300uL	0.00123					0.0012
Mercapto-ethanol 10uL	0.00088					0.0009
Ball bearings	0.015					0.015
T.E buffer 60uL	0.00000359					0.00000359
Blue tips	0.005828					0.0058
Tubes 1.5ml	0.01296					0.01296
Per single extraction						0.0472
For 200 extractions						**9.4416**

iv. Total costs

FOR 200 PLANTS	**Costs**
GROWTH	15.13
EXTRACTION	9.44
PCR	4223.38
TESTING GELS X30	74.1
SPREADEX (30 GELS) 6MKSX5CHR	1680
	6002.05

APPENDIX IV

DNA extraction procedure using the 'GenElute$^{TN'}$ plant genomic kit (SIGMA-Aldrich)

The plant tissue (100 – 150 mg) was ground in liquid nitrogen, then lysed with 350µl of lysis solution (Part A) and 50µl of lysis solution (Part B). The contents were mixed by inverting the tube and then incubated at 65°C for 10 minutes. About 130µl precipitation solution was added, then incubated on ice for 5 minutes. The debris was pelleted for 5 minutes and the clear supernatant was transferred to a blue filtration column and spun for 1 minute at full speed. 700µl binding solution was added to filtrate and mixed thoroughly by inversion. About 700µl of the mixture was transferred to binding column, spun for 1 min at 13,000 rpm and the flow-through was discarded. The procedure was repeated with the remainder of the mixture. The column was than transferred to a new collection tube.

The DNA was washed twice with 500µl wash solution. The column was transferred to a new collection tube and 100µl elution solution (pre-warmed to 65°C) was added, then spun for 1 min to collect the DNA.

The DNA concentrations were estimated by running the sample on 3% agarose gel alongside lamba standards of known concentrations (2, 4, 10 and 20 nm).

LIST OF REFERENCES

Ahn, S., Andeson, J. A., Sorrels, M. E. and Tanksley, S. D., 1993. Homoeologous relationships of rice, wheat and maize chromosomes. Mol. Gen. Genet. 241: 483-490.

Alonso-Blanco, C., El-Assal, S. E. D., Coupland, G. and Koornneef, M., 1998. Analysis of natural allelic variation at flowering time loci in Landberg *erecta* and Cape Verde Islands (Cvi) ecotypes of *Arabidopsis thaliana*. Genetics, 149: 749-764.

Alonso-Blanco, C. and Koorneef, M., 2000. Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. Trends Plant Science, 5: 22-29.

Alonso-Blanco, C., Bentsink, L., Hanhart, C. J., Vries H. B. and Koorneef, M., 2003. Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. Genetics, 164: 711-729.

Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, 408: 796-815.

Beckmann, J. S. and Soller, M., 1983. Restriction fragment length polymorphism in genetic improvement: methodologies, mapping and costs. Theor. Appl. Genet., 67: 35-43.

Beckman, J. S. and Soller, M., 1988. Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations. Theor. Appl. Genet., 76: 228-236.

Bohuon, E. J. R., Keith, D. J., Parkin, I. A. P., Sharpe, A. G. and Lydiate, D. J., 1996. Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. Theor. Appl. Genet., 93: 833-839.

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

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

Borevitz, J. O., Maloof, J. N., Lutes, J., Dabi, T., Redfern, J. L. et al., 2002. Quantitative trait loci controlling light and hormone response in two accessions of *Arabidopsis thaliana*. Genetics, 160: 683-696.

Bouchez, A., Hospital, F., Causse, M., Gallais, A. and Charcosset, A., 2002. Marker-assisted introgression of favourable alleles at quantitative trait loci between maize elite lines. Genetics, 162: 1945-1959.

Bowman, J. L. (ed.), 1994. *Arabidopsis*: An atlas of morphology and development. Springer-Verlag, New York. USA.

Bowman, J. L., Alvarez, J., Weigel, D., Meyerowitz, E. M. and Smyth, D. R., 1993. Genetic control of flower development in *Arabidopsis thaliana* by *APETALA 1* and interacting genes. Development, 119: 721-743

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

Brookes, A. J., 1999. The essense of SNP's. Gene, 234: 177-186.

