FINE MAPPING OF QTL AND MICROARRAY GENE EXPRESSION STUDIES IN ARABIDOPSIS USING STAIRS

By

SURIYA ARACHCHIGE CHANDRIKA NISHANTHI PERERA

A thesis submitted to

The University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

The University of Birmingham

December 2004

Abstract

QTL mapping with segregating populations results in poor map resolution which limits the applicability of mapped QTL in further research such as gene cloning. The current research project aimed mainly at developing STepped Aligned Recombinant Inbred Strains (STAIRS) covering the top region of chromosome 3 and demonstrating the feasibility of using STAIRS in high resolution mapping of QTL in *Arabidopsis*. The top region of chromosome 3 of *Arabidopsis* had been reported to house QTL related to flowering time. This region was first saturated with 24 polymorphic microsatellite markers and 23 narrow STAIRS were produced within the region via a marker-assisted backcross breeding programme using whole chromosome substitution lines. The analysis of QTL with the narrow STAIRS revealed a major pleiotropic QTL within 2-3 cM affecting flowering time, leaf number at day 20 and rosette and cauline leaf numbers at flowering. A second QTL with less but opposite effect on the same traits were located within 15-20 cM. The search for candidate genes within 2-3 cM of chromosome 3, to locate possible candidate genes revealed COL-2, CONSTANS-Like gene which affects flowering time. Microarray gene expression profiling was performed using the two genotypically closest lines which differ for flowering time to compare the two lines at the same chronological and physiological ages in two experiments respectively. The lists of differentially expressed genes were obtained from the two experiments. Differential expression was observed for the possible candidate gene in the latter experiment. The results emphasized the power of STAIRS in fine mapping of QTL and the possibility of using them in transcriptional profiling to study the expression of genes.

ACKNOWLEDGEMENT

I express my deep gratitude to Prof. Mike Kearsey, my supervisor, for his excellent guidance, valuable advice and enormous support throughout the project. I thankfully acknowledge the constructive comments of Dr. Harpal S. Pooni regarding the reports and data analysis.

I am indebted to Dr. Timothy Wilkes for his guidance in the laboratory, and constant support and encouragement all throughout the project. The immense assistance I received from Dr. Rachel Machado in conducting and analysing the microarraying experiments is gratefully acknowledged.

I would like to express my sincere gratitude to Ms. Sue Bradshaw for her generous help with the technical matters, the data collection and in numerous other tasks. I thankfully acknowledge the great help from the plant house staff specially Karen and Michael in growing Arabidopsis on a large scale.

I gratefully acknowledge the scholarship award for a PhD studentship from the ADB - Science and Technology Personnel Development Project in Sri Lanka. I am grateful to Dr. Mrs. C. Jayasekera, Director/CRI for her help with coordinating matters related to the scholarship award.

I am much obliged to Mr. Everard Jayamanne, Head of the Genetics and Plant Breeding Division of the Coconut Research Institute, Sri Lanka for his valuable guidance when a part of this research was carried out in Sri Lanka and for his enourmous support and encouragement throughout. I am thankful to my colleagues Champa and Lalith at CRI/SL for their support and advice on related issues. I greatly appreciate the help from Sandya, Nissanka and Ayodya at CRI/SL and all the staff of the Genetics & Plant Breeding division CRI/SL for helping me in various ways.

Many thanks to all my friends and fellow students Champa, Dave, Ankush , Bharti, Angela, Abdul, Hooman, Saeed, Cristina, Joana, Ali, Marianne, and Xiao for their friendship and inspiring lunch time conversations.

I am deeply indebted to my parents for their generous love and encouragement all throughout my life and, my brothers Sidath and Sudath and Sujeewa, Duleeka and Danushka for keeping me happy in our close family.

Affectionately

Dedicated to

My family

TABLE OF CONTENTS

Abstract ii

Acknowledgement iii

Table of Contents v

List of Tables xiii

List of Figures xv

CHAPTER 01

Introduction and Literature review **1**

1.1 Introduction 1

1.2 Objectives 4

1.3 Review of Literature 5

1.3.1 Arabidopsis thaliana as a model plant in basic research 5

1.3.2 Genetic Markers 7

1.3.3 Morphological Markers 8

1.3.4 Molecular markers 8

1.3.5 Advantages of Molecular Markers over Morphological markers 9

1.3.6 DNA Markers 10

1.3.7 Genetic Analysis in Plants 13

1.3.8 Qualitative traits 14

1.3.9 Quantitative traits and Quantitative trait loci (QTL) 14

1.3.10 QTL analysis 16

1.3.11 Gene mapping/QTL Mapping 17

1.3.12 Tracking QTL with single gene markers 18

1.3.13 QTL mapping with molecular markers 18

1.3.14 Types of maps 20

1.3.15 Mapping populations 22

1.3.16 Principles of genome mapping 24

1.3.17 Mapping software 27

1.3.18 Basic steps of a linkage analysis 28

1.3.19 Application of molecular markers and QTL analysis in Arabidopsis 30
1.3.20 Problems of QTL mapping with segregating populations 32
1.3.21 High resolution QTL mapping 34
1.3.22 Chromosome Engineering 36
1.3.23 Chromosome Substitution Strains and STAIRS in Arabidopsis 39
1.3.24 Microarray Gene Expression 40

CHAPTER 02
Microsatellites Marker Saturation of Top of Chromosome 3 in Arabidopsis 42
Abstract 42
2.1 Introduction 43
2.1.1 What are Microsatellites? 43
2.1.2 Isolation of SSRs 43
2.1.3 Detection of Microsatellite Polymorphism 44
2.1.4 Location of Microsatellites in Arabidopsis 45
2.1.5 Need for marker densification of chromosome 3 of Arabidopsis 45
2.2 Objectives 46
2.3 Materials and Methods 47
2.3.1 Location of Microsatellites and designing of flanking primers 47
2.3.2 Checking for polymorphism between Col and Ler and Optimizing PCR conditions 48
2.4 Results 49
2.4.1 Location of Microsatellites within the first 20 cM and the distal end of Chromosome 3 49
2.4.2 Designing flanking primers for the microsatellites 50
2.4.3 Optimizing PCR conditions for new markers and checking the polymorphism among accessions 50
2.5 Discussion 50

2.5.1 Factors considered in Primer Designing 50
2.5.2 The Types and Abundance of SSR on top of Chromosome 3 51
2.5.3 Marker Coverage of the Chromosome with Microsatellites 52

CHAPTER 03
Evaluation of Quantitative Traits in Arabidopsis Accessions Columbia and Landsberg 62
3.1 Background 62
3.2 Objectives 64
3.3 Materials and Methods 64
3.3.1 Plant house trial for scoring quantitative traits 64
3.3.2 Traits scored 64
3.3.3 Data analysis 65
3.3.4 Genotyping the lines 65
3.4 Results and Discussion 66
3.4.1 Analysis of Morphological Quantitative Traits 66
3.4.2 Genotyping the lines with microsatellite markers 67

CHAPTER 04
QTL Analysis Using Wide STAIRS of Chromosome 3 in Arabidopsis thaliana 70
Abstract 70
4.1 Introduction 71
4.2 Objectives 73
4.3 Materials and Methods 74
4.3.1 Genotyping wide STAIRS 74
4.3.2 Plant house experiment for scoring quantitative trait phenotypes 77
4.4 Results 79
4.4.1 Summary of the results of genotyping 79
4.4.2 Analysis of quantitative trait variation 79

4.4.3 Least Squares Model fitting for significantly different genotypes 80
4.4.4 Correlation coefficients 83
4.5 Discussion
4.5.1 Location of QTL for traits scored 84
4.5.2 Correlations among characters 85
4.5.3 Differences within lines of same STAIRS 86
4.5.4 STAIRS in QTL location 86
4.5.5 Conclusions 87
CHAPTER 05
Construction of Narrow STAIRS via a Marker Assisted Breeding Programme **101**
Abstract 101
5.1 Introduction 102
5.2 Objectives 105
5.3 Materials and Methods 106
5.3.1 Breeding programme 106
5.3.2 Genotyping the second BC progeny 108
5.3.3 Genotyping to identify homozygous SRLs (STAIRS) 110
5.3.4 Calculation of recombination frequencies 112
5.4 Results 114
5.4.1 Genotyping of the second BC progeny 114
5.4.2 Genotyping to identify homozygous single recombinants (STAIRS) 115
5.4.3 Calculation of recombination frequencies and genetic distances between markers 116
5.5 Discussion 117
5.5.1 Narrow STAIRS produced 117
5.5.2 Expected vs. observed recombination frequencies 118
5.5.3 Single recombinants outside the region of interest 119

CHAPTER 06

Fine Mapping of QTL Using Narrow STAIRS on top of Chromosome 3 in Arabidopsis **125**

Abstract 125

6.1 Introduction 126

6.1.1 Importance of fine mapping 126

6.1.2 Examples for different approaches of fine mapping 126

6.1.3 Zooming in on the top region of chromosome 3 for mapping of flowering related QTL 128

6.2 Objectives 128

6.3 Materials and Methods 129

6.3.1 Plant material and experimental design 129

6.3.2 Data analysis 130

6.3.3 Search for candidate genes 131

6.4 Results 131

6.4.1 Means of traits 131

6.4.2 Analysis of Variance and Least Squares Model Fitting 131

6.4.3 Pearson's Correlation Coefficients 135

6.5 Discussion and conclusions 135

6.5.1 Removal of two lines from the analysis 135

6.5.2 Reasons for significant differences within the STAIRS in marker bins 136

6.5.3 Difficulty of fitting models for certain traits 137

6.5.4 Location of QTL and comparison of the mapped QTL with results from the analysis of wide STAIRS 138

6.5.5 Resolution of the mapped QTL 140

6.5.6 Gene action 141

6.5.7 Search for candidate genes 141

6.5.8 Significance of QTL mapping with narrow STAIRS 141

CHAPTER 07

Analysis of Heterosis in Chromosome Substitution Strains in Arabidopsis thaliana **154**

Abstract 154

7.1 Introduction 155

7.1.1 What is heterosis? 155

7.1.2 Causes of heterosis 155

7.1.3 Use of heterosis in agriculture 156

7.1.4 Use of CSSs of Arabidopsis in studying heterosis 157

7.2 Objectives 157

7.3 Materials and Methods 158

7.3.1 Plant material 158

7.3.2 Traits scored 158

7.3.3 Determination of the presence of heterosis 158

7.3.4 Calculation of genetic components of means 159

7.4 Results 160

7.4.1 Detecting heterosis 160

7.4.2 Calculating genetic components of means and potence ratios 161

7.5 Discussion and conclusions 161

7.5.1 Traits showing heterosis 161

7.5.2 Inequality of population variances 162

7.5.3 Genetic components of the means and potence ratios 162

7.5.4 General comments 163

CHAPTER 08

Transcriptional Profiling of Two Narrow STAIRS Differing for Flowering-time with Microarray DNA Chips **170**

Abstract 170

8.1 Introduction 171

8.1.1 Microarrays 171

8.1.2 Implications of microarrays 171

8.1.3 Microarrays and gene expression 173

8.1.4 Microarrays in Arabidopsis 173

8.1.5 Design of arraying experiments 174
8.1.6 Basic steps in a microarray gene expression experiment 175
8.1.7 Microarray expression data analysis 176
8.1.8 Gene expression of two STAIRS differing in a small chromosome region 178
8.2 Objectives 178
8.3 Materials and Methods 178
8.3.1 Growing of plants for sample collection 178
8.3.2 Sample collection 179
8.3.3 Determination of time points for studying gene expression 179
8.3.4 Extraction and quantifying total RNA 180
8.3.5 The MWG oligonucleotide array 181
8.3.6 Experimental plan 181
8.3.7 cDNA synthesis and labelling 181
8.3.8 DNA microarray hybridisations 183
8.3.9 Slide washes 184
8.3.10 Array scanning and image analysis 185
8.3.11 Data analysis 185
8.3.12 Obtaining a list of genes within the region the two lines differ 187
8.4 Results 187
8.4.1 Gene lists at different stages 187
8.4.2 Differentially expressed genes 188
8.4.3 Genes list in the polymorphic region and the candidate genes 188
8.5 Discussion 189
8.5.1 Genes list in the polymorphic region for the two STAIRS 189
8.5.2 Differentially expressed genes 190
8.5.3 Gene expression of the possible candidate gene 191
8.5.4 Suggestions for further research 192
CHAPTER 09
Summary and Future Work **196**
REFERENCES 201
APPENDICES 213

Appendix 01: DNA Extraction Method 213
Appendix 02: Tables of ANOVAS of chapter 03 214
Appendix 03: ANOVAS for the quantitative traits in chapter 04 222
Appendix 04: Results of Model fitting for wide STAIRS in chapter 04 225
Appendix 5a : An Example of Least square model fitting from chapter 04 228
Appendix 5b: Line means for morphological traits of wide STAIRS 228b
Appendix 6a : ANOVAS for morphological traits in narrow STAIRS 229
Appendix 6b : Line Means for narrow STAIRS 229b
Appendix 07 : RNA extraction procedure 234
Appendix 08: Significant, differentially expressed gene lists from expression analysis (GeneSpring) experiment 235
Appendix 09 : Genes list within the polymorphic region (2-3 cM on top of chromosome 3) for the 2 STAIRS used in gene expression analysis 249

LIST OF TABLES

Table 2.1	The clones containing sequence repeats, the type and The number of each repeat and their cM distances and the physical base pair positions	53
Table 2.2	The number, clone, sequences and the product size of designed flanking primers for the microsatellites	55
Table 2.3	The polymorphisms for Col & L*er* and Nd the opt.ann. and Mg concentrations for microsatellites	57
Table 2.4	SSRs on top 20 cM of chromosome 3 of *Arabidopsis* and their polymorphism for Col and L*er*	59
Table 3.1	Microsatellite markers used in genotyping accessions Col-0 and Col-5 and their positions	68
Table 3.2	The means of the three lines for the 21 traits, standard errors of the means and significance levels from ANOVA	69
Table 4.1	Genotype scores of WIDE stairs for the nine Microsatellites Markers	89
Table 4.2	The means and the standard deviations of the traits scored for each genotype	90
Table 4.3	The model used for detecting and estimating genetic differences among wide STAIRS	91
Tables 4.4	The *significant* parameters and their estimated values, standard errors and normal deviate (c) for each of the traits	92
Table 4.5	Pearson correlation coefficients among traits	94
Table 5.1	The markers used for the calculation of recombination frequencies, recombinants and calculated genetic distances	120
Table 6.1	The twenty six distinct lines (STAIRS) introgressed regions of narrow STAIRS	143
Table 6.2	The model used for detecting and estimating genetic differences among narrow STAIRS	144
Table 6.3:	The genetic and physical distances of the regions explained in the model for wide STAIRS	145

Tables 6.4	The *significant* parameters and their estimated values, standard errors and normal deviate	146
Tables 6.5	The genotypes, number of individuals, observed means, calculated means and the expected means in narrow STAIRS	148
Table 6.6	Pearson's correlation coefficients between the traits	150
Table 7.1	Means and the standard errors of the means of narrow STAIRS	166
Table 7.2:	The mid parental values, additive (a_A) and dominance (d_A) genetic components of means, and the potence ratio (d_A/ a_A)	167
Table 8.1	The map lengths of each chromosome in *Arabidopsis*, total no. of genes present and the nos. of differentially expressed genes	193

LIST OF FIGURES

Figure 2.1	The coverage of markers outside the region of interest	60
Figure 2.2	The coverage of first 20 cM of chromosome 3 with microsatellite markers	61
Figure 4.1	CSSs of *Arabidopsis*	95
Figure 4.2	Structure of STAIRS	96
Figure 4.3	Stages of Fine mapping Using CSSs and STAIRS	97
Figure 4.4	Wide STAIRS: marker bins, genotypes,lines within Genotypes and introgressed and crossover regions	98
Figure 4.5	Graphical representation of means of traits for each genotype	99
Figure 5.1	Breeding programme used by Koumproglou et al. to produce CSSs and STAIRS	121
Figure 5.2	Genotypes of the Parents and the Progeny in the backcross in producing STAIRS	122
Figure 5. 3	The breeding programme adopted for the production of narrow STAIRS in chromosome 3	123
Figure 5. 4	Structure of STAIRS produced on top of chromosome 3	124
Figure 6.1	The morphological differences in two plants each from early flowering lines and late flowering lines.	151
Figure 6.2	Graphical representation of means of different traits for each genotype	152
Figure 7.1	Graphical representation of the means of different populations for each trait	168
Figure 8.1	The GeneSpring image of all the genes present in the expression analysis experiment 01	194
Figure 8.2	The GeneSpring scatter plot image of all the genes present in the expression analysis experiment 02	195

CHAPTER 01

INTRODUCTION AND REVIEW OF LITERATURE

1.1 Introduction

Knowledge about the genes controlling various economically important traits in plants is important in crop improvement programmes. The availability of genetic information about major and minor traits and their interactions increases the efficiency and probability of success in producing plants with desired attributes. Detailed genetic maps of organisms generate a vast amount of precise information which plant breeders can use in order to identify, manipulate and complement traits to their maximum advantage (Allen 1994).

The genes governing most of the economically important traits of eukaryotes, such as yield in crop plants and intelligence in humans are extremely complex. The phenotypes of such traits display continuous variation and are conditioned by allelic variation at several genetic loci, each with a relatively small effect compared with the effect of the environment. Such characters are often referred to as quantitative traits and their inheritance as polygenic. The individual loci controlling a quantitative trait are referred to as Quantitative trait loci or QTL (Tanksley 1993).

Traditionally, the genetic analysis of quantitative traits has been restricted to the statistical approaches of Biometry. However, these methods deal with the average effects of unresolved loci and provide little information about the number, location or relative effects of specific QTLs (Kearsey and Pooni 1996).

Early attempts to map the polygenic traits of interest were based on the linkage of such QTL to single gene morphological markers. But for most of the organisms of interest only a few single gene markers had been mapped and also most of the mapped ones were not suitable to

study QTLs. With the introduction of molecular markers in genome mapping, the studying of QTLs entered a new era of scientific research. The molecular markers have made it feasible to map and characterize the polygenes underlying quantitative traits in natural populations. The genetic dissection of a quantitative trait is termed QTL analysis, and is usually carried out with segregating populations using interval mapping or a related method (Vanooijen 1992).

The main problem with this approach is the difficulty of locating QTL with sufficient accuracy. Although the precision of this approach may be adequate for marker aided selection, it is far too low for detailed genetical analysis, chromosome walking or for map based cloning of the genes (Hyne and Kearsey 1995; Kearsey and Pooni 1996; Kearsey and Farquhar 1998). In order to achieve higher accuracy in locating QTL in segregating populations by this approach, the character needs to be highly heritable and often unrealistically large populations need to be scored (Kearsey and Farquhar 1998). Even under such conditions confidence intervals of the mapped QTL are seldom less than 5 cM, which is still too long a section of a chromosome for identifying candidate genes.

The more precise mapping of QTL requires the construction of chromosomes of defined constitution. This concept of engineered or 'designer' chromosomes allows more precise and reliable location of QTL. The basic principle underlying these approaches is to create two genotypes which are identical apart from a defined region on a particular chromosome (Howell, Marshall, and Lydiate 1996; Kearsey and Pooni 1996). The homozygous lines of different plants produced in such a way are called part-chromosome substitution lines. Ideally these part-chromosome substitution lines should carry a single defined segment of a novel (donor) genotype and have a genetic background of a distinct (recurrent) genotype.

In 2002 (Koumproglou et al. 2002)reported the production of whole chromosome substitution strains (CSSs) in the model plant *Arabidopsis thaliana* at the University of Birmingham. They also explained the production of a novel resource, STepped Aligned Inbred Recombinant Strains (STAIRS) of *Arabidopsis* and the usage of these resources in fine mapping of QTL. Furthermore, Koumproglou et al (2002) reported the presence of QTL for flowering time in *Arabidopsis* on the top 20 cM of chromosome three using these novel resources CSSs and wide STAIRS.

Arabidopsis thaliana, which is a member of the family *Brassicaceae*, has been chosen as the model plant in this study. *Arabidopsis* offers important advantages for basic research in genetics and molecular biology and has become the model plant in biological research at present (Meyerowitz and Sommerville 1994).

The first part of the current research involves the production of narrow STAIRS within the top region of chromosome 3 of *Arabidopsis* for fine mapping of QTL in that region with special emphasis on QTL for flowering time. The latter part of this research involves the gene expression profiling of the investigated QTL using micro-array biochip technology. Gene expression profiling has become an invaluable tool in functional genomics and at present microarray technology is the most favoured technology in carrying out gene expression studies under various conditions (Schena and Davis 2000).

1.2 Objectives

1) Production of narrow STAIRS on top of chromosome three in *Arabidopsis* using a marker assisted breeding programme.

2) Investigation for new polymorphic markers on the top region of chromosome three to facilitate the genotyping of the different generations of plants in producing narrow STAIRS.

3) Genotyping and QTL analysis of the existing wide STAIRS of chromosome three to validate the presence of QTL on the top region of chromosome three.

4) QTL analysis using narrow STAIRS on top of chromosome three to achieve fine mapping of relevant QTL.

5) Analysis of hybrid vigour related to individual chromosomes in *Arabidopsis* using the available CSSs, Col and the respective F_1s.

6) Gene expression profiling of the investigated QTL using DNA microarray biochip technology with special emphasis on flowering time QTL.

1.3 Review of Literature

1.3.1 *Arabidopsis thaliana* as a model plant in basic research

Arabidopsis thaliana is a small herb belonging to the family *Brassicaceae*. It occurs naturally throughout temperate regions of Europe, Asia and North Africa and has been widely introduced to other areas including North America and Australia (Wilson 2000). Although *Arabidopsis* has no economic value, it has become one of the most important model plants in biological research, because it offers certain important advantages for basic research in genetics and molecular biology (Meyerowitz and Sommerville 1994). These special features can be listed as,

- Small genome (125 Mb)
- Low DNA content with little interspersed repetitive DNA.
- A diploid with only 5 pairs of chromosomes.
- A rapid life cycle. (about 6 weeks from germination to mature seed when grown at 25^0C with 16 hours of day length.)
- Prolific seed production and easy cultivation in restricted space due to the comparatively smaller size of the plant.
- Normally self pollinating but easily cross-pollinated.
- Efficient transformation methods.
- Availability of a large number of genetic resources and abundance of mutants for genetic studies.
- Availability of the complete genome sequence information for ecotype Col. (www.arabidopsis.org)

These advantages make this small herb one of the most favoured choices among scientists for the use in molecular genetic research as a model plant. The purpose of an experimental

'model system' is to efficiently understand the organisms that are being modelled. *Arabidopsis* differs from several of the other important model systems in at least one key respect. *Arabidopsis* is closely related to the angiosperms; the species it models. All the angiosperms have evolved from a common ancestor within the last 150 million years. Because of this relatively recent evolution, the average *Arabidopsis* gene can be confidently expected to functionally replace a homolog in many other angiosperms. This implies that there are many opportunities to exploit detailed knowledge about *Arabidopsis* to gain insights into similar processes in other plants (Meyerowitz and Sommerville 1994).

Therefore, in addition to the ease of the use of *Arabidopsis* in molecular genetic research, there is a high potential for transferring the findings from *Arabidopsis* to other angiosperms and specially the plant species with economic importance (Rae, Howell, and Kearsey 1999). Comparisons of the *Arabidopsis* genomic sequence with sequences from other flowering plants have revealed that substantial collinearity exists between species in the arrangement of genes within chromosomal blocks. These homologies should prove to be of value in exploiting the *Arabidopsis* sequence to identify candidate genes in defined chromosomal regions within genomes that are less well characterized (Barnes 2002).

Over the years much research work has been carried out to investigate the genome collinearity between *Arabidopsis* and its closely related crop species *Brassica*. Fine scale *Arabidopsis* / *Brassica* comparative mapping demonstrated short-range collinearity between the genomes of *Arabidopsis* and *Brassica* proving the potential for identification the candidate genes in *Arabidopsis* homologous to genes controlling important agronomic traits in *Brassica* (Lagercrantz et al. 1996; Bohuon et al. 1998). Furthermore, eleven regions of conserved organization have been found between *Arabidopsis thaliana* and *Brassica oleracea* covering 158.2 cM (24.6 %) of *Arabidopsis* genome and 245 cM (28.9 %) of *Brassica* genome

(Kowalski et al. 1994). Kole *et al*. (Kole et al. 2001) reported that *VFR2,* one of the QTL controlling vernalization-responsive flowering time in *Brassica napus* is homologous to a region on chromosome 5 of *Arabidopsis* that contain several flowering time QTL. Mapping of QTL controlling flowering time in *Brassica napus* has identified genomic regions that contain homologues of the *CONSTANS (CO)* gene, which promotes flowering time in *Arabidopsis* (Robert et al. 1998).

Research has shown that sequence synteny exists between other crop species in addition to the genome sequence similarities between *Arabidopsis* and related species of Brassicaceae. Comparative analyses with the rice genome sequence revealed that a majority of the *Arabidopsis* components of the flowering pathway are present in rice (Izawa, Takahashi, and Yano 2003). In addition, functional analyses in rice demonstrated that key regulatory genes for flowering time (known as 'heading' in rice) are conserved between rice and *Arabidopsis* (Hayama and Coupland 2003). Furthermore, (van Nocker et al. 2000) identified a gene from maize that is closely related to the *Arabidopsis LUMNIDEPENDENS (LD),* which is a gene that encodes a potential transcriptional regulator that acts as a positive effecter of flowering. Ten flowering time gene homologues that have conserved, corresponding sequences in *Arabidopsis* were identified in Soya bean (Tasma and Shoemaker 2003).

Therefore, in summary the combination of the two factors - the ease of use of *Arabidopsis* in research and the transferability of the information from *Arabidopsis* to economically important plants (Fourmann et al. 2002) - has made *Arabidopsis* the most important model plant at present in genetic research.

1.3.2 Genetic Markers

The genetical analysis of variation or the underlying polymorphism involves observing the differences among individuals by various different methods. During the first half of the 20th

century major mutant variations, such as differences in morphology, anatomy or behaviour were used to identify the variation among individuals. Later, it became possible to observe the variation at the structure of polypeptide levels. Since 1980s methods have been developed to explore the variation at the fundamental level of DNA itself. These differences; varying from gross morphological changes to subtle differences in the DNA sequence can be used as genetic markers to identify chromosomal regions in genetic analysis studies (Kearsey and Pooni 1996; Peleman and van der Voort 2003).

1.3.3 Morphological Markers

Morphological markers were the only possible option as genetic markers in the early 19th century. Although morphological makers were comparatively easy to detect there were a number of disadvantages involved with them. To begin with, in any species the number of morphological differences that could serve as morphological markers were very limited. Also, they were affected by the environment and did not necessarily represent the true genetic variation. The majority of morphological markers exhibited themselves only at certain stages of the life cycle. As a result, the progress that could be made in genetic analysis using morphological markers was limited and very slow.

1.3.4 Molecular markers

A molecular marker is a readily identifiable protein or a piece of DNA typically detected as a band on a gel after electrophoresis. They show polymorphism that allows identification of different alleles at a genetic locus. The major molecular marker types include, iso-enzymes, RFLPs, RAPDs, SSRs, AFLPs and SNPs. Isoenzyme methods depend upon the electrophoretic separation of proteins in a non-denaturing gel, followed by enzyme specific staining which allows the visualization of bands of coloured reaction produce (Hamrick and

Godt 1990; Delourme and Eber 1992). This is a robust technique and the marker is co-dominant; i.e. both alleles can be scored in a heterozygote, but the number of markers is limited by the number of enzyme specific stains that are available (Kearsey and Luo 2003). The other markers listed above among molecular markers are based on the variation in DNA sequence level and display wide polymorphism in nature. The discovery of molecular markers, especially DNA markers, revolutionized the subject of genetics with their wide applicability and the huge advantages over morphological markers, which are listed below.

1.3.5 Advantages of Molecular Markers over Morphological Markers

1. Not affected by the environment
2. Focus on the variation on the genetic material itself.
3. Do not depend on plant being at a specific part of its life cycle.
4. Usually require only small amounts of tissue.
5. Display high degree of polymorphism.
6. Co-dominant inheritance detected by some.
7. Representative of all regions of the genome
8. Not pleiotropic.
9. Highly reproducible and transferable.

Due largely to the above advantages, molecular markers have been widely used in research related to many sub-disciplines of genetics such as genome mapping, genetic conservation studies, evolutionary discoveries etc. and also other disciplines in biology such as Pathology and Microbiology etc (Dekkers and Hospital 2002).

1.3.6 DNA markers

1.3.6.1 Restriction Fragment Length Polymorphism (RFLP)

Restriction Fragment Length Polymorphism (RFLP) was the first DNA marker system to be invented. RFLP is a method for detecting polymorphism between genotypes using the sizes of DNA restriction fragments as markers. The scoring of RFLP requires DNA sequences that can be used as probes and these probes are mostly clones from cDNA libraries of the relevant species. These probes are labelled and hybridised to homologous sequences on restricted-DNA fragments produced by the digestion of genomic DNA using a restriction enzyme such as *Eco*R1. These fragments are then separated by gel electrophoresis, transferred to a membrane and detected, usually, by auto-radiography. This is a robust and transferable technique in addition to being a co-dominant marker. The comparative disadvantages of RFLP are, the need for preliminary work in assembling probes, the requirement of relatively large amounts of DNA (about mg), the original requirement of radioactivity (although fluorescent labelling is now available) and the time consuming nature. RFLP was widely used in genetic studies (Kleinhofs et al. 1993; Du and Hart 1998) until the Polymerase Chain Reaction (PCR) based marker systems, which mitigate some of the disadvantages of RFLP, were developed.

1.3.6.2 Random Amplified Polymorphic DNA (RAPD)

Random Amplified Polymorphic DNA (RAPD) is a technique that amplifies anonymous stretches of DNA using arbitrary primers (Williams et al. 1990). RAPDs were the first invented PCR based marker system and is also known as AP-PCR: arbitrarily primed polymerase chain reaction; DAF: DNA amplification fingerprinting and MAAP: multiple arbitrary amplicon profiling. RAPD procedure involves extraction of genomic DNA, PCR using single short primer (usually 10-mers available in kits), electrophoresis in agarose gel

and DNA staining using Ethidium bromide. RAPD is a relatively simple procedure that does not involve radioactivity. Furthermore, there are an infinite number of marker bands available in RAPD, it requires very little preliminary work and needs very small amounts of DNA. On the other hand RAPDs are dominant markers and hence the heterozygosity cannot be differentiated from dominant homozygotes. In addition, the results are very sensitive to experimental conditions and hence difficult to transfer between laboratories. Yet, this molecular marker system was widely used in various researches about a decade ago.

1.3.6.3 Amplified Fragment Length Polymorphism (AFLP)

Amplified Fragment Length Polymorphism (AFLP) is a DNA marker that is heavily used in genetic studies at present. AFLP combines Restriction Fragment Length Polymorphism and PCR (Vos et al. 1995). This initially involves DNA extraction, and cutting DNA with two restriction enzymes (e.g. *Eco*R1, and *Mse*1) creating differing 'sticky ends'. Then different adapters, which are short double stranded DNA sequences, are added and this is called adapter ligation, followed by direct amplification of the fragments with primers specific to the two adapters. This pre-amplification leads to the amplification of every restriction fragment resulting in an uninformative smear of bands on a gel. However, in the final amplification only a subset of these bands is amplified due to the usage of primers carrying short extensions at their 3' end that result in selective amplification of available fragments. One of the primers used in amplification is usually radioactively labelled and the amplified products are run on poly-acrylamide sequencing gels. At the end, a large number of bands can be scored upon subjecting to autoradiography. At present the markers can be fluorescently labelled and automated DNA sequencers are used for the separation and scoring of amplification products. But AFLP is a dominant marker making it difficult to detect heterozygosity in addition to

being rather more complex than most of the other PCR based marker systems. AFLP has widely been used in molecular marker studies in plants, in mapping and especially in genetic diversity studies (Aggarwal et al. 2002; Mignouna et al. 2002)

1.3.6.4 Simple Sequence Repeats – Microsatellites (SSR)

Simple Sequence Repeats (SSR) are also known as Microsatellites, STMS - sequence tagged microsatellites, SSLP- simple sequence length polymorphism and SSRLP - simple sequence repeat length polymorphism. SSRs are a very widely used marker type and they rely upon the very high rate of polymorphism observed at microsatellite loci (Bell and Ecker 1994). These are tandem repeats of short units of usually one to four bases and are found to be highly abundant within eukaryotic genomes. In the microsatellite procedure, these simple-sequence repeats are amplified by PCR using locus specific, complementary primers that anneal to sequences flanking the repeat region. The amplification is followed by gel electrophoresis to detect alleles as bands differing in distance of mobility depending on the length of the SSR. The advantages of SSR are their co-dominant nature, the requirement of only small amounts of genomic DNA and the very high rate of available polymorphism. The possibility of automation is another advantage with the SSR markers (Macaulay et al. 2001). However, a large amount of preliminary work - i.e. DNA sequencing information around the microsatellite - is needed to define primers to begin routine microsatellite analysis. In addition, it may be necessary to use either complex polyacrylamide or expensive Spreadex gels in order to visualize polymorphism.

Microsatellites have been extensively used in molecular research over the last decade in many plant species including *Arabidopsis* (Bell and Ecker 1994; Loridon et al. 1998; Virk et al. 1999) in genetic mapping of *Arabidopsis* and a variety of other crops (Guadagnuolo et al. 2001; Sourdille et al. 2001; Song, Fickus, and Cregan 2002; Shen, Kong, and Ohm 2004)

Both AFLP and SSR marker systems can now be automated and multiplexed. By attaching fluorescent labelled primers fragments can be detected using automated DNA sequencers. If different fluorescent labels are attached to *individual* primers, amplification fragments for different loci can be distinguished in the same sample. This possibility for multiplexing, the capacity of many sequencers to analyse several 96-well plates of samples in a relatively short period of time and the use of liquid handling robots to set up PCRs has now facilitated high-throughput genotyping in commercial plant breeding programmes.

1.3.6.5 Single Nucleotide Polymorphisms (SNP)

SNP is the most recent molecular maker method to be used in genotyping. It investigates the polymorphisms at single nucleotide levels in organisms. SNP's are valuable genetic markers of human disease. They also comprise the highest potential density marker set available for mapping experimentally derived mutations in model organisms such as *Caenorhabditis elegans* (Wicks et al. 2001). The different alleles in SNP are scored using a sequencer followed by a primer extension protocol or by using mass spectrometry to distinguish allelic DNA fragments on the basis of their mass. The availability of DNA sequence of sets of loci is a pre-requisite in scoring SNP. This limits the applicability of high-throughput SNP marker system to organisms for which the sequence data are available. Thus, SNPs are widely used for genotyping in human (Gray, Campbell, and Spurr 2000; Immervoll et al. 2001); *Drosophila* (Berger et al. 2001); mouse (Lindblad-Toh et al. 2000) and also in plants such as *Arabidopsis* (Chicurel 2001; Hubley, Zitzler, and Roach 2003).

1.3.7 Genetic analysis in plants

Geneticists and plant breeders are interested in studying the functions of genes of interest in order to manipulate such genes to achieve the goals of plant and animal breeding and human disease control. Since the beginning of the discipline of Genetics, scientists have used a

reductionist approach to the genetical analysis of traits. In the early days, the inheritance and action of individual genes were the units of observation and in today's molecular era, methods have been developed to study the molecular structure of such genes (Kearsey and Pooni 1996). The variation that exists among individuals is the essence of genetic analysis of traits in organisms. Therefore, genetic analysis approaches entirely depend upon the genetical variability that causes distinct phenotypic differences among the individuals of a population. This genetic variability is a result of different allelic forms of genes. In the early days, the allelic differences could be observed only if they produced a change in structure, physiology or behaviour. Today, with the development of molecular genetics, such changes can be observed at the molecular level. This opens up unlimited possibilities for geneticists to achieve their specific goals.

1.3.8 Qualitative traits

The very pronounced, clear-cut phenotypic differences that can be observed among the individuals of a species are called qualitative differences; the individual phenotypes fall into discrete categories. Such qualitative differences are not greatly affected by the environment and they arise from major allelic differences at one or two genes. These genes are called major genes and such traits are termed qualitative traits (Kearsey and Pooni 1996). The genetical analysis of qualitative traits is not very complex because the classical Mendelian theory can directly be applied on them after scoring the phenotypes.

1.3.9 Quantitative traits and Quantitative Trait Loci (QTL)

However, most of the agriculturally important characters such as yield, quality, stress tolerance in plants, milk yield, growth, fitness in livestock and also traits such as intelligence, behaviour and personality etc. in humans, are quantitative in nature. These traits show a

continuous distribution often approximating a statistical normal distribution and they are greatly affected by the environment. Such traits are called quantitative or metrical traits and the observed variation is due to many naturally occurring polymorphic genes (Kearsey and Pooni 1996). Since there is more than one gene affecting such characters, with either increasing or decreasing effects on the trait, the ultimate expressed phenotype is the net effect of these polygenes plus the effect of environment (Mather and Jinks 1982). Unlike the major genes, the effect of a single gene on a quantitative trait may be small compared to the total effect of all the genes and thus the environment plays a major role in the ultimate expressed phenotype of a quantitative trait. The inheritance of quantitative traits is referred to as polygenic inheritance and the individual loci controlling a quantitative trait are referred to as QTL (Tanksley 1993).

The genetic analysis of quantitative traits is not as simple or straightforward as that of qualitative traits. That does not mean to indicate that quantitative traits do not follow the Mendelian theory; Thus far, the molecular basis underlying allelic variation at QTL is identical to the identified variation for simple Mendelian loci (Paran and Zamir 2003). They segregate independently at meiosis, sometimes exhibit linkage to other QTL, show some degree of dominance, (but perhaps not overdominance), and they could show gene interaction (Gregersen and Brehrens 2003; Kearsey and Luo 2003). But, because there are several different genes acting and perhaps interacting together to give the ultimate phenotype, it is not possible to follow the action of those genes simultaneously by using only the Mendelian theory. So the alternative approach has to rely upon statistics to describe the characteristics of continuous phenotypic distributions. This area of genetics is called quantitative or biometrical genetics.

In 1918 Fisher introduced the basic principles in biometrical genetics to explain the means and variances of a quantitative trait. Subsequently, these principles have been considerably developed by some quantitative geneticists (Mather and Jinks 1982; Falconer and Mackay 1996; Kearsey and Pooni 1996; Lynch and Walsh 1998).

1.3.10 QTL analysis

Quantitative genetics theory explains genetic models to describe genetic architecture of a quantitative trait. When two alleles segregate at each of multiple QTL affecting variation in the trait, individual genotypes are specified by the homozygous and heterozygous effects of each locus, and pair-wise and higher-order interactions (epistasis) between loci. However, in contrast to traits controlled by one or a few loci with large effects, variation in quantitative traits is caused by segregation at multiple QTL with individually small effects that are sensitive to the environment. Therefore, for complex traits, QTL genotypes cannot be determined from segregation of phenotypes in controlled crosses or pedigrees (Mackay 2001). Although individual QTL genotypes cannot be determined by the phenotype, the net effects of all loci affecting the trait can be calculated by partitioning the total phenotype variance into components of additive, dominance and epistatic genetic variance, variance of genotype-environment interactions and other environmental variance. These variances are specific to the population being studied, due to the dependence of the genetic terms on allele frequencies at the contributing loci, and real environmental differences between populations (Mackay 2001). Thus, until recently our knowledge about quantitative traits were restricted to these statistical approaches of Biometry such as heritability, response to selection, changes of mean on inbreeding etc. Over a long period this approach has helped immensely to improve the desired characters in animal and plant breeding (Walsh 2001). However, these inferences deal with the average effects of unresolved loci and provide little information about the number,

location or relative effects of specific QTL. It is a necessity today to identify and determine the properties of the individual genes governing variation in complex traits (Tanksley 1993).

1.3.11 Gene mapping/QTL Mapping

Gene mapping is the location of genes to positions on specific chromosomes (Kearsey and Pooni 1996). The phenomenon of genetic linkage between the genes located in the same chromosome is the basis in the development of genetic linkage maps. The mapping of chromosomal regions with genes affecting important qualitative and quantitative characters has become vital in plant breeding. Mapping and sequencing of plant genomes would help to understand the gene function, gene regulation and their expression (Mohan et al. 1997). Genome mapping will greatly accelerate the slower progress made with traditional breeding methods, in improvement of those complex characters such as grain yield, grain quality and drought resistance etc (Shoemaker 1994).

Easily recognizable genetic markers are needed to identify and locate different genes along the chromosomes. Such markers have to exist in two or more allelic forms in order to detect polymorphism between them and to be used in the mapping (Kearsey and Pooni 1996). In the early days, all the attempts at developing genetic linkage maps were based on morphological markers. Although the information on these conventional maps is important to know the location of genes corresponding to phenotypical traits, their usefulness is limited by the low number of morphological markers, which are available. In addition, the expression of morphological markers is affected by environmental conditions. The advent of biochemical markers, followed by the molecular or DNA markers marked the beginning of a more advanced era in genome mapping (Mohan et al. 1997).

1.3.12 Tracking QTL with single gene markers

The association of seed size in beans (a quantitatively inherited character) with seed-coat pigmentation (a discrete monogenic character) was the first report to indicate the linkage of a single gene to a quantitative character. After Sax reported this association in 1923 the scientists showed a greater interest in resolving the dilemma they faced with understanding the inheritance of quantitative traits. In 1961, (Thoday 1961)put-forward the idea of using single gene markers to systematically characterize and map individual polygenes controlling quantitative traits. Thoday's theory was that if the segregation of a single gene marker could be used to detect and estimate the effect of a linked polygene and if single gene markers were scattered throughout the genome of an organism, it should be possible to map and characterize all of the polygenes affecting a character (Thoday 1961).

Although theoretically this idea was flawless, widespread application of the theory proved to be difficult in practice. At the time, for most of the organisms only a very few monogenic markers had been mapped and, even with those individuals where a considerable number of morphological markers were mapped, the majority of them were not suitable to study quantitative traits. The reason for this was that most of the time the marker gene itself had a larger effect on the phenotype than the linked polygene thus preventing the detection or causing over/underestimation the effect of the detected polygene. There were also other problems such as dominance, epistasis and lack of polymorphism in natural populations etc (Tanksley 1993).

1.3.13 QTL mapping with molecular markers

With the advent of molecular markers, widespread mapping and characterizing of genes controlling quantitative traits in natural populations became a more feasible prospect. The

molecular markers show a number of advantageous properties over morphological markers in QTL mapping.

Molecular markers reveal sites of variation at the DNA sequence level. Unlike morphological markers, these variations may not show themselves in the phenotype, and each might be nothing more than a single nucleotide difference in a gene or a piece of repetitive DNA. Thus the problem of the marker gene having an effect of the phenotype of the quantitative trait under study of being affected by selection was largely overcome by the use of molecular markers.

Furthermore, the use of molecular markers proved to be a solution for the lack of polymorphism with morphological markers because the proportion of informative or segregating molecular markers can be high in crosses between individuals from the same or different populations.

DNA based markers solved the problem of limited marker abundance with the availability of different marker systems so that a full coverage of the whole genome can now be obtained. The publishing of molecular linkage maps covering the entire genomes for a vast number of organisms including humans is a result of this abundance of molecular markers covering entire genomes (Remington et al. 1999; Remington and Purugganan 2003; Wolf et al. 2004)

It is not possible to determine all possible genotypes by observing phenotypes in loci with dominant-recessive alleles but with co-dominant alleles there is a one to one relationship between the genotype and phenotype. Some of the marker systems such as SSR and RFLP are codominant markers making it possible to identify all the genotypes in a cross.

Finally, the use of molecular markers provided solutions to the problems that arose with morphological markers due to epistasis and pleiotropy; epistasis being the interaction between mono-allelic genes where one gene interferes with the phenotypic expression of another gene

and pleiotropy being the same gene involved in the phenotypic expression of more than one trait. The molecular markers normally do not exhibit epistasis or pleiotropy thus providing a virtually limitless number of segregating markers for QTL mapping (Tanksley 1993).

The development of linkage maps with large numbers of molecular markers has stimulated the search for methods to map genes involved in quantitative traits (QTLs). A promising method, proposed by (Lander and Botstein 1989), employs pairs of neighbouring markers to obtain maximum linkage information about the presence of a QTL within the enclosed chromosomal segment (Vanooijen 1992). QTL mapping with molecular markers offers a number of advantages, which were lacking in the traditional approaches of QTL analysis. The calculation of the number of QTL affecting a particular trait is one such advantage. The approach although not without limitations, is simply to add up the number of QTL detected in a particular study and to use that value as an estimate of the number of segregating polygenes affecting the character in that population (Tanksley 1993).

1.3.14 Types of maps

1.3.14.1 Genetic maps

Genetic maps are entirely based on recombination frequencies. It is an ordered set of loci with relative spacing determined from measured recombination frequencies. Thus, genetic maps plot the estimated arrangement of genes along a chromosome with a distance related to the number of recombinations occurring along the specific chromosome. The distance between two genes based on a genetic map is a function of the recombination frequency between the two genes measured in cM, and not the actual physical distance.

1.3.14.2 Physical maps

Unlike genetic maps, physical maps are based directly on the measurements of DNA structure in terms of the number of base pairs (in kilo bases). Physical maps focusing on the region of interest are a requirement to characterize and isolate genes of desirable characters. Yeast artificial chromosomes (YAC) and bacterial artificial chromosomes (BAC) and cosmids are used as vectors for cloning in order to produce physical maps of different organisms (Mozo et al. 1999; Terryn, Rouze, and Van Montagu 1999).

1.3.14.2 Framework maps

Framework maps are an arranged set of molecular markers along each chromosome based on recombination values from all pair-wise combinations of markers. Framework maps facilitate the location of important genes between specific markers along a chromosome and hence are useful in marker-assisted selection.

1.3.14.2 Consensus maps

Consensus maps are statistically generated by integrating the maps from different populations. The reliability of consensus maps depends on the availability of a subset of markers, which are common to all individual maps.

1.3.14.3 Other maps

There are sequence maps, which are sets of clones that have been sequenced or are chosen for sequencing. A function map consists of many functional genes for physiological traits, morphological traits, agronomic traits and structural genes and it could be used as a basic map to study genetic and physiological relationships in related crops. Synteny maps are co-linear maps of distantly related species for certain portions in the genomes. A reference map is a map that is used as a reference to merge other maps within the species and between related species. Candidate genes are major genes affecting a trait and also located in the vicinity of a

QTL responsible for the same trait and candidate maps are made out of candidate or structural and other genes (Etienne et al. 2002).

1.3.15 Mapping populations

1.3.15.1 Inbred populations

Selection of a mapping population is one of the most critical decisions in constructing a linkage map with DNA markers. A mapping population should comprise of the parents and their segregating population to score them accurately.

Segregating populations F_2s, backcrosses, Recombinant inbred lines are the commonly used mapping populations at present (Kuittinen, Mattila, and Savolainen 1997; Ponce, Robles, and Micol 1999; Kebede et al. 2001; Marquez-Cedillo et al. 2001). In selecting parents to construct a mapping population in inbreeding organisms, the parents should be genetically variable for the trait of interest. Usually, the parent strains will have different mean values for the trait, although it is not entirely necessary because the parent strains with the same phenotypic means can vary genetically due to complementary patterns of positive and negative allelic effects of QTL- gene dispersion (Mackay 2001). The selection of parents is followed by the derivation of a mapping population. Various possible mapping populations can be obtained:

(i) backcrossing the F_1 to one or both of the parents,

(ii) mating the F_1 *inter se* or selfing the F_1 to create an F_2 population,

(iii) constructing recombinant inbred lines (RILs) by breeding F_2 sub-lines to homozygosity over 7–8 generations by single seed decent (SSD),

(iv) production of double haploid lines from the F_1 or F_2.

Crosses between inbred lines generate maximum linkage disequilibrium between QTL and marker alleles, and ensure that only two QTL alleles segregate, with known linkage phase

(Mackay 2001). The segregating populations derived in such a way in inbred strains are usually assumed to be homozygous with different alleles at both QTL and genetic markers. In addition these populations have the advantages of having bi-allelic segregation at each of the loci; providing the full information about the linkage phase of genes at the marker allele and QTL; ease in creating a large full-sib family and versatile experimental design for both detecting marker-QTL linkage and estimating genetic parameters defining genetic effects at the QTL (Kearsey and Luo 2003).

The choice of a mapping population depends upon the biology of the organism and the power of the different methods under the heritability value of the trait of interest. Under the circumstance of low heritability, a large number of individuals of the same genotype need to be assessed and this is possible using RILs. RILs are a permanent genetic stock and they need be genotyped only once for the subsequent usage to map any number of traits. RILs allow more precision of mapping than a similar number of back cross or F_2 individuals due to the increased number of recombination events during their construction. However, it is more difficult and more time consuming to construct a RIL population than it is to do an F_2 or a BC population (Mackay 2001).

The next important point is the size of the mapping population from which the ultimate resolution of the map would be determined. However, the size of the mapping population will again depend on the choice of the mapping population. Yet, a mapping population with less than 50 individuals would not be sufficient for constructing a reliable map.

1.3.15.2 Outbred populations

Unlike inbred populations, it is not practicable to construct the above mentioned mapping populations in outbred species thus posing additional challenges in QTL mapping. As a result in QTL mapping ventures in outbred species, (e.g. most of the trees, humans etc.) the

information is restricted to existing pedigree populations. The problems involved in obtaining information from pedigree populations are that the family size is substantially smaller than the commonly used segregating populations and the direct unavailability of information about the linkage phase of genes at the marker loci and QTL. To solve these problems, respectively a large number of families are required and complicated statistical tools are needed for modelling the inheritance of genes within a multiple generation pedigree (Mackay 2001; Kearsey and Luo 2003).

1.3.16 Principles of genome mapping

1.3.16.1 Recombination frequency

Location of genes on chromosomes is based on the frequency of chiasmata formation and crossing over (as a result of crossing over two of the chromatids will have one end of maternal origin and the other of paternal origin) between genes. The genetic distance between any two genes is simply a function of the average number of crossovers that occur between them. Therefore, the number of crossovers between two loci in a mapping population can be used as a measure of the distance between them. The further apart the two loci are, the more likely there is to be a crossover between them. A chromosome may have two or more chiasmata and this number is roughly proportional to its length. The length of chromosome that, on average, has one chiasma is defined as having a genetic length of 50 centimorgans (cM).

It is not possible to count chiasmata between individual genes because the genes are not observable under a microscope. Therefore, recombination frequency (RF) is used as an indirect way to estimate the chiasma frequency (μ) because RF involves investigating the results of the genetic exchange (recombination of maternal and paternal alleles) caused by the chiasmata. If two genes were unlinked, we would expect there to be equal numbers of

recombinants and parental types and so the estimate of RF will be 0.5. Therefore, RF ranges from 0 - 0.5 (0 – 50%).

1.3.16.2 LOD threshold

The term LOD was developed through the maximum likelihood approach for linkage analysis by Mather. LOD means the log of the odds ratio and the odds ratio is the ratio between the probability of linkage of the two loci when they are linked (alternate hypothesis) versus the probability of linkage of the two loci when they are not linked (null hypothesis). If we put LOD as a function according to the above definition, LOD = log [L(x)/L(0.5)] is used (Risch 2001). In many analyses, a significance level of LOD>3 is appropriate as an acceptance level of linkage between two loci. The researcher must determine a threshold LOD value below which linkage is not considered significant (Churchill and Doerge 1994). As the LOD threshold is raised, fewer markers are assigned to linkage groups (i.e., independent loci), and many smaller linkage groups are identified.

1.3.16.2 Statistical methods for QTL mapping

Several statistical methods are available for data analysis for detecting marker-trait associations. The basic criterion for selecting a method is the number of markers that can be used to perform a test of significance.

Single Point Regression is the method used when information from one marker at a time is used. In this method the individuals are grouped into classes based on the marker genotype and the phenotypic mean of each class is obtained (Weller 1986; Beckmann and Soller 1988; Luo and Kearsey 1989; Luo and Woolliams 1993). The phenotypic means are then compared to see whether they differ significantly between marker alleles for any marker. A t-test can be performed to test for the significance in two marker classes (as in RILs or DHLs) or ANOVA or linear regression analysis when there are more than two classes as in F_2 population. The

drawbacks of single point analysis are that, it can not localize the exact position of QTL but only identify the marker(s) to which the QTL is closely linked and also the dependence of the detection of QTL on its distance from the marker and the magnitude of the effect of QTL.

Interval Mapping (IM) is a statistical method of locating QTL when information from two linked adjacent markers is used (Lander and Botstein 1989). This method is based on maximum likelihood (ML) approach, which is computationally more intensive because ML uses full information from the marker-trait distribution instead of the marker means. Further enhancing this method the LOD score is calculated across all the intervals to produce a likelihood ratio profile and the likelihood map can be constructed by plotting likelihood ratios against the map positions of the putative QTL. The QTL is then mapped to the position where the LOD score exceeds a specified significance level. CIs of the mapped QTL are defined by the map positions that are given by a decline in LOD score equal to 1 either side of the peak. The disadvantage of the Interval mapping is that it is computationally intensive and requires a complex software package. Furthermore, this method can produce 'ghost QTLs' by producing significant peaks when there are no actual QTL in the vicinity (Haley and Knott 1992; Martinez and Curnow 1992) due to the fact that IM does not perform a test of significance of the detected QTL (Doerge, Zeng, and Weir 1997). Also, the estimates of location and effect may be biased by the presence of more than one QTL on a chromosome because these QTL will affect the test statistics being calculated for each interval. Finally, the IM does not account for the effects of QTL that are located on other chromosomes because each test is confined to two markers at a time. Thus the observed residual variation for each interval increases and the power of the analysis is reduced.

Composite Interval Mapping (CIM) of QTL is a method in which selected markers are used as cofactors to separate linked QTL in addition to the adjacent markers in order to improve

the power and precision (Jansen 1993; Zeng 1993; Jansen 1994; Jansen and Stam 1994; Jansen 1996). With this method 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. (Kearsey and Hyne 1994) developed another multilocus approach that utilizes information from all the markers on a chromosome. This method is based on a regression approach to produce the best estimate using all marker means on a particular chromosome (Hyne and Kearsey 1995).

Multiple Interval Mapping (MIM) is yet another statistical method for QTL location (Kao, Zeng, and Teasdale 1999). This method uses multiple marker intervals simultaneously to fit multiple putative QTL directly. The MIM is based on Cockerham's model for interpreting genetic parameters and the method of maximum likelihood for estimating genetic parameters. This is a method that increases the precision and power of QTL mapping in addition to being able to calculate epistasis between QTL, genotypic values of individuals and heritabilities of quantitative traits.

However, it is difficult to separate two QTL located on the same chromosome with the above methods largely due to the high CIs of the mapped QTL. In order to reduce the CI it is necessary to score more recombinations which can only be done by increasing the population size and genotyping with four to five well-spaced markers along a chromosome, rather than increasing the number of markers along the chromosome (Darvasi et al. 1993; Kearsey and Pooni 1996).

1.3.17 Mapping software

The most widely used genetic mapping software is Mapmaker (Lander et al. 1987). Mapmaker has several routines to simplify multipoint analysis including an algorithm that quickly groups markers into likely linkage groups and another for guessing the best possible

order. Once a plausible order has been established, another algorithm compares the strength of evidence for that order compared to possible alternatives in a routine called 'ripple'. It enables the user to confirm the best order in a way that increases only arithmetically with increasing number of loci.

JoinMap (Stam 1993) is a computer package used to construct an integrated genetic map from different sets of mapping populations where population sizes can be limiting. It searches for an arrangement of like markers between previously established genetic maps and suggests a linear arrangement of all the markers dependent on recombination values JoinMap sequentially builds up a genetic map by comparing LOD values of an assigned marker to a constructed linkage group

1.3.18 Basic steps of a linkage analysis

1. **Constructing a mapping population**

   The selection of a mapping population is one of the most important considerations in a genome mapping project. The points discussed in the above section on mapping populations should be seriously considered in decision-making regarding the mapping population.

2. **Selection of Molecular Marker Techniques and Genotyping the mapping population**

   Selection of molecular marker techniques for genome mapping depends on the breeding habit & genome size of the organism, already known information on genome organization of the organism and funds & facilities available. Genotyping is the process that is used to distinguish individuals of the progeny who are homozygous for female parent (AA), homozygous for male parent (BB) and heterozygous for both parents (AB).

3. **Calculate recombination frequencies (RF) between all pairs of markers**

It is not possible to count chiasmata between individual genes. Therefore, RF is used as an optional way to estimate the chiasma frequency (μ).

4. **Position markers into groups.**

Markers in different groups are inherited independently while markers in the same group are likely to be inherited together.

5. **Judge the most likely order of markers within each group**

Marker order could be altered due to distorted genomic regions that result in a lack of ability for reliable interpretation to give a favoured order. Segregation distortion between markers is the deviation of segregation ratios from the expected Mendelian fractions.

6. **Estimate the distance between markers within each group to give a framework map**

A mapping function is necessary to construct the map derived from RF data because RF and chiasma frequency are not linearly related. Two main mapping functions, Haldane and Kosambi, are available for this application and also several other mapping functions have been proposed (Zhao and Speed 1996). Haldane mapping function assumes that there is no cross-over interference (which is the effect of one cross-over on the occurrence of a second cross-over along the same chromosome) which does not hold true over 15 cM. Kosambi mapping function is the commonly used mapping function, which takes the crossover interference into account.

7. **Score quantitative traits**

It is first necessary to obtain a progeny segregating for the character of interest to map QTL. The segregating population should then be raised using a proper experimental design and score the phenotype of the trait of interest.

8. QTL analysis to locate QTL in the framework map

When establishing a QTL map, it should have two essential steps namely, the mapping of the molecular markers and establishing the association of the traits (QTL) with markers. Accurate marker data and trait scoring data from a segregating population are vital for analyses (Brzustowicz et al. 1993).

9. Identify markers linked with QTL

The final step in locating QTL is to select an appropriate statistical method depending on the situation.

1.3.19 Application of Molecular markers and QTL analysis in Arabidopsis

The first comprehensive and internally consistent linkage map of *Arabidopsis* based on F_2/F_3 populations using exclusively morphological markers was published in 1983 (Koornneef et al. 1983). They assigned 76 loci to 5 linkage groups of *Arabidopsis* and calculated the total map length to be 430 cM. Isozyme variants in *Arabidopsis* have been used by (Abbott and Gomes 1989) in population surveys. Furthermore, (Chang et al. 1988) and (Nam et al. 1989) developed RFLP maps for *Arabidopsis*. (Reiter et al. 1992) developed a high-density genetic linkage map containing 252 RAPD markers using a RIL population. In total this map contained 320 marker loci (including previously mapped RFLP loci) and the total map length was determined to be 630 cM (which is slightly higher than previously observed). (Hauge et al. 1993) assembled an integrated genetic/RFLP linkage map of *Arabidopsis*, based on two independent sets of RFLP data and the mathematical integration of the independent data set was performed using the JOINMAP computer package. (Lister and Dean 1993) used RILs for mapping RFLP and phenotypic markers in *Arabidopsis*. The RILs in this study were generated from a cross between the two ecotypes Columbia and Landsberg *erecta*. The F_8 RIL population, which consisted of 300 individual F_2 seedlings, was widely used in linkage map

studies in *Arabidopsis*. After the discovery of SSR markers, 30 microsatellite loci were assigned to the linkage map of *Arabidopsis* (Bell and Ecker 1994). In this study the microsatellite polymorphism was detected in ecotypes Columbia and Landsberg, and the polymorphic markers were incorporated into the genetic linkage map of Lister and Dean using the software package MAPMAKER 3.0.

(Alonso-Blanco et al. 1998b) developed a map for Ler/Cvi RILs using AFLP markers. They have also analysed 395 AFLP markers in Lister & Dean Col/Ler RIL population and integrated them into the molecular map generated by Lister & Dean. They also integrated the two RIL maps through 49 common markers and demonstrated that segregating bands from a common parent can be compared between different populations. Subsequently numerous genetic and other maps have been developed for different populations with different markers and traits (Alonso-Blanco et al. 1998a; Alonso-Blanco et al. 1998b; Soppe et al. 2000; Vision, Brown, and Tanksley 2000; El-Assal et al. 2001; El-Assal et al. 2002).

With the start of the *Arabidopsis* Genome Initiative, (Mara et al. 1999) reported their work on fingerprinting more than 20,000 BACs from 2 publicly available libraries, in support of the genome sequencing effort. This was the first example of whole-genome random BAC fingerprint analysis of a eukaryote. With the invention of SNP marker technology, the first report of bi-allelic mapping in diploid genomes was on *Arabidopsis*. This study reported the construction of a bi-allelic genetic map using high-density oligonucleotide arrays to map a large number of SNPs (Cho et al. 1999). (Lukowitz, Gillmor, and Scheible 2000) discussed the opportunities available for high resolution mapping of QTL in *Arabidopsis* by positional cloning, making use of the information generated by the genome initiative. With yet another breakthrough technology, (Peters 2001) generated a physical Amplified Fragment Length Polymorphism map of *Arabidopsis*. They have positioned AFLP markers directly on the

genome sequence of *Arabidopsis* by combining gel based AFLP analysis with *in silico* restriction fragment analysis using the published genome sequence.

As discussed at the beginning of this chapter, when it dawned that information concerning the genetic control of basic biological processes in *Arabidopsis* can be transferable to other species, studies on comparative mapping of *Arabidopsis* with other plant species was started.

Extensive research has been carried out for QTL studies on many traits in *Arabidopsis* with the aid of molecular markers and using the maps. These studies include, research on genetic variation at marker loci in quantitative traits in natural populations (Kuittinen, Mattila, and Savolainen 1997); the molecular basis of quantitative genetic variation in central and secondary metabolism in *Arabidopsis* (Mitchell-Olds and Pedersen 1998); use of RILs for the genetic dissection of complex traits mainly flowering time (Loudet et al. 2002) and numerous other studies related to QTL analysis (Ponce, Robles, and Micol 1999; Ungerer et al. 2002; Kearsey, Pooni, and Syed 2003; El-Lithy et al. 2004).

Research on QTL governing flowering in *Arabidopsis* is one of the extensively studied area. These studies include the identification of loci responsible for flowering (such as *CONSTANS* and *FRIGIDA*), effects of vernalization on flowering time, genetic interactions related to flowering etc (Lee, Bleecker, and Amasino 1993; Clarke and Dean 1994; Koornneef et al. 1998a; Johanson et al. 2000; Hagenblad and Nordborg 2002; Ungerer et al. 2003; Noh et al. 2004).

1.3.20 Problems of QTL mapping with segregating populations

As discussed above, the most used mapping populations in QTL analysis at present are the segregating populations such as F_2s, backcrosses, recombinant inbred lines or doubled haploid lines. The power of detecting QTL with these populations is affected by a number of factors

such as heritability of the trait under study, size of the mapping population, magnitude of the individual QTL effect, method of analysis and significant threshold.

One of the problems with the use of such populations in QTL mapping is the difficulty of locating QTL for specific short regions in a chromosome. Normally, the confidence intervals associated for mapped QTL using segregating experimental populations are very rarely less than 5 cM, and often 30–50 cM (Tanksley 1993; Hyne et al. 1995; Darvasi 1997). About 30-50 cM along the length of a chromosome is a very long region compared with the total cM length of a chromosome and thus will accommodate a large number of genes. For example, in *Arabidopsis* an average chromosome is about 100 cM long, and a 5 cM interval could include 250 genes on average and more in regions of low crossing over (Koumproglou et al. 2002). Furthermore, the average size of intervals containing significant QTL from seven studies in *Drosophila* was 8.9 cM. There are at least 13,600 genes in the 120 Mb of *Drosophila* eukaryotic DNA and thus this region will encompass about 507 genes. This precision of mapped QTL may be adequate for marker assisted breeding but is far too low for detailed genetical analysis, chromosome walking or map-based cloning. In addition to the problem of higher CIs with the above methods of QTL mapping, it is particularly difficult to separate linked QTL, and this results in the underestimation of the number of actual QTL affecting a trait.

This low precision in locating QTL with segregating populations is mainly due to the relatively low levels of chiasma occurrence along a chromosome resulting in low recombination frequencies. The number of meioses, which produce information about RF depends on the number of individuals sampled in the segregating population. Therefore, in order to increase the precision of the location of the mapped QTL it is necessary to score a very large number of individuals in a population. On the other hand if the sample consists of a

low number of individuals it may result in bias estimation, especially if the RFs of the selected samples depart from the expected and also due to relatively low number of gamete forming meioses in a small sample (Kearsey and Pooni 1996). As a result greater precision in QTL mapping in segregating populations can be obtained only for traits which are highly heritable and with much larger population sizes that are often impracticable (Kearsey and Farquhar 1998).

1.3.21 High resolution QTL mapping

Mapping resolution is related to the size of the genomic region within which a QTL can be located and in order to achieve higher resolution it is necessary to successively reduce the region where the QTL can be located along the chromosome (Mackay 2001). The scale of LD between the gene corresponding to the QTL and flanking marker loci determines the size of the related genomic region. In principle, the QTL can be positioned relative to a pair of flanking markers if both (i) a high-density molecular marker map spanning the interval of the QTL and (ii) a population of recombinant genotypes with break points between adjacent markers are available. But in contrast to Mendelian loci, individual QTL are expected to have small effects which are sensitive to environment so that the phenotype of a single individual is not a reliable indicator of the QTL genotype. As a result, for high resolution QTL mapping it is necessary to have increased recombination in the QTL interval and the accurate determination of QTL genotype.

One way of achieving this is to use more advanced generations than the F_2 because the number of recombinations increases proportionately with the number of generations. However, the finest method available at present to magnify the effect of QTL is to construct chromosomes of defined constitution (Kearsey and Pooni 1996).

The difficulty in locating QTL in segregating populations is mainly due to the relatively low level of recombination. More precise mapping of QTL can be achieved only by specifically selecting for recombinational events in particular regions but this requires constructing chromosomes with defined constitution (Kearsey and Pooni 1996). One solution to this problem is to follow the initial, approximate QTL locations, based on the segregating populations, with a finer analysis using near isogenic lines (NILs). NILs can be derived by partially inbreeding recombinant inbred lines and then identifying individuals that are homozygous for all chromosomes except for a short segment. But, it may be difficult to find NILs for a particular chromosomal region of interest and a large number of pairs are needed to cover the whole genome (Kearsey and Pooni 1996).

Various methods have been suggested to achieve high-resolution genetic maps in different organisms (Jorgensen et al. 2002) for which populations with defined chromosomes cannot be developed. Association based fine mapping is an example for precise mapping that is used mostly in medical research to map susceptible genes for complex diseases using large numbers of SNP markers (Kaplan and Morris 2001). In plant populations various other methods such as recurrent selection and back cross schemes (RSB) have been investigated to achieve higher precision in QTL mapping (Luo, Wu, and Kearsey 2002; Ronin et al. 2003). (Ronin et al. 2003) proposed a selective 'Recombinant Genotyping' method to be applied when the target is a locus with QTL with a moderate or even a small substitution effect. (Mott et al. 2000) have shown the possibility of fine mapping QTL in genetically heterogeneous stocks with their studies on locating QTL for fearfulness of mice. They used a multipoint analysis method and QTLs could be fine mapped provided that the same QTL can be detected in the inbred populations and the heterogeneous stocks derived from the same inbred lines.

1.3.22 Chromosome Engineering

The basic principle in chromosome engineering approach is to create two genotypes, which are identical apart from a defined region on a particular chromosome. Then, any genetical differences in phenotype between these two genotypes must be due to genes in this defined region. The smaller the region, the more precisely the positions of these genes will be known (Kearsey and Pooni 1996; Gibson and Mackay 2002). This will facilitate the specific selection for recombinational events within the region of the QTL. Three different approaches of constructing designer chromosomes have been studied at present namely, chromosome substitution lines, near isogenic lines and STAIRS.

1.3.22.1 Substitution Lines

Ideally, substitution lines carry a single defined segment of a novel (donor) genotype and have a pure genetic background of a distinct (recurrent) genotype. Therefore, a complementary set of substitution lines making an overlapping library represents the whole donor genotype divided into a limited number of distinct segments. The substituted region of the genome can be a whole chromosome or a small region along the chromosome. The interactions between donor alleles are limited to those between genes on the same homozygous substituted tract, because the genetic background is constant. This simplifies the calculations of the significance and magnitude of the mean effects of each substituted tract (Howell, Marshall, and Lydiate 1996).

Substitution lines are produced via several rounds of backcrossing with selection. Eventually, individuals with a single segment of donor genotype in the genetic background of the recurrent parent will be isolated. Marker assisted selection helps in identifying the individuals with the desired genetic make up. Substitution lines are homozygous lines so they exhibit the advantage of being maintained as true breeding 'immortal' stocks.

Some organisms display specific genomic characters that help in constructing substitution lines. For example, in *Drosophila*, the male meiosis is achiasmate. *Drosophila* haploid karyotype contains only four chromosomes and thus the male gametes of an F_1 consist of all haploid combinations of four pairs of un-recombined chromosomes. Therefore the males can be used as vectors for cloning whole chromosomes and passing them to subsequent generations. In female *Drosophila*, no recombinations occur in chromosome four while the other three chromosomes display normal meiosis. These features in *Drosophila* can be used to easily substitute single whole chromosomes from one strain to another by backcrossing using the female parent as the recurrent parent.

Wheat, which is an allohexaploid, with 2n =6x=42 chromosomes has been derived from three ancestral diploids. Using these three genome homeologous groups, nullisomic and monosomic lines have been produced in wheat, first in the variety Chinese Spring and later with a range of modern commercial varieties. Wheat substitution lines have subsequently been produced using these lines specially, monosomic stocks (Korzun et al. 1997; Pestsova et al. 2000) and have used microsatellite markers to confirm the authenticity of chromosome substitution lines of wheat. The wheat substitution lines have subsequently been used for various QTL analysis studies such as, the control of agronomic traits and their stability across environments; (Chun, Yu, and Griffith 1998); the starch-pasting properties of wheat (Araki, Miura, and Sawada 2000); studies on host plant resistance (Castro et al. 2001) studies on bread making quality of wheat (Rousset et al. 2001; Amiour et al. 2002); and studies on heat stress and tolerance (Kocsy et al. 2004).

Libraries of substitution lines with overlapping introgressed segments were generated in Brassica (Howell, Marshall, and Lydiate 1996; Ramsay et al. 1996; Rae, Howell, and Kearsey

1999). (Burns et al. 2003) reported studies on QTL analysis of an inter-varietal set of substitution lines in *Brassica napus* for seed oil content and fatty acid composition.

Extensive QTL analysis studies mainly related to the location of QTL for fruit quality have been carried out with tomato substitution lines over the last few years (Ji and Chetelat 2003; Lecomte et al. 2004; Yates et al. 2004).

Substitution lines have been produced and used in the genetic analysis in mice (Nadeau and Frankel 2000). They reported research on construction, applications and advantages of CSSs compared with conventional crosses for detecting and analyzing QTLs in mice.

1.3.22.2 Near Isogenic Lines (NILs)

NILs are constructed by recurrent backcrosses to one parent accompanied by selection for the marker associated with the genotype of the non-recurrent strain. Yet another method of producing NILs is to inbreed F_2 successively by a single seed decent method. By about the 6^{th} generation the population will be homozygous except for one or a few very small regions. Then these individuals can be selfed and pairs of NILs related to each heterozygous region can be identified by molecular marker assisted genotyping. NILs have the advantages of being easier to construct and the capacity to produce quite small substitutions due to more rounds of recombinations than with substitution lines. But with NILs the ability to get NILs with a QTL region depends on chance rather than design. Furthermore, the genetic background of each pair of NILs is different to all the other NILs but this can be avoided by producing NILs by the back cross method rather than selfing an F_2. (Brouwer and St Clair 2004) reported studies on fine mapping of three QTL for late blight resistance in tomato using a set of NILs and sub-NILs.

1.3.22.3 Stepped Aligned Inbred Recombinant Strains (STAIRS)

STAIRS are a novel resource which is being produced and used to locate QTL to a highly precise narrow region in a chromosome. The objective of using STAIRS is to narrow down the region of QTL like NILs. But with STAIRS this is done in a more targeted and a systematic way. Availability of whole CSSs is a prerequisite to produce STAIRS (Kearsey and Luo 2003). At present such lines are available only for a few organisms namely, *Arabidopsis*, wheat, *Drosophila* and mice.

1.3.23 Chromosome Substitution Strains and STAIRS in *Arabidopsis*

Arabidopsis has a haploid karyotype of five chromosomes and accordingly five whole chromosome substitution strains (CSSs) can be produced. (Koumproglou et al. 2002) reported the production of CSSs in *Arabidopsis* for the genetic analysis of quantitative traits using a marker assisted breeding programme. *Arabidopsis* accessions Columbia and (Col) and Landsberg (L*er*) have been used as the recurrent parent and the donor parent respectively in this breeding programme because of the extensive existing genetic analysis information and the sequence information on these lines. In order to introgress whole chromosomes from the donor (L*er*) to the recipient (Col) by back crossing it is essential to select un-recombined, intact chromosomes.

STAIRS have been derived from CSSs for further progressing of the gene location. STAIRS consist of a large number of lines, each of which contains a homozygous chromosome with a single crossover in such a way that the chromosome contains Col genes at one end and L*er* genes at the other end. These homozygous lines are single recombinant lines (SRLs). When the SRLs for each chromosome are sequentially stacked, they show a step-like progression with each successive line having a little more L*er* chromosome. Therefore, they are called

STAIRS to reflect their structural relationship. The STAIRS for each CSS exist in two reciprocal forms depending on whether the donor chromosome extends from the top or the bottom of the chromosomes.

Fine mapping of QTL is facilitated in three steps using CSSs and STAIRS. Firstly QTL can be assigned to a particular chromosome using CSSs. Then with wide STAIRS the QTL can be located to a wider region of about 5-10 cM along a particular chromosome. The narrower STAIRS within that region then allow the QTL to be located to a region less than 1 cM allowing fine mapping of the QTL.

The previous research conducted on QTL analysis with CSSs and wide STAIRS have revealed the presence of significant QTL for flowering time on chromosome 3 in the region 0– 20 cM (Koumproglou et al. 2002).

1.3.24 Microarray Gene Expression

Gene expression profiling has become an invaluable tool in functional genomics. Since the mid 1990's DNA microarrays has emerged as the leading transcript profiling technology in the global analysis of biological systems (Peng et al. 2003; Cao et al. 2004).

Using mRNA of a given cell at a particular time, under a given set of conditions, DNA microarrays can provide a snapshot of the level of expression of all the genes in the cell. Such snapshots can be used to study fundamental biological phenomena such as development or evolution, to determine the function of new genes, to infer the role that individual genes may play in diseases, and to monitor the effect of drugs and other compounds on gene expression (Baldi and Hatfield 2002).

At present most of the research using DNA microarrays are related to understanding the gene expression patterns of human genes especially the genes related to diseases and the effect of

drugs on the changes of the expression of genes (Graeber and Eisenberg 2001; Jenssen et al. 2001; Cao et al. 2004; Grigoryev et al. 2004). However, the technology of DNA microarray has now been extended for the gene expression profiling of plants also, especially *Arabidopsis.*

1.3.24.1 Gene Expression Profiling in Arabidopsis

As discussed above, over the past decade, the methods for parallel studying of the expression of many genes have become increasingly sophisticated and high throughput in nature. The availability of whole genome sequence of *Arabidopsis* encourages the gene expression studies using microarrays. The DNA arrays constructed with oligonucleotides are becoming popular as a practical tool for quantitative monitoring of gene expression on a large scale. For *Arabidopsis*, the techniques that have been used have included membrane-spotted microarrays, cDNA glass microarrays, cDNA-AFLP, serial analysis of gene expression (SAGE) and oligonucleotide-based arrays (Aharoni and Vorst 2002; Donson et al. 2002).

Using microarrays, gene expression studies in which the NIL QTL (or the chromosome region defined by narrow STAIRS in our study) is compared to that from the parental line, can be performed to identify candidate genes. Several replicate lines should be used to control for biological variance and potential maternal effects. The conditions and tissue selected for RNA extraction for the gene expression must be chosen on the basis of the phenotype of the QTL (Borevitz and Chory 2004).

Given the very high rate of the spreading of the microarray technology it will not be long before this technology will be widely used for gene expression transcript profiling of the organisms for which a considerable amount of sequence data are available.

CHAPTER 02

MICROSATELLITES MARKER SATURATION OF TOP OF CHROMOSOME 3 IN ARABIDOPSIS

Abstract

With the major objective of the current research being fine mapping of QTL on the top of chromosome 3, it was a necessity to produce narrow STAIRS differing ideally by about 1 cM within the region of interest. The construction of STAIRS involves a marker-assisted backcross breeding programme, hence the availability of a large number of polymorphic markers along region of interest was a prerequisite for producing narrow STAIRS. During the early stages of the current research, CSSs of *Arabidopsis* have been produced using accession Col as the recurrent parent and accession *Ler* as the donor parent. Furthermore, there was an attempt to produce CSSs using accessions Col and Nd also. Within this chapter, the development of microsatellite markers that were polymorphic for accessions Col and L*er* is described along with the polymorphism of Col and Nd and L*er* and Nd data are available. Microsatellites were the first chosen marker system due to the availability of the whole genome sequence of *Arabidopsis* which makes marker designing an easy task, coupled with the ease of detection of polymorphism with microsatellites. In this project, all the SSRs present within the top 20 cM of chromosome 3 were located, flanking primers were designed for each SSR, the polymorphism of each primer pair was checked, and the PCR conditions were optimized for the polymorphic markers. A total of 29 SSRs were located within the top 20 cM of chromosome 3, out of which 24 proved to be polymorphic between Col and L*er*. This density of markers with an average interval of 0.8 cM was decided to be sufficient for the purpose of producing narrow STAIRS.

2.1 Introduction

2.1.1 What are Microsatellites?

Microsatellites are simple sequence repeats (SSRs) of 1-6 nucleotides. The length of a given repeat sequence varies greatly, with different alleles varying in the number of units of the repeat motif. The variability in the number of repeat units is typically the basis of observed polymorphism. Microsatellites appear to be ubiquitous in higher organisms, although their frequency varies among species; for example plant genomes contain relatively less simple repeats than the human genome. SSRs are abundant, are dispersed through-out the genome and show higher levels of polymorphism than other genetic markers.

The high degree of observed microsatellite polymorphism is reported to be the result of increased rates of sequence mutation affecting the number of repeat motifs present at an SSR locus, probably due to replication slippage or unequal crossing over (Edwards et al. 1996). They have the additional advantages of codominant inheritance, ease of detection and the potential for automation (Holton 2001). Due to all these positive factors, plant genotyping by the analysis of microsatellites has received considerable attention and SSRs are now been widely adapted as a molecular marker system (Stajner et al. 2005).

2.1.2 Isolation of SSRs

When the sequence of nucleotides around the SSR is not known the isolation of microsatellite loci can be a time consuming and expensive process. Many approaches have been described to isolate SSRs and the flanking sequences for primer designing. Genomic clones containing SSRs can be isolated by screening with labeled oligonucleotides containing the desired repeat sequences. Microsatellites may be obtained by screening sequences in databases or by screening libraries of clones. Database searching is the least costly in terms of time and

resources for obtaining microsatellite loci providing there are sufficient entries in the database. This limitation will change in the future for many organisms as large sequencing and expressed sequence tag (EST) projects generate large databases (Maguire 2001). In one study which compared the variability characteristics of microsatellites obtained from EST databases with those obtained from genomic libraries, it was found that database derived microsatellites had lower values than genomic library microsatellites (Cho et al. 2000). The frequency of microsatellites is reported to be significantly higher for the ESTs i.e. transcribed sequences, than the random genomic sequences for many plant species including *Arabidopsis* (Powell, Machray, and Provan 1996; Morgante, Hanafey, and Powell 2002).

(Powell, Machray, and Provan 1996) presented standard methods for the isolation of SSRs from clones. The basic steps they presented can be listed as below.

- The creation of a small insert genomic library
- Library screening by hybridization
- DNA sequencing of positive clones.
- Primer design and PCR analysis
- Identification of polymorphisms

To improve the efficiency of SSR isolation, SSR-enriched libraries have been developed using a variety of methods with selection either before or after library construction.

2.1.3 Detection of Microsatellite Polymorphism

Variations in the length of tandem repeats can be identified by amplification of the region containing the repeat via PCR using primers designed to the regions flanking individual SSRs. Size polymorphisms which result from differences in the number of repeats can be detected by gel electrophoresis. Agarose, PAGE, denaturing PAGE are the normally used gel systems. Ethidium bromide staining is commonly used to detect PCR products in non-denaturing gels.

However, accurate sizing is difficult with agarose when the difference between alleles is only a few base pairs, in which case the high resolution but expensive ready made gels such as spreadex need to be used. PAGE is another gel system that facilitates the differentiation of very small allele differences; even as small as single nucleotide differences. The SSR products run on PAGE are detected by silver staining, radio- labelling or fluorescent labelling. The advantage of fluorescent labelling is that it allows automated band reading and multiplexing, resulting in high throughput microsatellite analysis.

2.1.4 Location of Microsatellites in Arabidopsis

With the completion of the *Arabidopsis* Genome Initiative, *Arabidopsis* became the first plant to have the entire genome sequenced. Now is a good time to capitalize on the achievements of the *Arabidopsis* Genome Initiative. There is a large number of DNA markers listed in the *Arabidopsis* database for anyone searching for molecular markers (www.arabidopsis.org). In addition, with the whole genome sequence publicly available it is an easy task to search for those microsatellites within any region of interest on a particular chromosome. The number of known markers that are already listed may not be sufficient for a particular research project for which a very large number of closely positioned markers is a necessity. But then the published genome sequence can be searched in the above database in order to obtain microsatellites, followed by designing flanking primers and the polymorphism checked for the accessions in question.

2.1.5 Need for marker densification of chromosome 3 of Arabidopsis

One of the major aims of the current research is to produce narrow STAIRS within the top 20 cM region of chromosome 3 of *Arabidopsis*. These STAIRS should contain varying lengths of

donor introgressed regions, beginning from the top region of chromosome 3, in a pure genetic background of the recurrent parent. The available CSSs of *Arabidopsis* have been produced using the accession Columbia (Col) as the recurrent parent and accession Landsberg *erecta* (L*er*) as the donor parent. The production of STAIRS in chromosome 3 involves crossing the CSS of chromosome 3 of *Arabidopsis* to the recurrent parent, further backcrossing of the resultant progeny followed by selfing of those individuals with the desired genetic make-up. Because this is a marker assisted breeding programme, the populations have to be screened at all stages with molecular markers. The genotyping will facilitate the identification of individuals with the desired genetic make up to be carried down to the next stage of the breeding programme. Therefore, the availability of a large number of closely and evenly spaced molecular markers polymorphic to Col and L*er*, along the region of interest in the chromosome is a prerequisite in producing narrow STAIRS. In addition to that, fairly uniformly distributed markers with intervals varying from about 15-20 cM were needed along the rest of the chromosome 3. So, where there is not sufficient coverage with available markers or when the previously listed markers were not informative, new markers were synthesized outside the region of 0-20 cM. In addition to the above mentioned two accessions, the *Arabidopsis* accession Niederzenz (Nd) has also been used initially in the production of CSSs. So whenever possible Nd also was checked to obtain polymorphism among all the three accessions Col, L*er* and Nd.

2.2 Objectives

- To locate all the simple sequence repeats within the first 20 cM of chromosome 3 and additional SSRs at ~20 cM intervals along the rest of the chromosome 3.
- Design flanking primers for the listed SSRs.