Burn, J. E., Smyth, D. R., Peacock, W. J. and Dennis, E. S., 1993. Genes conferring late flowering in *Arabidopsis thaliana*. Genetica, 90: 147-155.

Burns, M. J., 1997. Quantitative trait loci mapping in Arabidopsis – Theory and practice. PhD thesis. The University of Birmingham, Birmingham, UK.

Burns, M. J., Barnes, S. R., Bowman, J. G., Clarke, M. H. E., Werner, C. P. and Kearsey, M. J., 2003. QTL analysis of an intervarietal set of substitution lines in *Brassica napus*: (i) Seed oil content and fatty acid composition. Heredity, 90: 39-48.

Burr, B., 2001. Some concepts and new methods of molecular mapping in plants. In: R. L. Phillips and I. K. Vasil (eds.). DNA-based Markers in Plants, 2nd Edition, pp1-8. Kluwer Academic Publishers, Netherlands.

Camargo, L. E. A. and Osborn, T. C., 1996. Mapping loci controlling flowering time in *Brassica oleracea*. Theor. Appl. Genet., 92: 610-616.

Camargo, L. E. A., Savides, L., Jung, G., Nienhuis, J. and Osborn, T. C., 1997. Location of the self-incompatibility locus in an RFLP and RAPD map of *Brassica oleracea*. J. Heredity, 88: 57-59.

Cheung, W. Y., Champagne, G., Hubert, N. and Landry, B. S., 1997. Comparison of the genetic maps of *Brassica napus* and *Brassica oleracea*. Theor. Appl. Genet. 94: 569-582.

Cho, R. J., Mindrinos, M., Richards, D. R., Sapolsky, R., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T. L. and Stammers, M., 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nat. Genet., 23: 203-207.

Churchill, G. A. and Doerge, R. W., 1994. Emperical threshold values for quantitative trait mapping. Genetics, 138: 963-971.

Clarke, J. H. and Dean, C., 1994. Mapping FRI, a locus controlling flowering time and vernalization response in *Arabidopsis thaliana*. Mol. Gen. Genet., 242: 81-89.

Clarke, J. H., Mithen, R., Brown J. K. M. and Dean, C., 1995. QTL analysis of flowering time in *Arabidopsis thaliana*. Mol. Gen. Genet., 248: 278-286.

Coupland, G., 1995. Genetic and environmental control of flowering time in *Arabidopsis*. Trend Genet., 11: 393-397.

Darvarsi, A. and Soller, M., 1994. Optimum spacing of the genetic markers for determining linkage between marker loci and quantitative trait loci. Theor. Appl. Genet. 89: 351-357.

Darvasi, A. Weinreb, A., Minke, V., Weller, J. and Soller, M., 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics, 134: 943-951.

Doerge, R. W., 2002. Multifactorial genetics: mapping and analysis of quantitative trait loci in experimental populations. Nature Reviews Genetics, 3: 43-52.

Doerge, R. W., Zeng, Z. B. and Weir, B. S., 1997. Statistical issues in the search of genes affecting quantitative traits in experimental populations. Statistical Science, 12: 195-219.

Dorn, L. A. and Mitchell-Olds, T., 1991. Genetics of *Brassica campestris*. 1. Genetic constrains on evolution of life-history characters. Evolution, 445: 371-379.

Dudley, J. W., 1993. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Science, 33: 660-668.

Edwards, M. D. and Page, N. J., 1994. Evaluation of marker-assisted selection through computer simulation. Theor. Appl. Genet., 88: 376-382

Edwards, M. D., Stuber, C. W. and Wendel, J. F., 1987. Molecular-marker-facilitated investigation of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics, 116: 113-125.

El-Assal, S. E. D., Alonso-Blanco, C., Peeters, A. J. M., Raz, V. and Koorneef, M., 2001. A QTL for flowering time in Arabidopsis reveals a novel allele *CRY2*. Nat. Genet., 29: 435-440.

Falconer, D. S. and Mackay, T. F. C., 1996. Introduction to Quantitative Genetics. 4th Edition. Longman Group Ltd, England.