- Check the polymorphism of the identified microsatellites specifically for the accessions Col and L*er* and also the accession Nd of *Arabidopsis*.
- Optimize the PCR conditions i.e Magnesium ion concentration and annealing temperature for the new markers.

2.3 Materials and Methods

2.3.1 Location of Microsatellites and designing of flanking primers

The *Arabidopsis* Information Resource (TAIR) database (www.arabidopsis.org) was searched in order to obtain the microsatellites over 30 nucleotides long. All the clones spanning the top region from 0–20 cM of chromosome 3 were checked to locate the microsatellites. In addition, two microsatellites at 80 cM distance and also three microsatellites at the most distal end of chromosome 3 at 99 cM were located.

The sequences of each of the detected microsatellites, along with flanking sequence on either side of each microsatellite, were applied to a primer designing software programme: Primer 3.0 – primer designing programme (www.dur.ac.uk). From the primers designed for each region, a suitable primer pair for each SSR repeat was selected based on the following criteria,

1. Length of the primer to be from 17 – 24 bases, ideally about 21 bases.
2. Melting temperature (Tm) to be about 60^0C.
3. Primers for both forward and backward strands having approximately similar Tm.
4. The primer to be as close to the SSR repeat as possible.
5. Primer to have approximately 50: 50, AT: GC ratio.
6. Fragment size of the PCR product to be between 150-250 bp.

2.3.2 Checking for polymorphism between Col and L*er* and Optimizing PCR conditions

2.3.2.1 Extraction of DNA

Seeds of the three accessions of Columbia, Landsberg and Niederzenz of *Arabidopsis* were sown in 5-inch pots having John Innes compost mixture. The pots were placed in a glasshouse at 25 ^{0}C and a photoperiod of 16 hours.

When the plants are about 2 weeks old, leaf samples were collected for the DNA extraction. The DNA extraction was done using GenElute™ Plant Genomic DNA kit (Sigma) according the manufacturer's guidelines.

2.3.2.2 Polymerase Chain Reaction (PCR)

PCR were designed in such a way as to observe both the polymorphism for different accessions and to get the optimum PCR conditions for each marker pair from the same set of reactions.

A range of three annealing temperatures and three $MgCl_2$ concentrations were tested for each primer pair with each accession in order to observe respectively the optimum PCR conditions for each marker and the polymorphism among the tested accessions. The PCRs were designed as written below.

In order to get the optimum annealing temperature, 3 temperatures were tested.

(The lowest melting temperature of the two strands – (3^{0}C, 4^{0}C and 5^{0}C.).

In order to get the optimum salt concentration 1.5 mM, 2.5 mM and 3.5 mM $MgCl_2$ concentrations in the PCR reaction with each temperature level.

DNA of accessions Col, L*er* and Nd separately as template with each temperature level and each Mg ion concentration.

Concentrations of 10 µL PCR reactions were,

2 µl template DNA (concentration 20 ng/µl), 1 µl 10 x PCR buffer, $MgCl_2$ separate reactions with (1.5 mM, 2.5 mM. And 3.5 mM), 0.4 µl forward primer (10 pM/µl), 0.4 µl reverse primer (10 pM/µl), 0.4 µl 5 mM dNTPs, 0.08 µl Taq polymerase (5u/µl) (Bioline) and sterile distilled water to make a total volume of 10µl.

The PCRs were run in a robocycler (gradient 96 STRATAGENE) with the following thermal cycles and duration:

1 x (94^{o}C, 3 min; yy ^{o}C, 30 s"; 72 ^{o}C, 1 min)

30 x (94^{o}C, 30 s"; yy ^{o}C, 30 s"; 72 ^{o}C, 1 min)

1 x (94^{o}C, 30 s"; yy ^{o}C, 30 s"; 72 ^{o}C, 5 min)

yy ^{o}C = Melting temperature - 3 ^{o}C , 4 ^{o}C and 5 ^{o}C for each primer pair.

2.3.2.3 Gel electrophoresis

The PCR products were run firstly on 1.2% Agarose gels and post eletrophoretically stained with Ethidium bromide before observing under UV lights. When the polymorphism was not clear with Agarose the products were run on spreadex, (Spreadex EL 400 in a Submerged Gel Eletrophoresis Apparatus – SEA 2000 Elchrom Scientific) followed by cyber green staining to detect small base pair differences that differ by about 6-18 bp. The gel images were scored for polymorphism among the accessions, the optimum annealing temperature and the optimum $MgCl_2$ concentration for each primer pair.

2.4 Results

2.4.1 Location of Microsatellites within the first 20 cM and the distal end of Chromosome 3

A total of 29 sets of tandem repeats were located in the clones within the first 20 cM of chromosome 3, out of which microsatellites markers have been already published for 2

repeats - nga172 and nga162 (Bell and Ecker 1994). The clones within which the mcrosatellites were located, the orientation of the clones, the repeat type and the number, the cM distance and the base pair position of each microsatellite are listed in table 2.1.

In table 2.1 repeats from 1 to 29 are located on the top 20 cM while the repeats from 30-35 are located at the distal end of the chromosome. The distal end markers were needed to provide a sufficient and uniform coverage even outside the region of interest.

2.4.2 Designing flanking primers for the microsatellites

The synthesized primer sequences for each microsatellite are given in table 2.2

2.4.3 Optimizing PCR conditions for new markers and checking the polymorphism among accessions

Optimized PCR conditions for new markers and the polymorphisms among Col and L*er* (and for Nd when the data are generated) are presented in table 3.1. Optimized PCR conditions are presented only for the polymorphic markers.

2.5 Discussion

2.5.1 Factors considered in Primer Designing

Several factors were taken into consideration in designing flanking primers. The length of primer sequences was kept within 17-24 bases because, in general, the optimum length of primers for microsatellites is about 21 bases. Shorter sequences may result in non-specific binding, so amplifying regions outside the microsatellite and hence 21 bases is considered to be ideal.

The primers were synthesized to get a higher melting temperature about 60^0C. Normally the annealing temperature of primers are optimum about 3-5^0C lower than the melting temperature to prevent unspecific binding which occurs at lower annealing temperatures.

High melting temperatures facilitate running the PCR at higher annealing temperatures thus preventing non specific binding of primers at lower temperatures. Primers with both forward and reverse strands having approximately similar melting temperatures were selected for the same reason.

The primers as close to the SSR as possible were selected to ensure that the microsatellite is properly amplified, preventing amplification elsewhere in the genome when the PCR is run. The primers were designed to have approximately 50:50 AT:GC ratios to enable the primer to be stable and function properly so that the PCR amplification will be optimum.

2.5.2 The Types and Abundance of SSR on top of Chromosome 3

Out of the 29 tandem repeats identified within the first 20 cM, twenty four proved to be polymorphic for Col and L*er,* although with some markers the difference between the two accessions was very too small for visualisation with agarose gel systems. Four markers were monomorphic for Col and L*er* and the amplification was not observed with one marker due to improper synthesis with one of the strands. Because there was sufficient marker coverage at this region the faulty strand was not synthesized again. No comparison is made with the polymorphism for Nd with the other two accessions because the data is not complete for Nd.

Within the first 20 cM of chromosome 3 of *Arabidopsis,* only single nucleotide or di-nucleotide repeats were observed to occur among the SSRs accessed. AT is the most abundant SSR occurring 12 times out of a total of 29 (41.4%) followed by TA (24.1%). On the whole the bases A and T are the bases constructing most of the SSRs at this region of the chromosome. The SSRs within the top 20 cM of chromosome 3 and their polymorphisms for Col & L*er* are given in table 2.4.

2.5.3 Marker Coverage of the Chromosome with Microsatellites

A sufficient coverage was obtained with microsatellite markers both within the region of interest spanning 0-20 cM and also along the rest of the chromosome. Figure 2.1 indicates the marker coverage outside the 0-20 cM of chromosome 3.

There were a total of 25 polymorphic SSR markers within the first 20 cM of chromosome 3 resulting in an average interval of 0.8 cM between two adjacent markers. This dense marker coverage was required to produce narrow STAIRS within the top region of chromosome 3 with STAIRS varying from 1 cM distance. However, no polymorphic SSRs were found from 15-20 cM. The microsatellite found at 18 cM was not polymorphic for Col and L*er*. The marker coverage on top of chromosome 3 is presented in figure 2.1 with already known markers in bold letters.

Table 2.1 The clones containing sequence repeats, the type and the number of each repeat and their cM distances and the physical base pair positions.

Number	Clone	Orientation	Repeat type & number	Position (cM)	Position (bp)
1	F14 P3	3'-5'	AT(15)	2	465556-465585
2	F16B3	5'-3'	A(31)	2	526232-526262
3	T17B22	3'-5'	TA(15)	3	739768-739797
4	T12J13 a	3'-5'	AT(17)	4	893383-893416
5	T12J13 b	3'-5'	AT(15)	4	888441-888370
6	F20H23	3'-5'	TA(17)	4	942687-942720
7	T11I18	5'-3'	AT(16)	4	1058841-1058872
8	T6K12	3'-5'	TA(16)	5	
9	T12H1	5'-3'	GA(21)	6	1514904-1514945
10	F22F7	3'-5'	AT(15)	7	1604912-1604941
11	F18C1	3'-5'	A(39)	8	1648756-1648794
12	F20O10	5'-3'	AT(16)	8	-
13	F24F17	5'-3'	AT(16)	9	1815811-1815842
14	F5E6 a	3'-5'	A(33)	9	2003288-2003320
15	F5E6 b	3'-5'	T(33)	9	2020849-2020881
16	F5E6 c	3'-5'	AT(20)	9	2052699-2052738
17	F3E22	3'-5'	AG(25)	10	2132851-2132900
18	T1B9 a	3'-5'	AT(17)	10	2311870-2311903
19	T1B9 b	3'-5'	AG(21)	10	-
20	MLP3	5'-3'	TA(17)	11	2420316-2420349
21	F17O14	5'-3'	AT(27)	13	2630718-2630769
22	F8A24	3'-5'	T(34)	13	3046095-3047128
23	T22K18	3'-5'	AG(31)	14	3098809-3098882
24	T7M13	3'-5'	TA(23)	15	3403322-3403367
25	F9F8	3'-5'	AT(25)	15	3430538-3430587
26	F11B9	5'-3'	TA(17)	18	3527927-3527953

Table 2.1. Continued.

Number	Clone	Orientation	Repeat type and number	Position (cM)	Position (bp)
27	MDC11	5'-3'	TA (17)	20	4295627-4295660
28	MDC16	5'-3'	AG(21)	20	4608319-4608361
29	MAG2	5'-3'	AT(24)	20	4728485-4728532
30	T16K5a	5'-3'	AT(25)	80	18443110-18443164
31	T17J13 B	5'-3'	GA(32)	99	23042062-23042123
32	MAA21	5'-3'	TA(17)	101	23001543-23001576

Table 2.2 The number, clone, sequences and the product size for Col in designed flanking primers for the microsatellites

No:	Clone	Forward Primer	Reverse Primer	Product size bp(Col)
1	F14 P3	TGAAAATCTGCAACGTCAGC	TTGACGCTCAAGCTGCTCTA	192
2	F16B3	CAAGGTCATTTGGGATGAGG	GGTTTAAGACCGGAACAACG	207
3	T17B22	ACTGCTGGATCCCACATGAT	TTTGGTCGAAACTCGAAAGC	229
4	T12J13A	TGCTCACTGCAGTATAAAACCTG	GTCGCCGATTATGATTGTCC	209
5	T12J13B	GTAAAACGCTTAGTGATCTGATTTT	GGAGAGCAGGAATCGACAAG	213
6	F20H23	ACGAGCGTATCAATGGGAAG	CAGACGCAATGTACGCATTT	199
7	T11I18	CGGTAGAGACAAGGTGTTGGA	CGTGGACCTCTTGATACATTTTG	221
8	T6K12	AGCGGTAGAGACAAGGTGTTG	CGTGGACCTCTTGATACATTTTG	221
9	T12H1	TGCAAGCAACAGACTTTGAA	CGATTCATGCAGGTTTCTGA	215
10	F22F7	GCTCCGTCTCCAATGAACA	TGCCACCTTCCTTAAATTGG	238
11	F18C1	AAAACCGGTTGTGTAATGAAAA	TCTGCGTTGAATTGTCAACAT	173
12	F20O10	TTGAAAACGGGTTTTGTTAAAAG	GTGCGATCCTTTTCACGTTT	204
13	F24F17	AAAGTAGTTTGCATGGGACGA	TTGAAAACGGGTTTTGTTAAAAG	204
14	F5E6 A	TTTTTCCCATGTGGTGAGAG	TTGCACCAAACCAAGTAACAA	209
15	F5E6 B	TGCCAAATTCCAACAAATCTT	GCGTTTGAATTGAATCTAGGG	224
16	F5E6 C	TGGCAAGATCTCCTTCACAA	GCCATAGGGTTTAGGGTTCA	230
17	F3E22	AACCTGATCGGGATTGACAG	TCGCTTCCATTTCTTCTTCC	235
18	T1B9 A	GCCTGGAAACAAACTTCAGC	TGAAGTTGGTGCTTAGAGTTGC	193
19	T1B9 B	GCCAACTTGGGTTTTCCATA	TTTCATCTCTCCACGTTCCTTT	200
20	MLP3	TGATGATGACGACGACAATG	AGCCGACGAGAGACAGAGAA	242
21	F17O14	GGAATGAACCCGCTTAAAAA	TGATTCTGTTCCTAATAAACCCATAA	281
22	F8A24	CGGCAACTAGCAAGAGACG	TCGAGCTTTTAAATTTTGTTACCC	201
23	T22K18	GATCGTGGGTTACCTTTTGG	TGTATCAAGAGCAATATCAAGA GCA	199

Table 2.2 Continued.

Number	Clone	Forward Primer	Reverse Primer	Product size bp Col
24	T7M13	CAAAAGTCGGATCCACAACC	TGATTTTGATGCAGCCGTTA	182
25	F9F8	GCTGCTTAACTTGTGAGTGGTC	TTTGGTCGTATCCTCTTATCGAA	203
26	F11B9	TCATCATCATCATCAACCATCA	TGAATTCACCGGAAAGAGTTG	218
27	MDC11	CTTCTCGGGGTTCTCCTCTT	AGGGAAACCCCTGTTTTACG	225
28	MDC16	GGAGTGGCCTCGTGTAGAGA	GCTCCTGAGTTTCGGACAGA	187
29	MAG2	GGTTGTGGTCGCTAGCAAAT	AAGGTCACAGAAAATGAATACCG	238
30	T16K5a	TGTCATCTGCTTATGGCAAAC	AGCTGATGAGTGGATGTGGA	-
31	T17J13 b	GTGGGTCCCTTGGATTTCTT	AAGGAAGCGAATCCGAAAA	-
32	MAA21	GGCCATTCTCAAACACA	CCATACTCCACCAGACC	-

Table 2.3 The polymorphisms for Col & L*er* and Nd (where data were generated), the optimum annealing temperatures and Mg concentrations for each microsatellite. (P=polymorphic ; M = monomorphic ; NA = did not amplify)

No.	Clone	Polymorphism			Opt. Ann. T. ^{0}C	Opt. Mg. Conc. mM
		Col & Ler	Col & Nd	Ler & Nd		
1	F14 P3	P	-	-	58	1.5
2	F16B3	P	P	P	58	1.5
3	T17B22	P	M	P	59	2.5
4	T12J13a	NA	NA	NA	-	-
5	T12J13b	P	M	P	58	1.5
6	F20H23	P	M	P	55	2.5
7	T11I18	P	P	P	59	2.5
8	T6K12	P	P	P	59	1.5
9	T12H1	P	M	P	52	1.5
10	F22F7	M	M	M	-	-
11	F18C1	P	-	-	57	2.5
12	F20O10	P	P	-	56	1.5
13	F24F17	P	P	P	57	1.5
14	F5E6 a	M	M	M	-	-
15	F5E6 b	P	-	-	57	1.5
16	F5E6 c	P	-	-	58	2.5
17	F3E22	P	M	P	59	2.5

Table 2.3. continued.

No.	Clone	Polymorphism			Opt. Ann. T. ^{0}C	Opt. Mg. Conc. mM
		Col & Ler	Col & Nd	Ler & Nd		
18	T1B9 a	P	-	-	55	1.5
19	T1B9 b	P	-	-	56	1.5
20	MLP3	P	-	-	58	1.5
21	F17O14	P	-	-	57	3.5
22	F8A24	M	-	-	-	-
23	T22K18	P	P	M	59	2.5
24	T7M13	P	P	P	58	1.5
25	F9F8	P	-	-	57	2.5
26	F11B9	M	P	P	56	1.5
27	MDC11	P	P	P	58	1.5
28	MDC16	P	P	P	59	1.5
29	MAG2	P	P	M	60	2.5
30	T16K5a	P	-	-	57	2.5
31	T17J13 b	P	M	P	57	2.5
32	MAA21	-	NA	NA	-	-

Table 2.4 : SSRs on top 20 cM of chromosome 3 of *Arabidopsis* and their polymorphism for Col and L*er*. (NA = did not amplify)

SSR	No. occurrences	no. polymorphic	no. monomorphic	NA
AT	12 (41.4%)	10	1	1
TA	7 (24.1%)	6	1	-
AG	4 (13.8%)	4	0	-
A	3 (10.3%)	2	1	-
T	2 (6.7%)	2	0	-
GA	1 (3.4%)	1	0	-
Total	29	25	3	1

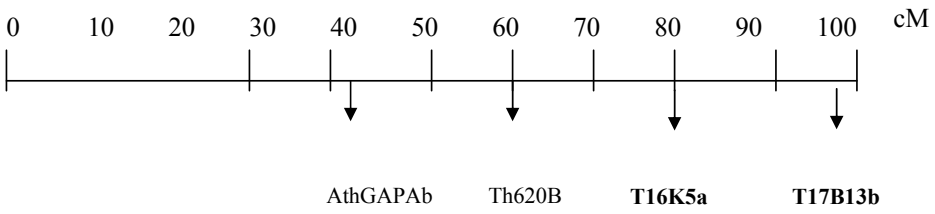


Figure 2.1: The coverage of markers outside the region of interest.

Newly synthesized markers are in bold letters.

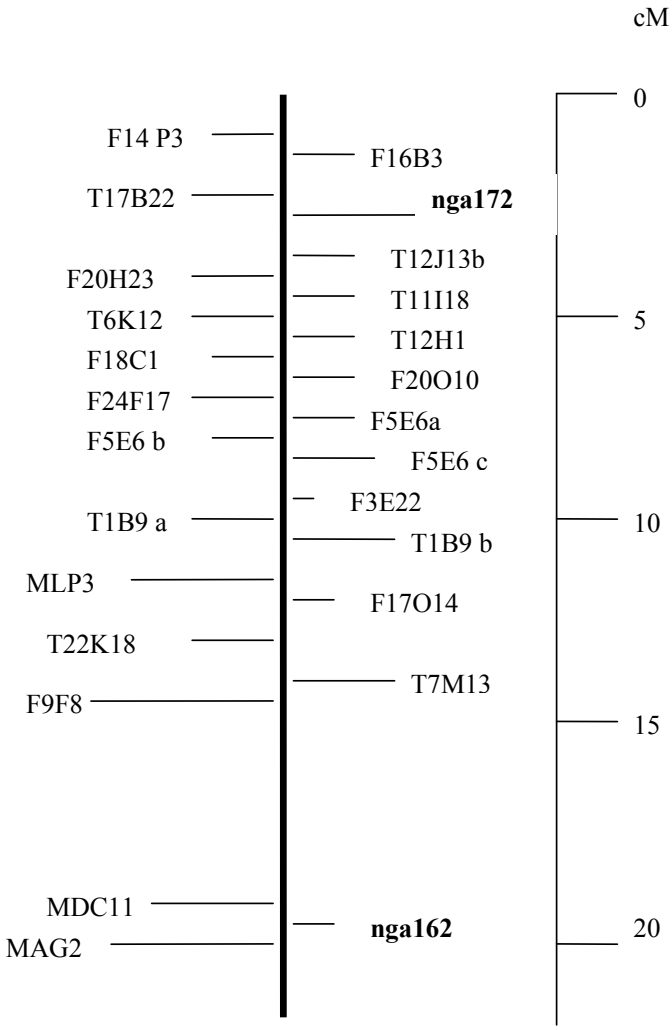


Figure 2.2: The coverage of first 20 cM of chromosome 3 with microsatellite markers.

Already known microsatellite markers are in bold letters.

CHAPTER 3

EVALUATION OF QUANTITATIVE TRAITS IN *ARABIDOPSIS* ACCESSIONS COLUMBIA AND LANDSBERG

3.1 Background

The availability of genetic variation is a prerequisite for the functional analysis of genomes. At present, the functional analysis of *Arabidopsis* genes, mainly of qualitative nature is based largely on the phenotypic characterization of mutants. These mutants have been selected by forward and reverse genetics in a few laboratory 'wild type' genotypes. The generally used inbred strains for such studies are Columbia (Col), Landsberg *erecta* (L*er*) and Wassilewskija (Ws); which were originally collected from the wild by the pioneers of *Arabidopsis* research. As an alternative to using laboratory induced mutants the naturally occurring variation in these ecotypes can be made use of in dissecting complex traits of *Arabidopsis* (Alonso-Blanco and Koornneef 2000).

At the earlier stages, the phenotypic characterization of ecotypes collected from the wild from different geographical regions revealed considerable genetic variation among them (Redei 1970). *Arabidopsis* is predominantly a self-pollinating plant and thus, most of the collected plants represent inbred lines that are practically homozygous. These wild homozygous lines are called ecotypes, which is a term used to define distinct races of a species genetically adapted to particular habitats. At present, these laboratory experimental lines are referred to as different accessions, which is a term often used in germplasm collection to refer to a plant genotype of a species collected at a specific location (Alonso-Blanco and Koornneef 2000).

Accessions Col and L*er* have been used in our current research of producing CSSs and STAIRS in *Arabidopsis*. These two lines were the obvious candidates because of the availability of extensive existing genetic analysis and the sequence information about these

two lines. There are several different lines within the accession Columbia, named Col-1, Col-0, Col-2, Col-3, Col- 4, Col- 5 etc. (www.arabidopsis.org). Col-1 is the wild type Columbia while Col ecotypes used separately in different laboratories all over the world have been given different number prefix and are maintained in *Arabidopsis* stock centres as bulk lines or single seed descent lines. All these lines have been collectively referred to as Columbia or Col in reporting research. Landsberg *erecta* displays considerable morphological variation compared with Columbia and hence the lines Col and L*er* have been extensively used in QTL analysis with segregating populations (Jansen et al. 1995; Mitchell-Olds and Pedersen 1998; Swarup et al. 1999).

Accession Col has been used as the recurrent parent in producing *Arabidopsis* CSSs with L*er* being used as the donor or the introgressed parent. In QTL analysis, sufficient genetic variability should be present between the two selected parents for an efficient analysis. While there was considerable genetic variability between Col and L*er*, different lines of Col are expected to be morphologically and genetically similar. However, it was observed that Col-0 and Col-5 differ in morphology with respect to the density of trichomes on the leaves. The leaves of Col-0 were densely hairy compared to the relatively very smooth leaf blades of Col-5. There were no other morphological differences that could be observed with the naked eye between these two Col lines. It was considered to be worthwhile to assess the quantitative trait variation, if there is any, between the two Col lines and these two lines together against L*er* because both Col-0 and Col-5 have been used as the recurrent parent and L*er* as the donor introgressed parent in producing CSSs and STAIRS. In this chapter the evaluation of morphological QTL traits and genotyping marker data with the accession Col-0 and Col-5 against L*er* lines that have been used in our laboratory is explained.

3.2 Objectives

- To identify the morphological differences between the ecotypes Col and L*er* and to check for the morphological similarity/ dissimilarity if there is any between Col-0 and Col-5 for important quantitative traits.
- To compare molecular genotypic differences between Col-0 and Col-5 using polymorphic microsatellite markers for Col and L*er*.

3.3 Materials and Methods

3.3.1 Plant house trial for scoring quantitative traits

A total of 120 pots comprising forty pots each from the three *Arabidopsis* lines, Col-0, Col-5 and L*er*-0 were sown. For each different line seeds from different replicate mother plants were used. In total five seed parents for Col-0, three seed parents for Col-5 and three seed parents for L*er*-0 were used. The growth medium was John Innes compost mixture filled into 5 cm pots. The growth conditions were maintained at 24^0C temperature, long day conditions of 16-hour photoperiod at a light intensity of 150-350 μE m^{-2} s^{-1}. Watering was done regularly at daily intervals.

Three seeds were placed in a strait line along the diameter of each pot and they were placed in a completely randomised experimental design surrounded by a peripheral guard row. The germination of the three seeds in each pot was recorded separately and later, thinning down was done leaving a single plant in each pot.

3.3.2 Traits scored

The plants were scored for 21 morphological characters at different time points. They are listed below. All the time points are the number of days from sowing unless otherwise stated.

- Germination time.
- Rosette leaf number at days 20, 25, 30 and at flowering.
- Cauline leaf number at days 20, 25, 30 and at flowering.
- Rosette width at days 20, 25, 30 and at flowering.
- Height at days 20, 25, 30, 35, 40 and at flowering.
- Date of flowering - number of days from germination to the opening of the first flower.

3.3.3 Data analysis

The data were analysed in MINITAB version 16 by using the General Linear Models Procedure (GLIM) for the analysis of variance. The analysis was conducted for all the three lines together and then using only Col-0 and Col-5. In ANOVA the lines and the tubes (denoting different mother plants) within lines were considered as the sources of variance and the factor 'tube within line' was considered as a random factor.

3.3.4 Genotyping the lines

Genotyping Col-0 and Col-5 was performed with 25 microsatellite markers which were known to be polymorphic for Col and L*er*. Most of the markers were on chromosome three while some were on other chromosomes. The DNA extraction protocol explained in appendix 01 was used for the extraction of genomic DNA. Genomic DNA extracted from Col-0, Col-5 and L*er* were used as templates and 10 µl PCR reactions were done using the same PCR protocol explained in length in Chapter 5. Non-denaturing polyacrylamide gel electrophoresis was performed and the gels were post-electrophoretically stained with Ethidium bromide and the bands scored under UV light. The microsatellite marker list is given in table 3.1

3.4 Results and Discussion

3.4.1 Analysis of Morphological Quantitative Traits

The means of the three lines for the 21 characters scored, standard errors of the means and significance levels (when a significant difference is observed) from ANOVA are presented in table 3.2. The Analysis of Variance tables resulted from GLIM procedure in MINITAB are given in appendix 02.

The lines and the replicates of seed parent were considered as the sources of variation in Analysis of Variance. The replicates of seed parents within a line were considered random factors while the lines were regarded as fixed factors.

The analysis of the two Col lines revealed that the lines are not significantly different for any of the characters measured except for rosette width. Rosette width was measured at four different time points and has shown to be significantly different at all the four time points between Col-0 and Col-5. Other than this difference observed in diameter of the rosette of the plant and the density of the trichomes that was already known, the two Col lines do not differ for any of the quantitative morphological traits measured. However, the replicates (consisting of different seed parents) were significantly different for some of the traits scored (e.g. germination time, RW day 20, RLN day 25, RLN at flowering) indicating significant maternal effects at germination and continuing longer with rosette leaf number.

All the traits scored, except the time taken for germination, cauline leaf number at day 30 and at flowering and height at day 30, were significantly different between Col and L*er* as revealed by analysing all the 3 lines together.

It was observed that Landsberg *erecta* grows faster than Columbia in height at the initial stages, and reaches its maximum height sooner. In comparison the initial growth rate of Columbia in height is slower than Landsberg, but later reaches a maximum height which is

considerably higher than Landsberg. As a result, in the middle stages around day 30, the heights between Col and L*er* do not differ significantly. Similarly, the cauline leaf number is significantly different between Col and L*er* at the initial stages but in latter stages, around the time of flowering, becomes non significant.

The results can be made use of as a guideline in determining time points for each specific character in future morphological trait scoring experiments. Furthermore the results highlight the importance of replication in order to account for the maternal effects. This is especially relevant in scoring for traits of early vegetative growth.

3.4.2 Genotyping the lines with microsatellite markers

Col-0 and Col-5 were monomorphic for all the markers tested indicating no genetic differences between the two at the microsatellite loci while Col and L*er* were all polymorphic.

Table 3.1 Microsatellite markers used in genotyping accessions Col-0 and Col-5 and their positions.

Marker name	Marker position	
	chromosome	cM distance
nga59	1	2.9
nga280	1	72.9
nga1145	2	9.6
nga361	2	63.02
F16B3	3	2
T17B22	3	3
F20H23	3	4
T11I18	3	5
T6K12	3	6
T12H1	3	7
F18C1	3	8
F5E6 b	3	9
T1B9 a	3	10
T1B9 b	3	11
F8A24	3	14
F9F8	3	15
F11B9	3	18
MDC11	3	20
T16K5a	3	80
T17J13 B	3	99
T18A10	4	1
T5J17	4	119
MED24	5	7.4
MM19	5	116.9

Table 3.2: The means of the three lines for the 21 characters scored, standard errors of the means (in parenthesis) and significance levels from ANOVA.

(NS = non significant)

Trait	Means (SE mean)			Sig. levels; all lines	Sig. levels; 2 Col lines
	Col-0	Col-5	Ler-0		
germination	4.210 ± (0.086)	4.368 ± (0.096)	4.25 ± (0.073)	NS	NS
Rosette leaf number					
At day 20	10.711± (0.141)	10.842 ± (0.208)	8.333 ± (0.120)	>0.001	NS
At day 25	12.421 ± (0.212)	13.158 ± (0.218)	7.083 ± (0.156)	0.006	NS
At day 30	12.289 ± (0.206)	12.632 ± (0.221)	6.000 ± (0.199)	>0.001	NS
At flowering	12.316 ± (1.185)	13.211 ± (0.220)	7.611 ± (0.184)	0.002	NS
Cauline leaf number					
At day 20	1.526 ± (0.188)	1.447 ± (0.206)	2.278 ± (0.141)	0.016	NS
At day 25	2.895 ± (0.140)	2.737 ± (0.184)	5.528 ± (0.476)	0.001	NS
At day 30	22.45 ± (1.02)	20.47 ± (1.22)	17.55 ± (0.543)	NS	NS
At flowering	3.18 ± (0.067)	3.31± ± (0.09)	4.22 ± (1.51)	NS	NS
Rosette width					
At day 20	57.29 ± (1.76)	50.45 ± (1.41)	48.00 ± (1.17)	0.003	0.014
At day 25	75.37 ± (2.56)	64.84 ± (2.29)	57.67 ± (1.22)	0.005	0.045
At day 30	98.84 ± (2.13)	81.24 ± (1.76)	55.25 ± (1.17)	>0.001	0.001
At flowering	82.63 ± (1.74)	73.05 ± (1.58)	51.53 ± (1.43)	>0.001	0.032
Height					
At day 20	6.263 ± (0.428)	6.000 ± (0.500)	13.72 ± (1.19)	0.001	NS
At day 25	55.74 ± (4.44)	54.58 ± (6.10)	86.69 ± (4.72)	0.006	NS
At day 30	251.29 ± (9.85)	224.4 ± (12.7)	200.56 ± (3.78)	NS	NS
At day 35	404.95 ± (8.54)	377.92 ± (8.16)	277.28 ± (5.85)	>0.001	NS
At day 40	446.50 ± (7.90)	428.55 ± (6.95)	295.40 ± (4.63)	>0.0001	NS
At flowering	71.26 ± (2.93)	78.47 ± (2.60)	50.14 ± (1.86)	0.002	NS
Flowering time	25.737 ± (0.255)	26.237 ± (0.323)	23.389 ± (0 .166)	0.001	NS

CHAPTER 4

QTL ANALYSIS USING WIDE STAIRS OF CHROMOSOME 3 IN *ARABIDOPSIS THALIANA*

Abstract

A study was conducted to confirm the genotypes of 16 lines known to be wide STAIRS of chromosome 3 of *Arabidopsis*. The second part of the experiment involved the QTL analysis of the wide STAIRS for twelve quantitative traits by scoring the phenotypes of STAIRS grown in a controlled environment. The major objective was to identify the regions of interest that contain QTL related to flowering time, in chromosome 3 of *Arabidopsis* in order to proceed to develop narrow STAIRS for fine mapping of the particular trait. The genotyping with nine microsatellite markers identified the sixteen lines as belonging to eight steps of STAIRS within chromosome 3. The analysis of variance in MINITAB revealed significant differences for between STAIRS for all the quantitative traits scored. It was possible to identify 21 QTLs along chromosome 3 affecting the 12 traits measured by least squares model fitting. The region spanning 0- 20 cM proved to be a QTL hotspot which housed QTL for all the traits scored for in the experiment. The experiment provided conclusive evidence for the decision to proceed for the construction of narrow STAIRS within the top 20 cM of chromosome 3 for fine mapping QTL especially related to flowering time.

4.1 Introduction

Quantitative traits are typical of commercially important traits in crop plants and live-stock as well as in vital traits in humans. Until the invention of molecular markers quantitative trait analysis methods were limited to the approaches of biometry. By employing analysis methods of quantitative traits in biometry, considerable progress was made in advancing knowledge of the genetics of traits, predicting response to selection followed by crop improvement in agriculture (Kearsey and Farquhar 1998). However, with the approaches of biometry alone, it was not possible to unravel the 'black box' nature of the genetic architecture of quantitative traits. As a result vital information such as the number of genes that affect the trait, the effects of each gene or their locations in the genome remained unknown to the scientist until the development of molecular marker systems in quantitative trait analysis (Dekkers and Hospital 2002).

The advent of molecular markers facilitated the wide spread adaptation of QTL mapping and provided a better approach for understanding the complex quantitative traits. The use of polymorphic molecular markers in segregating populations such as F_2s, DHLs or RILs enables the construction of the framework map, around which the QTL can be located. Yet, there are several important drawbacks with this approach of locating QTL, particularly regarding the fine mapping. The main problems associated with conventional QTL mapping methods for precise QTL location are: the very wide confidence intervals, failure to give a precise location for QTL, underestimation of the number of QTLs for a particular trait, overestimation of their effects, the necessity for the character to be highly heritable and the requirement for very large experimental populations.

A solution to this problem is to use substitution lines for locating QTL. Ideally, substitution lines carry a single defined segment of a donor genotype in a pure genetic background of a

recurrent genotype. In *Arabidopsis* the whole chromosome substitution lines are available from which the wide STAIRS have been derived. To make a complete set, five whole chromosome substitution lines with respect to the haploid number of *Arabidopsis* are needed (Figure 4.1). For each CSS the respective set of STAIRS can be constructed with the number of STAIRS depending upon the precision with which the QTLs need to be located (Koumproglou et al. 2002).

STAIRS are homozygous single recombinant lines. They have the recurrent parent genotype at one end and donor genotype at the other end of the particular chromosome due to the occurrence of the single recombination, along the chromosome. When all the STAIRS of a chromosome are stacked together a stepwise ladder-like progression results (Figure 4.2).

STAIRS are constructed by means of several rounds of back crossing of each CSS to the recurrent parent, using a marker assisted breeding programme. Genotyping the progeny at each stage of the breeding programme assists in selecting desirable individuals to be carried down to the subsequent stage and ultimately to verify the authenticity of the introgression region of the STAIRS constructed. Determination of the width of each step in STAIRS depends upon the density of the coverage of the chromosome with polymorphic genetic markers. The width of each step in the ladder-like progression when the STAIRS are sequentially stacked, gives an indication of the level of precision in QTL mapping using STAIRS. The wider STAIRS help locating the QTL to a wider length of a chromosome while the narrower STAIRS facilitate more precise map location. Thus, fine mapping of QTL can be achieved in three steps using the resources CSSs and STAIRS by growing these lines in a controlled environment using an appropriate experimental design and scoring for quantitative phenotypes (Koumproglou et al. 2002) (Figure 4.3).

- Using the CSSs respective QTLs can be located to particular chromosome/chromosomes.
- With wide STAIRS the region of QTL can be identified to a region of about 5-10 cM along a particular chromosome.
- With narrow STAIRS within the region of interest resulting from analysis with wide STAIRS, QTL can theoretically be located to a region less than 1 cM.

Koumproglou et al used the accession Columbia (Col) as the recurrent parent and accession Landsberg *erecta* as the introgressed donor parent in producing CSSs of *Arabidopsis*. They also constructed several wide STAIRS in chromosome 3 and reported the presence of significant QTL for flowering time, rosette leaf number and plant height at day 35 at the top region of chromosome 3, spanning 0 – 20 cM by the statistical analysis of data from a field trial under controlled environment. The current study explains the further verification of the genotypes of wide STAIRS using microsatellite markers and the genetic analysis of quantitative data obtained from a plant house trial using the STAIRS used by Koumproglou et al. together with several additional STAIRS in chromosome 3 of *Arabidopsis*.

4.2 Objectives

- To verify the genotypes of sixteen wide STAIRS in chromosome 3 of *Arabidopsis*.
- To grow the wide STAIRS in a growth chamber under controlled environment to score the phenotypes of the quantitative traits.
- To perform genetic analysis of quantitative data for QTL mapping in chromosome 3 of *Arabidopsis*.
- To verify the region of interest for certain traits in chromosome 3 to proceed to the production of narrow STAIRS.

4.3 Materials and Methods

4.3.1 Genotyping wide STAIRS

4.3.1.1 Molecular markers used

Sixteen lines known to be STAIRS of chromosome 3 of *Arabidopsis* from Koumproglou et al. (2002) were genotyped to verify the location of recombination using nine microsatellite markers positioned along the chromosome 3. Five out of the nine markers used were newly developed microsatellite markers during this research project (explained in chapter 2) while four markers were published earlier (Bell and Ecker 1994).

The names of the nine primers are F16B3, nga172, T22K18, nga162, MAG2, AthGapAB, Th620B, T16K5a and T17J13b which are positioned respectively at 2cM, 6.9 cM, 14 cM, 21 cM, 21 cM, 44 cM, 59 cM, 88 cM and 99 cM along chromosome 3. Out of the above markers the primer sequences, magnesium ion concentration and the annealing temperatures in the PCR for the newly developed markers (F16B3, T22K18, MAG2, T16K5a and T17J13b) are given in Chapter two. The corresponding details for the previously published markers are as follows.

1) **Marker nga172 :**

Forward primer sequence	AGCTGCTTCCTTATAGCGTCC
Reverse primer sequence	CATCCGAATGCCATTGTTC
Annealing temperature	52^0C
Magnesium ion concentration	3.5mM

2) Marker nga162 :

Forward primer sequence	CATGCAATTTGCATCTGAGG
Reverse primer sequence	CTCTGTCACTCTTTTCCTCTGG
Annealing temperature	54^0C

3) Marker AthGAPAB :

Forward primer sequence	CACCATGGCTTCGGTTACTT
Reverse primer sequence	TCCTGAGAATTCAGTGAAACCC
Annealing temperature	54^0C

4) Marker TH620B

Forward primer sequence	CAGAAATCGACGTCGATACGAA
Reverse primer sequence	GGGCAGAGAGAACTAAAAAAGC
Annealing temperature	51^0C

4.3.1.2 Extraction and quantification of DNA.

DNA of the sixteen lines was extracted from two week old plants grown in a growth room (16 hour day length, 24^0C) using a modified C-TAB DNA extraction procedure which is given in appendix 01.

DNA quantification was done using spectrophotometer (Eppendorf Bio Photometer) and the dilution done with TE buffer pH 8.0, to obtain a final concentration of 50ng/µl.

4.3.1.3 Polymerase Chain Reaction (PCR)

Ten µl PCR reactions were set up in 0.5 mL PCR tubes with the following ingredients and quantities.

Template DNA 50ng/µl, 2 µl,10x PCR buffer (Moltaq)* 1 µl, 50 mM $MgCl_2$ 0.2µl, mM dNTPs,0.4 µl,Forward primer (10pM/µl), 0.4 µl,Reverse primer (10pM/µl) 0.4 µl, Taq DNA polymerase (Moltaq) (5u/µl) 0.08 µl,Sterile distilled water to volumerize to 10 µl.

* Moltaq PCR buffer contained 1.5 mM $MgCl_2$ and therefore only 0.5 µl of $MgCl_2$ was added additionally into the PCR reaction.

A drop (20 µl) of mineral oil was added on top of the PCR mixture to prevent evaporation from the tube during the PCR cycles. The reactions were run in a thermal cycler (HYBAID Omnigene) using the following amplification cycles.

1 x (94^{o}C, 5 min; yy ^{o}C, 30 s"; 72 ^{o}C, 1 min)
30 x (94^{o}C, 30 s"; yy ^{o}C, 30 s"; 72 ^{o}C, 1 min)
1 x (94^{o}C, 30 s"; yy ^{o}C, 30 s"; 72 ^{o}C, 5 min)
min = minutes; s''= seconds ; yy = annealing temperature for each primer.

4.3.1.4 Gel electrophoresis

The electrophoresis of PCR products was done on non-denaturing polyacrylamide (AccuGel 29:1 by National Diagnostics) gels. The gels were poured into sealed glass plates at a final concentration of 8%. To achieve 8% final concentration the reagents were mixed in the following volumes.

AccuGel mixture 29:1 Acrylamide : Bisacrylamide 10 mL ,10x buffer TBE 5 mL,TEMED (N,N,N',N'-Tetramethyl diamine (Sigma)30 µL, Ammonium per sulphate (30%) 500 µL and distilled water to make up to a final volume of 50 mL.

The gels were left to polymerise for thirty minutes after pouring before setting up the gel running apparatus (Vertical Electrophoresis Unit V-10WCDC- Geneflow Ltd.) and loading 5 µl of the amplified PCR reaction to which 3 µl of loading buffer had been added. The gels were electrophoresed for 60 – 90 minutes depending upon the fragment size, at a constant voltage of 120 volts using a power supply unit (Bio-RAD Power PAC).

Upon the completion of the gel run, the two glass gel plates were carefully separated to get the gel on one plate, followed by staining in 2 µl/100ml Ethidium Bromide solution for 30 minutes and detection and scoring of bands under UV light.

4.3.2 Plant house experiment for scoring quantitative trait phenotypes

4.3.2.1 Experimental design and growth conditions.

A total of five hundred and sixty plants consisting of thirty five individuals from each of sixteen lines were grown in a completely randomised experimental design surrounded by a peripheral guard row. Each plant was grown in a 5-inch pot filled with John Innes (number 01) compost mixture. Three seeds were placed in each pot at sowing in order to allow for non-germination, and ten days after germination the plants were thinned down to one plant per pot. The environment in the growth chamber was controlled, and maintained at 16-hour photoperiod and 24^0C temperature.

The phenotypic traits, germination time (dates from sowing to germination), leaf number at day 20 (LN-20), Rosette width at day 20 (RW-20), height at day 30 (H-30), height at day 35 (H-35), flowering time (FT), height at flowering (HFT), rosette width at flowering (RWF), rosette leaf number at flowering (RLNF), cauline leaf number at flowering (CLNF), height at day 45 (H-45) and height at day 54 (H-54) were scored for each plant. All the length measurements were taken in millimetres and the flowering time was measured as the number of days from germination to the opening of the first flower.

4.3.2.2 Analysis of phenotypic data

Analysis of Variance (ANOVA)

The data were analysed in MINITAB version 16 using General Linear Models Procedure which accommodates the analysis of variance with missing values. For ANOVA all the lines having the same genotype for all the marker bins were considered as common STAIR genotypes. A hierarchical ANOVA was performed considering genotypes and the lines within respective genotypes as the sources of variation. The genotypes were considered as fixed factors and the lines within genotypes were considered as random factors. Because the time

for germination was revealed to be significantly different among the genotypes and lines within genotypes it was used as a covariate in the ANOVA for all the other traits to account for the variation this will impose on the other traits scored.

Correlation among traits

The correlation coefficients between various traits were calculated by Pearson's Correlation method in MINITAB to calculate the inter-trait relationships.

Least Square Model Fitting

Weighted least squares model fitting was performed for significantly different traits among STAIR genotypes as revealed by the ANOVA, in order to locate the QTLs to specific marker bins in chromosome 3. Results of the ANOVA were combined with genotypic data to construct a model in locating QTLs by the least squares model fitting. Weights were calculated (weight = n /variance) for the means of the STAIR genotypes for each of the significantly different traits and the least squares model fitting was done. Models which are adequate as identified by the non significant Chi-squared values for the models with the least number of significant parameters were chosen as the best fit model for each trait. The regions containing QTL and the effects of substitution of Columbia alleles by Landsberg alleles were calculated for each QTL region.

4.4 Results

4.4.1 Summary of the results of genotyping

Out of the sixteen different lines fourteen were found to be STAIRS within seven bins while two were CSS3 having intact L*er* genotype along the whole length of chromosome 3. The genotype of each line for each marker is given in table 4.1

Based on the genotypes fourteen lines were grouped into seven STAIRS demarcated by different marker bins. In this grouping the lines having the same genotypes for all the markers are genetically similar for all the bins and considered to form one STAIR genotype. The graphical representation of these STAIR genotypes showing Col and L*er* regions and the regions of crossovers, marker positions and the number of replicated lines within each genotype are given in figure 4.4.

4.4.2 Analysis of quantitative trait variation

Analysis of variance by GLIM procedure revealed statistically significant differences among the genotypes for all the characters scored. Lines within genotypes proved to be significantly different for all the traits except for leaf number and rosette width at day 20 and height at day 35. When the lines within STAIRS showed significant differences in ANOVA, the mean squares of the lines within STAIRS were considered as variations in calculating the weights for model fitting. The quantitative traits analysed and their means and standard deviations are given in table 4.2. Graphical representations of trait means after adjusting for the covariate, germination time are given in figure 4.5

4.4.3 Least Squares Model fitting for significantly different genotypes

The model used for detecting and estimating genetic differences among STAIRS is given in table 4.3. According to this model the marker demarcated regions are as follows in genetic distances.

Region a1 = 0 -14 cM

Region a2 = 14 -20 cM

Region a3 = 20 – 44 cM

Region a4 = 44- 60 cM

Region a5 = 60 – 80 cM

Region a6 = 80 – 100 cM

The combination of the results of the ANOVA and the least square model fitting for each significant character are discussed below in identifying possible QTL regions that explain the differences among genotypes. The ANOVAS are given in appendix 03, the tables of observed means, expected means and weights in appendix 04 and an example of a model fitting result sheet in appendix 05.

1) Germination time

Germination time was highly significantly different both among the genotypes and the lines within genotypes in ANOVA. The best-fit model with a chi-squared value of 7.3099 with 5 d.f. identified regions a1 and a5 as having significant QTL. When Landsberg substitutes Columbia in region a1, it increases the germination time by 1.90 days. Similarly when L*er* substitutes Col in region a5 the germination is delayed by 2.08 days.

2) Leaf number at day 20

Statistically significant differences were observed among genotypes in ANOVA. However, it was not possible to fit a model that describes the variation satisfactorily with the least squares model fitting.

3) Rosette width at day 20

The ANOVA revealed significant differences among genotypes for rosette width at day 20. Significant QTL were identified by the best-fit model with non significant chi-squared (χ^2 = 3.49, 5 d.f) in regions a1 and a2. The substitution of Col by L*er* in region a1 decreases the RW-20 by 5.3 mm while the substitution of Col by L*er* in region a2 increases the said trait by 4.5 mm.

4) Height at day 30

The best fit model (χ^2 = 5.356, 4 d.f) uncovered three QTLs in regions a1, a2 and a3 to account for the statistically significant differences shown in the ANOVA. In region a1 substitution of Col by L*er* decreases the trait by 0.69 mm. The same substitution in region a2 increases the height at day 30 by 1.1 mm while in region a3 L*er* alleles in place of Col increases the height at day 30 by 0.52 mm.

4) Height at day 35

The ANOVA has shown significant differences among genotypes for height at day 35. The best fit model (χ^2 = 4.045, 4 d.f) revealed three QTLs with similar effects to those QTLs found for height at day 30 which is the same trait scored five days earlier. The significant QTL in region a1 showed a decreasing substitution effect of Col by L*er* by 15.642 mm. The substitution of Col by L*er* in region a2 increases the height at this time point by 15.95 mm. The same substitution in region a3 increases the phenotype by 1.72 mm.

5) Flowering time

Very highly significant differences were observed among genotypes for flowering time in ANOVA. ($p > 0.0001$) The differences were attributed to a single QTL region within a2 by the best fit model ($\chi^2 = 9.2519$, 6 d.f) The substitution of Col alleles in region a1 by L*er* alleles causes delaying in flowering time by 10.767 days under these experimental conditions.

6) Height at flowering time

ANOVA for height at flowering time showed significant differences among the STAIRS ($p = 0.001$). QTL that affect on this trait were located in regions a1, a3 and a6 by the best-fit model. ($\chi^2 = 4.5219$, 4 d.f) Substitution of Col in regions a1 and a6 has decreasing effects on height at flowering by 19.536 mm and 12.36 mm respectively. In contrast the substitution of Col by L*er* in region a3 causes an increase in the height of the plant at flowering by 8.92 mm.

7) Rosette leaf number at flowering time

Very highly significant differences were observed in the ANOVA ($p>0.0001$) among genotypes for rosette leaf number at flowering. This variation was attributed to QTL present in the region a1 by the best-fit model ($\chi^2 = 7.042$, 6 d.f). Accordingly, the substitution of Col in this region by L*er* has an increasing effect on the phenotype by 7.214 mm

8) Cauline leaf number at flowering

The cauline leaf number at flowering showed significant differences among genotypes ($p=0.01$). The least squares model-fitting procedure resulted in a best-fit model ($\chi^2 = 5.0135$, 6 d.f) that attributes the variation among genotypes to QTL present in the region a1. When Landsberg substitutes Columbia in region a1 the cauline leaf number at flowering increases by 0.606.

9) Rosette width at flowering

The ANOVA revealed significant differences for rosette width at flowering (p=0.002). This variation was attributed to QTL in region a1 in chromosome 3, by the best-fit model. (χ^2 = 6.331, 6 d.f) The expected means of the best-fit model revealed that the substitution of Col alleles by L*er* alleles in region a1 has an increasing effect on rosette width at flowering by 27.855 mm.

10) Height at day 45

The variation among STAIRS genotypes was very highly significant as shown by ANOVA. The best fit model (χ^2 = 8.839, 5 d.f) identified two regions of QTL that affect this trait at this time point, region a1 and region a2. The substitution of Col in region a1 by L*er* decreases the phenotype by 145.2 mm, while the same substitution in region a2 increases the phenotype by 48.045 mm.

11) Height at day 54

The variation among genotypes was very highly significant as revealed by the ANOVA (p>0.0001). The best fit model with a chi squared value of 8.038 with 5 d.f revealed two significant regions for QT, region a1 and region a2. The substitution of Col alleles in region a1 by L*er* alleles has a decreasing effect on final height (time point, day 54) by 166.5 mm. In contrast the same substitution in region a3 has an increasing effect on the said phenotype by 33.23mm.

4.4.4 Correlation coefficients

Strong positive correlations were found between the traits; FD and RLF (0.8); FD and RW (0.7) while negative correlations were observed between FD and H45 (-0.7); FD and H54 (-0.8); H54 and RLF (-0.7). Apart from these strong correlations all characters except for a few,

displayed some degree of correlation either positive or negative. The correlation coefficients between all the traits and the P values are given in table 4.5.

4.5 Discussion

4.5.1 Location of QTL for traits scored

Significant QTLs were located along chromosome 3 for all the traits scored except for leaf number at day 20. Two QTLs were found for germination time in regions a1 and a2. However, it has been observed that factors such as age and quality of the seed as affected by the growth environments of the mother plants, also play a major role in addition to the genetic factors in determining the time taken for germination.

For the vegetative growth measurement leaf number at day 20, it was difficult to fit a model although regions a1, a2 and a5 showed signs of significant QTLs.

For rosette width two QTL were located in regions a1 and a2 at day 20 and only the QTL in region a1 was located at flowering. For height QTLs were located in regions a1, a2 and a3 at time points day 30 and 35. However, at the later stages in day 45 and day 54 only two QTLs were located in regions a1 and a2 at day 45 and a1 and a3 at day 54. It is possible that on the two later time points of height measurements one QTL was not observed under the conditions of this experiment in each time point. In contrast to the QTLs for height in the so far mentioned time points a QTL was located in region a6 in addition to the QTL in region a1 for the height at flowering time.

For flowering time a single QTL region was located in a1. Similarly for the three flowering time associated traits, rosette width, rosette leaf number and cauline leaf number at flowering a single QTL was located in region a1.

The maximum number of QTL located in this study to affect a single trait is three. Three QTLs were located to affect plant height in three time points the character was scored.

A total of 21 QTLs were detected along chromosome 3 of *Arabidopsis* in this study for eleven of the traits scored for. The most observable fact was that region a1 contained significant QTL for all of the eleven traits. Thus out of 21 QTL eleven were located in region a1. This indicates pleiotropy or tight linkage between genes controlling the above traits.

Four QTLs each were located in regions a2 and a3. No QTLs were found to affect any of the scored traits in region a4 while a QTL each were located in regions a5 and a6. The lower density of the QTLs in regions a4, a5 and a6 may simply be due to the fact that there are no additional QTLs located in these regions that affect the traits under study. Or else, if in fact there are QTLs in these regions, they may have very small effects which were not detected in the current experiment or they may be in repulsion negating the effect of genes with positive and negative effects on the trait.

With a Col/L*er* RIL mapping population, QTL controlling flowering time has previously been mapped to a similar region (Jansen 1996). Furthermore,(Koornneef et al. 1998b) reported QTL on top of chromosome 3 using a L*er* x Col RIL population. Koumproglou et al. (2002) using more or less the same STAIRS located QTL for flowering time in similar regions on top of chromosome 3 for flowering time, rosette leaf number and height at day 35 with the same positive and negative Col/L*er* substitution effects.

4.5.2 Correlations among characters

Most of the traits showed significant positive or negative correlation with other traits scored. This is because all the traits scored are growth related characters and they are therefore interrelated to certain degrees. The strongest correlations were observed for flowering time with three other traits measured. Flowering time thus showed strong positive correlations with rosette leaf number (0.808) at flowering and rosette width (0.742) at flowering. This is due to the obvious fact of increased vegetative growth with the delay in flowering. Furthermore,

flowering time was negatively correlated with height at day 45 (-0.788) and height at day 54 (-0.847). This can be explained in relation to the growth pattern of *Arabidopsis,* because the plant remains a rosette until the time of bolting which precedes flowering. With flowering, the plant develops the main stem and after flowering, the plants show rapid shoot elongation. Therefore when the time for flowering is higher, the two post-flowering height measurements remain lower. Strong negative correlation was observed between rosette leaf number and height at day 54 (-0.747) as well. Rosette leaf number is a trait related to the vegetative growth of the plant while height indicates post vegetative growth of plants. So at this particular time point, the two traits display negative relationship between them.

4.5.3 Differences within lines of same STAIRS

For some traits significant differences were observed between the lines within the same STAIR genotype. This was mostly evident for the traits RW20 Hd-35, HF and H-54 in the STAIRS containing the recombination within the region 21-44. The STAIRS in this study were demarcated by the availability of the molecular markers and thus the lines within a genotype are not necessarily true duplicates of the same genotype. The lengths of crossover regions are as high as 10 - 15 cM in most of the STAIRS and this region may include QTL affecting certain traits. This can lead to significant differences within lines of a particular STAIR genotype. In order to avoid this, more polymorphic genetic markers need to be identified thereby reducing the crossover region to a minimum. This will facilitate better model fitting that will result in identifying additional QTLs if there are any.

4.5.4 STAIRS in QTL location

Koumproglou et al (2002) reported the first demonstration of QTL mapping using STAIRS. This study is a further demonstration of the QTL location using STAIRS and further validated

the findings of them. Many of the STAIRS used in the current study are the same as STAIRS used by Koumproglou et al As they discussed STAIRS are a permanent resource and can be maintained true to type by selfing and the same lines can be used to verify the results of the same study as in this case or else can be used to map different QTLs which show polymorphism within Col and L*er*.

Koumproglou et al (2002) identified the top region in chromosome 3 of *Arabidopsis* to carry significant QTL affecting flowering time and related characters. Using more or less the same STAIRS in chromosome 3, this study further affirmed the findings of Koumproglou et al and confirmed that the top region of chromosome 3 carries significant QTL for several traits including the important trait, flowering time.

Out of the three steps of fine QTL mapping using STAIRS, this study has demonstrated the feasibility up to the wide STAIR stage. The region of QTL location is still 10-15 cM and in order to achieve fine mapping it is needed to proceed to the third stage of this procedure, which is the use of narrow STAIRS.

4.5.5 Conclusions

This study has amply demonstrated the power of STAIRS in locating QTL. In addition, the findings of Koumproglou et al (2002) to carry QTL for several traits including flowering time were further validated.

The major aim of this study was to re-ascertain the region of interest to proceed to the fine mapping of QTL with special relevance to QTL affecting flowering time.

It can finally be concluded that the evidence suggests flowering time QTL in region a1 which is 0-14 cM of chromosome 3 with the possibility of the extension of the region to 0–20 cM depending upon where the recombination is located in STAIR no.4. So the region especially

with the flowering time and related traits can be concluded as the top region of chromosome 3 extending from 0- 20 cM.

This study thus indicates that in order to fine map the QTLs related to flowering time it is necessary to proceed to the construction of narrow STAIRS within the top 20 cM of chromosome 3 of *Arabidopsis*.

Table 4.1 : Genotype scores of wide STAIRS for the nine microsatellite markers.

(11 = Col homozygous; 22 = L*er* homozygous)

Line	Markers									Line by L*er* introgression (cM)
	1	2	3	4	5	6	7	8	9	
1	11	11	11	11	11	11	11	22	22	80-100 A
2	11	11	11	11	11	11	11	22	22	80-100 B
3	11	11	11	11	11	11	22	22	22	60-100 A
4	11	11	11	11	11	22	22	22	22	44-100 A
5	11	11	11	11	11	22	22	22	22	44- 100 B
6	11	11	11	11	11	22	22	22	22	44- 100 C
7	11	11	11	11	11	22	22	22	22	44- 100 D
8	11	11	11	22	22	22	22	22	22	20-100 A
9	22	22	22	22	22	22	22	22	22	CSS3
10	22	22	22	22	22	22	22	22	22	CSS3
11	22	22	22	22	22	11	11	11	11	0-20 A
12	22	22	22	22	22	11	11	11	11	0-20 B
13	22	22	22	22	22	11	11	11	11	0-20 C
14	22	22	22	22	22	11	11	11	11	0-20 D
15	22	22	22	22	22	22	11	11	11	0-44 A
16	22	22	22	22	22	22	22	22	11	0-80 A

Table 4.2: The means and the standard deviations of the traits scored for each genotype (gen=genotype)

Trait	Means and standard deviations							
	gen. 1	gen. 2	gen.3	gen. 4	gen. 5	gen. 6	gen. 7	gen. 8
GT	5.19 ± 0.43	6.09 ± 0.79	5.85 ± 1.55	5.17 ± 0.38	7.73 ± 1.68	5.52 ± 0.74	5.11 ± 0.32	7.16 ± 1.74
LN- 20	7.01 ± 0.75	6.92 ± 1.08	7.19 ± 0.87	7.50 ± 0.51	6.68 ± 0.78	6.87 ± 0.60	6.96 ± 0.43	6.79 ± 1.32
RW -20	16.02 ± 2.80	15.67 ± 3.67	16.29 ± 3.76	20.62 ± 17.27	14.30 ± 3.22	15.52 ± 3.96	16.27 ± 1.93	15.37 ± 4.41
H -30	3.74 ± 0.92	3.74 ± 0.79	4.03 ± 1.52	5.23 ± 2.35	4.23 ± 0.89	6.89 ± 0.95	4.82 ± 1.14	5.05 ± 1.37
H – 35	9.91 ± 6.40	7.51 ± 2.24	11.31 ± 10.76	24.07 ± 24.09	8.62 ± 1.50	6.89 ± 1.42	6.54 ± 1.15	9.45 ± 4.35
FT	40.62 ± 2.82	41.43 ± 4.15	40.63 ± 3.15	39.17 ± 2.61	49.47 ± 5.59	53.06 ± 7.58	48.76 ± 5.70	49.76 ± 6.43
HF	38.55 ± 13.96	31.63 ± 11.85	45.40 ± 17.26	52.44 ± 15.12	26.76 ± 8.01	30.96 ± 14.49	35.70 ± 11.71	39.73 ± 13.57
RLF	13.69 ± 1.57	13.98 ± 1.34	13.87 ± 1.90	13.93 ± 1.95	19.23 ± 4.94	22.85 ± 6.94	17.91 ± 2.68	20.04 ± 5.96
CLF	2.85 ± 0.47	2.72 ± 0.63	2.86 ± 0.48	2.92 ± 0.51	3.23 ± 0.63	3.55 ± 0.97	3.29 ± 0.62	3.62 ± 0.84
RWF	67.22 ± 9.20	73.02 ± 9.60	70.74 ± 9.52	71.19 ± 8.20	106.20 ± 18.11	92.74 ± 21.58	101.67 ± 15.85	102.51 ± 21.48
H - 45	122.77 ± 53.66	92.45 ± 62.39	122.52 ± 61.98	166.43 ± 53.75	35.59 ± 20.86	13.64 ± 26.10	15.35 ± 23.28	36.54 ± 37.92
H - 54	296.22 ± 55.45	293.05 ± 81.70	300.28 ± 66.09	310.55 ± 44.91	136.46 ±100.56	100.57 ± 98.86	213.88 ± 79.37	139.86 ±106.12

Table 4.3 : The model used for detecting and estimating genetic differences among wide STAIRS. (-) denotes regions of introgressed L*er* donor alleles and (+) denotes regions of Col alleles.

STAIR*S*	*Substituted region (cM)*	*m*	*a1*	*a2*	*a3*	*a4*	*a5*	*a6*
1	80-100	1	1	1	1	1	-1	-1
2	60-100	1	1	1	1	-1	-1	-1
3	44-100	1	1	1	-1	-1	-1	-1
4	20-100	1	1	-1	-1	-1	-1	-1
5	CSS3	1	-1	-1	-1	-1	-1	-1
6	0-20	1	-1	-1	1	1	1	1
7	0-44	1	-1	-1	-1	1	1	1
8	0-80	1	-1	-1	-1	-1	-1	1

Tables 4.4: The significant parameters and their estimated values, standard errors and normal deviate (c) for each of the traits following least-squares model fitting. All 'c' values greater than 1.96 were taken as significant.

Table 4.4.1 Germination time

Parameter	Estimate	Standard Error	C
m	5.522	0.122	45.232
a1	-0.951	0.155	-6.120
a5	-1.040	0.166	-6.270

Table 4.4.2 Rosette Width day 20

Parameter	Estimate	Standard Error	C
m	15.737	0.247	63.825
a1	2.633	0.487	5.409
a2	-2.250	0.490	-4.594

Table 4.4.3 Height day 30

Parameter	Estimate	Standard Error	C
m	4.078	0.084	48.799
a1	0.346	0.171	2.025
a2	-0.551	0.170	-3.239
a3	-0.257	0.084	-3.064

Table 4.4.4 Height at day 35

Parameter	Estimate	Standard Error	C
m	8.912	0.394	22.631
a1	7.821	0.805	9.715
a2	-6.475	0.801	-8.083
a3	-0.860	0.395	-2.179

Table 4.4.5 Flowering time

Parameter	Estimate	Standard Error	C
m	45.919	0.453	101.366
a1	-5.383	0.453	-11.884

Table 4.4.6 Height at flowering time

Parameter	Estimate	Standard Error	C
m	38.244	1.223	31.105
a1	9.768	2.037	4.795
a3	-4.462	1.265	-3.527
a6	6.179	2.202	2.806

Table 4.4.7 Rosette leaf number at flowering

Parameter	Estimate	Standard Error	C
m	17.450	0.520	33.540
a1	-3.607	0.520	-6.932

Table 4.4.8 Cauline leaf number at flowering

Parameter	Estimate	Standard Error	C
m	3.155	0.050	62.829
a1	-0.303	0.050	-6.035

Table 4.4.9 Rosette width at flowering

Parameter	Estimate	Standard Error	C
m	84.072	1.680	50.035
a1	-13.928	1.680	-8.289

Table 4.4.10 Height at day 45

Parameter	Estimate	Standard Error	C
m	69.811	3.500	19.951
a1	72.600	6.896	10.529
a2	-24.022	6.930	-3.466

Table 4.4.11 Height at day 54

Parameter	Estimate	Standard Error	C
m	212.657	7.460	28.508
a1	83.250	7.518	11.074
a3	-16.614	7.541	-2.203

Table 4.5 Pearson correlation coefficients among traits. (Only significant correlations are given.

	LN20	d.g	RW-20	H-30	H-35	FT	HF	RLF	CLF	RWF	H45
d.g	-0.636										
	0.000										
RW20	0.542	-0.419									
	0.000	0.000									
H-30	0.277	-0.220	0.260								
	0.000	0.000	0.000								
H-35	0.315	-0.203	0.317	0.643							
	0.000	0.000	0.000	0.000							
FD	-0.516	0.378	-0.328	-0.192	-0.377						
	0.000	0.000	0.000	0.000	0.000						
HF	0.290	-0.254	0.230	0.243	0.391	-0.376					
	0.000	0.000	0.000	0.000	0.000	0.000					
RLF	-0.292	0.243	-0.186	-0.102	-0.285	0.808	-0.331				
	0.000	0.000	0.000	0.021	0.000	0.000	0.000				
CLF	-0.140	-	-0.098	-	-0.158	0.559	-	0.502			
	0.001		0.026		0.000	0.000		0.000			
RWF	-0.278	0.271	-0.167	-	-0.274	0.742	-0.346	0.686	0.552		
	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000		
H-45	0.510	-0.356	0.397	0.278	0.556	-0.788	0.507	-0.615	-0.371	-0.622	
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
H-54	0.463	-0.347	0.282	0.126	0.250	-0.847	0.400	-0.747	-0.463	-0.666	0.703
	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Cell Contents: Pearson correlation
P value

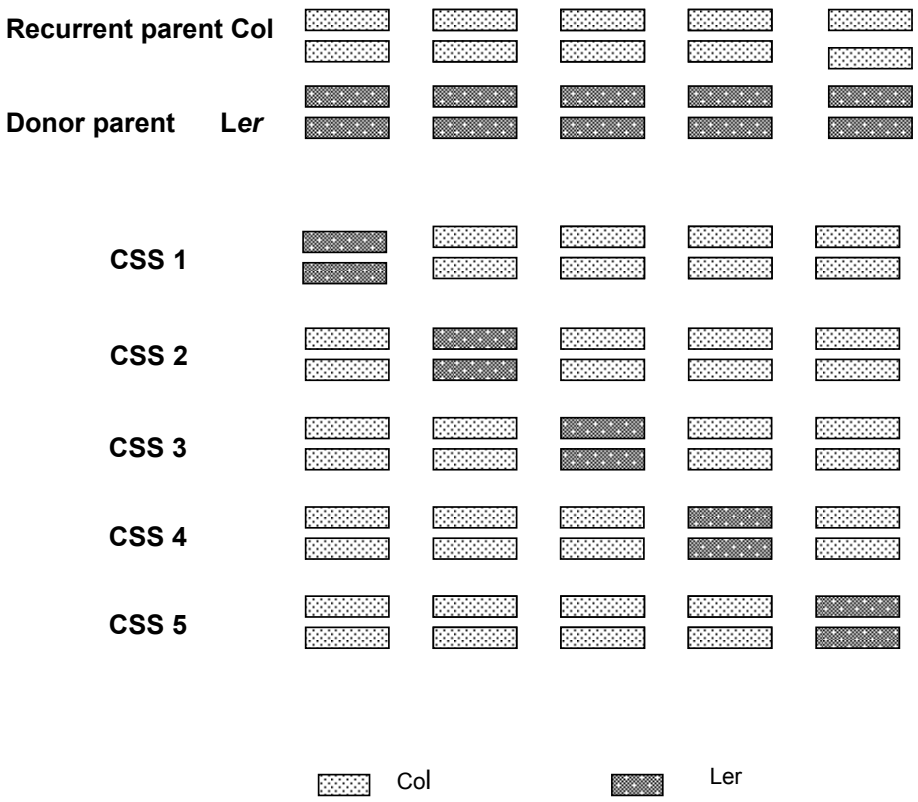


Figure 4.1: CSSs of *Arabidopsis*

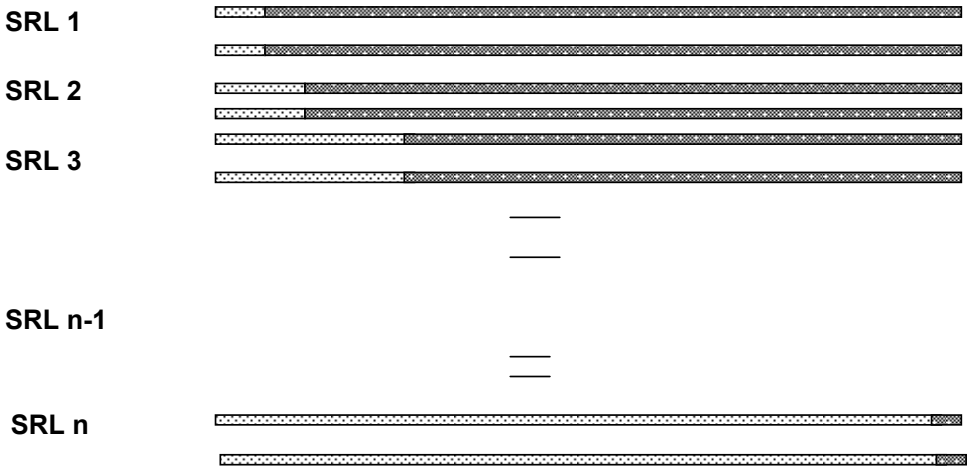


Figure 4.2: Structure of STAIRS

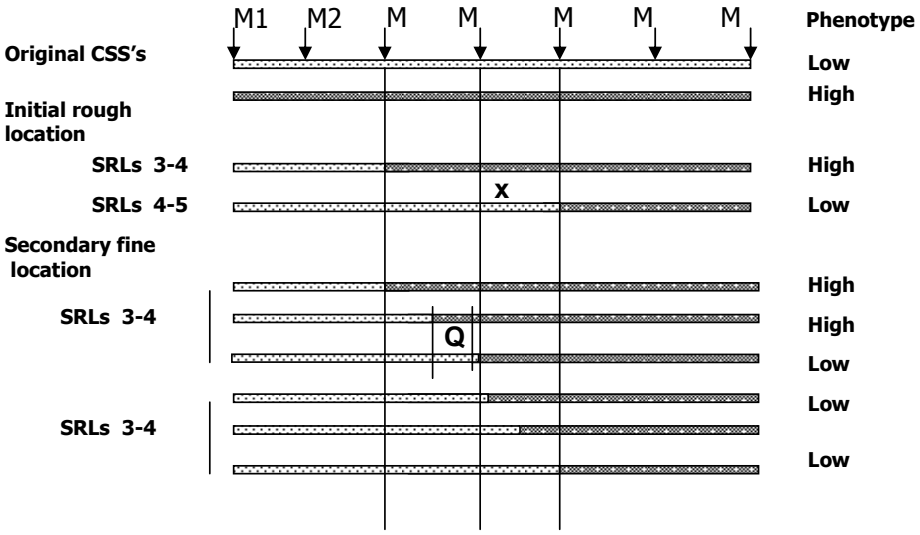


Figure 4.3 : Stages of Fine mapping Using CSSs and STAIRS

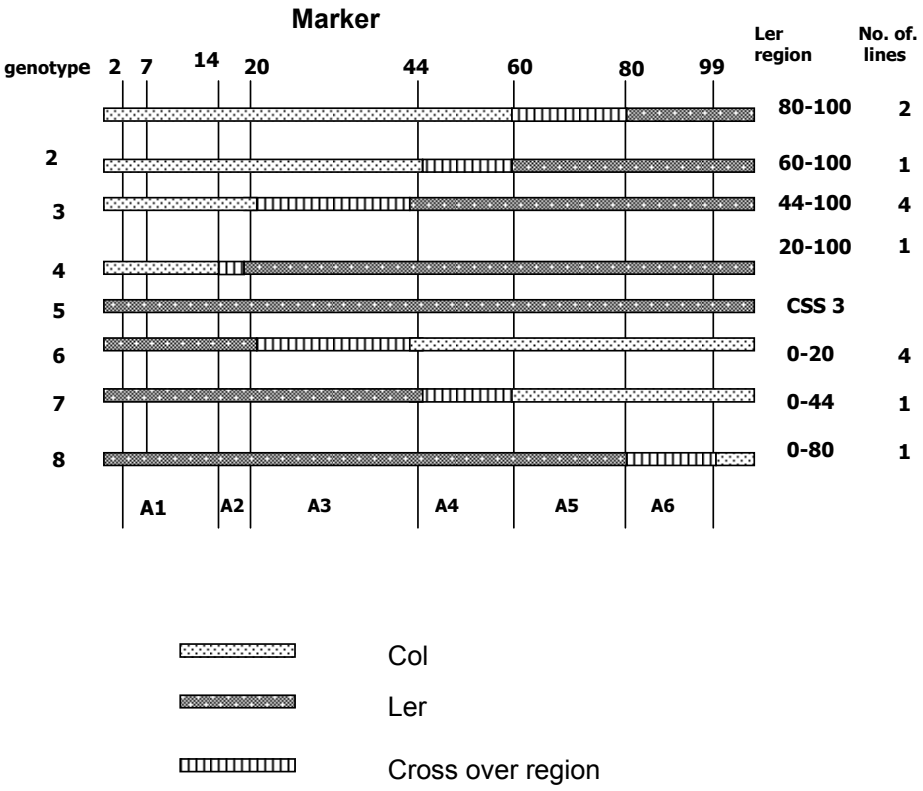
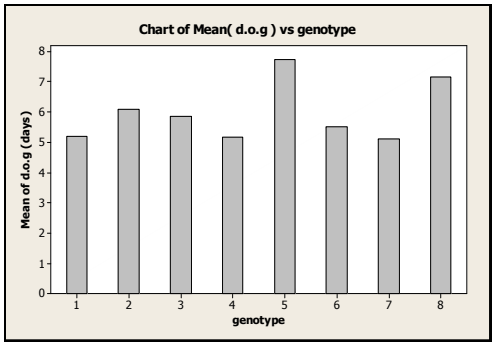
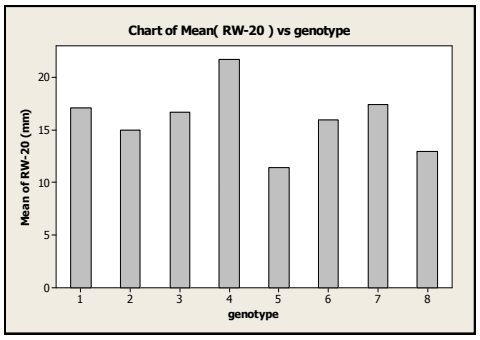
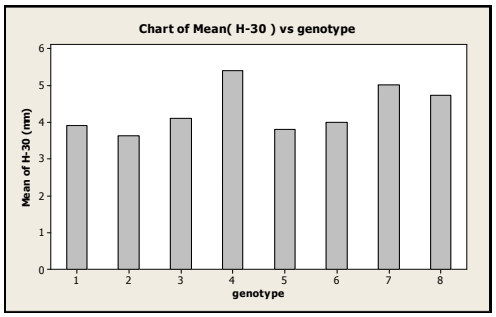


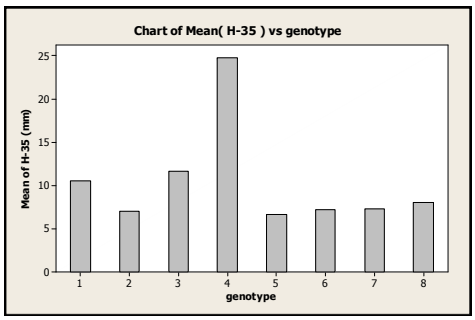
Figure 4.4 : Wide STAIRS: marker bins, genotypes,lines within genotypes and introgressed and crossover regions

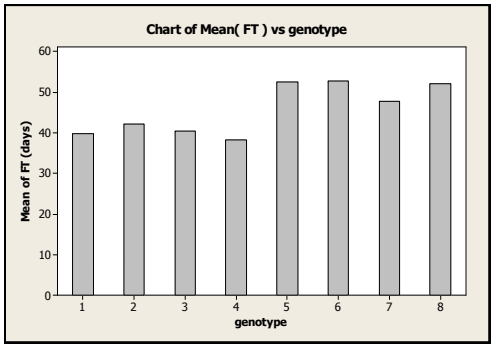
Figure 4.5 Graphical representation of means of traits for each genotype

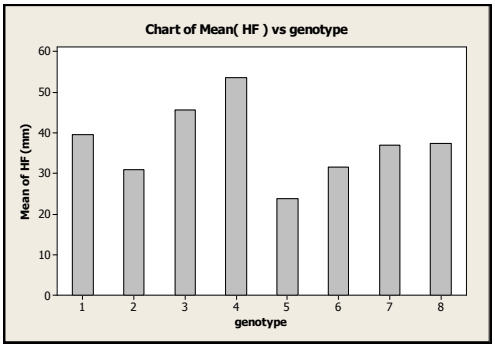












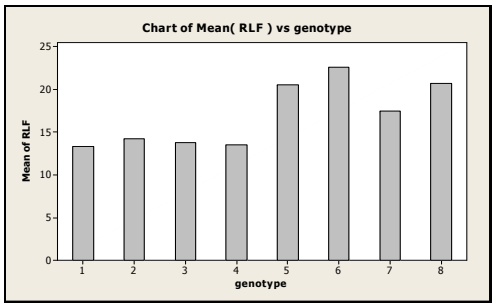


Chart of Mean(RLF) vs genotype
Mean of RLF
genotype
0
5
10
15
20
25
1
2
3
4
5
6
7
8

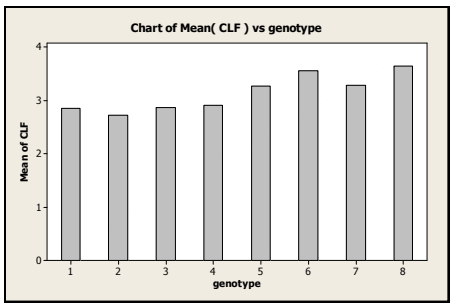


Chart of Mean(CLF) vs genotype
Mean of CLF
genotype
0
1
2
3
4
1
2
3
4
5
6
7
8

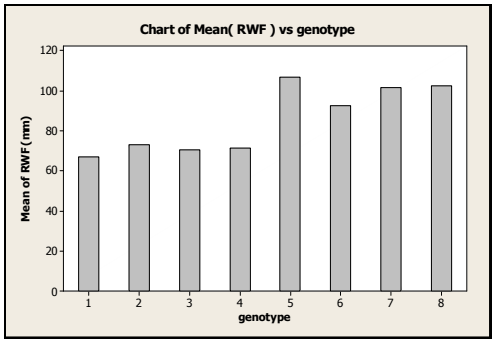


Chart of Mean(RWF) vs genotype
Mean of RWF (mm)
genotype
0
20
40
60
80
100
120
1
2
3
4
5
6
7
8

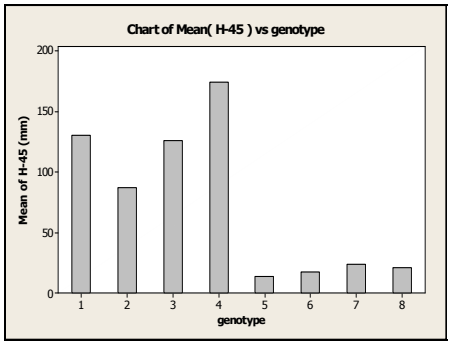


Chart of Mean(H-45) vs genotype
Mean of H-45 (mm)
genotype
0
50
100
150
200
1
2
3
4
5
6
7
8

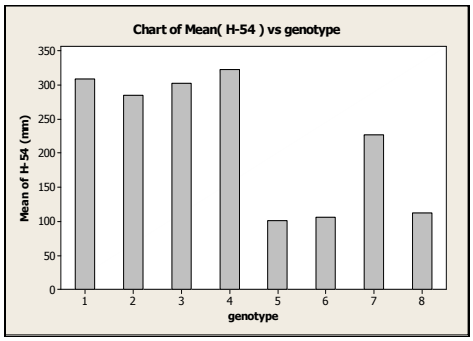


Chart of Mean(H-54) vs genotype
Mean of H-54 (mm)
genotype
0
50
100
150
200
250
300
350
1
2
3
4
5
6
7
8

CHAPTER 5

CONSTRUCTION OF NARROW STAIRS IN *ARABIDOPSIS* VIA A MARKER ASSISTED BREEDING PROGRAMME

Abstract

A breeding programme was initiated with a view of producing narrow STAIRS within the top twenty cM of chromosome 3 of *Arabidopsis thaliana*. The aim was to construct about twenty STAIRS each differing from the next by one cM regions of introgression along the targeted region of the chromosome 3. The starting material was the back crossed progeny of chromosome substitution strain three (CSS3) and repeated back crossing and subsequent selfing were done to produce the required single recombinant lines. At all stages of the breeding programme microsatellite markers and one CAPS marker were used to identify the desired individuals to be carried down to the subsequent stage of the breeding programme and the rest to be discarded. In addition to producing the STAIRS, the genotyping data were used to calculate the recombination frequencies and genetic distances between five marker loci along chromosome 3 of *Arabidopsis*. In the end twenty-three narrow STAIRS within the top twenty cM were constructed achieving the major aim of the current project.

5.1 Introduction

Construction of a mapping population is an important aspect of gene mapping. Factors such as the size of the mapping population for reliable gene mapping and the precision of the mapped genes are affected by the nature of the mapping population. The most widely used mapping populations at present in QTL mapping are segregating populations such as backcrosses, RILs and DHLs. As discussed in chapter one, mapping with these populations does not identify map locations of desired genes with sufficient precision for positional gene cloning.

As an alternative method, substitution lines can be constructed using appropriate parents that show polymorphism for the traits of interest and can be used in QTL mapping. This procedure involves two parents providing donor and recipient genotypes in the substituted line. The size of the region of substitution can vary from being a whole chromosome to a small region of a chromosome. The use of substitution lines helps locate QTL with greater precision provided that a library of overlapping substitution lines with sufficient coverage of the genome is available. The number of regions that can be examined this way can be very large depending on the size of the substitutions. Therefore, a more practical approach would be to focus attention on particular chromosomal sections revealed by previous marker based studies on segregating populations (Kearsey and Pooni 1996).

The breeding programme to produce these substitution lines involves crossing the two parents to obtain an F_1 and then repeated backcrossing to the recurrent parent in order to introgress the defined regions from the donor parent into the recipient parent. Once an individual with a desired genotype is found, the chromosome with the introgression should be made homozygous by selfing. These homozygotes can be maintained as true breeding, permanent stocks.

The breeding programme to produce substitution lines should be assisted by genetic markers for the identification of individuals with the desired genetic make up, in order to proceed with those

individuals and to discard the rest. Thus the construction of substitution lines is also limited by the availability of polymorphic molecular markers to provide coverage of reasonable density of the entire genome of the relevant organism. The use of co-dominant molecular markers such as microsatellites makes the identification of heterozygotes an easy task. But in instances, where dominant markers need to be used, it is important to have the dominant alleles in the non-recurrent parent in order to identify the relevant genotypes that are heterozygous for the donor introgressed region.