Fan, J. B., Chen, X., Halushka, M. K., Berno, A., Huang, X., Ryder, T., Lipshutz, R. J., Lockhart, D. J. and Chakravarti, A., 2000. Parallel genotyping of human SNPs using gene rich high-density oligonucleotide tag arrays. Genome Res., 10: 853-860.

Fehr, W. R., 1987. Principles of cultivar development, Vol. 1, Theory and Technique. Macmillian Publishing Company, New York.

Gerhold, D., Rushmore, T. and Caskey, T., 1999. DNA chips: promising toys have become powerful tools. Techniques, 24: 168-173.

Gibson, G., 2002. Microarrays in ecology and evolution: a preview. Molecular ecology. 11:1-17.

Gimelfarb, A. and Lande, R., 1994a. Simulation of marker-assisted selection in hybrid populations. Genetical Research, 63: 39-47.

Godwin, I. D., Mace, E. S. and Nurzuhairawaty, 2001. Genotyping pacific Island Taro (*Colocasia esculenta* (L.) Schott). In: R. J. Henry (ed.). Plant genotyping – The DNA fingerprinting of plants, 109-124. CABI Publishing, UK.

Grant, D., Cregan, P. and Shoemaker, R. C., 2000. Genome organisation in dicots: Genome duplication in Arabidopsis and synteny between soybean and Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America, 197: 4168-4173.

Guo, H., Yang, H., Mockler, T. C. and Lin, C., 1998. Regulation of flowering time by Arabidopsis photoreceptors. Science, 279: 1360-1363.

Guo, H. W., Duong, H., Ma, N. and Lin, C. T., 1999. The *Arabidopsis* blue light receptor cryptochrome2 is a nuclear protein regulated by a blue light-dependent post-transcriptional mechanism. Plant J., 19: 279-287.

Haldane, J. B. S., 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. J. Heredity, 8: 299-309.

Haley, C. S. and Knott, S. A., 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity, 69: 315-324.

Hardy, O. J., Charbonnel, N., Fréville, H. and Heuertz, M., 2003. Microsatellite allele size: a simple test to assess their significance on genetic differentiation. Genet., 163:1467-1482.

Hartl, D. L., 2000. A primer of population genetics. 3rd Edition. Sinauer Associates, Inc. Sunderland. USA. Pp1-58.

Heslop-Harrison, J. and Maluszynska, J., 1994. Molecular cytogenetics of Arabidopsis. In: E. M. Meyetowitz and C. R. Somerville (eds.). Arabidopsis, 63-87. Cold Spring Harbor Laboratory Press, USA.

Holton, T. A., 2001. Plant genotyping by analysis of microsatellites. In: R. J. Henry (ed.). Plant genotyping – The DNA fingerprinting of plants, 15-27. CABI Publishing, UK.

Howell, E. C., Barker, G. C., Jones, G. H., Kearsey, M. J., King, G. J. et al., 2002. Integration of the cytogenetic and genetic linkage maps of *Brassica oleracea*. Genetics, 161: 1225-1234.

Hulbert, S. H., Ritchter, T. E., Axtell, J. D. and Bennetzen, J. L., 1990. Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proc. Natl. Acad. Sci., USA. 87: 4251-4255.

Hyne, V., 1995. QTL analysis in segregating populations. PhD thesis. School of Biological Sciences. University of Birmingham.

Hyne, V. and Kearsey, M. J., 1995. QTL analysis: further uses of 'marker regression'. Theor. Appl. Genet., 91: 471-476.

Hyne, V., Kearsey M. J., Pike D. J. and Snape, J. W., 1995. QTL analysis: unreliability and bias in estimation procedures. Mol. Bred., 1: 273-282.

Irish, V. F. and Sussex, I. M., 1990. Function of the *apetala-1* gene during Arabidopsis floral development. Plant Cell, 2: 741-753.

Jansen, R. C., 1993. Interval mapping of multiple quantitative trait loci. Genet.135: 205-211.

Jansen, R. C., 1996. Complex plant traits: Time for polygenic analysis. Trends in Plant Science, 11: 89-94.

Jansen, R. C. and Stam, 1994. High resolution of quantitative traits into multiple loci via interval mapping. Genetics, 136: 1447-1455.