QTL mapping using NILs is another method used to obtain higher resolution of the mapped QTL. NILS are generated by inbreeding an F_1 for several generations and eventually identifying individuals that are entirely homozygous except for one or two marker loci, at an advanced generation. These are then selfed to produce two different isogenic lines, which contain either one of the alleles that were heterozygous. NILs are essentially substitution lines but involve only very small regions of substitution. The problem associated with mapping using NILs is the requirement of a large number of lines to cover the entire genome. The development of such a vast number of lines involves an enormous amount of work. Yet even with the use of substitution lines and NILs the CI of the location of a QTL cannot be reduced much below 2-5 cM (Kearsey 2002).

A more gene-targeted approach for fine mapping of QTL is the use of STAIRS in organisms for which such chromosome engineering activities are feasible. STAIRS are a genetic resource that facilitate very fine mapping of QTL down to 1 cM or 0.5 cM depending on the availability of markers.

Breeding for such designer chromosomes is practically limited to organisms that are amenable to both inbreeding and out-crossing and possess only a manageable number of haploid chromosomes. The model dicot *Arabidopsis* is the first plant to be used for the production of

STAIRS (Koumproglou et al. 2002). The availability of a large number of molecular markers makes *Arabidopsis* an appropriate plant for chromosome engineering. Furthermore, the availability of whole genome sequence information facilitates the development of new PCR based molecular markers as needed in the breeding programme.

One of the major advantages of QTL mapping using STAIRS is the ability to zoom in on a specific region of a chromosome known to house QTLs, as revealed by the use of wide STAIRS. This involves the construction of narrow STAIRS within the region of interest. Komproglou et al (2002) produced and analysed QTL in wide STAIRS of *Arabidopsis* to report the presence of several QTL including QTL for flowering time at the top of chromosome 3. Following the research of Koumproglou et al, the current research involves the development of narrow STAIRS within the top 20 cM of *Arabidopsis* using a marker assisted back cross breeding programme.

The ladder like progression of sequentially stacked STAIRS can occur in two orientations depending on the genotypes of proximal and distal ends of the chromosome. The two orientations are to have the donor genotype at the top end extending towards the distal end that should contain the recurrent genotype as a result of a single recombination along the chromosome or vice-versa. It is useful to have STAIRS in both orientations in QTL analysis with wide STAIRS to identify the initial rough location of QTL. In the later stage to verify the fine position of QTL within a particular wide STAIR a large number of narrow STAIRS need to be produced within the particular region of interest and it is sufficient and less complicated to have STAIRS in one orientation.

The production of STAIRS of a particular chromosome using the CSS of the relevant chromosome as starting material involves several backcrosses to the recurrent parent followed by selfing to get homozygosity. The first BC progeny of the CSS to the recurrent parent can be regarded as an F_1 for the particular chromosome of interest, because of the heterozygosity of that

particular chromosome in a pure genetic background of the recurrent parent. The second BC progeny consist of non-recombinants and single, double or triple recombinants. At this stage genotyping the progeny helps to identify those individuals with a single recombination to proceed with the breeding programme and the rest to be discarded. When the selected single recombinants from the second BC progeny are selfed, the resultant progeny of each selected plant is a segregating population. The genotyping of this progeny allows the identification of homozygous single recombinants with the donor introgression which are the STAIRS.

The breeding programme used by Koumproglou et al to obtain CSSs and wide STAIRS is illustrated in figure 5.1 and the current breeding programme to obtain narrow STAIRS is illustrated in figures 5.2 and 5.3.

To ensure a sufficient number of narrow STAIRS at the end of the programme, a large number of second BC individuals are needed. On average, about 50% of the individuals are single recombinants at the second BC stage. When the interest is to have single recombinants of a particular orientation within a certain part of the chromosome it is necessary to calculate the size of the second BC population assuming that the recombinations occur in equal frequency along different regions of the chromosome.

5.2 Objectives

1) To produce a total of about twenty lines of narrow STAIRS within the top 20 cM on chromosome 3 of *Arabidopsis thaliana* using a marker assisted breeding programme.
2) To calculate recombination frequencies and genetic distances between five marker loci along chromosome 3 of *Arabidopsis*.

5.3 Materials and Methods

5.3.1 Breeding programme

5.3.1.1 Starting material

The starting material of the breeding programme in the current project was the F_1 seeds of CSS3 (Chromosome Substitution Strain of chromosome 3 of *Arabidopsis*). F_1 seeds are the seeds obtained by backcrossing CSS3 to the recurrent parent Columbia via a cross pollinating programme.

5.3.1.2 Emasculation and cross pollination

Ten plants each of recurrent parent Col and the above mentioned F_1 seeds were grown under the same growth conditions and media described in chapter 04. When the plants have bolted and the inflorescence developed (about 3-4 weeks old) the buds that are about to be opened of the Col plants were emasculated. Emasculation was performed by removing the four anthers carefully without causing any mechanical damage to the stigma, observing under a stereoscope.

This was followed by applying pollen from the flowers of F_1 plants to the stigmas of the emasculated Col flowers. The pollination was repeated on three successive dates in each flower to ensure successful pollination. The emasculation and pollination procedure was continued when the new flower buds were about to open, until a sufficient number of successful pollinations were done to yield over 1000 back crossed seeds.

5.3.1.3 Number of BC progeny needed

The aim of the project was to produce about 20 STAIRS within the top 20 cM of chromosome 3. The orientation of the lines should be to have the donor introgression at the top end of the chromosome. Considering the chromosome to be 100 cM long, the top region we are interested in is 1/5 of the chromosome. Assuming that the recombination frequency is equal along the chromosome and 50 % single recombinants, then (50/100)*(20/100)*100 = 10% of a population

will be single recombinants at the top end. But because we are searching for the recombinants of a particular orientation, only a half of 10 %, i.e. 5% of the total population will have the desired genetic make up for this project. With the expected percentage of 5% progeny of the desired genetic make up to get 20 such individuals a minimum of 400 second BC progeny were needed.

5.3.1.4 The second back cross (BC) Progeny

The seeds of this second BC progeny were collected and 1320 of these seeds were sown again under the same conditions. When the plants were about 2 weeks old, leaf samples were collected from each plant separately for the extraction of DNA for genotyping. The plants were then selfed to obtain the selfed progeny, among which homozygote single recombinant lines - which are the STAIRS – may be present.

The genotyping results of the second BC progeny were used to identify individuals that satisfy three selection criteria, namely; single recombinants, the recombination to occur within the top 20 cM of chromosome 3 and having the donor genotype L*er* at the top end of the chromosome and Col at the distal end of the chromosome.

5.3.1.5 Search for homozygote SRLs (STAIRS)

At the second backcross stage, the selected individuals are heterozygous for the desired recombinants. In order to identify the desired homozygotes (L*er* at top end and Col at the distal end) 30 selfed seeds each of selected second back cross progeny were grown under the same growth conditions. The total number of plants raised was 750 (25 selected BC individuals x 30 seeds per selected individuals).

Again when the plants were 2 weeks old leaf samples were collected from each plant separately to be used for DNA extraction for genotyping and the plants were selfed to obtain seeds of STAIRS.

5.3.2 Genotyping the second BC progeny

5.3.2.1 Extraction of DNA and the molecular markers used

DNA was extracted from 1320 samples using the modified CTAB DNA extraction procedure given in appendix 1.

Sixteen mirosatellite markers along chromosome 3 were used for the genotyping of the BC progeny. They are, in order of position along chromosome 3, as follows:

F16B3 (2 cM), T17B22 (3 cM), T11I18 (4 cM), T6K12 (5 cM), T12H1 (6 cM), nga172 (7 cM), F24F17 (9 cM), F3E22 (10 cM), T7M13 (15 cM), F9F8 (15 cM), MDC11 (20 cM), nga162 (21 cM), AthGAPAb (44 cM), Th620B (60 cM), T16K5a (80 cM) and T17J13b (99 cM)

Primer sequences and optimum PCR conditions for these microsatellite markers are given in chapters 2 and 3.

5.3.2.2 Scheme used for Genotyping the BC progeny

Two easily scorable microsatellite markers (sub terminal marker nga172 & terminal marker T17J13b) were selected for the initial genotyping of all the 1320 individuals. The scores of genotypes of these two markers were used to select the single recombinants, which were heterozygous for one marker and Col homozygous for the other marker. Although the selection was for single recombinants the individuals with 3 or 5 recombinants (if there were any) would also be included in the selected category because the exclusion of such individuals could not be done at this stage.

In the subsequent stage those selected individuals that were the single recombinants along with heterozygous non recombinants were genotyped with 4 more microsatellite markers spaced evenly along the middle portion of the chromosome: [nga162 (20.56 cM), AthGAPAb (43.77 cM), Th620B (59.1 cM), T16K5a (80 cM)]. The results were used to identify single recombinants

having the recombination within the top 20 cM of chromosome 3 and also having L*er* donor genotype at the top end and recurrent Col genotype at the distal end of the chromosome.

The third stage of genotyping involved scoring for closely positioned ten markers within the top 20 cM for the accurate identification of the point of recombination in the selected progeny from the 'stage-two' genotyping. The following fine mapping markers were used: [T17B22 (3 cM), T11I18 (4 cM), T6K12 (5 cM), T12H1 (6 cM), nga172 (7 cM), F24F17 (9 cM), F3E22 (10 cM), T7M13 (15 cM), F9F8 (15 cM), MDC11 (20 cM)].

5.3.2.3 PCR Amplification and Gel Electrophoresis

PCR protocol

10 µl PCR reactions were used for the genotyping. Each reaction consisted of,

2 µl DNA (concentration 20 ng/ul)

1 µl 10 x PCR buffer

1 ul 25mM $MgCl_2$

1 ul forward primer (10 pM/ul)

1 ul reverse primer (10 pM/ul)

0.4 ul 5 mM dNTPs

0.2 ul Taq polymerase (5u/ul)

3.2 ul sterile distilled water.

The reactions were done in TECHNE genius thermocycler as,

1 x (94^{o}C, 3 min; yy ^{o}C, 30 s"; 72 ^{o}C, 1 min)

30 x (94^{o}C, 30 s"; yy ^{o}C, 30 s"; 72 ^{o}C, 1 min)

1 x (94^{o}C, 30 s"; yy ^{o}C, 30 s"; 72 ^{o}C, 5 min)

yy ^{o}C = annealing temperature for each primer.

Gel Electrophoresis

Polyacrylamide gel electrophoresis was practised to separate the PCR DNA products. The Polyacrylamide gel consisted of 6% (w/v) Acrylamide/Bis (19:1), & 6 M urea, 1 x TBE, TEMED and Ammonium persulphate. *BIORAD* Sequi-gen Sequencing Cell apparatus was used for running the gel.

Gel Staining

Silver staining was practised to visualize DNA bands on the sequencing gel. The following staining protocol was practised.

1) Fixing the bands in the fixing solution (0.005% glacial acetic acid) for 15 minutes with gentle shaking.
2) Washing the gel thoroughly with distilled water for 2-3 minutes with shaking.
3) Staining solution (0.0018 w/v Silver nitrate) for 20 minutes with gentle shaking.
4) A quick wash to remove the excess Silver from the gel.
5) Developing solution (0.007 w/v Sodium hydroxide and 0.005 Formaldehyde.) for a few minutes until the bands appear.
6) Stopping solution (0.005% glacial acetic acid) if needed only.

5.3.3 Genotyping to identify homozygous SRLs (STAIRS)

At the first stage, DNA was extracted (procedure in appendix 01) from eight individuals out of the thirty plants grown per each selected progeny. If the desired homozygotes were not identified within the eight plants, more genotyping was performed on the remaining twenty-two plants until the relevant SRL was obtained. Genotyping was practised in four different stages, proceeding with the selected individuals from each stage to the subsequent stage.

Scheme of genotyping

In stage one of genotyping three microsatellite markers located within the top 20 cM [F16B3 (2 cM), T12H1 (6 cM), MDC11 (20 cM)] were used to score the genotypes of eight individuals of each of the selected progeny. The individuals that were L*er* (donor) homozygous for the first marker (F16B3) and either L*er* or Col homozygous for the other two markers were chosen for further genotyping.

In the second stage, genotyping was practised with four evenly spaced markers [AthGAPAb, Th620B , T16K5a and T17J13b] along the rest of the chromosome 3 to check the selected progeny for Col homozygosity for all the four markers.

The third stage of genotyping involved the fine mapping markers along the top twenty cM of chromosome 3. The twenty microsatellite markers used and their positions are as follows:

[T17B22 (3 cM), T12J13b (4 cM), F20H23 (4 cM), T11I18 (4 cM), T6K12 (5 cM), T12H1 (6 cM), F22F7 (7 cM), F20O10 (8 cM), F24F17 (9 cM), F5E6 a (9 cM), F5E6 c (9 cM), F3E22 (10 cM), T1B9 a (10 cM), MLP3 (11 cM), F17O14 (13 cM), T22K18 (14 cM), T7M13 (15 cM), F9F8 (15 cM), MDC11 (20cM), MAG2 (20 cM)].

The results of the third stage of genotyping were used to identify the region of recombination as closely as possible and define the region of introgression in STAIRS.

The fourth and final stage of genotyping was aimed at determining the purity of the STAIRS with respect to the other four chromosomes in *Arabidopsis*. The markers used and their positions are as follows. All the markers listed below are microsatellite markers except DHS1, which is a CAPS marker.

Chromosome 01

nga392 (2.9 cM), nga59 (41.6 cM), T27K12Sp6 (61.2 cM), nga280 (83.8 cM), nga111 (115.55 cM)

Chromosome 02

nga1145 (9.6 cM), MSF3A (35.04), nga361 (63.02 cM), nga168 (73.7 cM), T9J23 (92 cM),

Chromosome 04

T18A10 (1cM), nga8 (26.56 cM), FCA9 (54.8 cM), F25O24 (72.35 cM), nga1139 (83.4 cM), DHS1 (108.5 cM)

Chromosome 05

MED24 (7.4 cM), nga249 (23.7cM), nga139 (50.4 cM), nga76 (68.4 cM), MJB21A (89.5 cM), MM19 (116.9 cM)

The selected lines were checked for Col homozygosity for all the above listed markers.

PCR amplification and Gel electrophoresis

PCR protocols and the non-sequencing polyacrylamide gel electrophoresis with Ethidium bromide staining is the same as described in chapter 04.

5.3.4 Calculation of recombination frequencies

Recombination frequencies between the markers were calculated with the genotyping data from the second BC progeny. The results of the two terminal markers were used to calculate recombinants and non-recombinants. The subsequent genotyping with four internal markers facilitated the calculation of recombination frequencies between the marker pairs.

The parents in the breeding programme at this stage were pure breeding Col and heterozygous F_1 with regard to chromosome 3. Theoretically half of the population are expected to be non-recombinants and the other half to be recombinants, with part of the chromosome being homozygous Col and the rest to be heterozygous. There would be double and even triple recombinants among the recombinants. The resultant non recombined progeny would be Col homozygotes and the heterozygotes. The single recombinants can either contain donor genotype at the proximal or distal end of the chromosome. This is illustrated in figure 5.2. The double

recombinants also would be of two orientations namely; those that arise from gametes that contain L*er* at both ends of the chromosome with a middle portion of Col and vice versa due to two recombination events occurring along the chromosome.

So when considering the population to consist of non-recombinants, single recombinants and double recombinants there were six types of progeny that could have been observed namely;

1. Non-recombinants that were Col homozygotes
2. Non-recombinants that were heterozygotes
3. Single recombinants, which are heterozygous at the top end of the chromosome
4. Single recombinants, which are heterozygous at the distal end of the chromosome
5. Double recombinants, which are heterozygous at both ends and homozygous Col in the middle of the chromosome
6. Double recombinants which are homozygous Col at both ends and heterozygous in the middle of the chromosome.

The genotypic scores for the two end markers were available for all the six categories listed above. However due to selective genotyping at the later stage groups 1 and 6 were identified as one group because the data for internal markers were not available for these two groups.

Theoretically the ratio between groups 2 and 5 should be the same as the ratio between the groups 1 and 6. Therefore the known ratio of individuals in groups 2 and 5 was applied to identify the numbers of individuals in groups 1 and 6 separately.

Afterwards, the percentages of recombinants, single recombinants and the double recombinants were calculated and the recombination frequencies between markers nga 162, nga172, AthGAPAB, Th620B, T16K5a and T17J13b were calculated. The recombination frequencies were converted to genetic distances using Haldane's mapping function, {-ln (1-2R) }50 where R is the proportion of recombinants.

5.4 Results

The cross-pollination programme of the F1 of CSS3 to the recurrent parent Col yielded over 1500 seeds of second BC progeny.

5.4.1 Genotyping of the second BC progeny

A total of 1320 individuals were grown and genotyped in the second BC progeny. However, there were some missing values for some of the progeny for certain markers due to failures in DNA extractions, PCR amplification or gel electrophoresis. Due to the large scale of the genotyping it was difficult to repeat the failed reactions and thus the selection of individuals were based on those progeny for which the complete data were available.

5.4.1.1 Genotyping with two terminal markers

The genotyping results with the two terminal markers revealed the proportions of non-recombinants and single recombinants in the population. The non-recombinants were identified by the scores of Col homozygosity or heterozygosity for both the terminal markers scored. The single recombinants in the progeny were revealed by the genotype scores of Col homozygosity for one of the markers and the heterozygosity for the other terminal marker. Out of the 1320 individuals, grown and genotyped, scores for both of the markers were available on 1149 individuals while either one or both the scores were missing for 171 individuals.

A total of 619 individuals of the progeny were non-recombinants while 503 were identified as single recombinants. A total of 348 of the non-recombinants were Col homozygotes while 271 were heterozygotes. Out of the single recombinants 245 were heterozygotes at the top end of the chromosome while 232 were heterozygotes at the distal end of the chromosome. There were a total of 53 double recombinants in the population.

The selection at this stage was for the single recombinants. Although the required single recombinants for the construction of STAIRS were progeny which are heterozygotes at the top end of the chromosome, all the single recombinants were chosen to be genotyped in the next stage with the four internal markers.

5.4.1.2 Genotyping with the internal markers

A total of 659 individuals were genotyped with the four selected internal microsatellite markers. These data, combined with the scores for the two terminal markers, were used to calculate the recombination frequencies and cM distances between markers.

At this stage, the progeny which had a single recombination within the top 20 cM of chromosome 3 were selected to be genotyped with 10 microsatellite markers and there were 26 individuals that satisfied the selection criteria. At the third stage of genotyping, the second BC progeny, the region of recombination of the selected 26 individuals at the above stage were located.

5.4.2 Genotyping to identify homozygous single recombinants (STAIRS)

In the scheme of genotyping for identifying the STAIRS, the first step was to look for the desired homozygotes within the first 20 cM based on the scores of three markers. Out of the eight progeny genotyped from each of the twenty-six selfed individuals, the expected homozygous individuals were found for twenty two lines. Out of the four lines remaining, one homozygote was identified by genotyping an additional five progeny. However, for the three further remaining lines the expected homozygotes could not be isolated even after genotyping the whole thirty progeny which were grown. Because there were other lines identified within the same marker bins, no further attempts were made to recover these three lines.

The region of recombination of the selected 23 lines, were identified by genotyping with 21 closely positioned markers within the top 20 cM. These data were used to define the region of introgression as accurately as possible. The 23 lines could be grouped into nine marker-

demarcated bins. There were 2 lines in bin 1, 2 lines in bin 2, 1 line in bin 3, 8 lines in bin 4, 2 lines in bin 5, 1 line in bin 6, 1 line in bin 7, 3 lines in bin 8 and 2 lines in bin 9. For some of the lines more than the one desired homozygote has been identified within the first eight progeny and those lines have been kept as true replicates of the particular lines. This structure of the 22 STAIRS is illustrated in figure 5.4.

The other two stages of genotyping at this generation involved checking for pure recurrent genotype of Col along the rest of the chromosome 3 and in the remaining four chromosomes in *Arabidopsis*. All the twenty-two selected lines were observed to be homozygous Col for the tested markers, which was the expected result.

5.4.3 Calculation of recombination frequencies and genetic distances between markers.

Out of the total population of 1320, clear genotypic data for the two terminal markers were available for1149 individuals. Out of the six categories of the progeny respective numbers were available for four categories and the total number for the other two categories were available. The six categories are shown diagrammatically in figure 5.2.

The numbers observed for categories two to five (heterozygous non-recombinants, single recombinants 1 and 2 and double recombinant 1) in order were 271, 232, 245, and 20. The total for categories one (homozygous single recombinants) and six (double recombinants 2) were 368. The ratio of categories 2 to 5 (271: 20) was applied to proportionately divide the categories 1 to 6. This resulted in 348 individuals in category one and 33 individuals in category 6.

The total number of non recombinants and recombinants observed were 619 (53.9%) and 530 (46.1%) respectively as against the expected ratio of 1:1. The Chi-squared test showed that the single recombinants are significantly slightly less than the expected. ($\chi 2_{(1\ d.f)} = 6.894$) (P=0.01-0.001)

Out of the non recombinants there were 348 (56.22%) Col homozygotes and 271 (43.78%) heterozygotes. Here again the expected ratio was 1.1 and the observed heterozygous single recombinants were significantly less as revealed by Chi-squared test. ($\chi 2_{(1\ d.f)} = 9.578$) (P=0.01-0.001)

The total number of single recombinants observed were 477, out of which 245 (51.36) were heterozygous at the top end while 232 (48.64%) were heterozygous at the distal end of the chromosome. There was no significant difference between the observed and expected as revealed by the Chi-squared value. ($\chi 2_{(1\ d.f)} = 0.354$)

The observed recombination frequencies, calculated genetic distances between the markers are listed below in table 5.1

5.5 Discussion

5.5.1 Narrow STAIRS produced

A total of 23 lines have been produced achieving the goal of about 20 lines. However, when these lines were grouped into marker demarcated bins they grouped into nine bins out of a possible number of 20 bins demarcated by the available markers, leaving some empty bins. Because each of the 23 lines came from a different back cross individual, the exact point of recombination varies even among the lines within a bin. Therefore, lines within a bin are not genetically identical and, had intermediate markers been available, they could have been assigned to separate marker bins. But, for some lines, true replicates - which are the progeny of the same BC individual - were found, when eight progeny out of a single selected BC individual were genotyped to identify the homozygote SRLs.

Out of the possible number of 20 bins most bins carried 0, 1 or 2 lines, 1 bin carried 3 lines while 8 different lines were located in bin number 4. The reason for the higher number of recombinants

in bin number 4 indicates that the relevant region is a hotspot for recombination. On the other hand no lines could be identified in the region of 15-20 cM, which is a long region considering the total region of 20 cM. Out of a total of 1320 second BC individuals 171 were discarded after the initial genotyping due to missing scores leaving 1149 individuals. The expected percentage of desired single recombinants at the top region was 5% of the population. So the expected number of SRLs were 57 but the observed number was only 26 (23 lines plus the 3 lines discarded at a later stage). This indicates that the recombination frequencies are not equal over the whole length of the chromosome. However, the elimination of individuals at later genotyping stages also may have caused a reduction in the expected number because once no clear scores were observed these were discarded without dwelling further on those individuals. This was necessary to save time.

5.5.2 Expected vs. observed recombination frequencies

In general 50 % of gametes in meioses and thus the resultant population are expected to be recombinants while the rest to be non-recombinants. The results show that the expected percentage of single recombinants is less than the expected. Furthermore, out of the non-recombinants, a higher proportion of Col homozygotes were observed than the heterozygotes. Because the recurrent background is Col this suggests that Col gametes preferred the assortment with Col gametes rather than the L*er* gamete although literature does not provide information to support this hypothesis.

The calculated genetic distances between marker loci are more or less the same except in two occasions. The difficulty of scoring the marker Th620B may have added a slight bias and also the initial lower than expected recombinants ratios may account for the slight changes from the expected genetic distances.

5.5.3 Single recombinants outside the region of interest

Out of a total of 476 single recombinants identified among the second BC progeny, only 26 lines which were homozygotes at the top end were made use of in the current project to produce STAIRS. The remaining 450 single recombinants theoretically can be used to produce a large number of STAIRS covering the entire length of chromosome 3 with both possible orientations of the donor introgressed region. The selfed seeds of all these lines have been stored and the regions of recombination have been identified to assign them to wide STAIRS. Now it is a matter of identifying homozygotes from relevant individuals to produce STAIRS all along chromosome 3. It will still be a massive task if all the possible narrow STAIRS are to be produced but if the interest is to produce wide STAIRS this can be achieved quite comfortably. As a subsequent step more markers can be developed within a region of interest and can refer back to the stored seeds to look for narrow STAIRS within the defined region. Therefore, the stored seeds of the identified single recombinants will be of immense importance both in producing wide and narrow STAIRS in future projects involving fine mapping of QTL in chromosome 3 of *Arabidopsis*.

Table 5.1. The markers used for the calculation of recombination frequencies, number of individuals for which there were data, number of recombinants observed, calculated genetic distances in cM (by Haldane's mapping function) and the previously known genetic distances among the markers.

Between markers	No. of individuals	No. of recombinants	RF (%)	Obs. Genetic distance cM	Published genetic distance cM
nga172-nga162	1025	118	11.51	13.06	13.65
nga162-AthGAPAB	1021	159	15.57	18.66	23.21
AthGAPAB-Th620B	1035	124	11.98	13.69	15.33
Th620B-T16K5a	1033	152	14.70	17.40	20.90
T16K5a- T17J13b	1038	164	15.80	18.99	19.00

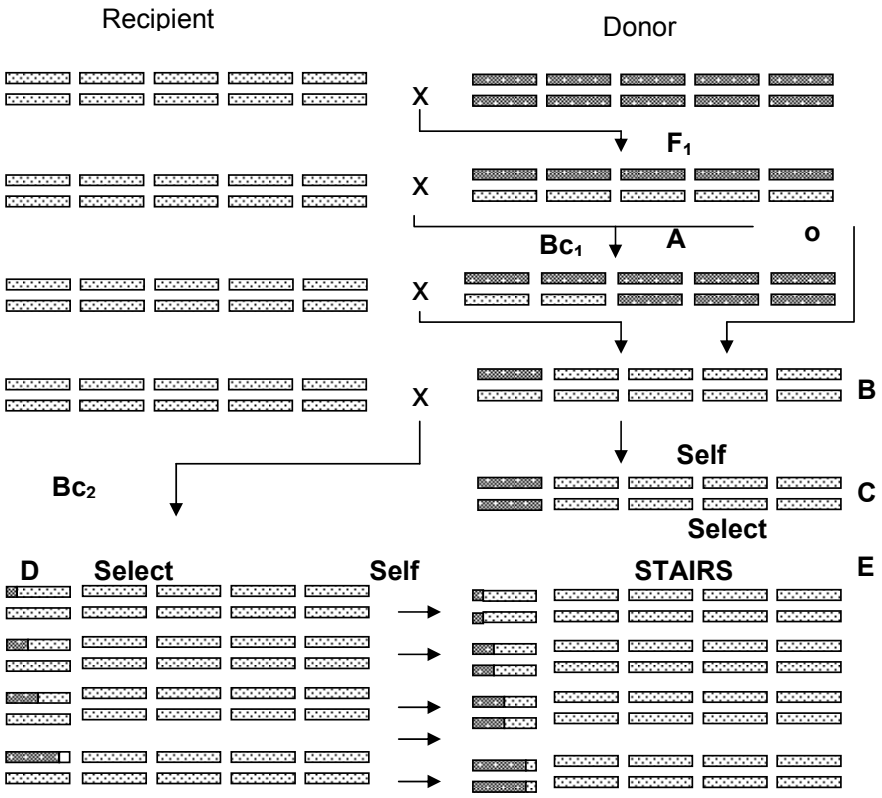


Figure 5.1 Breeding programme used by Koumproglou *et al*. to produce CSSs and STAIRS.

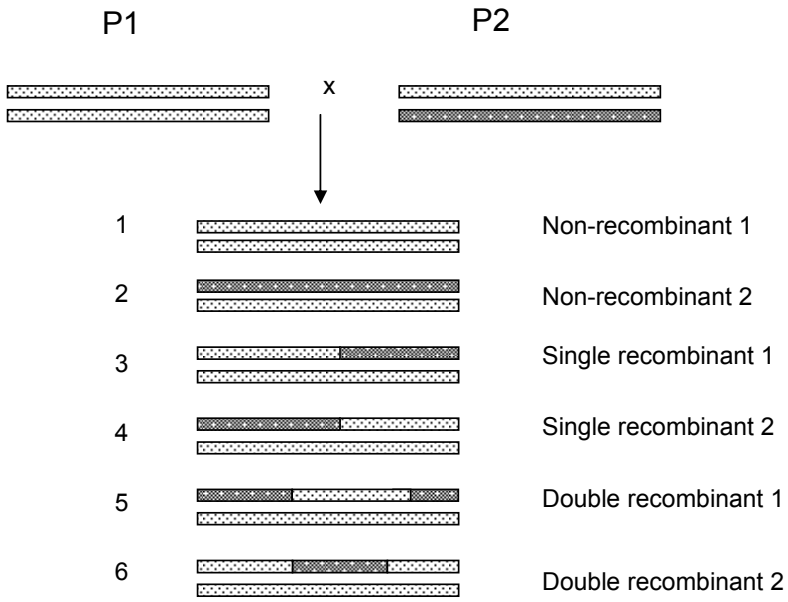


Figure 5.2 Genotypes of the Parents and the Progeny in the Backcross in producing STAIRS

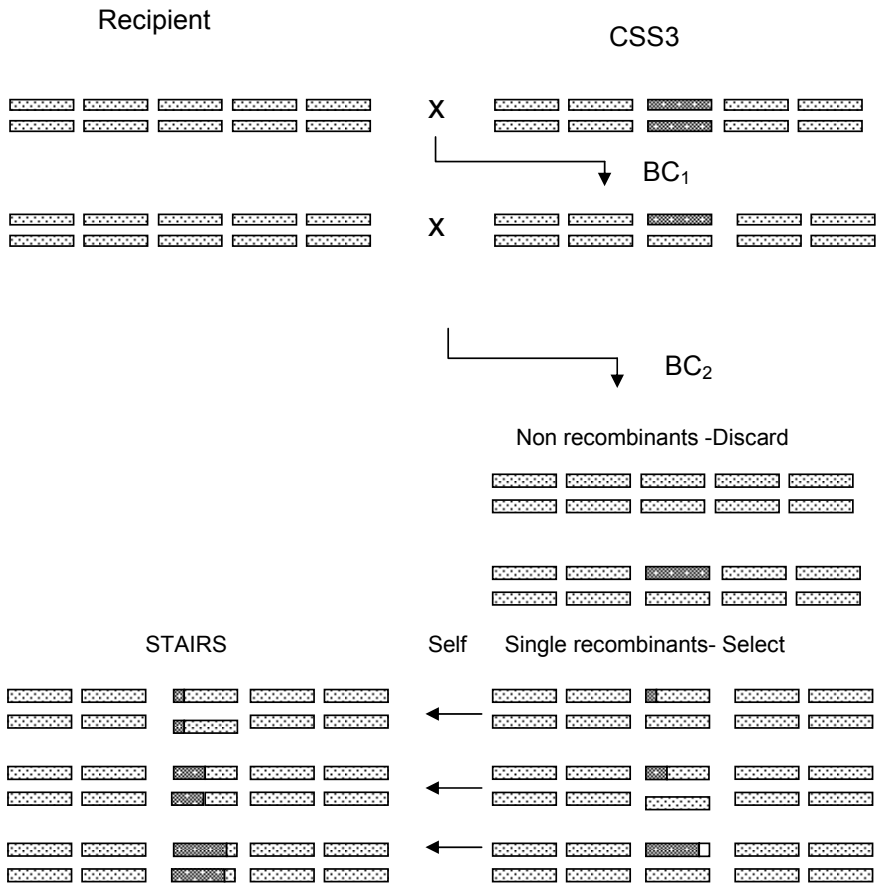


Figure 5.3: The current breeding programme adopted for the production of narrow STAIRS in chromosome 3.

Deleted: 2

Deleted: ¶

Deleted: for

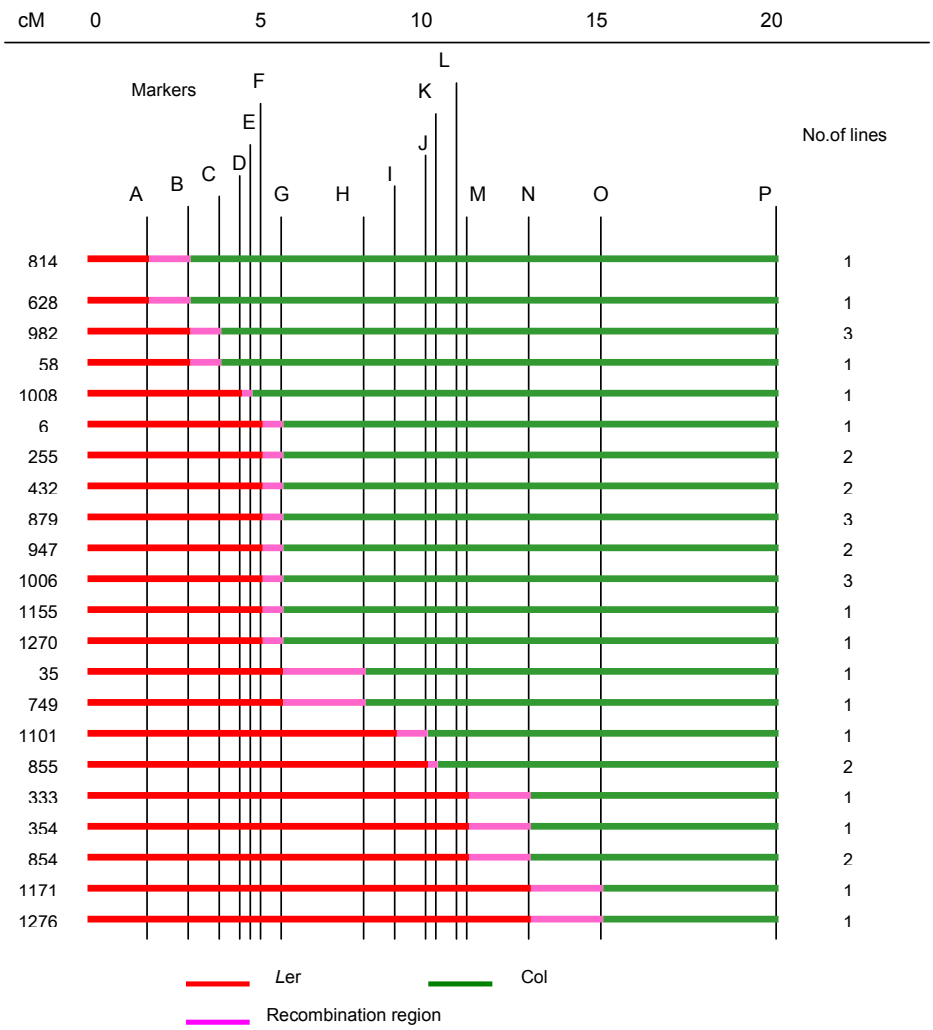


Figure 5.4: Structure of STAIRS produced on top of chromosome 3

CHAPTER 06

FINE MAPPING OF QTL USING NARROW STAIRS ON TOP OF CHROMOSOME 3 IN ARABIDOPSIS

Abstract

A plant house experiment was conducted using narrow STAIRS with the objective of fine mapping QTL with special relevance to flowering time on top of chromosome 3. In order to achieve this, thirty plants each of twenty four narrow STAIRS which were grouped into thirteen marker bins spanning the top 20 cM of chromosome 3, were grown in a controlled environment room in a completely randomized experimental design. Flowering time, along with flowering time related traits such as cauline and rosette leaf numbers at flowering time, rosette width and height at different time points, were measured and QTL analysis was performed on the scored data by ANOVA and least squares model fitting.

A major QTL for flowering time was located within 2-3 cM on chromosome 3, while QTL with less effect were located within 15-20 cM. QTL for flowering related traits cauline and rosette leaf number were also located in the same region i.e. 2-3 cM and was assumed to be due to pleiotropic effects of the same gene that controls flowering time rather than tight linkage between different genes. In a search for candidate genes for flowering time within this region using the TAIR database, the gene AT3G02380 which is listed as a locus homologous to CONSTANS proved to be a very strong candidate. This experiment amply illustrated the power of STAIRS for fine mapping of QTL by narrowing down the region of QTL to 1 cM and identifying candidate genes.

6.1 Introduction

6.1.1 Importance of fine mapping

Fine mapping of QTL is of utmost importance to answer many of the questions that arise in theoretical and applied quantitative genetics. Fine mapping of QTL governing a particular trait permits the better estimation of the number of genes in a particular region of the genome affecting a character. It will also facilitate the determination of genes that have major effects on the trait as opposed to those that show minor effects. Such information will be valuable to plant breeders for making decisions on selection strategies and for the scientists in medical profession for finding drugs for genetic diseases (Kearsey 2002).

Fine mapping of QTL also sheds light on the nature of dominance and epistatic properties of the relevant genes and their interaction with the environment. Such information can be used in exploiting heterosis (hybrid vigour) , in understanding the basis for inbreeding depression in producing new varieties in plant breeding (Coors and Pandey 1999) and also in pharmacogenomics.

Furthermore, the fine location of genes is essential in carrying out advanced searches on genes to understand the nature of the genes. The genes can be structural or regulatory; allelic variants of well-known genes or open reading frames for which no function has yet being assigned. Fine mapping helps in answer these questions and to unambiguously identify individual polygenes so as to carry out sequence and transcriptional level research (Kearsey 2002).

6.1.2 Examples for different approaches of fine mapping

Ronin et al. (2003) proposed a method called ‘Selective Recombinant Genotyping’ (SRG) for high resolution mapping of a QTL that has previously been mapped to a known confidence interval. SRG is a three-staged procedure that depends on the availability of a large mapping population. In stage one, the population is phenotyped and a proportion, *P*, of the high and low

tails is selected. In stage two, the selected individuals are genotyped for a pair of markers flanking the target C.I., and a group of individuals carrying recombinant chromosomes in the target interval are identified. Stage three involves the genotyping of recombinant individuals for a set of markers spanning the target C.I. Based on their simulation studies Ronin et al (2003) concluded that, standard error of the QTL (SEQTL) location decreased when QTL effect or population size increased (a fact that applies to most of the fine mapping approaches); an increase in the selected proportion in the tails beyond 0.25 had only a negligible effect on the SEQTL; increased marker spacing in the target interval had a very powerful effect on SEQTL (reducing down to 0.29 cM) at a given population size and QTL effect.

Recurrent selection backcross (RSB) is a method proposed by Wright (1952) for genetic analysis of quantitative traits. RSB is a breeding program where the backcrossing and selecting individuals at each stage, is repeatedly performed up to several generations. The selection in RSB is based entirely on the trait phenotype. Hill (1998) has formulated mathematical theory for isolating QTL with large effects using RSB. Later on Luo, Wu, and Kearsey (2002) developed an exact theoretical prediction of mean and variance of heterozygosity at a marker locus linked to one or two QTL with any degree of recombination for any number of generations of the RSB scheme. This theory provides a theoretical basis for RSB based QTL mapping. Luo et al concluded that the mapping strategy they developed facilitated fine mapping up to 1 or 2 cM.

In the current study we focus our attention on fine mapping of QTL via a chromosome engineering method described by Koumproglou et al. (2002). As discussed in previous chapters, STAIRS is a powerful resource for fine mapping of QTL and the mapping procedure has already been discussed in three stages in previous chapters of this thesis.

6.1.3 Zooming in on the top region of chromosome 3 for mapping of flowering related QTL

The top region of chromosome 3 has been reported to carry QTL related to flowering time and several flowering time related QTL by Koumproglou et al (2002) and further verified in a previous chapter in this thesis. Koumproglou et al identified QTL for flowering time, height and rosette leaf number on the top region of chromosome 3. In the current research QTL have been identified for all of the above characters as well as cauline leaf number. It is expected to achieve much higher resolution for the locations of the above mapped QTL in the current study using narrow STAIRS that differ by about 1 cM on average.

6.2 Objectives

1. To carry out a trial in a controlled-environment room for scoring growth related quantitative traits in narrow STAIRS at the top of *Arabidopsis* chromosome 3.
2. To fine map growth related QTL located on top of chromosome 3 with special relevance to QTL controlling flowering time.
3. To identify candidate genes within those regions to which the QTL mapped using the TAIR database.

6.3 Materials and Methods

6.3.1 Plant material and experimental design

Twenty-four narrow STAIRS spanning the top twenty cM of chromosome 3 together with Columbia and CSS3 were used in the trial. Out of these lines, 23 of the narrow STAIRS were produced in the current research while Koumproglou et al had produced the remaining one in 2002.

The 24 STAIRS could be grouped into twelve marker-demarcated bins/genotypes. Some of the different STAIRS that originated from different back-cross parents and thus had crossovers at different positions between two markers, had to be included in the same marker-demarcated bin because the bins could be identified only with the available markers. Out of the 24 STAIRS, some lines contained two or more true replicates of the STAIRS that originated from the same back-cross parent, thus increasing the total number of lines to 36. All the STAIRS have a L*er* introgression at the top region of chromosome 3 with a single recombination along the top 20cM of chromosome 3.

At the completion of constructing the narrow STAIRS within the top 20 cM, seeds were available from only one mother plant of every line. Therefore, a seed multiplication programme was carried out prior to the trial in order to obtain seeds from at least ten different mother plants for each line so as to provide replicated families for each line to account for any maternal effects.

Thirty individuals each from 37 lines (a total of 1110 plants) were grown in a completely randomised experimental design for the scoring of quantitative traits. The 30 individuals for each line consisted of six replicated families (i.e. from 6 mother plants) each having five individuals. The details of the lines used in the trial are given in table 6.1. The pot sizes, sowing of seeds, plant growth conditions and the scoring methods are the same as described in Chapter 04. The characters scored are the same except that an additional trait, bolting time was recorded in the

current experiment. Bolting time was defined as the number of days from the germination to the first date the plant produced the shoot from the rosette.