Jansen, R. C., Van-ooijen, J. W., Stam, P., Lister, C. and Dean, C., 1995. Genotype-by-environmental interaction in genetic mapping of multiple quantitative trait loci. Theor. Appl. Genet., 91: 33-37.

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., et al. 2000. Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science, 290: 344-347.

Jiang, C. and Zeng, Z. B., 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics, 140: 1111-1127.

Juenger, T., Purugganan, M. and Mackay, T. F. C., 2000. Quantitative trait loci for floral morphology in *Arabidopsis thaliana*. Genetics, 156: 1379-1392.

Kao, C. H., Zeng, Z. B., and Teasdale, R. D., 1999. Multiple interval mapping for quantitative trait loci. Genetics, 152: 1203-1216.

Kearsey, M. J., 1998. The principles of QTL analysis (a minimal mathematics approach). J. Experimental Botany, 49: 1619-1623.

Kearsey, M. J. and Farquhar, G. L., 1998. QTL analysis in plants; where are we now? Heredity, 80: 137-142.

Kearsey, M. J. and Hyne, V., 1994. QTL analysis: a simple 'marker regression' approach. Theor. Appl. Genet., 89: 698-702.

Kearsey, M. J. and Pooni, H. S., 1996. The Genetical Analysis of Quantitative Traits. Chapman and Hall, London, UK.

Keim, P., Diers, B. W., Olson, T. C. and Shoemaker, R. C., 1990. RFLP mapping in soybean: associations between marker loci and variation in quantitative traits. Genetics, 126: 735-742.

Kennard, W. C., Slowm, M. K., Figdore, S. S. and Osborn, T. C., 1994. Genetic analysis of morphological variation in *Brassica oleracea* using molecular markers. Theor. Appl. Genet., 87: 721-732.

Knapp, S. J., 2001. Mapping quantitative trait loci. In: R. L. Phillips and I. K. Vasil (eds.), DNA-based Markers in Plants, 2nd Edition, 59-99. Kluwer Academic Publishers, Netherlands.

Knapp, S. J., Bridges, W. C. and Birkes, D., 1990. Mapping quantitative trait loci using molecular marker linkage maps. Theor. Appl. Genet., 79: 583-592.

Kochert, G., 1994. RFLP Technology. In; R. L. Phillips and I. K. Vasil (eds.). DNA-based markers in plants, 8-35. Kluwer Academic Publishers, Netherlands.

Kochert, G., Halward, T., Branch, W. D. and Simpson, C. E., 1991. RFLP variability in peanut (*Arachis hypogea* L.) cultivars and wild species. Theor.Appl.Genet., 81: 565-570.

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

Koornneef, M., 1994. Arabidopsis genetics. In: E. M. Meyetowitz and C. R. Somerville (eds.). Arabidopsis, 89-120. Cold Spring Harbor Laboratory Press, USA.

Koorneef, M., Blankestijn-de Vries H., Hanhart C. J., Soppe W. and Peeters, A. J. M., 1994. The phenotype of some late-flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the Landsberg *erecta* wild-type. Plant J., 6: 911-919.

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

Koumproglou, R., 2002. Single recombinant substitutions and QTL analysis in *Arabidopsis thaliana*. PhD. thesis. The University of Birmingham, UK.

Koumproglou, R., Wilkes, T. M., Townson, P., Wang, X. Y., Beynon, J., Pooni, H. S., Newbury, H. J. and Kearsey, M. J., 2002. STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. The Plant J., 31: 1-12.

Kowalski, S. P., Lan, T., Feldmann, K. A. and Paterson, A. H., 1994a. Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveal islands of conserved organization. Genetics, 138: 499-510.

Kowalski, S. P., Lan, T., Feldmann, K. A. and Paterson, A. H., 1994b. QTL mapping of naturally-occurring variation in flowering time of *Arabidopsis thaliana*. Mol. Gen. Genet., 245: 548-555.

Ku, H. M., Vision, T., Liu, J. P. and Tanksley, S. D., 2000. Comparing sequenced segments of the tomato and Arabidopsis genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. Proceedings of the National Academy of Sciences of the United States of America, 97: 9121-9126.

Kuittinen, H., Sillanpää, M. J. and Savolainen, O., 1997. Genetic basis of adaptation: flowering time in *Arabidopsis thaliana*. Theor. Appl. Genet., 95: 573-583.

Kumar, L. S., 1999. DNA markers in plant improvement: An Overview. Biotechnology advances, 17: 143-182.

Lagercrantz, U. and Lydiate, D. J., 1996. Comparative genome mapping in *Brassica*. Genetics, 144: 1903-1910.

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

Lan, T. H. and Paterson, A. H., 2000. Comparative mapping of quantitative trait loci sculpting the curd of *Brassica oleracea*. Genetics, 155: 1927-1954.

Lan, T. H., DelMonte, T. A., Reischmann, K. P., Hyman, J., Kowalski, S., McFerson, J., Kresovich, S. and Paterson, A. H., 2000. An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Research, 10: 776-788.

Lan, T. H. and Paterson, A. H., 2001. Comparative mapping of the QTLs determining the plant size of *Brassica oleracea*. Theor. Appl. Genet., 103: 383-397.

Lande, R., 1992. Marker-assisted selection in relation to traditional methods of plant breeding. In: H. T. Stalker and J. P. Murphy (eds.). Plant Breeding in the 1990s. pro. Symp. C. A. B. International, Wallingford. UK.

Lande, R. and Thompson, R., 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics, 124: 743-756.

Lander, E. S. and Botstein, D., 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genet., 118: 185-198.

Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E. and Newbury, L., 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics, 1: 174-181.

Lange, C. and Whittaker, J. C., 2001. On prediction of genetic values in marker-assisted selection. Genetics, 159: 1375-1381.

Larsson, A. S., Landberg, K. and Meeks-Wagner, D. R., 1998. The *TERMINAL FLOWER2* (*TFL2*) gene controls the reproductive transition and meristem identity in *Arabidopsis thaliana*. Genetics, 149: 597-605.

Lee, I., Aukerman, M. J., Gore, S. L., Lohman, K. N., Michaels, S. D. et al., 1994a. Isolation of *LUMINIDEPENDENS*: a gene involved in the control of flowering time in *Arabidopsis*. Plant Cell, 6: 75-83.

Lee, I., Michaels, S. D., Masshardt, A. S. and Amasino, R. H., 1994b. The late-flowering phenotype of *FRIGIDA* and mutations in *LUMINIDEPENDENS* in the Landsberg *erecta* strain of *Arabidopsis*. Plant J., 6: 903-909.

Lee, I., Bleecker, A. and Amasino, R. M., 1993. Analysis of naturally occurring late flowering in *Arabidopsis thaliana*. Mol. Gen. Genet., 237: 171-176.

Lukens, L., Zou, F., Lydiate, D., Parkin, I. And Osborn, T., 2003. Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. Genetics, 164: 359-372.

Luo, Z. W., Thompson, R. and Woolliams, J. A., 1997. A population genetic model of marker-assisted selection. Genetics, 146: 1173-1183.

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

Mackay, T., 2001. The genetic architecture of quantitative traits. Annu. Rev. Genet., 35: 303-339.

Mansur, L. M., Lark, K. G. and Oliveira, A., 1993. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). Theor. Appl. Genet., 86: 907-913.

Martinez, O. and Curnow, R. N., 1992. Estimating the location and the size of the effects of the quantitative trait loci using flanking markers. Theor. Appl. Genet., 85: 480-485.

McGrath, J.M. and Quiros, C. F., 1991. Inheritance of isozyme and RFLP markers in *Brassica campestris* and comparison with *B. oleracea*. Theor. Appl. Genet., 82: 668-673.

Mei, R., Galipeau, P. C., Prass, C., Berno, A., Ghandour, G., Patil, N., Wolff, R. K., Chee, M. S., Reid, B. J., and Lockhart, D. J., 2000. Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. Genome Res., 10: 1126-1137.

Meyerowitz, E. M., 1994. Structure and organization of the *Arabidopsis thaliana* nuclear genome. In: E. M. Meyetowitz and C. R. Somerville (eds.). Arabidopsis, 21-36. Cold Spring Harbor Laboratory Press, Cold spring Harbor, NY.