6.3.2 Data analysis

Descriptive statistics were calculated for the traits for all the lines separately, for marker-demarcated bins/genotypes. Hierarchical Analysis of variance was performed in MINITAB version 14 using the General Linear Models Procedure which accounts for the missing data. Genotypes (as demarcated by marker bins), lines and replicates of lines were considered as sources of variation in ANOVA. When some of the sources of variation were non-significant such items were pooled in the ANOVA. Genotypes were regarded as fixed factors while, lines and replicated lines were considered as random factors within genotypes. Germination time was used as a covariate in the ANOVA when it proved to be significant.

After observing the means and the results of ANOVA, two lines were removed from the analysis for interpreting the data in a more meaningful way. The least squares model fitting was carried out to obtain the QTL location for the characters that showed significant differences among marker bins. Adjusted means after calculating the effect of covariate for each genotype were used in the model fitting. When the replicated lines within a genotype were significant in ANOVA, the MS related to replicated lines within genotype was used in calculating the weights. Otherwise the error MS was used for the calculation of weights for model fitting. Appropriate models for each trait were identified by the significant parameter levels and non significant chi squared values of the fitted models.

After locating QTL within marker bins the substitution effects of Col with L*er* were calculated for each locus. The model fitted is given in table 6.2 and the cM distances of each region in the model are given in table 6.3

In order to observe the inter-relationships among characters, Pearson's correlation coefficients were calculated in MINITAB between each pair of traits.

6.3.3 Search for candidate genes

After mapping QTL, TAIR database (www.arabidopsis.org) was searched in order to identify possible candidate genes specially for flowering time and flowering related characters within the QTL regions.

6.4 Results

6.4.1 Means of traits

The observation of means for each line and genotype indicated very high variation among the genotypes and lines. From the previous studies it was known that Col was early flowering and CSS3 was late flowering and the same pattern was observed in this trial. Out of the two lines in marker bin 1 one line was early flowering while the other was as late flowering as the rest of the lines. In order to avoid complications in the analysis and to locate QTL this late flowering line was removed from further analysis. Similarly, out of the three replicated lines of the STAIRS no. three in the marker bin 2 one line was early flowering and the rest were late flowering. This line also was removed from further analysis in order to get clear regions for QTL for the traits that showed significant differences among genotypes.

The graphical representation of the means of each of the marker bins are given in figure 6.1

6.4.2 Analysis of Variance and Least Squares Model Fitting

The ANOVAS are given in the appendix 06. The significant parameters and the results of model fitting and the observed and expected means, weights of the lines are given in tables 6.4 and 6.5 respectively.

6.4.2.1 Germination time

The observed variation in germination time was non-significant among the marker-demarcated genotypes. However, narrowly significant differences were observed among lines within genotypes (P=0.047) and replicated lines within STAIRS (P=0.018). No model fitting was carried out because the model fitting was performed only if the marker-demarcated genotypes were observed to be significantly different.

6.4.2.2 Leaf number day 20 (LN-20)

The variation in leaf number at day 20 was observed to be highly significant (P=0.001) among genotypes and showed significant differences (P=0.007) among replicated lines within genotypes. The best fitting model (χ^2=12.45, 10 d.f) revealed QTL in regions a2 and a10. Substitution of Col alleles by L*er* alleles in region a1 reduces the leaf number at this stage by 1.42 while the same substitution in region a10, increases the leaf number by 0.50.

6.4.2.3 Rosette width day 20 (RW-20)

RW-20 showed very highly significant differences both among genotypes and replicated lines within genotypes (P>0.0001). However, a significant model could not be fitted to the observed means because the model that was most close had a narrowly significant chi-squared value (χ^2=20.91, 9 d.f). However, the region a2 showed significant differences for this trait as well.

6.4.2.4 Bolting time

Very highly significant differences (P>0.0001), both among genotypes and replicated lines within genotypes were observed in ANOVA for the trait bolting time. The best-fit model (χ^2=14.437, 9 d.f) identified one region with QTL of major effect and two more regions containing QTL with minor effects. The allelic substitution of Col by L*er* in region a2 increases the bolting time by

9.84 days. The same substitution in region a6 decreases the bolting time by 1.62 days while the L*er* alleles in place of Col in region a10 further decreases the trait time by 3.33 days.

6.4.2.5 Flowering time

The ANOVA for flowering time showed very highly significant differences (P>0.0001) among genotypes and replicated lines within genotypes. Three regions containing QTL were identified by the best-fit model (χ^2=16.558, 9 d.f). The substitution of Col by L*er* in regions a2, a6 and a10, delays flowering time by 10.5 days, and advances flowering time by 1.6 days and by 3.6 days respectively.

6.4.2.6 Height at flowering

The variation in genotypes did not prove to be statistically significant in ANOVA. The STAIRS showed very highly significant differences (P>0.0001) in one-way ANOVA. However, no model fitting could be carried out due to the non-significance of marker-demarcated genotypes.

6.4.2.7 Rosette width at flowering (RWF)

Similar to height at flowering, rosette width at flowering also did not show significant differences among marker-demarcated genotypes in ANOVA. However, the STAIRS and replicated lines within STAIRS showed significant differences by the general linear model procedure. However, no model fitting could be conducted to reveal QTL for the available variation among STAIRS for this particular trait.

6.4.2.8 Rosette leaf number at flowering (RLNF)

Both the genotypes and the replicated lines within genotypes were revealed to be very highly significantly different (P> 0.0001) in ANOVA for RLNF. The best-fit model (χ^2=4.851, 9 d.f) in the least squares model fitting identified three regions containing QTL affecting the trait. Region a2 contained QTL with the highest effect, showing a Col by L*er* substitution effect of increasing

the rosette leaf number by 6.312. The substitution of Col by L*er* in regions both a5 and a10 had decreasing effects on the phenotype, respectively by 1.086 and 3.313 leaves.

6.4.2.9 Cauline leaf number at flowering (CLNF)

The observed variation in CLNF proved to be statistically significant both among genotypes (P=0.007) and replicated lines within genotypes (P=0.009). The best-fit model (χ^2=5.801, 9 d.f) revealed three regions, a2, a6 and a10 containing QTL for flowering time. The substitution of Col by L*er* increased CLNF by 3.194 in region a2, decreased by 0.963 in region a6 and further decreased by 1.132 in region a10.

6.4.2.10 Height at day 30 (H-30)

Analysis of variance revealed very highly significant (P>0.0001) differences among genotypes for H-30. However, a significant model to fit the data could not be identified due to the significant chi-square values of the models fitted. The regions a1, a2 and a11 consistently showed significant QTL but due to the inadequacy of the fitted models, QTL regions and their substitution effects could not be explicitly decided.

6.4.2.11 Height at day 36 (H-36)

ANOVA revealed very highly significant differences (P>0.0001) among genotypes and replicated lines among genotypes for the trait height at day 36. However, similar to the previous height measurement at day 30, a significant model could not be fitted to uncover the QTL regions. Once again for height regions a1, a2 and a11 were significant for the presence of QTL but the models were not significant. As a result, the QTL regions and their substitution effects could not be definitively identified.

6.4.2.12 Final height – day 54

The final height measured at day 54 did not prove to be significantly different among genotypes. Therefore, no model fitting was carried out for this particular trait.

6.4.3 Pearson's Correlation Coefficients

The correlations between the characters were similar to the observations with the experiment for the analysis of wide STAIRS explained in chapter 04. The additional trait that was scored in this experiment bolting time showed very high positive correlations with flowering time as expected. Apart from that flowering time and also bolting time were positively correlated with rosette and cauline leaf number at flowering time and negatively correlated with all the height measurements. Early vegetative growth measurements rosette width and leaf number were positively correlated. The correlation coefficients for the traits are given in table 6.6.

6.5 Discussion and conclusions

6.5.1 Removal of two lines from the analysis

After the initial analysis line no.2 in bin one and one line out of the three replicated lines in bin 2, STAIRS no.3 were removed from further analysis. Out of the two lines in bin 01, one line was very clearly early flowering as indicated by the means while the other was late flowering (means were 30.48 and 40.29 days). This difference of the flowering time was attributed to the presence of QTL within the particular bin. The two STAIRS in the bin are two different STAIRS that originated from two different back cross individuals. As a result each of them contains recombination at two different positions along the chromosome within the bin. The recombination for the late line could be further down along the chromosome after the position of the QTL locus resulting in this line being late flowering. Inclusion of this line in further analysis

would have complicated the finding of QTL region and the substitution effects. So the line two in bin one was removed from further analysis.

The other line which was removed from further analysis was included in bin two. There were two different STAIRS included in this marker-demarcated bin. The removed line was one of the three replicated lines that originated from the same back cross parent. All the other lines included in this marker bin were late flowering while the line removed was observed to be early flowering. The reason for this line to be early flowering while the others were late flowering cannot be explained as clearly as the above line. But one of the possible reasons is as follows. Even the replicated lines within a certain single recombinant line that originated from the same back cross parent undergo another round of selfing for them to be made homozygous. During this process it is possible for the chromosomes to recombine at the heterozygous region and some of the heterozygosity can still remain in the selected lines. This heterozygosity can remain unnoticed by a certain difficult marker and this may account for this specific line being early flowering. This is especially relevant considering that this particular line is in the adjacent bin to the bin with the QTL.

The second possible reason is that there may be a certain heterozygous region or regions in the recurrent genetic background that went unnoticed in genotyping even though the markers were fairly evenly spaced throughout the genome. Whichever is the reason for this unexpected behaviour, this line also had to be removed from further analysis for the same reasons explained above.

6.5.2 Reasons for significant differences within the STAIRS in marker bins

In the ANOVA statistically significant differences have been identified even within the duplicated lines within marker bins. As has been explained in the previous paragraph and in chapter 04, this can be explained by the fact that, even within the STAIRS in a marker bin,

recombination can occur at different points because they arise from different backcross individuals.

Even for the replicated lines that originated from the same BC parent, recombination between chromosomes can occur at different locations within the heterozygous region in the subsequent selfing, resulting in genetically different individuals. If these changes occur at regions that house even minor QTL for the trait concerned, the lines within marker demarcated bins can display statistically significant differences. However, these differences can sometimes cause problems in the analysis if they are very high, as happened in the current analysis where two lines had to be removed from.

6.5.3 Difficulty of fitting models for certain traits

Fitting a significant model was difficult for some traits such as height, even though the observed variation among the genotypes was very highly significant. For these traits, the fitted models had a slightly significant chi-squared, which indicated that the model was not sufficient to explain the means, even though very highly significant QTL regions could be identified. One reason for this can be the presence of QTL with minor effects that went unaccounted for in certain marker bins. As explained earlier, the 24 different STAIRS were grouped into 13 marker bins. In the model fitting all the STAIRS within a marker bin are pooled to form the collective marker-bin/genotype mean. This may lead to the non-identification of some QTL regions with minor effects and the models becoming inadequate. In order to avoid this, more genetic markers need to be developed within those marker bins which show significant differences among STAIRS and they need be grouped into new marker bins. However, this is a very time consuming task and may not be cost and time effective when such QTL have only very small effects on the phenotype of the trait.

Furthermore, if the trial is repeated with the same lines and same marker information it will still be possible to precisely map QTL for these traits because it is common in QTL mapping for some

of the QTL effects to be hidden in certain experiments and surface in a repeated experiment. This is due to the fact that environments play a major role in the phenotypic expression of QTL unlike major or qualitative traits.

6.5.4 Location of QTL and comparison of the mapped QTL with results from the analysis of wide STAIRS

Two regions of QTL were identified for flowering time within the first 20 cM of chromosome 3. A QTL with a major effect on flowering time was located within 2-3 cM on chromosome 3. In addition to this a second QTL with less effect on flowering time was located within 15-20 cM. In the previous experiment with wide STAIRS the first QTL was spotted between 0-14 cM while the second minor QTL was not located.

QTL identified for bolting time were the same as for flowering time and because bolting precedes flowering this may be the same gene rather than a separate QTL.

Two QTL were located within the first 20 cM for the pre-flowering vegetative character leaf number at day 20. The first QTL for this trait was located within 2-3 cM while the second QTL was located in region a10, which is within 15-20 cM. In the previous experiment with wide STAIRS a model could not be fitted due to the significant chi-square values of the fitted models. However, the best model (which was very close to being adequate) showed QTL in regions 0-14 cM and 14 – 20 cM which agrees perfectly with the results of the current experiment.

QTL for rosette width could not be unambiguously located during the current experiment at the two time points the trait was measured - day 20 and at flowering. In the previous experiment two QTL were located for this particular trait at regions 0-14 cM and 14-20 cM. The means of the trait for these two time points give supporting evidence for the earlier finding, but the available

data do not provide sufficient information to limit the QTL to finer positions than previously identified.

Three QTL regions were located for the traits rosette and cauline leaf numbers at flowering. The major QTL for both the traits were fine mapped to 2-3 cM while a QTL with less effect on the trait was located within 15-20 cM. Apart from these two a third QTL was located for rosette leaf number in the region 6-9 cM, while a third QTL was located for cauline leaf number at flowering in the region a6, which spans 9-10 cM. In the previous experiment with wide STAIRS a major QTL was located within the first 15 cM for both of the traits. The current experiment provided conclusive data to fine map the major QTL in addition to identifying two minor QTL regions to affect the trait.

The current research failed to locate QTL for height at any of the time points the trait was measured although significant differences were observed among the STAIRS. Although the model was narrowly inadequate there was a highly significant QTL present in region a2, which is from 2-3 cM. Other than that major QTL, two minor yet significant QTL were located in regions from 0-2 cM and 21-44 cM. In the previous experiment with wide STAIRS, apart from the major QTL from 0-14 cM, additional QTL were located for height in 14-20 cM, 20-44 cM and 80-100cM. The current study does not permit the identification of QTL beyond 44 cM due to the lack of STAIRS covering that part of the chromosome. Although the model was narrowly inadequate the current experiment supports the findings of the previous experiment of having QTL from 0-14 and 21-44 cM for height.

Therefore, although significant models could not be fitted to certain traits in the current experiment, the findings agree perfectly with the previous experiment results. Both the locations of QTL for each trait and the substitution effects observed were in perfect agreement with the

data obtained from the analysis of wide STAIRS although some of the minor QTLs could not be definitively located in either one or the other experiment.

Apart from the QTL mapping, correlations between the traits observed in the two trials are the same as expected.

Major QTLs were located in the region 2-3 cM for flowering time and flowering related characters cauline leaf number and rosette leaf number. Although adequate QTL could not be fitted for height, there was very strong evidence to suggest QTL were located within this region as well. In addition to this region the second major region to house QTL within the top of chromosome 3 was the region from 15-20 cM. The region 15-20 cM contained QTL with smaller but opposite effects on the trait to the QTL in the previous region.

6.5.5 Resolution of the mapped QTL

The objective of the current experiment was to achieve fine mapping of QTL with special relevance to flowering time using narrow STAIRS. The major QTL for flowering time on top of chromosome 3 could be mapped to 2-3 cM with a C.I of 1 cM. This very high map resolution is very rarely achieved by QTL mapping with conventional segregating populations.

The second QTL mapped to the region of 15-20 cM. Five cM region for the location of QTL is high when compared with 1 cM in the previous QTL. Yet, in conventional terms mapping QTL to a 5 cM region is still achieving high resolution. This QTL could have been mapped to a finer position if more narrow STAIRS within this region could have been isolated in the breeding programme.

6.5.6 Gene action

In the earlier research with wide STAIRS, it was suggested that QTL located in the top region of chromosome 3 affecting flowering time and related traits may be due to tightly linked genes or to pleiotropy of the same gene. Now, the analysis with narrow STAIRS gives strong evidence for pleiotropy rather than tight linkage between the genes because the major QTL to affect flowering time and related traits mapped to the same region that is, 2-3 cM. Therefore, it is highly likely to be the same gene that controls flowering time that affects the other flowering related traits as well.

6.5.7 Search for candidate genes

From the TAIR web site, a total of 100 loci have been identified in the region 2-3 cM. Out of the 100genes, the locus AT3G02380, which is named COL-2 with TAIR accession numbers 3698228 and 4370595) was identified as a very strong candidate gene to affect flowering time. This gene has been listed as homologous to the flowering time gene CONSTANS but the exact role in flowering has not yet been demonstrated (Koornneef et al. 1998b).

6.5.8 Significance of QTL mapping with narrow STAIRS

The current experiment has amply displayed the power of STAIRS in fine mapping of QTL. However, in order to achieve this very high resolution in QTL mapping with STAIRS it is essential to produce STAIRS with sufficiently narrow steps. As an example in the current research there were no lines having the recombination within 15-20 cM. Considering the very large size of the population that was used in constructing these narrow STAIRS it was expected to find STAIRS differing in about 1 cM all along the chromosome. The non-availability of single recombinants along certain parts of the chromosome gives evidence for the non-uniformity of

recombination along the chromosome. On the other hand, had there been any QTL in marker bin three, which differs by only 0.5 cM, even greater resolution could have been achieved in mapping. Furthermore there are eight different STAIRS within marker bin 04, which spans from 5-6 cM. Any QTL that may be present in this region could be mapped to a region much less than 1 cM provided that more markers can be developed within this region. With the availability of genetic markers such as single nucleotide polymorphisms identifying more polymorphic markers within a region of interest should not be an unachievable target. Finally, constructing STAIRS narrow enough involves an enormous amount of work. But once these single recombinant lines are produced and genotyped they can be maintained as permanent resources and can be used in any project in locating polymorphic QTL.

Table 6.1: The twenty six distinct lines (STAIRS) used in the experiment, introgressed regions of STAIRS, categories of marker bins of each line, regions of recombination of each line and the number of replicated lines for each of the STAIRS.

STAIR no. (line)	Introgressed region (cM)	Marker bin/genotype	Region of recombination (cM)	No. of replicated lines
1 (814)	0-2	1	2-3	1
2 (628)	0-2	1	2-3	1
3 (982)	0-3	2	3-4	3
4 (58)	0-3	2	3-4	1
5 (1008)	0-4	3	4-4	1
6 (5)	0-5	4	5-6	1
7 (255)	0-5	4	5-6	2
8 (432)	0-5	4	5-6	2
9 (879)	0-5	4	5-6	3
10 (947)	0-5	4	5-6	2
11 (1006)	0-5	4	5-6	2
12 (1155)	0-5	4	5-6	1
13 (1270)	0-5	4	5-6	1
14 (35)	0-6	5	6-9	1
15 (749)	0-6	5	6-9	1
16 (1101)	0-9	6	9-10	1
17 (855)	0-10	7	10-10	2
18 (354)	0-11	8	11-13	1
19 (303)	0-11	8	11-13	1
20 (854)	0-11	8	11-13	2
21 (1171)	0-11	8	11-13	1
22 (1276)	0-11	9	11-13	1
23 (410)	0-15	10	15-20	1
24 (0-21)	0-21	11	21-44	2
25 CSS3	-	12	-	1
26 COL	-	-	-	1

Table 6.2: The model used for detecting and estimating genetic differences among narrow STAIRS. (-) denotes regions of introgressed L*er* donor alleles and (+) denotes regions of Col alleles.

Bin	m	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12
Col	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	-1	1	1	1	1	1	1	1	1	1	1	1
2	1	-1	-1	1	1	1	1	1	1	1	1	1	1
3	1	-1	-1	-1	1	1	1	1	1	1	1	1	1
4	1	-1	-1	-1	-1	1	1	1	1	1	1	1	1
5	1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1
6	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1
7	1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1
8	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1
9	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1
10	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1
11	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
CSS3	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Table 6.3: The genetic and physical distances of the regions explained in the model.

Region	cM distance	Physical distance (kbp)
a1	0-2	0-526
a2	2-3	526-740
a3	3-4	740-888
a4	4-4	888-1058
a5	5-6	1060-1514
a6	6-9	1514-2020
a7	9-10	2020-2311
a8	10-10	2311-2200
a9	11-13	2420-2630
a10	15-20	3430-4608
a11	21-44	-
a12	44-100	-

Tables 6.4: The *significant* parameters and their estimated values, standard errors and normal deviate (c) for each of the traits following least-squares model fitting. All 'c' values greater than 1.96 were taken as significant.

Table 6.4.1: Leaf number- day 20

Parameter	Estimate	Standard Error	C
m	10.576	0.134	78.649
a2	0.709	0.115	6.175
a10	-0.247	0.080	3.076

Table 6.4.2: Bolting Time

Parameter	Estimate	Standard Error	C
m	28.686	0.551	52.033
a2	-4.920	0.477	10.312
a6	0.810	0.249	3.259
a10	1.663	0.372	4.474

Table 6.4.3: Flowering time

Parameter	Estimate	Standard Error	C
m	32.021	0.529	60.572
a2	-5.261	0.458	11.493
a6	0.800	0.239	3.344
a10	1.815	0.357	5.090

Table 6.4.4: Rosette leaf number at flowering

Parameter	Estimate	Standard Error	C
m	16.048	0.584	27.492
a2	-3.158	0.507	6.224
a5	0.543	0.253	2.145
a10	1.657	0.384	4.315

Table 6.4.5: Cauline leaf number at flowering

Parameter	Estimate	Standard Error	C
m	4.332	0.412	10.502
a2	-1.597	0.357	4.479
a6	0.482	0.188	2.565
a10	0.566	0.280	2.022

Tables 6.5: The genotypes, number of individuals, observed means, calculated means and the expected means for each trait resulted from least squares model fitting.

Table: 6.5.1 Leaf number day 20

Genotype	*N*	*Observed mean*	*Weight*	*Expected mean*
1	30	11.034	10.997	11.038
2	25	10.883	9.164	11.038
3	90	9.548	32.991	9.620
4	30	9.769	10.997	9.620
5	417	9.666	152.859	9.620
6	60	10.006	21.994	9.620
7	30	9.640	10.997	9.620
8	60	9.333	21.994	9.620
9	120	9.528	43.988	9.620
10	60	9.404	21.994	9.620
11	30	10.000	10.997	10.115
12	60	9.888	21.994	10.115
13	30	10.682	10.997	10.115

Table: 6.5.2 Bolting time

Genotype	*N417*	*Observed mean*	*Weight*	*Expected mean*
1	30	25.380	0.654	26.239
2	25	27.270	0.545	26.239
3	89	36.310	1.941	36.079
4	30	35.510	0.654	36.079
5	417	35.980	9.095	36.079
6	60	36.710	1.309	36.079
7	30	34.510	0.654	34.458
8	60	34.160	1.309	34.458
9	119	34.040	2.595	34.458
10	60	35.560	1.309	34.458
11	30	32.540	0.654	31.133
12	60	32.140	1.309	31.133
13	30	27.710	0.654	31.133

Table: 6.5.3 Flowering time

Genotype	*N*	*Observed mean*	*Weight*	*Expected mean*
1	30	28.380	0.712	29.376
2	25	30.570	0.593	29.376
3	89	40.160	2.111	39.898
4	30	39.470	0.712	39.898
5	411	39.820	9.749	39.898
6	57	40.280	1.352	39.898
7	30	38.330	0.712	38.298
8	60	37.920	1.423	38.298
9	118	37.880	2.799	38.298
10	60	39.480	1.423	38.298
11	30	36.020	0.712	34.668
12	60	35.750	1.423	34.668
13	30	31.150	0.712	34.668

Table: 6.5.4: Rosette leaf number at flowering

Genotype	*N*	*Observed mean*	*Weight*	*Expected mean*
1	30	14.290	0.585	15.090
2	25	16.050	0.488	15.090
3	88	21.810	1.716	21.406
4	29	21.390	0.566	21.406
5	409	21.320	7.978	21.406
6	57	20.460	1.111	20.320
7	30	19.300	0.585	20.320
8	59	20.410	1.151	20.320
9	117	20.120	2.282	20.320
10	60	21.000	1.170	20.320
11	29	18.270	0.566	17.007
12	60	17.190	1.170	17.007
13	30	15.420	0.585	17.007

Table: 6.5.5 Cauline leaf number at flowering

Genotype	*N*	*Observed mean*	*Weight*	*Expected mean*
1	30	3.500	1.175	3.782
2	25	4.120	0.979	3.782
3	87	7.490	3.408	6.976
4	28	6.071	1.097	6.976
5	406	6.814	15.903	6.976
6	55	7.820	2.154	6.976
7	30	5.833	1.175	6.013
8	58	6.071	2.272	6.013
9	117	5.777	4.583	6.013
10	59	6.515	2.311	6.013
11	28	5.143	1.097	4.881
12	60	5.150	2.350	4.881
13	30	4.100	1.175	4.881

Table 6.6: Pearson's correlation coefficients between the traits. (Only significant correlations are given.) Correlation coefficients over (+/-) 0.5 are in bold.

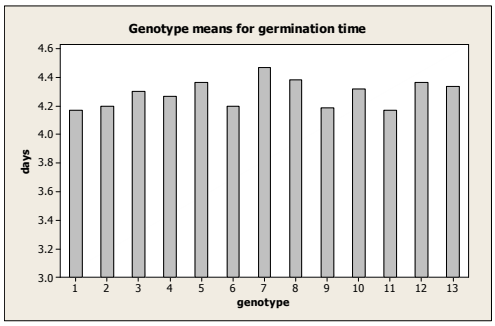
Pearson Correlations Coefficients:

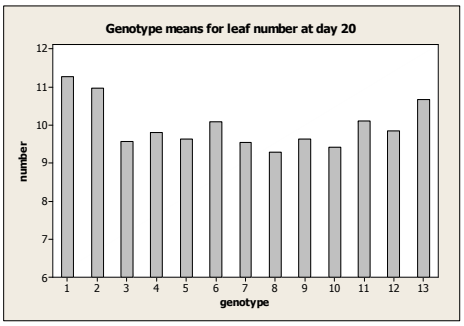
	Germ.	LN-20	RW-20	BT	FT	HF	RWF	RLNF	CLNF	H-30	H-36
LN20	-0.357										
	0.000										
RW20	-0.441	**0.692**									
	0.000	0.000									
BT	0.175	-0.441	-0.460								
	0.000	0.000	0.000								
FT	0.143	-0.394	-0.422	**0.928**							
	0.000	0.000	0.000	0.000							
HF	-0.063	0.083	0.092	0.104	0.207						
	0.034	0.005	0.002	0.000	0.000						
RWF	-	-	0.061	-0.034	0.028	-0.069					
	-	-	0.041	0.258	0.346	0.020					
RLNF	0.087	-0.238	-0.221	**0.557**	**0.581**	-	0.138				
	0.004	0.000	0.000	0.000	0.000	-	0.000				
CLNF		-0.136	-0.126	**0.573**	**0.596**	0.438	-	0.331			
		0.000	0.000	0.000	0.000	0.000	-	0.000			
H30	-0.100	0.345	0.408	**-0.545**	**-0.524**	0.071	-	-0.355	-0.184		
	0.001	0.000	0.000	0.000	0.000	0.019	-	0.000	0.000		
H36	-0.108	0.387	0.436	**-0.775**	**-0.746**	0.091	-	**-0.510**	-0.312	**0.802**	
	0.000	0.000	0.000	0.000	0.000	0.002	-	0.000	0.000	0.000	
H	-0.106	0.191	0.250	**-0.617**	**-0.570**	-0.181	0.228	-0.280	-0.484	0.212	0.302
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

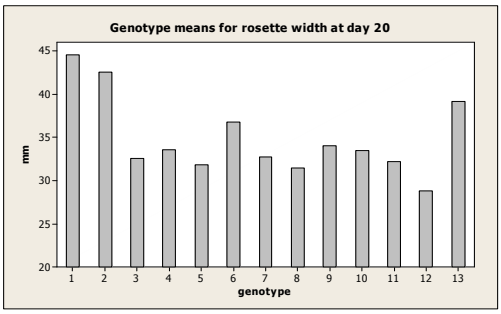
Cell Contents: Pearson correlation
P-Value

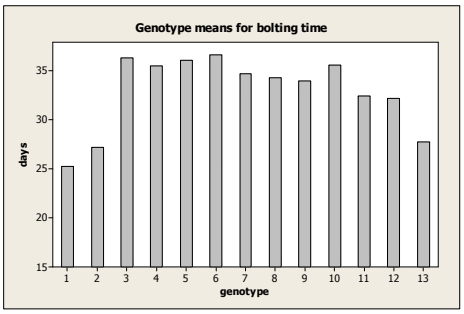
Figure 6.1 The morphological differences in two plants each from early flowering lines (on the left) and late flowering lines (on the right) 30 days after sowing.

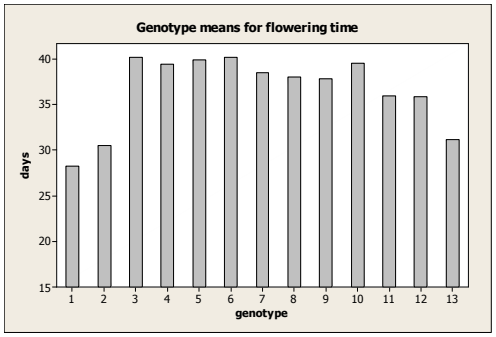
Figure 6.2 Graphical representation of means of different traits for each genotype

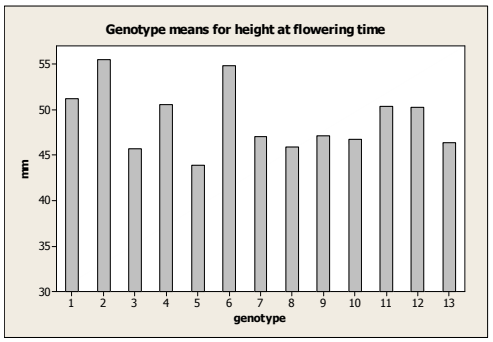


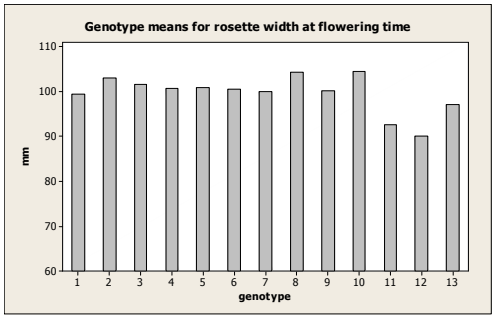


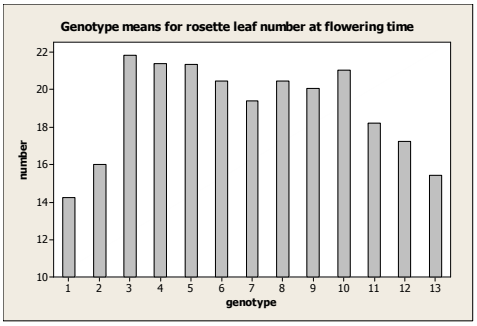


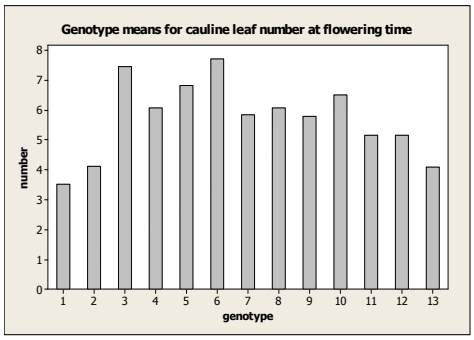


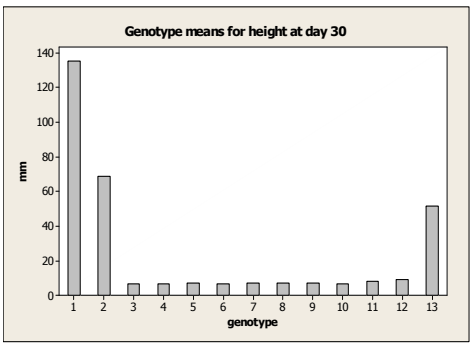


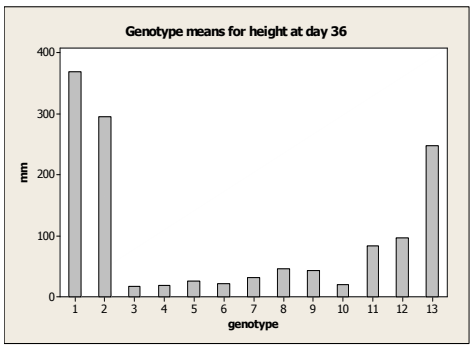


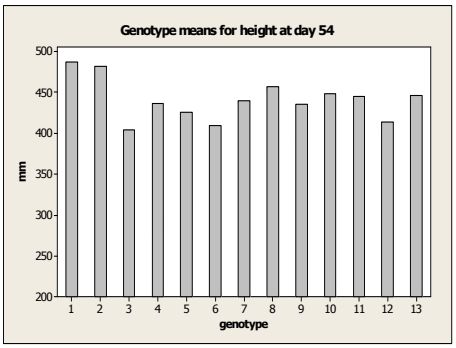

Genotype means for rosette width at flowering time
mm
genotype


Genotype means for rosette leaf number at flowering time
number
genotype


Genotype means for cauline leaf number at flowering time
number
genotype


Genotype means for height at day 30
mm
genotype


Genotype means for height at day 36
mm
genotype


Genotype means for height at day 54
mm
genotype

CHAPTER 07

ANALYSIS OF HETEROSIS IN CHROMOSOME SUBSTITUTION STRAINS IN *ARABIDOPSIS THALIANA*[1]

Abstract

An experiment was conducted to investigate the hybrid vigour related to several morphological traits in *Arabidopsis thaliana* on an individual chromosome basis. Whole chromosome substitution strains (CSSs) of chromosome three, four and five of *Arabidopsis*, recurrent parent Columbia and the three relevant F_1 progenies were grown in a completely randomised experimental design and several morphological traits were scored. The means of the populations were calculated and the presence or absence of heterosis for each trait on a per-chromosome basis was determined. The traits rosette width at day 20, flowering time, height at flowering time, rosette leaf and cauline leaf number at flowering displayed significant better parent heterosis in different chromosomes. The genetic components of the means and the potence ratios indicated those parents having more dominant alleles in the cross in each chromosome. The analysis of the results indicated the appropriateness of CSSs in studying the phenomenon of heterosis in a micro scale rather than the whole genome level.

[1] Andrew Bennet helped in carrying out the plant house trial and did all the data scoring as part of his MSc research project from May to July 2004.

7.1 Introduction

7.1.1 What is heterosis?

Heterosis is the superiority of an F_1 hybrid over the mean performance of its better-inbred parent. This phenomenon is also known as hybrid vigour and has been widely used in crop and livestock breeding in agriculture and in evolutionary genetics (Stuber et al. 1992; Xiao et al. 1995).

However, the term heterosis is also used occasionally to describe the F_1 exceeding the mid-parental value of a particular cross or to the hybrid progeny of a top-cross between several inbred lines exceeding the average performance of the parental inbreds. Such heterosis is referred to as mid-parent heterosis while the superiority of the hybrid over the better parent is termed better parent heterosis (Kearsey and Pooni 1996).

7.1.2 Causes of heterosis

7.1.2.1 Heterosis in the absence of epistasis

In the absence of gene interactions and other complicating genetic relationships between loci, heterosis or hybrid vigour is displayed when the average dominance is greater than the degree of gene dispersion. As a result, if the genes are dispersed in the two parents it requires only very little dominance at individual genes to produce quite considerable heterosis. This also implies that characters that are controlled by many genes, such as fitness or yield, would require only very small amounts of dominance to produce very major hybrid superiority without the need to invoke over-dominance (Kearsey and Pooni 1996).

7.1.2.2 Heterosis in the presence of epistasis

When epistasis, i.e. the interaction between genes is present heterosis can be caused by dominance complementation of genes. This theory indicates that in the presence of complementary epistasis, dominant increasing alleles at each of the two genes have a proportionately greater effect when they occur together than would be expected from their individual effects (Kearsey and Pooni 1996).

Apart from the dominant complementation of genes, yet another hypothesis as to the cause of heterosis states that the over-dominance also may contribute to heterosis in some instances. Yet evidence for actual over-dominance remains scarce.

7.1.3 Use of heterosis in agriculture

Most of the practical interest in heterosis centres on the breeding of cultivated plants. Heterosis for size, vigour or yield is most evident in outbreeding crops and over the years plant breeders have exploited the hybrid vigour in producing mostly high yielding cultivars of crops. Maize is a very good example in which heterosis has long been exploited in the production of high yielding F_1 seeds in commercial quantities.

The experiments with the cultivated crop wheat have provided evidence to support the hypothesis that heterosis in wheat is due to dominance complementation with linkage and interaction of alleles (Pickett and Galwey 1997). The same theory was shown to apply for the heterosis in maize in the experiments conducted by (Sprague 1983).

(Xiao et al. 1995) demonstrated that over-dominance is not a major cause of heterosis for yield in a cross between the two subspecies of rice, because there was no correlation between most traits and overall genomic heterozygosity. In this experiment heterozygotes were never superior to both homozygotes in analysis of QTL and some F_8 inbred lines were actually superior to the F_1 for all traits evaluated.

7.1.4 Use of CSSs of *Arabidopsis* in studying heterosis

Despite the exploitation of heterosis in improving desirable traits in crop and animal breeding the underlying genetic basis of heterosis is still unclear. The advances of molecular genetics in the recent past, paves the way for the study of molecular basis related to heterosis. An organism such as *Arabidopsis* to which the most advanced molecular techniques can be applied, is a suitable candidate to be used in such a study. The availability of whole chromosome substitution strains (CSSs) in *Arabidopsis* facilitates the investigation of heterosis on an individual chromosome basis rather than the whole genome basis.

When using CSSs in studying heterosis, the relevant CSS and the recurrent parent are considered as the inbred parents. A cross between these two parents will generate an F_1 progeny in which the particular chromosome is entirely heterozygous in a pure homozygous genetic background of the recurrent parent. The genetic analysis of desired traits in these three populations (two parents and the F_1) facilitates the studying of heterosis on an individual chromosome basis.

7.2 Objectives

1) To detect the presence or absence of heterosis for certain morphological traits in *Arabidopsis* CSSs.
2) To calculate the additive and dominance genetic components of means of the above traits.

7.3 Materials and Methods

7.3.1 Plant material

Seeds of CSSs of *Arabidopsis* related to chromosomes 3, 4 and 5, (CSS2-5), the relevant F_1 progeny of CSSs (obtained by backcrossing each CSS to the recurrent parent Columbia (CSS3-F_1, CSS4- F_1 and CSS5- F_1) along with Col (the recurrent parent in substitution strains) were used in the experiment. CSS-1 and CSS2-F_1 were not available therefore the experiment had to be limited to chromosomes 3, 4 and 5.

Twenty-five plants of each line (a total of 175 plants) were grown in a completely randomised experimental design under controlled environmental conditions explained in chapter 04.

7.3.2 Traits scored

The traits germination time (GT), bolting time (BT) (days from germination to bolting), flowering time (FT) (days from germination to opening of the first flower) leaf number and the rosette width at day 20 from sowing (LN-20 and RW 20), height, rosette width, rosette and cauline leaf number at flowering time (HF, RWF, RLNF and CLNF) were scored for each plant.

7.3.3 Determination of the presence of heterosis

Descriptive statistics were calculated for all the traits for the lines Col, CSS3, CSS4, CSS5, CSS3- F_1, CSS4- F_1 and CSS5- F_1. The means of Col, each CSS and the relevant F_1 were compared for each trait to detect whether the F_1 means were either lower or above the means of the parents. When the mean value of the F_1 was either below or above the means of both the parents, Student's t test was performed to compare the means of the F_1 and the parent that has a mean closer to the F_1 value. The results of this test helped determine whether the observed difference between the means was statistically significant. If it was, the trait was assumed to

show heterosis or hybrid vigour. When germination-time was shown to have an effect on the observed variation among the lines (as displayed by ANOVA using germination time as a co-variate) the adjusted means with the covariate were used in calculating the heterosis.

In addition, homogeneity tests (Bartlett's test) for the variance between the three populations in each chromosome were performed for each trait to confirm the uniformity of environmental variation for each trait.

7.3.4 Calculation of genetic components of means

The mid parental value m, the additive and dominance components (a_A and d_A respectively) were calculated for each trait for chromosomes 3, 4 and 5 using the following formulae.

$$m = (P_1 + P_2) / 2$$

$$a_A = (P_1 - P_2) / 2$$

$$d_A = F_1 - (P_1 + P_2) / 2$$

P_1 = mean of Col

P_2 = mean of CSS

F_1 = mean of F_1

The potence ratio (d_A / a_A) was then calculated for each trait in each chromosome.

7.4 Results

7.4.1 Detecting heterosis

7.4.1.1 Means

The means for the traits (germination time, rosette width and leaf number at day 20, flowering time, height at flowering time, rosette and cauline leaves at flowering time) are given in table 7.1

Out of all the traits scored for each of the chromosome, GT (chr. 3), LN-20 (chr.3, chr.5), RW-20 (chr.3, chr.5), HF (chr.3), RLNF (chr.3, chr.5), RWF (chr.3, chr.5) and CLNF (chr.3, chr.5) showed F_1 means higher than the means of the parent with greater mean values. The traits RW-20 (chr.4), BT (chr.4), HF (chr.4), RLNF (chr.4) RWF (chr.4) and CLNF (chr.4) showed means less than the means of the parent with lower mean values.

7.4.1.2 Student's t test

The results of the t-tests conducted to determine whether the observed mean differences were significantly different indicated the presence or absence of heterosis of the above listed traits on the relevant chromosomes. Accordingly, significant positive heterosis was observed for the traits LN-20 (chr.5), RW-20 (chr.3), RW-20 (chr.5), HF (chr.3), RLNF (chr.5) and CLNF (chr.3). Also out of the traits listed above significant negative heterosis was observed for RLNF (chr.4).

7.4.1.3 Bartlett's homogeneity test for variance

Test for equal variance in chromosome three, showed that the variation among the three populations (parent 01–Col, parent 02- CSS and F_1) differed for germination time ($P>0.0001$) and rosette width at flowering ($P=0.024$) while for all the other traits the observed variation among the populations was uniform.

The variation among the three populations in chromosome four were uniform except for the traits, germination time ($P=0.041$) and rosette width at flowering ($P=0.02$). In chromosome five in addition to germination time ($P=0.001$) and rosette width at flowering ($P>0.0001$), the two

traits leaf number (P>0.0001) and rosette width at day 20 (P>0.012) proved to have significant different variations while the variance in other traits were uniform.