Michaels, S. D. and Amasino, R. M., 1999. Flowering locus C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell, 11: 949-956.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T. G., Yano, M., Bhatia, C. R. and Sasaki, T., 1997. Genome, mapping, molecular markers and marker-assisted selection in crop plants. Mol. Breed., 3 (2): 87-103.

Moreau L., Charcosset, A, Hospital, F. and Gallais, A., 1998. Marker-assisted selection efficiency in populations of finite size. Genetics, 148: 1353-1365.

Ollivier, L., 1998. The accuracy of marker-assisted selection for quantitative traits with a linkage equilibrium. Genet., 148: 1367-1372.

Osborn, T. C., Kole, C., Parkin, I. A. P., Sharpe, A. G., Kuiper, M. and Lydiate, D. J., 1997. Comparison of flowering time genes in *Brassica rapa*, *Brassica napus* and *Arabidopsis thaliana*. Genetics, 146: 1123-1129.

Paran, I. and Michelmore, R. W., 1993. Development of reliable PCR-based markers linked to downey mildew resistance genes in lettuce. Theor. Appl. Genet., 85: 985-993.

Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E. and Tanksley, S. D., 1988. Resolution of quantitative traits into Mendelian factors, using a complete linkage map of restriction fragment length polymorphisms. Nature, 335: 721-726.

Paterson, A. H., Damon, S., Hewitt, J. D., Zamir, D., Rabinowitch, H. D., Lincoln, S. E., Lander, E. S. and Tanksley, S. D., 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. Genetics, 127: 181-197.

Peeters, A. J. M. and Koorneef, M., 1996. Genetic variation in flowering time in *Arabidopsis thaliana*. Cell & Development Biology, 7: 381-389.

Pérez-Pérez, J. M., Serrano-Cartagena, J. and Micol, J. L., 2002. Genetic analysis of natural variations of *Arabidopsis thaliana* vegetative leaves. Genetics, 162: 893-915.

Poduska, B., Humphrey, T., Redweik, A. and Grbic, V., 2003. The synergistic activation of *FLOWERING LOCUS C* by *FRIGIDA* and a new flowering gene *AERIAL ROSETTE 1* underlies a novel morphology in *Arabidopsis*. Genetics, 163: 1457-1465.

Poehlman, J. M. and Sleper, D. A., 1995. Breeding Field Crops. 4th Edition. Iowa State University Press, Ames, Iowa. USA.

Putterill, J., Robson, F., Lee, K., Simon, R. and Osborn, T. C., 1995. The *CONSTANS* gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell, 80: 847-857.

Quiros, G. F., 2001. DNA-based marker maps of *Brassica*. In: R. L. Phillips and I. K. Vasil (eds.). DNA-based markers in plants, 2nd Edition, 201-237. Kluwer Academic Publishers. Netherlands.

Rae, A. M., 2000. The quantitative genetics of *Brassica oleracea*. PhD. thesis. The University of Birmingham, Birmingham, UK.

Rae, A. M., Howell, E. C. and Kearsey, M. J., 1999. More QTL for flowering time revealed by substitution line in *Brassica oleracea*. Heredity, 83: 586-596.

Rafalski, J. A., 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science, 162: 329-333.

Ramsay, L. D., Jennings, D. E., Bohuon, E. J. R., Arthur, A. E., Lydiate, D. J., Kearsey, M. J. and Marshall, D. F., 1996. The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative loci. Genome, 39: 558-567.

Ribaut, J. M. and Hoisington, D., 1998. Marker-assisted selection: new tools and strategies. Trends in plant science, 3 (6): 236-239.

Rieseberg, L. H., Archer, M. A. and Wayne, R. K., 1999. Transgressive segregation, adaptation and speciation. Heredity, 83: 363-372.

Robles, P. and Micol, J. L., 2001. Genome-wide linkage analysis of Arabidopsis genes required for leaf development. Mol. Genet. Genomics, 266: 12-19.

Rogstad, N. Y., 1993. Surveying plant genomes for variable number of tandem repeat loci. In: Zimmer, E. A., White T. J., Cann R. L., Wilson, A. C. (eds.). Methods in Enzymology, 278-279. (Molecular Evolution: Producing the Biochemical Data). Academic Press.

Rossberg, M., Theres, K., Acarkan, A., Herrero, R., Schmitt, T., Schumacher, K., Schmitz, G. and Schmidt, R., 2001. Comparative sequence analysis reveals extensive microcollinearity in the lateral supressor regions of tomato, Arabidopsis, and Capella genomes. Plant Cell, 13: 979-988.

Sambrook, J., Fritsch, E. F., and Maniatis, T., 1989. A Laboratory Manual. 2nd Edition. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Salinas-Garcia, G., 1996. Mapping quantitative trait loci controlling traits in *Brassica napus*. PhD thesis, University of Birmingham.

Sax, K., 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics, 8: 552-560.

Sebastian R. L., Howell, E. C., King, G. J., Marshall D. F. and Kearsey, M. J., 2000. An integrated AFLP and RFLP *Brassica oleracea* linkage map from two morphologically distinct doubled-haploid mapping populations. Theor. Appl. Genet., 100: 75-81.

Sebastian, R. L., Kearsey, M. J. and King, G. J., 2002. Identification of quantitative trait loci controlling developmental characteristics of *Brassica oleracea* L. Theor. Appl. Genet., 104: 601-609.

Shannon, S. and Meeks-Wagner, D. R., 1993. Genetic interactions that regulate inflorescence meristem development in *Arabidopsis*. Plant Cell, 3: 877-892.

Slocum, M. K., Figdore, S. S., Kennard, W. C., Suzuki, J. Y. and Osborn, T. C., 1990. Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. Theor. Appl. Genet, 80: 57-64.

Soller, M. and Brody, T., 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet., 47: 35-39.

Song, K., Osborn, T. C. and Williams, P. H., 1990. *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs): 3. Genome relationships in *Brassica* and related genera and the origin of *B. oleracea* and *B. rapa* (syn. campestris). Theor. Appl. Genet., 79: 497-506.

Stam, P., 1993. Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. The Plant Journal, 3(5): 739-744.

Staub, J. E. and Serquen, F. C., 1996. Genetic markers, map construction and their application in plant breeding. Hort. Sci., 31: 729-741.

Stratton, D. A., 1998. Reaction norm functions and QTL-environmental interactions for flowering time in *Arabidopsis thaliana*. Heredity, 81: 144-155.

Stuber, C. W. and Edwards, M. D., 1986. Genotypic selection for improvement of quantitative traits in corn using molecular marker loci. Pro. Annu. Corn Sorghum Res. Conf., 41: 70-83.

Stuber, C. W., Lincoln, S. E., Wolff, D. W., Helentjaris, T. and Lander, E. S., 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics, 132: 823-839.

Suarez-Lopez, P., Wheatley, K., Robson, F., Onoughi, H., Valverde, F. et al. 2001. *CONSTANS* mediates between the circadian clock and the control of flowering time in *Arabidopsis*. Nature, 410: 1116-1120.

Tadege, M., Sheldon, C. C., Helliwell, C. A., Stoutjesdijk, P., Dennis, S. and Peacock, W. J., 2001. Control of flowering time by *FLC* orthologues in *Brassica napus*. The Plant J., 28(5): 545-553.

Tanksley, S. D., 1993. Mapping polygenes. Annu. Rev. Genet., 27: 205-233.

Tanksley, S. D., Medina-Filho, H. and Rick, C. M., 1982. Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Heredity, 49: 11-25.

Tanksley S. D., Young, N. D., Paterson, A. H. and Bonierbale, M. W., 1989. RFLP mapping in plant breeding: new tools for an old Science. Bio/technology, 7: 257-264.

Tanksley, S. D., Ganal, M. W., Prince, J., Devicente, M., Bonierbale, M., et al., 1992. High density molecular linkage maps of the tomato and potato genomes. Genetics, 132: 1141-1160.

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

U, N., 1935. Genome-analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jnp. J. Genet., 7: 389-452.