7.4.2 Calculating genetic components of means and potence ratios

The calculated genetic components of means and the potence ratios based on the three population means for each chromosome for every trait are presented in table 7.2. The dominance genetic components of the means were greater than the additive genetic components of the mean for several traits indicating (pseudo) over-dominance when each chromosome is considered as a whole.

The dominance components of means were greater than the additive components of the means (in either direction) for the traits GT, LN-20, RW-20, HF, RLNF, CLNF and RWF in chromosome three. This was true with the traits RW-20, BT, HF, RLNF, CLNF and RWF in chromosome four and LN-20, RW-20, RLNF, CLNF and RWF in chromosome five.

7.5 Discussion and conclusions

7.5.1 Traits showing heterosis

Positive heterosis (F_1 mean higher than the mean of the higher parent) was observed for trait LN-20 in chromosome 05, RW-20 in chromosomes 03 and 05, HF in chromosome 03, RLNF in chromosome 05, CLNF in chromosome 03 and RWF in chromosome 05. Negative heterosis (F_1 mean lower than the mean of lower parent) was observed for RLNF in chromosome 04.

Out of the above traits, the population variances were not homogeneous for RW-20, LN20 and RWF in chromosome 05. As a result it can be concluded that heterosis has been confirmed for RW-20, HF and CLNF in chromosome three, RLNF in chromosome 04 and RLNF in

chromosome five. Out of these the heterosis observed for RLNF was observed to be lower than the lower parent value.

7.5.2 Inequality of population variances

The three populations P_1, P_2 and F_1 are all non segregating populations. Thus there are no any genetic parameters that contribute to the observed variation. Because all the three generations are grown under the same environment the variability that is observed among the populations should be equal provided that the experiment is adequately randomised and there is no interaction between the genotypes and the environment.

However, in this experiment the variation for germination time and rosette width at flowering proved to be non-homogeneous for the populations in all the three chromosomes. Germination time is a trait that has frequently been observed to show maternal effects, therefore, the variance differences in this trait can probably be attributed to maternal effects rather than the genotype environment interaction.

The trait RWF, which showed non-homogeneous variance across all the three chromosomes tested may be due to the differential interaction of the genotypes with the environment.

Inequality in variances has been observed in populations of chromosome 05 for LN-20, RW-20 and as well. Although no definitive conclusion can be drawn as to the reasons for this, genotype–environment interaction can be suggested as a reason. It is also noteworthy that this inequality in variances for these two additional traits has been observed only in chromosome 05.

7.5.3 Genetic components of the means and potence ratios

The additive and dominant components of the means helps determining the presence or absence of dominance for the traits tested. Potence ratio gives an indication about the parent having the most dominant alleles and is therefore more potent in the cross. This ratio cannot be equated with

the true dominance ratio that can be obtained only when dominance is uni-directional and there is complete association of alleles in the parents.

When considering the increasing and decreasing effects of alleles, Col carried increasing alleles for LN20, RW-20 and HF while L*er* alleles had increasing effects on the traits BT and flowering related traits RLNF and CLNF in chromosome three. Similarly the observation of the genetic components of the means and the potence ratio revealed that Columbia contained more dominant alleles in chromosome three for the traits LN-20. RW-20 and HF while, L*er* contained more dominant alleles for RLNF and CLNF.

In chromosome four, Col alleles had increasing effect on traits LN-20, andRW-20 while L*er* alleles had increasing effects on traits BT, HF, RLNF, CLNF and RWF. The very highly negative potence ratio of RW-20 in chromosome four is attributed to the very narrow difference between the parental means. In addition the potence ratio revealed that L*er* contained more dominant alleles for the traits HF, RLNF, CLNF and RWF.

With regard to chromosome five, Col alleles were increasing the trait in LN-20, RW-20, BT, FT and RWF and the potence ratio indicated that Col alleles were more dominant in traits LN-20, RW-20, BT. L*er* alleles were observed to be increasing the traits HF and RLNF in chromosome five. The very high potence ratio for the trait RLNF in chromosome five was due to the very narrow difference in parental means.

7.5.4 General comments

If the populations were available in F_1 (Col x Ler) and CSS1, CSS1- F_1 and CSS2- F_1 a comprehensive analysis on heterosis on each chromosome level and on the basis of the whole genome would have been carried out. Furthermore, in this experiment only three basic generations were available. Had there been more basic generations including segregating populations such as F_2 and backcrosses, far more detail could have been extracted from the

experiment. Also, the unavailability of any further basic generations beyond F_1 prevented tests for determining the presence or absence of complicating factors such as epistasis. Hence it was assumed that complicating gene interactions such as epistasis do not occur which may not hold true in all cases.

Hybrid vigour has been reported in *Arabidopsis* for height at flowering in a far more detailed study using sixteen basic generations of the same ecotypes Col and L*er* (Kearsey, Pooni, and Syed 2003). They had scored more or less the same traits but had reported the findings on a whole genome basis so that the results cannot be strictly comparable. In the current experiment we have observed heterosis for more traits but on an individual chromosome basis with a fewer number of generations. But, as the results of the current study indicate, there are positive and negative effects of different chromosomes on the same trait. The whole genome analysis will reveal the net effect of all the effects. This stresses the importance of studying heterosis at a more micro level so that the breeding programmes can be designed to keep the desirable effects and remove disadvantageous effects.

Syed and Chen (2004) used Col and L*er* RILs population in explaining heterosis in *Arabidopsis* through molecular marker genotypes, heterozygosity and genetic interactions. They scored similar morphological traits and observed high F_1 performance when the RILs were back crossed to Col or L*er* irrespective of L*er* being the low performing parent. However, in this study the performance of RILs for most of the traits remained within mean values of the two parents ruling out dominance complementation.

The use of CSS in the study of heterosis helps investigating the phenomenon of hybrid vigour at a more micro level rather than the whole genome level. This would help in resolving so far unanswered questions about the genetic basis of heterosis. A better experiment with a few more

basic generations as discussed above will be of immense value in order to broaden the knowledge regarding the important phenomenon of heterosis.

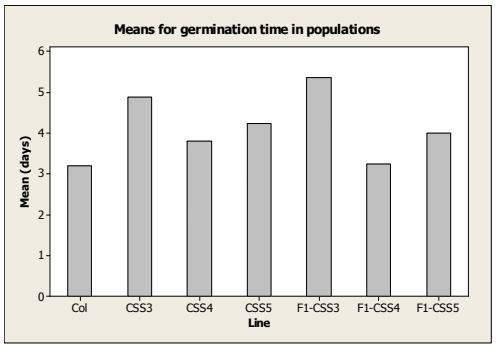
Table 7.1: Means and the standard errors of the means (in parenthesis) of each line for each of the traits scored. The means of the F_1s that exceed the parents, in either direction, are underlined while those showing significant heterosis are underlined and given in bold letters.

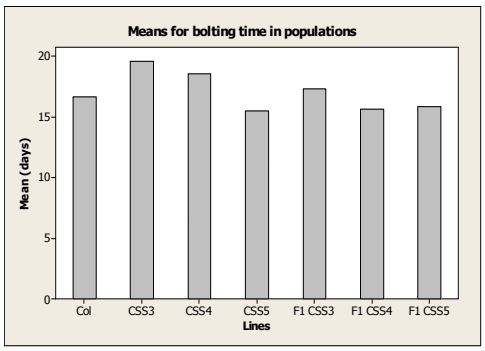
Line	GT	LN-20	RW-20	BT	FT	HF	RLNF	CLNF	RWF
Col	3.200	13.420	47.490	16.600	21.170	49.080	22.080	7.000	73.560
	(0.100)	(0.447)	(2.053)	(0.252)	(0.310)	(4.630)	(1.230)	(0.510)	(4.560)
CSS3	4.750	12.880	44.640	19.545	23.890	40.550	23.730	8.364	84.910
	(0.296)	(0.445)	(2.045)	(0.314)	(0.453)	(2.780)	(1.470)	(0.509)	(3.840)
CSS3-F1	4.864	14.580	**55.660**	17.318	21.900	**61.710**	26.120	**9.941**	91.710
	(0.249)	(0.463)	(2.128)	(0.304)	(0.367)	(3.920)	(1.110)	(0.433)	(3.800)
CSS4	3.545	13.110	47.370	18.522	24.37	67.000	24.000	11.250	85.820
	(0.157)	(0.441)	(2.028)	(0.416)	(0.378)	(5.670)	(1.310)	(0.413)	(2.900)
CSS4-F1	3.208	13.390	46.200	15.640	21.900	48.080	**16.380**	6.160	65.460
	(0.104)	(0.445)	(2.946)	(0.416)	(0.367)	(3.930)	(1.150)	(0.610)	(4.160)
CSS5	4.240	12.680	41.560	15.440	20.220	72.950	22.136	7.000	64.090
	(0.210)	(0.432)	(1.983)	(0.332)	(0.321)	(4.410)	(0.825)	(0.354)	(1.970)
CSS5-F1	4.000	**15.890**	**60.310**	15.826	20.290	68.090	**26.170**	8.261	**90.410**
	(0.225)	(0.440)	(2.022)	(0.241)	(0.311)	(5.540)	(1.130)	(0.480)	(2.500)

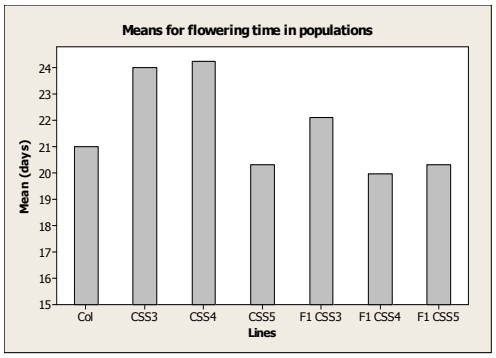
Table 7.2: The mid parental values, additive (a_A) and dominance (d_A) genetic components of means, and the potence ratio (d_A/ a_A) for each trait in each chromosome

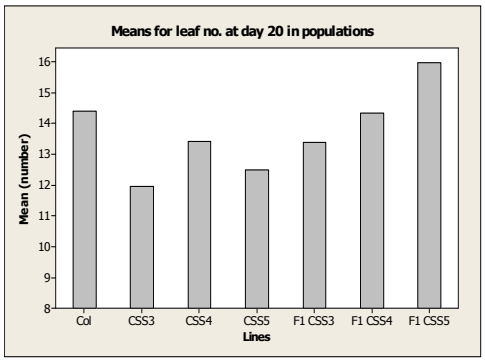
	GT	LN-20	RW-20	BT	FT	HF	RLNF	CLNF	RWF
Chr.03									
m	3.985	13.150	46.065	18.075	22.530	44.815	22.905	7.682	79.235
a_A	-0.775	0.270	1.425	-1.473	-1.360	4.265	-0.825	-0.682	-5.675
d_A	0.889	1.430	9.595	-0.754	-0.630	16.895	3.215	2.259	12.475
d_A/ a_A	-1.147	5.296	6.733	0.512	0.460	3.960	-3.897	-3.312	-2.198
Chr.04									
m	3.373	13.265	47.430	17.561	22.770	58.040	23.040	9.125	79.690
a_A	-0.172	0.155	0.060	-0.961	-1.600	-8.960	-0.960	-2.125	-6.130
d_A	-0.165	0.125	-1.230	-1.921	-1.325	-9.960	-6.660	-6.465	-14.23
d_A/ a_A	0.954	0.806	-20.50	1.999	0.828	1.112	6.938	3.042	2.321
Chr.05									
m	3.720	13.050	44.525	16.020	20.695	61.015	22.108	7.000	68.825
a_A	-0.520	0.370	3.165	0.580	0.475	-11.94	-0.028	0.000	4.735
d_A	0.280	2.840	15.785	-0.194	-0.405	7.075	4.062	1.261	21.585
d_A/ a_A	-0.538	7.676	4.987	4.653	-0.853	0.593	145.07	0.000	4.559

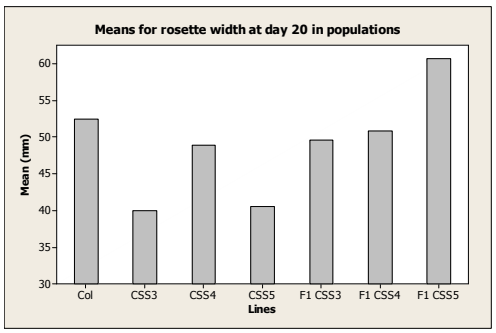
Figure 7.1 Graphical representation of the means of different populations for each trait

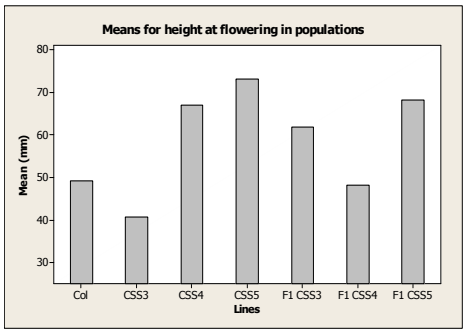


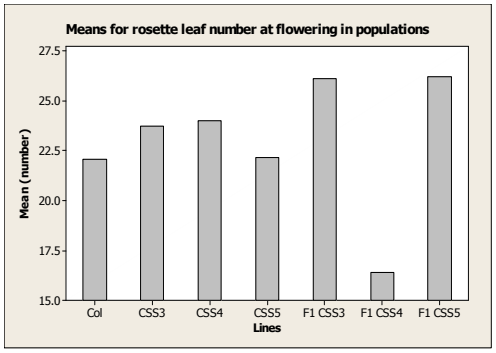


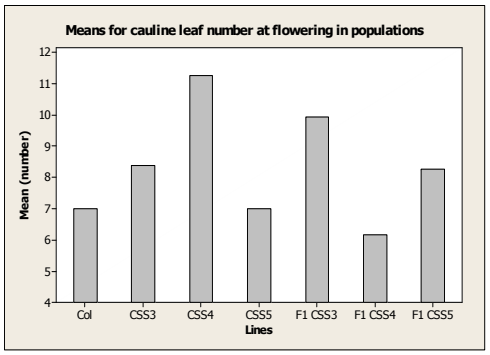


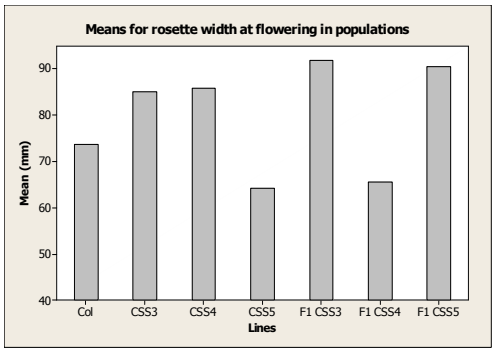







Means for rosette leaf number at flowering in populations
Mean (number)
27.5
25.0
22.5
20.0
17.5
15.0
Col
CSS3
CSS4
CSS5
F1 CSS3
F1 CSS4
F1 CSS5
Lines


Means for cauline leaf number at flowering in populations
Mean (number)
12
11
10
9
8
7
6
5
4
Col
CSS3
CSS4
CSS5
F1 CSS3
F1 CSS4
F1 CSS5
Lines


Means for rosette width at flowering in populations
Mean (mm)
90
80
70
60
50
40
Col
CSS3
CSS4
CSS5
F1 CSS3
F1 CSS4
F1 CSS5
Lines

CHAPTER 08

TRANSCRIPTIONAL PROFILING OF TWO NARROW STAIRS DIFFERING FOR FLOWERING-TIME WITH MICROARRAY DNA CHIPS

Abstract

DNA microarray biochip technology facilitates the genome-wide study of gene expression in a single experiment. The availability of the whole genome sequence and data makes *Arabidopsis* a very good candidate to be used in such transcriptional profiling experiments. The current research describes the gene expression study using two narrow STAIRS that differ by a maximum of 2 cM and also polymorphic for flowering time. Two experiments were conducted to observe the differential expression of the genes. The first compared the two STAIRS at the same chronological age before bolting and the second to compare the two STAIRS at more or less the same physiological age 4 days before mean bolting time. Respectively 190 genes and 586 genes were significantly differentially expressed in experiments 01 and 02. A possible candidate gene for flowering time, which is present within the polymorphic region of the 2 lines, was differentially expressed in experiment 02. The results also indicated that more genes are differentially expressed in the two lines at the same physiological age compared to the same chronological age.

8.1 Introduction

8.1.1 Microarrays

Microarrays are miniature arrays of gene fragments attached to glass chips. These biochips consist of an orderly arrangement of tens to hundreds of thousands of unique DNA molecules (probes) of known sequence. Microarrays are used to examine gene activity and identify gene mutations, using a hybridization reaction between the sequences on the array and a fluorescent sample. After hybridization the chips are read with high-speed fluorescent detectors and the intensity of each spot is quantified. The location and the intensity of each spot reveal the identity and amount of each sequence present in the sample. The data are then mined and modelled using the tools of computational biology. Because thousands of gene fragments can be present on a single microarray, data for entire genomes can be acquired in a single experiment (Schena and Davis 2000).

8.1.2 Implications of microarrays

A number of array-based technologies have been developed over the last few years. Out of these, DNA arrays designed to determine gene expression levels in living cells have received the most attention. DNA arrays allow simultaneous measurements of thousands of interactions between m-RNA-derived target molecules and genome derived probes and thus are rapidly producing enormous amount of raw data never before encountered by biologists. At the most basic level, DNA arrays provide a snapshot of all the genes expressed in a cell at a given time. Gene expression is the fundamental link between genotype and phenotype and as a result DNA arrays are bound to play a major role in understanding biological processes and systems ranging from gene regulation, through development to evolution, and to diseases from simple to complex (Baldi and Hatfield 2002).

DNA arrays are also playing a significant role in the annotation of gene function, which is a fundamental task of the genomics era. The identification of the function of a gene or protein depends on many things including structure, expression levels, cellular localization and functional neighbours that are often co-regulated in a biochemical pathway. Thus, establishing the function of new genes cannot depend on sequence analysis alone but requires considerable additional sources of information including phylogeny, environment, molecular and genomic structure and metabolic and regulatory networks. The DNA array data need be integrated with sequence data, structure and function data, pathway data, phenotypic and clinical data etc. for combining and correlating the diverse data sets along multiple dimensions and scales. Basic research in bioinformatics must deal with these issues of systems and integrative biology in a situation where the amount of data is growing exponentially (Baldi and Hatfield 2002).

In the genomic era it is anticipated that DNA array technologies will assume an increasing role in the investigation of evolution, directly by the study of mRNA levels in organisms that have faster generation times and indirectly by giving a better understanding of regulatory circuits and their structure. On the medical side, DNA arrays will help obtain a better understanding of complex issues concerning human health and disease, creating new diagnostic tools.

Until recently, molecular biologists have predominantly concentrated on single-gene/single-protein studies and cell biologists have been studying mRNA and protein levels during development via *in situ* hybridization that allows the study of expression of individual genes in multiple tissues. As we enter into the genomics era the basic paradigm is shifting from the study of single variable systems to the study of complex interactions. This shift is greatly assisted by the DNA arrays because this technology provides the facility to follow the expression levels of all the genes in the cells of a given tissue at a given time.

However, while the DNA array will be an important tool in achieving many goals in biological and biomedical sciences, it needs to be emphasized that this technology is still at an early stage of development. To a certain extent it is still cluttered with heterogeneous technologies and data formats as well as basic issues of noise, fidelity, calibration and statistical significance, issues which are being sorted out (Baldi and Hatfield 2002).

8.1.3 Microarrays and gene expression

Over the past decade, methods for studying the expression of many genes in parallel have become increasingly sophisticated and high throughput in nature. For *Arabidopsis*, techniques that have been used have included membrane-spotted macro-arrays, cDNA glass micro-arrays, cDNA-AFLP (amplified fragment length polymorphism), SAGE (serial analysis of gene expression), and oligonucleotide-based arrays (reviewed by (Aharoni and Vorst 2002; Donson et al. 2002)).
With the development of array technology, the use of arrays in the medical field especially with relation to research on various aspects of cancer studies was started (Chung et al. 2001; Graeber and Eisenberg 2001; Gerhold, Sellix, and Freeman 2002).

8.1.4 Microarrays in *Arabidopsis*

Arabidopsis with its availability of considerable amount of sequence information, even before the completion of the whole genome sequence, became a very good candidate for the application of DNA array based technologies. (Hanano et al. 2002) analysed the gene expression in *Arabidopsis* by array hybridization with genomic DNA fragments aligned along chromosomal segments for functional genomic studies. One well-proven oligonucleotide-based array technology for *Arabidopsis* is that established by Affymetrix in which each gene on the array is represented by a set of oligonucleotides (termed probes) that are tiled across the gene along with a corresponding set of single-mismatch probes. Before the completion of the *Arabidopsis* genome sequence

Affymetrix (in collaboration with Syngenta) developed a custom array representing about 8000 genes (AG array; Zhu and Wang, 2000) and, after the completion of the genome sequence, produced Affymetrix whole genome array (ATH1 gene array) which was more robust and reproducible than the comparable spotted cDNA arrays (Redman et al. 2004).

With the availability of various *Arabidopsis* arrays several groups started transcription profiling expression analyses on various themes. (Seki et al. 2002) reported the monitoring of expression profiles of 7000 *Arabidopsis* genes under, cold and high-salinity stresses using a full length cDNA microarray. In another study the orchestrated transcription of key pathways by the circadian clock in *Arabidopsis* was explored using high-density oligonucleotide microarrays to examine more than 8000 genes on the array (Harmer et al. 2000). After the completion of the sequencing of the whole genome of *Arabidopsis,* (Redman et al. 2004) described the development and evaluation of an *Arabidopsis* whole genome (Affymetrix) probe array representing approximately 24000 genes. (Hanano et al. 2002) applied genomic DNA array technology to analyse transcripts produced by 10 different regions of *Arabidopsis* chromosome five that together encompass 819 kb. They have applied a gene prediction algorithm and sequence similarity search and identified 249 putative protein-coding regions in these regions of chromosome five. Gene expression profiling of the tetrapyrrole metabolic pathway in *Arabidopsis* with a mini array system was reported by (Matsumoto et al. 2004) Expanding the array technology beyond *Arabidopsis,* (Taji et al. 2004) reported the studies on comparative genomics in salt tolerance between *Arabidopsis* and *Arabidopsis* related halophyte salt cress using *Arabidopsis* microarray.

8.1.5 Design of arraying experiments

A single set of microarray experiments often generates a vast amount of data**.** The microarray data glut can be more easily managed by proper prior design of experiments. Microarray life

cycle should begin with a clearly defined biological question and proceed to sample preparation, microarray reaction, microarray detection, and data analysis and modelling to form a new biological question (Schena and Davis 2000). (Yang and Speed 2002) described in length the design issues for cDNA microarray experiments under three categories namely, scientific, logistic and other factors. Furthermore, (Churchill 2002; Kerr et al. 2002) stressed the importance of adapting proper experimental designs and discussed fundamental issues related to experimental design in order to get results amenable to statistical analysis that mainly apply to two colour data.

8.1.6 Basic steps in a microarray gene expression experiment

Sample preparation and labelling is the first step in a microarray gene expression experiment. The extraction of RNA from the tissue of interest should precede the construction of labelled cDNA from the extracted RNA. The labelling method generally used at present is fluorescent labelling with the two dyes Cy3 (excited by a green laser) and Cy5 (excited by a red laser). Most commonly two samples, one labelled with each dye are hybridised to the arrays to allow the measurement of both the samples. The most common method of making labelled cDNA is direct incorporation by reverse transcriptase by priming the mRNA with a poly-T primer.

The next step in expression analysis is the hybridisation in which the DNA probes on the glass and the labelled DNA (or RNA) target form heteroduplexes via Watson-Crick base-pairing. This procedure is affected by conditions such as temperature, humidity, salt concentrations, formamide concentrations and the volume of target solution etc.

Hybridisation is followed by slide washing, firstly to remove excess hybridisation solution from the array and secondly to increase the stringency of the experiment by reducing cross-hybridisation. Washing can be done either in a low salt solution or at a high temperature.

Image acquisition is the final step of the laboratory procedure in a microarray gene expression experiment. The slide is placed in a scanner and the heteroduplexes in the array that contain the

fluorescent dyes are obtained in an image using appropriate light wavelengths. However, there are important considerations in scanning such as pixel size, spot size of the laser etc (Stekel 2003).

8.1.7 Microarray expression data analysis

8.1.7.1 Image processing

The first step of data analysis is the conversion of the image into numerical information using computer algorithms known as 'feature extraction software'. In some of the tailor made arrays, such as Affymetrix platforms, the image processing algorithms have been integrated into the genechip experimental process. However, in pin-spotted arrays the end user has a wide range of choices as to how to process the image including feature extraction, identifying feature positions, identifying feature-comprising pixels etc. (Stekel 2003).

8.1.7.2 Normalisation

The next step in the analysis of array data is the normalisation which is aimed at resolving the systematic errors and bias introduced by the microarray experimental platform. Normalisation procedures begin with data cleaning and transformation.

The microarray data, which are generated by feature extraction software, are in the form of text files. To ensure the high quality of the data, data cleaning and transformation such as removing flagged features (such as bad features, negative features, dark features and manually flagged features) or re-examining the flagged features need be done. Background subtraction is the second step in data cleaning. The background signal represents the non-specific hybridisation of labelled target to the glass and the natural fluorescence of the glass itself. Data transformation algorithms convert the raw intensities to log-intensities before proceeding with the analysis. The objectives of this practice are to get a reasonably even spread of features across the intensity

range, to get constant variability at all intensity levels, to make the experimental errors approximately normal and to make the distribution of intensities approximately bell shaped.

The second step, which applies only to two colour data, is intra-array normalisation. This allows a valid comparison of Cy3 and Cy5 channels, by eliminating four sources of systematic bias (1) differential incorporation of the two dye labels into DNA of different abundance, (2) the different emission responses of the two dyes to the excitation laser, (3) differential measurement of the two colours by the photomultiplier and (4) the level differences in the slide itself. The removal of systematic bias is achieved by various approaches: linear regression of Cy5 against Cy3; linear regression of log ratio against average intensity and non-linear (lowess) regression of log ratio against average intensity; and by two dimensional or block by block lowess regression to correct spatial bias.

The last step of normalisation is the 'inter-array normalisation' to compare samples hybridised to different arrays. Box plots can be used to visualise several distributions simultaneously and scaling, centering and distribution normalisation methods can be used to make data from different arrays comparable (Stekel 2003).

8.1.7.3 Analysis of differentially expressed genes

The analysis of differential gene expression of genes helps to identify relationships between genes or samples and facilitates classification of genes based on expression measurements. Most of the analytical tools are based on fundamental concepts such as statistical inferences, hypothesis testing, P values and independence testing. Statistical analysis methods such as Classical *t* tests, non parametric statistics (clustering, principal component analysis), ANOVA and general linear models etc. can be used to perform comprehensive analysis of expression array data (Stekel, 2003).

8.1.8 Gene expression of two STAIRS differing in a small chromosome region

Two Single Recombinant Lines (SRLs) known to differ in a small length of chromosome can be used in a microarray experiment to observe the differential expression of the genes in that region. During this investigation, two SRLs were identified that differ by a maximum of 2 cM and which carry QTL for flowering time and several other related traits. It is possible to examine the differential expression of the genes at any given time in the two lines by microarray gene expression analysis using these two SRLs. Furthermore, the expression pattern of the candidate genes of flowering time related QTL can be studied in such an experiment.

8.2 Objectives

- To grow plants of the two closest SRLs from STAIRS that differ for flowering time in order to collect tissue samples at various time points.
- To compare the expression of genes of the two closest SRLs, at the same chronological age, which differ for the trait flowering time.
- To compare the expression of genes of the two lines at the same physiological age, i.e. on average when the flowering time genes are activated for both the lines.
- To obtain a gene list in the region 2-3 cM of chromosome three and observe the expression of these genes and the candidate genes in the two SRLs.

8.3 Materials and Methods

8.3.1 Growing of plants for sample collection

Four hundred and twenty plants of the STAIRS no. 01 (which is early flowering) and 630 plants from the STAIRS no.02 (which is late flowering) were grown in a completely randomised design surrounded by a guard row. These 2 STAIRS were selected based on the results of the QTL

analysis experiment in narrow STAIRS. STAIRS 1 and 2 were the genetically closest lines yet differed for the trait flowering time. The polymorphic region for these two lines was only a maximum of 2 cM within the top region of chromosome 3.

The plant growth conditions are the same as described in chapter 04. An additional 840 *Arabidopsis* plants were grown at the same time as replacements in order to provide a similar environment to all the plants by filling the gaps created by the removal of plants for sampling. One plant in every pot that germinated within a period of 48 hours was kept and the others were discarded.

8.3.2 Sample collection

The sampling was started from the 12^{th} day from sowing and continued at daily intervals for 21 days. Plants of SRLs no. 01 and 02 (early-flowering and late flowering respectively) were sampled from 12-25 days and 12-32 days respectively. From each line 30 plants were sampled at each time point in three replicates comprising 10 plants each.

The above-ground part of each plant was sampled as a whole and the 10 plants in each replicate were collected separately. The samples were frozen in liquid nitrogen immediately after harvesting and stored in a freezer at -70^0C.

8.3.3 Determination of time points for studying gene expression

The two lines that were used for sample collection differed for the trait flowering time. The transcription profiling experiment was planned with a view to observing differential gene expression for this particular trait along with other gene expression differences. No data related to the age of the plant at which the particular gene is expressed were available. Therefore, it was decided to use the samples 4 days before the mean bolting time for the early flowering line and

compare the late flowering line at the same chronological age in experiment 01. A second experiment was planned to compare the 2 lines more or less at the same physiological age, namely 4 days before the mean bolting time for each line.

8.3.4 Extraction and quantifying total RNA

Total RNA was extracted from the samples of both the lines 21 days after sowing and samples of line two 29 days after sowing using the Qiagen RNeasy Plant Mini Kit. Not more than 50 mg of plant material were used in each extraction, as overloading the columns would significantly reduce both the yield and quality (Qiagen 2001). The method of extraction followed that described in the Qiagen RNeasy Plant Mini Kit Handbook (Qiagen 2001) with a few exceptions regarding pooling samples of the same replicate at the homogenised lysate stage. Because there were 10 plants per replicate and 3 replicates per line at each time point, the extraction of RNA from each plant separately would have involved a very large number of sample extractions. Pooling replicate samples together at a later stage would have followed this. Thus, in order to reduce the number of samples to be extracted, pooling equal quantities of material from five plants within a replicate was practised at the homogenised lysate stage. The protocol of total RNA extraction is given in appendix 07.

The extracted total RNA samples were tested using RNA nano LabChip kit on an Agilent 2100 Bioanalyzer (Agilent technologies, 2004) to assess the concentration and quality of the RNA that had been extracted. RNA quality assessment has been identified as one of the most critical elements in order to obtain meaningful gene expression measurements. Gene expression can only be measured if intact RNA is used for the experiment.

8.3.5 The MWG oligonucleotide array

A set of *Arabidopsis* oligonucleotides were obtained from MWG and the arrays were printed in-house at The University of Birmingham by Dr. Timothy Wilkes and Dr. Rachel Machado. The MWG *Arabidopsis* oligonucleotide array includes 24,960 different 50-mer oligonucleotides consisting of distinct features of all the known open reading frames. Each array slide has the whole set of oligonucleotides printed in duplicate across 48 sub-arrays. In addition, there were 23 Lucidea scorecard controls and two landing lights (M13 universal primer labelled with Cy3) across each array, to clearly identify positions. Each array slide had a total of 51, 120 spots including all the oligonucleotides and the controls.

8.3.6 Experimental plan

In experiment 01 the two lines were compared on the same microarray slide at the same chronological age, which is 4 days before mean bolting time of the early-flowering line. The idea behind this was that the flowering time gene may likely be activated in the early-flowering line, while it may not yet be activated in the late-flowering line. Because there were 3 replicates for each line at each time point and Cy-3 and Cy-5 dye swap was adopted for each slide there were a total of 6 (3 replicates x 2 dye swap) for experiment 01.

In experiment 02 the early-flowering line at 4 days before the mean bolting time was compared with the late-flowering line at 4 days before the mean bolting time of that line. As with experiment 01, there were a total of 6 slides in experiment 02.

8.3.7 cDNA synthesis and labelling

8.3.7.1 First strand synthesis

Total RNA samples were converted to single stranded cDNA in 12 separate micro-centrifuge tubes relating to 12 samples for each experiment. Each tube contained 1 μg of total RNA from 2

samples per replicate, 2 µl of the relevant control spike (either test or reference - the control spike provides a method of assessing the success of the labelling reaction), 0.5 µl oligo dT-V (2 µg/µl) (Sigma-Aldrich) and ddH2O (DEPC/RNAase free water) to make up to a total volume of 24.5 µl. Each micro-centrifuge tube was incubated, in a PCR machine, for 10 minutes at 70 °C then transferred immediately to ice. Briefly after, 8 µl 5x Superscript buffer (Life technologies), 4 µl 0.1M DTT (Sigma Aldrich), 2 µl 10 mM dNTPs (Bioline Ltd.) and 1.5 µl Superscript II RT (reverse transcriptase enzyme) at 200 U/µl (Life technologies) was added to each tube, mixed thoroughly and incubated at 42 °C for 1 hour. Then, 0.5 µl RNaseH at 2 U/µl (Promega) was added and mixed by flicking and pulsing briefly. Samples were then incubated at 37 °C for 30 minutes using a PCR machine.

8.3.7.2 First strand clean-up

The Qiagen QU7 Quick PCR purification kit (Qiagen 2002) was used to clean up each sample after the first strand synthesis. This protocol is designed to purify single- or double-stranded DNA fragments from PCR and other enzymatic reactions. 5 volumes of Buffer PB were added to 1 volume of the sample and mixed. To bind the DNA, the sample was applied to the centre of a QIAquick spin column and centrifuged for 60 seconds. The flow-through was discarded. 0.75 ml Buffer PE was added to the QIAquick column and centrifuged for 60 seconds. The flow-through was discarded and the QIAquick column was centrifuged for a further 60 seconds. The QIAquick column was then placed in a clean 1.5 ml micro-centrifuge tube. The DNA was eluted by adding 30 µl Buffer EB (10 mM Tris-Cl, pH 8.5) to the centre of the QIAquick membrane, left to stand for 1 minute and then centrifuged for 1 minute.

8.3.7.3 Second strand synthesis

Labelling reactions with the fluorescent labels Cyanine-3 (Cy3) and Cyanine-5 (Cy5) were carried out at the second strand synthesis stage. Cy-3 and Cy-5 labels are used together as they

possess reasonably high incorporation efficiencies with reverse transcriptase, good photo-stability and yield, and absorb and emit light at distinct and separable wavelengths (Aharoni, 2001). 28 µl of the first strand product, 4 µl of klenow buffer (Bioline), 1 µl of Hexanucleotide primer (3 mg/ml) (Sigma-Aldrich) were mixed in a micro-centrifuge tube and incubated at 90°C for 2 minutes, then left to cool to room temperature for 5 minutes. Following this 4 µl 10x dNTPs (low dCTP: all 0.25 mM, except dCTP which was 0.09 mM) (Bioline), 1 µl Cy3 dCTP (25 nMole) or Cy5 dCTP (25 nMole) and 2 µl Klenow (5 U/µl)(Bioline), were added to each reaction tube and incubated overnight at 37 °C in a PCR machine.

8.3.7.4 Second strand clean-up

Prior to the second strand clean up, the two reactions that were going to be used to hybridise the same slide (2 lines, one labelled with Cy3 and the other labelled with Cy5) were combined.

The volume in each tube was corrected to 100 µl, by the addition of TE buffer pH8.0. The reactions were then cleaned up using the Qiagen QIA quick PCR clean up kit (Qiagen) as described above (in Section 8.3.7.2) for first strand clean-up. All the samples were then dried in a speed vacuum for 20 minutes at 60 °C. The pellets were resuspended in 20 µl DEPC H_2O.

8.3.8 DNA microarray hybridisations

8.3.8.1 Pre-Hybridisations

Before hybridisation of the labelled samples to the arrays, the slides were incubated for 3 hours at 42 °C in pre-hybridisation solution (25% (v/v) formamide, 5x SSC, 0.1 % (w/v) SDS and 1% (w/v) BSA). When the slides were removed from the pre-hybridisation solution they were each dipped briefly in dH_2O, then in EtOH (Absolute), transferred to a 50 ml Falcon tube and dried by centrifugation at 2000 rpm for 10 minutes.

8.3.8.2Hybridisations

Before the samples were applied to the arrays, a probe cocktail was made up for each of them. This included (final concentration of each component in brackets following it), 10 µl sample (probe mix), 10 µl formamide (25 % v/v), 10 µl 20x SSC (5x), 0.4 µl 10 % SDS (0.1 % w/v), 8 µl poly A at 2 µg/µl (0.4 µg/µl), 1.6 µl yeast tRNA at 25 µg/µl (1 µg/µl), giving a total volume of 40 µl. This probe cocktail was mixed and heated to 95 °C for 3 minutes, then centrifuged for 30 seconds at 15000 rpm. Each probe cocktail was then pipetted one at a time, onto an array cover slip and an array slide dropped gently (array side down) onto the cover slip, in order to avoid air bubbles. Slides were then carefully placed on two toothpicks in a box containing a piece of tissue paper (dampened with 5x SSC). These boxes were then wrapped in foil, as they were light sensitive, and incubated overnight at 42 °C.

8.3.9 Slide washes

Care was taken to ensure that at each stage of the washing the falcon tubes containing the slides were kept in the dark, because of the sensitivity of the labelled probe to the light. Each slide needed to be washed in a low salt solution in order to remove all unincorporated hybridisation solution. Slides were washed individually in different 50 ml falcon tubes for each wash. Firstly slides were dipped into pre-warmed (42 °C) 2x SSC and 0.1 % SDS, until the cover slips dropped off and transferred to another falcon tube with the same wash solution for 5 minutes at 42 °C. Slides were then transferred to a fresh tube containing 0.1x SSC and 0.1 % SDS at room temperature for 5 minutes. This last wash step was repeated four times, each for 5 minutes at room temperature, with fresh wash solution each time. Each slide was then quickly dipped into two washes of SDW and then finally into EtOH (Absolute) once. Each array was then placed

into a fresh 50 ml falcon tube and dried by centrifugation at 2000 rpm, at 20 °C for 10 minutes. Slides were then kept in the dark at room temperature and scanned on the same day.

8.3.10 Array scanning and image analysis

Arrays were scanned using an Axon scanner, using GenePix 3.0 image software. A preview scan was first performed and the hardware settings were optimised at this stage. The hardware settings for each scan were saved individually as a .gps (GenePix Settings) file. Once the array has been scanned the image was saved as a 16-bit multi-image TIFF file.

The software uses the following definitions for image analysis. Each spot (oligonucleotide) is a feature, each spot is detected by a feature indicator and a collection of feature-indicators is termed a block. At the time of slide printing, a GenePix Array List (GAL) file was created. This was used to generate a grid of blocks to overlay the image of the array, so that each spot, and its respective intensity could be accurately identified. The features were aligned using the software followed by manual observation of each block to call any of the spots that had been missed. The results were then saved as gpr (GenePix Results) files.

8.3.11 Data analysis

GeneSpring (Silicon Genetics) software was used for the analysis of the expression data. The results of the processed images in the form of gpr files were loaded into GeneSpring and 2 experiments were set-up to analyze the 6 slides of each experiment separately.

Experimental parameters were set up as dye swap, developmental stage (as days before bolting) and chronological age (as days after sowing).

Normalisations of data were done using;

i) Data transformation : Dye swap

ii) Intensity dependent Lowess normalisation (with 40% smoothing and 0.01 cut off margin)

Quality control of data was performed using clustering condition tree and Pearson correlation and then using principal component analysis.

Slide 01 in experiment 01 and slides 09 and 12 in experiment 02 proved to have > 50 % of the intensity values zero or negative and also these slides clustered separately in quality control tests. Therefore, they were eliminated from further analysis in order to achieve reliable results. Because the experiment was replicated sufficiently the removal of these slides did not have a detrimental effect on the analysis. After the removal of these slides, the intensity dependent Lowess normalization could be successfully performed for both of the experiments.

A default experiment interpretation was set up using log of ratio, with stage of development as the non-continuous parameter. A ratio (signal/control) interpretation was also set up. Filtering was performed on the all genes (minus control first, to determine those genes that were present or marginal in at least 2 out of 5 slides in experiment 01, and 2 out of 4 slides in experiment 02. Filtering on error was then done to obtain the lists of genes whose expression falls within 2 standard deviations of the mean. The genes which passed these filters were considered to be 'reliable genes'. Another filter, based on expression levels, was then performed on these 'reliable genes' to remove non-changing genes. A final filter was performed on the changing 'reliable gene' list to obtain a gene list based on a confidence limit of 0.1. The final gene lists were obtained using the TAIR database annotated descriptions.

8.3.12 Obtaining a list of genes within the region the two lines differ

The maximum region for which the 2 lines used in the arraying experiment could differ was approximately 2 cM at the top of chromosome 3 between the AGI clone F16B3 and T17B22. A list of all the open reading frames present within this region was obtained from the TAIR website. This gene list was compared with gene lists obtained from the expression analysis to determine those genes differentially expressed between the 2 lines at the particular time points the analysis was carried out.

8.4 Results

8.4.1 Gene lists at different stages

8.4.1.1 Experiment 01

The MWG microarray slides contained 24960 genes excluding controls. When the all genes (minus controls) list was filtered, 17036 genes were present or marginal on at least 2 slides out of 5 (according to the flag information). When the above list of 17036 genes was filtered to identify those genes whose expression levels fell within 2 standard deviations of the mean, 6892 genes passed the filter. These were named the 'reliable genes' and further filtering on expression level, to remove the non-changing genes from the reliable genes list, resulted in 3977 genes out of 6892 passing the filter. At the final filter, using a confidence interval of 90%, a total of 190 genes out of 3977 passed the filter.