Ungerer, M. C., Halldorsdottir, S. S., Modliszewski, J. L., Mackay, F. C. and Puruggana, M. D., 2002. Quantitative trait loci for inflorescence in *Arabidopsis thaliana*. Genetics, 160: 1133-1151.

Van Ooijen, J. W., 1992. Accuracy of mapping quantitative trait loci in autogamous species. Theor. Appl. Genet., 84: 803-811.

Van Ooijen, J. W. and Voorrips, R. E., 2001. Joinmap Version 3.0, software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands.

Van der Schaar, W., Alonso-Blanco, C., Leonl., Kloosterziel, K. M., Jansen, R. C., van Ooijen, J. W. and Koorneef, M., 1997. QTL analysis of seed dormancy in Arabidopsis using recombinant inbred lines and MQM mapping. Heredity, 79: 190-200.

Vodenicharova, M., 1982. Use of proteins as molecular-genetic markers in plants. Genet. Sel., 22: 269-277.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Vandelee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Jurper, M. and Zabeau, M., 1995. AFLP – a new technique for DNA fingerprinting. Nucleic Acids Research, 377: 495-500.

Wang, G. L., Mackill, D. J., Bonman, J. M., McCouch, S. R. and Nelson R. J., 1993. RFLP mapping of genes conferring complete and partial resistance in a rice cultivar with durable resistance to blast. In: Jacobs T., Parlevliet J. E. (eds). Durability of disease resistance, 209-215. Kluwer Academic Publishers, The Netherlands.

Wang, R., Stec, A., Hey, J., Lukens, L. and Doedley, J., 1999. The limits of selection during maize domestication. Nature, 398: 236-239.

Warwick, S. I. and Black, L. D., 1991. Molecular systematics of *Brassica* and allied genera (subtribe Brassicinae, Brassiceae) – Chloroplast genome and cytodeme congruence. Theor. Appl. Genet., 82: 81-92.

Weber, W. E. and Wricke, G., 1994. Genetic Markers in Plant Breeding. Paul Parey Publishers. Berlin and Hambury.

Weinig, C., Ungerer, M. C., Dorn, L., A., Kane, N. C., Toyonaga, Y., Halldorsdottir, S. S., Mackay, T. F. C., Purugganan, M. D. and Schmitt, J., 2002. Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. Genetics, 162: 1875-1884.

Weising, K., Kaemmer, D., Ramser, J., Bierwerth, S. and Kahl, G., 1992. Plant DNA fingerprinting with simple repetitive oligonucleotides. DNA polymorphisms in eukaryotic genomes. Adv. Mol. Genet., 5: 135-156.

Weising, K., Winter, P., Huttel, B. and Kahl, G., 1998. Microsatellite markers for molecular breeding. J. Crop Prod., 11: 113-143.

Weller, J. I., Soller, M. and Brody, T., 1988. Linkage analysis of quantitative traits in an interspecific cross of tomato (*L. esculentum* x *L. pimpinellifolium*) by means of genetic markers. Genetics, 118: 329-339.

Westman, A. L. and Kresovich, S., 1997. Use of molecular marker techniques for description of plant genetic variation. In: Callow, J. A., Ford-Lloyd, J. A. and H. J. Newbury (eds), Biotechnology and Plant Genetic Resources: Conservation and Use, pp. 9-48. CAB International, Oxford. UK.

Whittaker, J. C., Curnow, R. W., Haley, C. S. and Thompson, R., 1995. Using marker-maps in marker-assisted selection. Genet. Res., 66: 255-265.

Xie, G. and Xu, S., 1998. Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. Heredity, 80: 489-498.

Young, N. D., 1994. Plant gene mapping. Encyclopedia of Agricultural Science, 3: 271-282.

Young, N. D., 2001. Constructing a plant genetic linkage map with DNA markers. In: R. L. Phillips and I. K. Vasil (eds.), DNA-based markers in plants, 31-47. Kluwer Academic Publishers, Netherlands.

Zeng, A. B., 1994. Precision mapping of quantitative trait loci. Genetics, 136: 1457-1468.

Zhang, W. and Smith, C., 1992. Computer simulation of marker-assisted selection utilising linkage disequilibrium. Theor. Appl. Genet., 83: 813-820.