8.4.1.2 Experiment 02

The initial filter applied to the genes in experiment 02 showed that 16114 out of 24960 genes were present or marginal on at least 2 out of 4 slides (according to the flag information). The second filter, to obtain genes within 2 standard deviations of the mean, resulted in 10118 genes out of 16114 passing the filter. In the subsequent filtering step, to remove the genes that were non-changing between the two lines from the reliable genes, 4573 genes out of 10118 passed the

filter. Filtering using a confidence interval of 90% revealed that 586 genes were differentially expressed between the 2 lines.

8.4.2 Differentially expressed genes

Two genes out of a total of 190 (Expt. 01) and 3 genes out of a total of 586 (Expt. 02) differentially expressed genes lie within the polymorphic region of the two lines. A comparison of the number of differentially expressed genes in relation to the length of the chromosomes and the number of genes present in each chromosome are given in table 8.1. In experiment 01 the two differentially expressed genes were a gene coding for aspartyl protease family protein (TIAR accession number At3g02740; p=0.042) and a putative UTP glucose – phosphate uridylyl transferase gene (TAIR accession number At3g03250; p=0.0226). The 3 genes differentially expressed in the experiment 2 were a putative calmodulin gene (TAIR accession number At3g03000; p=0.0161), a gene coding for aspartyl protease family protein (TAIR accession number At3g02740; p=0.042) and zinc finger protein CONSTANS-LIKE 2 (COL2) which is identical to putative flowering-time gene CONSTANS (TAIR accession number At3g02380). The two differentially expressed genes lists are given in appendix 08.

8.4.3 Genes list in the polymorphic region and the candidate genes

There were a total of 100 loci and 126 distinct gene models in the polymorphic region of the 2 lines. A locus is defined as the genomic sequence corresponding to a transcribed unit (e.g. At2G03340) in the genome. A gene model is defined as any description of a gene product from a variety of sources including computational prediction, mRNA sequencing, or genetic characterization. In TAIR, many gene models can exist for a given locus, therefore a search for a gene may result in multiple hits for the same gene name (www.arabidopsis.org).

Observation of gene models among the 100 loci revealed a possible candidate gene for flowering time namely, CONSTANS-LIKE 2 (COL2) which is homologues to putative flowering-time gene CONSTANS (TAIR accession number At3g02380). Out of the 2 experiments, this candidate gene was 7 times down regulated in experiment 02. Apart from this particular possible candidate gene a few differentially expressed genes, which may affect flowering time, were observed in experiment 02. These include a CONSTANS-like protein-related gene which contains similarity to photoperiod sensitivity quantitative trait locus (Hd1) GI:11094203 from (Oryza sativa); similar to Zinc finger protein CONSTANS-like 15 (SP:Q9FHH8- *Arabidopsis thaliana*) TAIR accession no. At1g07050 (p=0.0116). A second gene, auxin-responsive family protein similar to indole-3-acetic acid induced protein ARG7 (SP:P32295) (Vigna radiata) (At1g20470; p= 0.0957) and auxin efflux carrier family protein (At1g71090; p=0.0899), may be affecting cell growth and expansion and so causing or affecting flowering.

The genes list within the polymorphic region, and the two lists of genes which are differentially expressed in the two experiments are given in appendix 09.

8.5 Discussion

8.5.1 Genes list in the polymorphic region for the two STAIRS

The genes list in the polymorphic region of the 2 STAIRS, which extends for approximately 2 cM, included 100 loci with 126 distinct gene models. This region included the maximum region the two lines can differ by, depending on the positions of the markers for which the lines were genotyped. The most recent figures indicate that about 30000 genes are present in the genome of *Arabidopsis* (www.arabidopsis.org). The sum of the length of 5 chromosomes in *Arabidopsis* is 597 cM (www.arabidopsis.org) and, if it is assumed that the genes are uniformly distributed throughout the whole length of 5 chromosomes of *Arabidopsis,* there are approximately 50 genes

per cM. Therefore 100 loci within a 2 cM length of a chromosome is the expected number of loci within this region. The candidate gene list is given in appendix 08.

8.5.2 Differentially expressed genes

Experiment 01 of transcript profiling tested the 2 lines at the same chronological age (that is 4 days before mean bolting time for the early line). Experiment 02 tested the 2 lines 4 days before the mean bolting time for each of the lines (that is, at the same physiological age). Because there was no prior information about when the genes relating to flowering time were activated, 4 days prior to mean bolting was tested on a random basis to observe whether the flowering time candidate genes that are present within the region are activated at this time point. The time points of the later experiment were intended to be a comparison of the gene expression of the 2 lines at the same physiological age. However, this may not strictly hold true because the same chronological number of days before mean bolting for the 2 lines may not be the same physiologically.

In experiment 01 there were 6892 reliable genes, of which 3977 (58%) were differentially expressed between the two lines and a total of 190 genes (2.7%) were expressed under 0.1 confidence limits. In experiment 02, the number of reliable genes was 10118, of which 4573 (45%) were differentially expressed and 586 (5.8%) were expressed under 0.1 confidence limits. These numbers indicate that at each stage there were more genes expressed and/or differentially expressed between the 2 lines in experiment 02. In experiment 02 both the lines were compared at more or less at the same physiological age and from the results it is apparent that this is a period where the differential expression of genes related to various biochemical and physiological processes can be observed.

Of the differentially expressed genes, all but 2 genes in experiment 01 and 3 genes in experiment 02 lie outside the polymorphic region of the 2 lines. The genetic theory explains that the genes do

not act in isolation. There are various interactions among genes and the activation of some may be triggered by others, either on the same or on separate chromosomes, which is explained by the *cis* and *trans* regulation phenomenon of genes. Thus, genes within a certain region in a chromosome can trigger the expression of genes elsewhere in the chromosomes in various bio-chemical and physiological pathways. Therefore, it is possible that some of the differentially expressed genes outside the polymorphic region of the 2 STAIRS are affected by the genes within the polymorphic region. An understanding of the action and interaction of genes and gene function are needed to unravel such complicated interrelationships but there was insufficient time during this project to explore these issues any further.

The proportion of differentially expressed genes is higher within the polymorphic region of the 2 STAIRS than elsewhere in the genome although it is a small number, 2 and 3 genes respectively in experiments 01 and 02. In experiment 1 this proportion is well over 2 fold while it is somewhat less in experiment 02. However it is not possible to count the number of genes located outside the polymorphic region but whose expression is affected by the genes within the polymorphic region.

8.5.3 Gene expression of the possible candidate gene

The possible candidate gene which is related to flowering-time gene CONSTANS was not included in the list of genes which are differentially expressed in experiment 01. This indicates that if in fact this is the gene affecting flowering time within the region of interest, it is not activated at the time point 4 days before the mean bolting time of the early line. The differential expression of this gene in experiment 02 indicates that the 2 lines are not physiologically the same in terms of the activity of this gene at the same number of days prior to bolting. However, it is very interesting to note that this possible candidate gene appeared in the list of only 3 genes which are differentially expressed and also lie within the polymorphic region. The evidence

suggests that this particular candidate gene is expressed in the early-flowering line 4 days before the mean bolting time for that line.

Out of a total of 100 loci within the polymorphic region of the 2 STAIRS used in the current experiment, there can be other candidate genes affecting flowering time, for which the function has not yet been identified. In such a situation, the expression of such genes cannot be associated with the flowering time from the data of this study alone.

8.5.4 Suggestions for further research

When a candidate gene affecting a particular trait has been identified, RT-PCR techniques can be applied to determine the expression of such genes (Petersen *et al.* 2003). This will be especially relevant in the current research to identify the time of the activation of the possible flowering-time candidate gene COL2. The plant tissue samples are stored at -70^0C from all the stages of plant growth and total RNA can be extracted from these samples for use in RT-PCR. Due to time limitations this was not done during this investigation, but it is a possible way of identifying the time of the activation of the relevant gene and also the time related polymorphism of the gene.

Table 8.1: The map lengths of each chromosome in *Arabidopsis*, total no. of genes present and the nos. of differentially expressed genes, the ratio of the map length to the no. of genes differentially expressed and the percentage of differentially expressed genes from the total number of genes present with comparison of the polymorphic region between the two STAIRS.

Exp.	Chrm.	Map length (cM) (A)	Total no. of genes	No. of genes differentially Expressed (B)	B/A	genes expressed as a % from total
1	1	135	6325	58	0.429	0.917
1	2	97	4089	22	0.226	0.538
1	3	101	5023	35	0.346	0.696
1	4	125	3844	29	0.232	0.754
1	5	139	5679	35	0.251	0.616
	unknown	-	-	11	-	-
	Total	597	-	190	0.318	0.77612
	poly.region	2	100	2	1.000	2
2	1	135	6325	147	1.088	2.324
2	2	97	4089	81	0.835	1.981
2	3	101	5023	104	1.029	2.070
2	4	125	3844	89	0.712	2.315
2	5	139	5679	131	0.942	2.306
	unknown	-	-	34	-	-
	Total	597	-	586	0.981	2.348
	Int.region	2	100	3	1.500	3

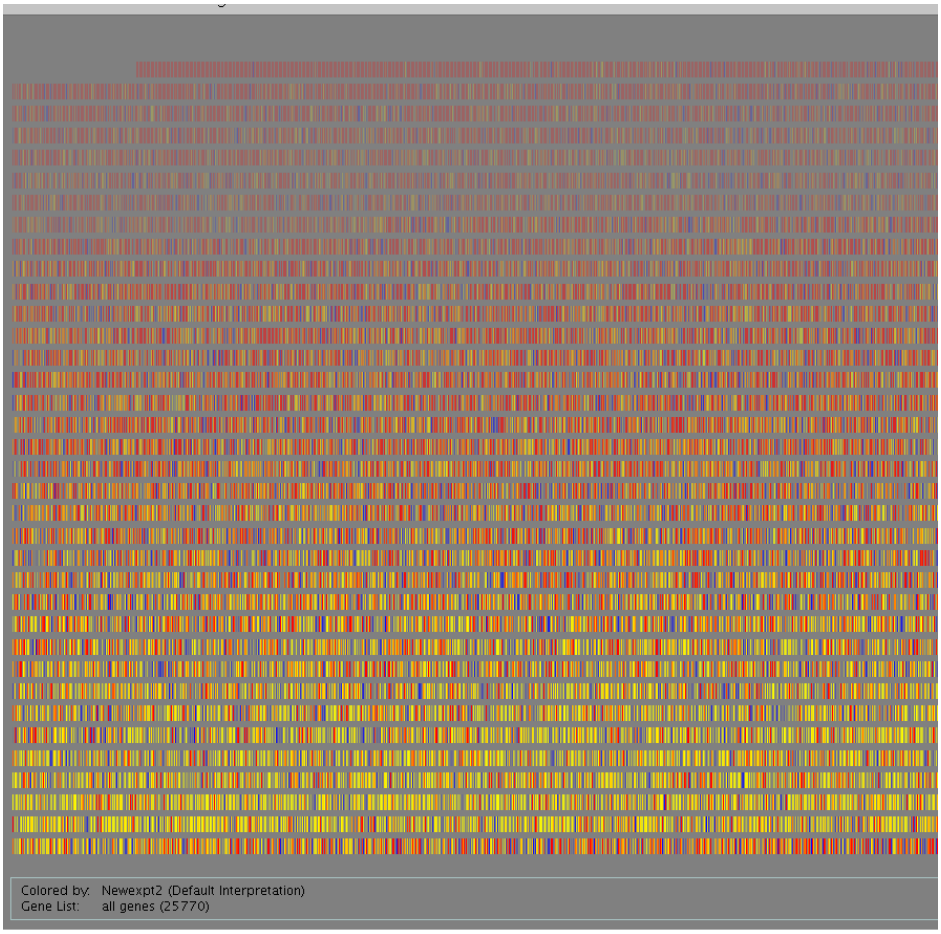


Figure 8.1 The GeneSpring image of all the genes present in the expression analysis experiment 01.

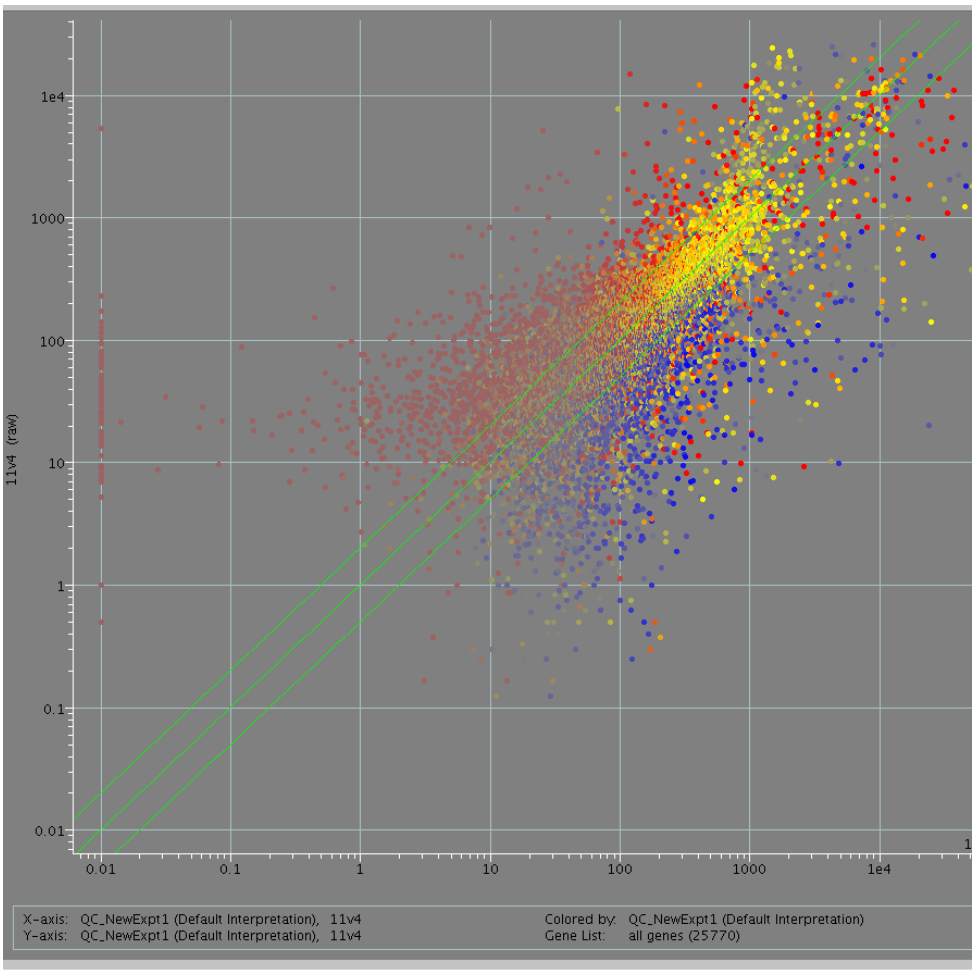


Figure 8.2 The GeneSpring scatter plot image of all the genes present in the expression analysis experiment 02.

CHAPTER 09

SUMMARY AND FUTURE WORK

The first part of this research focussed on the production of a relatively novel resource population called STepped Aligned Recombinant Inbred Strains (STAIRS) for fine mapping of quantitative trait loci in *Arabidopsis thaliana.* Following the findings of Koumproglou et al. (2002) who located QTL related to flowering time on the top region of chromosome 3 using wide STAIRS (also confirmed in this project), the current project aimed at the production of narrow STAIRS within the top of chromosome 3.

STAIRS are produced via a backcross breeding programme using whole chromosome substitution lines. The breeding programme is assisted with molecular marker genotyping at all stages. Thus the marker saturation of the top region of the chromosome 3 was a necessity and this was achieved by the development of 24 polymorphic microsatellites markers within the top region of chromosome 3. A total of 23 narrow STAIRS spanning the top 20 cM were produced in the current breeding programme while over 400 heterozygous single recombinants which have been selfed to identify more narrow STAIRS also were identified. These STAIRS were assigned to marker demarcated bins with the microsatellites genotyping.

QTL analysis using narrow STAIRS revealed high resolution map location for QTL related to flowering time, leaf number at day 20 and rosette and cauline leaf numbers at flowering time. The major QTL for these traits were located within 2-3 cM while minor QTLs were located between 15-20 cM. It was possible to achieve a resolution as high as 1 cM for the main QTL that affected all these traits. In addition it was possible to distinguish pleiotropy of this particular QTL for flowering time and related traits from the alternative possibility of tight linkage among several genes. This analysis thus amply demonstrated the high resolution power of STAIRS in mapping QTL and also the possibility of unveiling some of the complicated gene actions.

The STAIRS produced are a universal and permanent resource in locating QTL related to various traits which are polymorphic between the ecotypes Col and L*er*. The seeds of these STAIRS will be publicly available through National Arabidopsis Stock Centre in the UK and already two research groups are using these STAIRS in locating QTL related to Carbon assimilation rates (USA) and leaf movement (The Netherlands) in *Arabidopsis*.

The search for candidate genes for flowering time within the region resulted in one possible candidate gene CONSTANS like gene, *COL2* (TAIR accession no. At3g02380). The transcriptional profiling experiments with DNA microarray technology using two genetically closest yet early and late flowering lines, revealed differentially expressed genes at two particular stages within the 2 lines. This included differential expression of the particular candidate gene also in one experiment.

During the last decade molecular genetic approaches have been applied to study and understand the control of flowering time in *Arabidopsis*. These studies indicate that multiple environmental and endogenous inputs regulate the flowering. The molecular genetic dissection of flowering time control in *Arabidopsis* has identified an integrated network of pathways that quantitatively control the timing of this developmental switch. Out of these, genetic pathways that describe the effects of cold temperature, photoperiod and light quality on flowering time have been defined. However, the molecular basis of other responses such as the promotion of flowering by age and ambient temperature are yet to be understood (Simpson and Dean 2002).

The acceleration of flowering by a long period of cold temperature is known as vernalization. A vernalization requirement is a reproductive strategy adopted by many species and bred into several crops to ensure they overwinter vegetatively and flower in the favourable conditions in spring. In *Arabidopsis*, this has been mapped as a monogenic trait with *FRI* alleles which code a

protein to promote the accumulation of *FLC* mRNA. This represses the floral transition to override the otherwise favourable conditions to flowering (Simpson and Dean 2002).

Photoperiod and light quality control is a fundamental feature of seasonal progression that affects flowering time. An acceleration of floral transition in response to long days has been observed in *Arabidopsis*. Duration of the day and night is measured by the circadian clock (oscillator) which controls many aspects of plant biology in addition to flowering time (Hayama and Coupland 2004). The link between the oscillator and flowering time is considered as *CONSTANS (CO)* which is a transcription factor with 2 B-box type zinc fingers. The analysis of mutant and transgenic plants strongly suggests that the regulation of *CO* expression and activity is important for photoperiod flowering (Putterill, Laurie, and Macknight 2004). It has been observed that loss-of-function *CO* mutants flower late in inductive long days but like wild type in short days, whereas ectopic overexpression of *CO* promotes early flowering independently of the day length.

In addition to being controlled by the clock, *CO* expression is also modulated by day length. In long day periods, *CO* mRNA abundance is high at the end and the beginning of the photoperiod, but in short days, peak *CO* abundance occurs in darkness. Therefore it has been suggested that CO may function in an output pathway that integrates day length perception and time-keeping mechanisms to promote flowering (Simpson and Dean 2002).

The autonomous pathway comprises of a group of 6 genes (*FCA*, *FY*, *FLD*, *FVE*, *FPA* and *LD*) and is related to the internal signals that regulate flowering. Mutants of the autonomous pathway are late-flowering in long days and short days. This can be overcome by vernalization or growth in far-red enriched light. Because the mutants are thus not defective in their ability to respond to day-length and vernalization signals, it has been suggested that they might respond to internal developmental signal (Putterill, Laurie, and Macknight 2004). Autonomous pathway components

normally function to limit the accumulation of *FLC* mRNA and which accounts for their late flowering mutant phenotype (Simpson and Dean 2002).

Recent work suggests complex regulatory mecahanism among the genes that affect flowering time and also the involvement of microRNAs in regulating flowering time. MicroRNAs are approximately 22 bp non-coding RNAs and they are negative regulators of gene expression in eukaryotes (Putterill, Laurie, and Macknight 2004).

The Arabidopsis gene *CONSTANS-LIKE 1*(*COL1*) and *CONSTANS-LIKE 2*(*COL2*) are predicted to encode zinc finger proteins with about 67% amino acid identity to the protein encoded by the flowering-time gene *CONSTANS*. Ledger et al. (2001) reported that the circadian clock regulates expression of *COL1* and *COL2* with a peak in transcript levels around dawn. They have analyzed the expression of *COL1* and *COL2* in transgenic plants and showed that both *COL1* and *COL2* had little effect on flowering time although the overexpression of COL1 shortened the period of two distinct circadian rhythms ahffecting light input pathways. Ledger et al. (2001) concluded stressing the need for establishing the functions of *COL1* and *COL2* more clearly.

In the current research a thorough analysis of the data obtained from the 2 microarraying experiments needs still be done. The possible links of the genes expressed in the expression analysis experiments to those involved in the known flowering time control pathways should be explored. Due to the time limitation this analysis was restricted to obtaining differentially expressed genes lists at the moment but the further analysis on individual gene basis should be carried out. Quantitative RT-PCR with specific primers can be performed to identify the time of gene expression of the particular candidate gene for flowering time. Furthermore, the findings from the microarray expression analysis experiments can be validated by QRT-PCR technique.

More genome wide gene expression analysis studies are possible with the stored tissues collected at many stages of growth of the plants from the 2 genotypically closest STAIRS which show

polymorphism for flowering time. These expression profiles of the whole genome will generate a vast amount of valuable information when combined with the data from various biochemical and physiological pathways.

REFERENCES

Abbott, R. J., and M. F. Gomes. 1989. Population Genetic-Structure and Outcrossing Rate of Arabidopsis-Thaliana (L) Heynh. Heredity **62**:411-418.

Aggarwal, R. K., V. V. Shenoy, J. Ramadevi, R. Rajkumar, and L. Singh. 2002. Molecular characterization of some Indian Basmati and other elite rice genotypes using fluorescent-AFLP. Theoretical and Applied Genetics **105**:680-690.

Aharoni, A., and O. Vorst. 2002. DNA microarrays for functional plant genomics. Plant Molecular Biology **48**:99-118.

Allen, F. L. 1994. Usefulness of Plant Genome Mapping to Plant Breeding *in* P. M. Gresshoff, ed. Plant Genome Analysis. CRC Press.

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

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

Alonso-Blanco, C., A. J. M. Peeters, M. Koornneef, C. Lister, C. Dean, N. van den Bosch, J. Pot, and M. T. R. Kuiper. 1998b. Development of an AFLP based linkage map of Ler, Col and Cvi Arabidopsis thaliana ecotypes and construction of a Ler/Cvi recombinant inbred line population. Plant Journal **14**:259-271.

Amiour, N., J. Jahier, A. M. Tanquy, H. Chiron, and G. Branlard. 2002. Effect of 1R(1A), 1R(1B) and 1R(1D) substitution on technological value of bread wheat. Journal of Cereal Science **35**:149-160.

Araki, E., H. Miura, and S. Sawada. 2000. Differential effects of the null alleles at the three Wx loci on the starch-pasting properties of wheat. Theoretical and Applied Genetics **100**:1113-1120.

Baldi, P., and G. W. Hatfield. 2002. DNA Microarrays and Gene Expression.

Barnes, S. 2002. Comparing Arabidopsis to other flowering plants. Current Opinion in Plant Biology **5**:128-133.

Beckmann, J. S., and M. Soller. 1988. Detection of Linkage between Marker Loci and Loci Affecting Quantitative Traits in Crosses between Segregating Populations. Theoretical and Applied Genetics **76**:228-236.

Bell, C. J., and J. R. Ecker. 1994. Assignment of 30 Microsatellite Loci to the Linkage Map of Arabidopsis. Genomics **19**:137-144.

Berger, J., T. Suzuki, K. A. Senti, J. Stubbs, G. Schaffner, and B. J. Dickson. 2001. Genetic mapping with SNP markers in Drosophila. Nature Genetics **29**:475-481.

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

Borevitz, J. O., and J. Chory. 2004. Genomics tools for QTL analysis and gene discovery. Current Opinion in Plant Biology **7**:132-136.

Brouwer, D. J., and D. A. St Clair. 2004. Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub- NILs. Theoretical and Applied Genetics **108**:628-638.

Brzustowicz, L. M., C. Merette, X. Xie, L. Townsend, T. C. Gilliam, and J. Ott. 1993. Molecular and Statistical Approaches to the Detection and Correction of Errors in Genotype Databases. American Journal of Human Genetics **53**:1137-1145.

Burns, M. J., S. R. Barnes, J. G. Bowman, M. H. E. Clarke, C. P. Werner, and M. J. Kearsey. 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.

Cao, W. X., C. Epstein, H. Liu, C. DeLoughery, N. X. Ge, J. Y. Lin, R. Diao, H. Cao, F. Long, X. Zhang, Y. D. Chen, P. S. Wright, S. Busch, M. Wenck, K. Wong, A. G. Saltzman, Z. H. Tang, L. Liu, and A. Zilberstein. 2004. Comparing gene discovery from Affymetrix GeneChip microarrays and Clontech PCR-select cDNA subtraction: a case study. Bmc Genomics **5**:art. no.-26.

Castro, A. M., S. Ramos, A. Vasicek, A. Worland, D. Gimenez, A. A. Clua, and E. Suarez. 2001. Identification of wheat chromosomes involved with different types of resistance against greenbug (Schizaphis graminum, Rond.) and the Russian wheat aphid (Diuraphis noxia, Mordvilko). Euphytica **118**:321-330.

Chang, C., J. L. Bowman, A. W. Dejohn, E. S. Lander, and E. M. Meyerowitz. 1988. Restriction Fragment Length Polymorphism Linkage Map for Arabidopsis-Thaliana. Proceedings of the National Academy of Sciences of the United States of America **85**:6856-6860.

Chicurel, M. 2001. Faster, better, cheaper genotyping. Nature **412**:580-582.

Cho, R. J., M. Mindrinos, D. R. Richards, R. J. Sapolsky, M. Anderson, E. Drenkard, L. Dewdney, T. L. Reuber, M. Stammers, N. Federspiel, A. Theologis, W. H. Yang, E. Hubbell, M. Au, E. Y. Chung, D. Lashkari, B. Lemieux, C. Dean, R. J. Lipshutz, F. M. Ausubel, R. W. Davis, and P. J. Oefner. 1999. Genome-wide mapping with biallelic markers in Arabidopsis thaliana. Nature Genetics **23**:203-207.

Cho, Y. G., T. Ishii, S. Temnykh, X. Chen, L. Lipovich, S. R. McCouch, W. D. Park, N. Ayres, and S. Cartinhour. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (Oryza sativa L.). Theoretical and Applied Genetics **100**:713-722.

Chun, J. U., X. M. Yu, and M. Griffith. 1998. Genetic studies of antifreeze proteins and their correlation with winter survival in wheat. Euphytica **102**:219-226.

Chung, G. G., E. Provost, E. P. Kielhorn, L. A. Charette, B. L. Smith, and D. L. Rimm. 2001. Tissue microarray analysis of beta-catenin in colorectal cancer shows nuclear phospho-beta-catenin is associated with a better prognosis. Clinical Cancer Research **7**:4013-4020.

Churchill, G. A. 2002. Fundamentals bof experimental design for cDNA microarrays. Nature Genetics supplement **32**:490-495.

Churchill, G. A., and R. W. Doerge. 1994. Empirical Threshold Values for Quantitative Trait Mapping. Genetics **138**:963-971.

Clarke, J. H., and C. Dean. 1994. Mapping Fri, a Locus Controlling Flowering Time and Vernalization Response in Arabidopsis-Thaliana. Molecular & General Genetics **242**:81-89.

Coors, J. G., and S. Pandey. 1999. The Genetics and Exploitation of Heterosis in Crops.

Darvasi, A. 1997. The effect of selective genotyping on QTL mapping accuracy. Mammalian Genome **8**:67-68.

Darvasi, A., A. Weinreb, V. Minke, J. I. Weller, and M. Soller. 1993. Detecting Marker-Qtl Linkage and Estimating Qtl Gene Effect and Map Location Using a Saturated Genetic-Map. Genetics **134**:943-951.

Dekkers, J. C. M., and F. Hospital. 2002. The use of molecular genetics in the improvement of agricultural populations. Nature Reviews Genetics **3**:22-32.

Delourme, R., and F. Eber. 1992. Linkage between an Isozyme Marker and a Restorer Gene in Radish Cytoplasmic Male-Sterility of Rapeseed (Brassica-Napus L). Theoretical and Applied Genetics **85**:222-228.

Doerge, R. W., Z. B. Zeng, and B. S. Weir. 1997. Statistical issues in the search for genes affecting quantitative traits in experimental populations. Statistical Science **12**:195-219.

Donson, J., Y. W. Fang, G. Espiritu-Santo, W. M. Xing, A. Salazar, S. Miyamoto, V. Armendarez, and W. Volkmuth. 2002. Comprehensive gene expression analysis by transcript profiling. Plant Molecular Biology **48**:75-97.

Du, C., and G. E. Hart. 1998. Triticum turgidum L. 6A and 6B recombinant substitution lines: extended linkage maps and characterization of residual background alien genetic variation. Theoretical and Applied Genetics **96**:645-653.

Edwards, K. J., J. H. A. Barker, A. Daly, C. Jones, and A. Karp. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques **20**:758-&.

El-Assal, S., C. Alonso-Blanco, A. J. M. Peeters, V. Raz, and M. Koornneef. 2002. A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2 (vol 29, pg 435, 2001). Nature Genetics **30**:123-123.

El-Assal, S. E. D., C. Alonso-Blanco, A. J. M. Peeters, V. Raz, and M. Koornneef. 2001. A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2. Nature Genetics **29**:435-440.

El-Lithy, M. E., E. J. M. Clerkx, G. J. Ruys, M. Koornneef, and D. Vreugdenhil. 2004. Quantitative trait locus analysis of growth-related traits in a new Arabidopsis recombinant. Plant Physiology **135**:444-458.

Etienne, C., C. Rothan, A. Moing, C. Plomion, C. Bodenes, L. Svanella-Dumas, P. Cosson, V. Pronier, R. Monet, and E. Dirlewanger. 2002. Candidate genes and QTLs for sugar and organic acid content in peach Prunus persica (L.) Batsch. Theoretical and Applied Genetics **105**:145-159.

Falconer, D. S., and T. Mackay. 1996. Introduction to Quantitative Genetics.

Fourmann, M., P. Barret, N. Froger, C. Baron, F. Charlot, R. Delourme, and D. Brunel. 2002. From Arabidopsis thaliana to Brassica napus: development of amplified consensus genetic markers (ACGM) for construction of a gene map. Theoretical and Applied Genetics **105**:1196-1206.

Gerhold, L. M., M. T. Sellix, and M. E. Freeman. 2002. Antagonism of vasoactive intestinal peptide mRNA in the suprachiasmatic nucleus disrupts the rhythm of FRAs expression in neuroendocrine dopaminergic neurons. Journal of Comparative Neurology **450**:135-143.

Gibson, G., and T. F. C. Mackay. 2002. Enabling population and quantitative genomics. Genetical Research **80**:1-6.

Graeber, T. G., and D. Eisenberg. 2001. Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles. Nature Genetics **29**:295-300.

Gray, I. C., D. A. Campbell, and N. K. Spurr. 2000. Single nucleotide polymorphisms as tools in human genetics. Human Molecular Genetics **9**:2403-2408.

Gregersen, P. K., and T. W. Brehrens. 2003. Fine mapping them phenotype in autoimmune disease: the promise and pitfalls of DNA microarray technologies. Genes and Immunity **4**:175-176.

Grigoryev, D. N., S. F. Ma, R. A. Irizarry, S. Q. Ye, J. Quackenbush, and J. G. N. Garcia. 2004. Orthologous gene-expression profiling in multi-species models: search for candidate genes. Genome Biology **5**:art. no.-R34.

Guadagnuolo, R., D. Savova-Bianchi, J. Keller-Senften, and F. Felber. 2001. Search for evidence of introgression of wheat (Triticum aestivum L.) traits into sea barley (Hordeum marinum s.str. Huds.) and bearded wheatgrass (Elymus caninus L.) in central and northern Europe, using isozymes, RAPD and microsatellite markers. Theoretical and Applied Genetics **103**:191-196.

Hagenblad, J., and M. Nordborg. 2002. Sequence variation and haplotype structure surrounding the flowering time locus FRI in Arabidopsis thaliana. Genetics **161**:289-298.

Haley, C. S., and S. A. Knott. 1992. A Simple Regression Method for Mapping Quantitative Trait Loci in Line Crosses Using Flanking Markers. Heredity **69**:315-324.

Hamrick, J. L., and M. J. W. Godt. 1990. Allozyme diversity in plants *in* B. A. H. D, C. M. T, K. A. L, and W. B. S, eds. (Plant population genetics, breeding and genetic resources.

Hanano, S., M. Amagai, T. Kaneko, C. Kuwata, T. Tsugane, N. Sakurai, Y. Nakamura, D. Shibata, and S. Tabata. 2002. Analysis of gene expression in Arabidopsis thaliana by array hybridization with genomic DNA fragments aligned along chromosomal regions. Plant Journal **30**:247-255.

Harmer, S. L., L. B. Hogenesch, M. Straume, H. S. Chang, B. Han, T. Zhu, X. Wang, J. A. Kreps, and S. A. Kay. 2000. Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science **290**:2110-2113.

Hauge, B. M., S. M. Hanley, S. Cartinhour, J. M. Cherry, H. M. Goodman, M. Koornneef, P. Stam, C. Chang, S. Kempin, L. Medrano, and E. M. Meyerowitz. 1993. An Integrated Genetic Rflp Map of the Arabidopsis-Thaliana Genome. Plant Journal **3**:745-754.

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

Hayama, R., and G. Coupland. 2004. The molecular basis of diversity in the photoperiodic flowering responses of Arabidopsis and rice. Plant Physiology **135**:677-684.

Hill, W. G. 1998. Selection with recurrent backcrossing to develop congenic lines for quantitative trait loci analysis. Genetics **148**:1341-1352.

Holton, T. A. 2001. Plant Genotyping by Analysis of Microsatellites. Pp. 15-27 *in* R. J. Henry, ed. Plant Genotyping - The DNA fingerprinting of Plants.

Howell, P. M., D. F. Marshall, and D. J. Lydiate. 1996. Towards developing intervarietal substitution lines in Brassica napus using marker-assisted selection. Genome **39**:348-358.

Hubley, R. M., E. Zitzler, and J. C. Roach. 2003. Evolutionary algorithms for the selection of single nucleotide polymorphisms. Bmc Bioinformatics **4**:art. no.-30.

Hyne, V., and M. J. Kearsey. 1995. Qtl Analysis - Further Uses of Marker Regression. Theoretical and Applied Genetics **91**:471-476.

Hyne, V., M. J. Kearsey, D. J. Pike, and J. W. Snape. 1995. Qtl Analysis - Unreliability and Bias in Estimation Procedures. Molecular Breeding **1**:273-282.

Immervoll, T., S. Loesgen, G. Dutsch, H. Gohlke, N. Herbon, S. Klugbauer, A. Dempfle, H. Bickeboller, J. Becker-Follmann, F. Ruschendorf, K. Saar, A. Reis, H. E. Wichmann, and M. Wjst. 2001. Fine mapping and single nucleotide polymorphism association results of candidate genes for asthma and related phenotypes. Human Mutation **18**:327-336.

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

Jansen, R. C. 1993. Interval Mapping of Multiple Quantitative Trait Loci. Genetics **135**:205-211.

Jansen, R. C. 1994. Controlling the Type-I and Type-Ii Errors in Mapping Quantitative Trait Loci. Genetics **138**:871-881.

Jansen, R. C. 1996. Complex plant traits: Time for polygenic analysis. Trends in Plant Science **1**:89-94.

Jansen, R. C., and P. Stam. 1994. High-Resolution of Quantitative Traits into Multiple Loci Via Interval Mapping. Genetics **136**:1447-1455.

Jansen, R. C., J. W. Vanooijen, P. Stam, C. Lister, and C. Dean. 1995. Genotype-by-Environment Interaction in Genetic-Mapping of Multiple Quantitative Trait Loci. Theoretical and Applied Genetics **91**:33-37.

Jenssen, T. K., A. Laegreid, J. Komorowski, and E. Hovig. 2001. A literature network of human genes for high-throughput analysis of gene expression. Nature Genetics **28**:21-+.

Ji, Y., and R. T. Chetelat. 2003. Homoeologous pairing and recombination in Solanum lycopersicoides monosomic addition and substitution lines of tomato. Theoretical and Applied Genetics **106**:979-989.

Johanson, U., J. West, C. Lister, S. Michaels, R. Amasino, and C. Dean. 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science **290**:344-347.

Jorgensen, P., B. Nelson, M. D. Robinson, Y. Q. Chen, B. Andrews, M. Tyers, and C. Boone. 2002. High-resolution genetic mapping with ordered arrays of Saccharomyces cerevisiae deletion mutants. Genetics **162**:1091-1099.

Kao, C. H., Z. B. Zeng, and R. D. Teasdale. 1999. Multiple interval mapping for quantitative trait loci. Genetics **152**:1203-1216.

Kaplan, N., and R. Morris. 2001. Issues concerning association studies for fine mapping a susceptibility gene for a complex disease. Genetic Epidemiology **20**:432-457.

Kearsey, M. J. 2002. QTL Analysis: Problems and (Possible) Solutions. Pp. 45-58 *in* M. S. Kang, ed. Quantitative Genetics, Genomics and Plant Breeding.

Kearsey, M. J., and A. G. L. Farquhar. 1998. QTL analysis in plants; where are we now? Heredity **80**:137-142.

Kearsey, M. J., and V. Hyne. 1994. Qtl Analysis - a Simple Marker-Regression Approach. Theoretical and Applied Genetics **89**:698-702.

Kearsey, M. J., and Z. W. Luo. 2003. Mapping, characterization and deployment of quantitative trait loci. Pp. 1-29 *in* H. j. Newbury, ed. Plant molecular breeding.

Kearsey, M. J., and H. S. Pooni. 1996. The Genetical Analysis of Quantitative Traits. Chapman & Hall.

Kearsey, M. J., H. S. Pooni, and N. H. Syed. 2003. Genetics of quantitative traits in Arabidopsis thaliana. Heredity **91**:456-464.

Kebede, H., P. K. Subudhi, D. T. Rosenow, and H. T. Nguyen. 2001. Quantitative trait loci influencing drought tolerance in grain sorghum (Sorghum bicolor L. Moench). Theoretical and Applied Genetics **103**:266-276.

Kerr, M. K., C. A. Afshari, L. Bennett, P. Bushel, J. Martinez, N. J. Walker, and G. A. Churchill. 2002. Statistical analysis of a gene expression microarray experiment with replication. Statistica Sinica **12**:203-217.

Kleinhofs, A., A. Kilian, M. A. S. Maroof, R. M. Biyashev, P. Hayes, F. Q. Chen, N. Lapitan, A. Fenwick, T. K. Blake, V. Kanazin, E. Ananiev, L. Dahleen, D. Kudrna, J. Bollinger, S. J. Knapp, B. Liu, M. Sorrells, M.

Heun, J. D. Franckowiak, D. Hoffman, R. Skadsen, and B. J. Steffenson. 1993. A Molecular, Isozyme and Morphological Map of the Barley (Hordeum-Vulgare) Genome. Theoretical and Applied Genetics **86**:705-712.

Kocsy, G., G. Szalai, J. Sutka, E. Paldi, and G. Galiba. 2004. Heat tolerance together with heat stress-induced changes in glutathione and hydroxymethylglutathione levels is affected by chromosome 5A of wheat. Plant Science **166**:451-458.

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

Koornneef, M., C. Alonso-Blanco, H. Blankestijn-de Vries, C. J. Hanhart, and A. J. M. Peeters. 1998a. Genetic interactions among late-flowering mutants of Arabidopsis. Genetics **148**:885-892.

Koornneef, M., C. Alonso-Blanco, A. J. M. Peeters, and W. Soppe. 1998b. Genetic control of flowering time in arabidopsis. Annual Review of Plant Physiology and Plant Molecular Biology **49**:345-370.

Koornneef, M., J. Vaneden, C. J. Hanhart, P. Stam, F. J. Braaksma, and W. J. Feenstra. 1983. Linkage Map of Arabidopsis-Thaliana. Journal of Heredity **74**:265-272.

Korzun, V., A. Borner, A. J. Worland, C. N. Law, and M. S. Roder. 1997. Application of microsatellite markers to distinguish inter- varietal chromosome substitution lines of wheat (Triticum aestivum L). Euphytica **95**:149-155.

Koumproglou, R., T. M. Wilkes, P. Townson, X. Y. Wang, J. Beynon, H. S. Pooni, H. J. Newbury, and M. J. Kearsey. 2002. STAIRS: a new genetic resource for functional genomic studies of Arabidopsis. Plant Journal **31**:355-364.

Kowalski, S. P., T. H. Lan, K. A. Feldmann, and A. H. Paterson. 1994. Qtl Mapping of Naturally-Occurring Variation in Flowering Time of Arabidopsis-Thaliana. Molecular & General Genetics **245**:548-555.

Kuittinen, H., A. Mattila, and O. Savolainen. 1997. Genetic variation at marker loci and in quantitative traits in natural populations of Arabidopsis thaliana. Heredity **79**:144-152.

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

Lander, E., J. Abrahamson, A. Barlow, M. Daly, S. Lincoln, L. Newburg, and P. Green. 1987. Mapmaker a Computer Package for Constructing Genetic-Linkage Maps. Cytogenetics and Cell Genetics **46**:642-642.

Lander, E. S., and D. Botstein. 1989. Mapping Mendelian Factors Underlying Quantitative Traits Using Rflp Linkage Maps. Genetics **121**:185-199.

Lecomte, L., V. Saliba-Colombani, A. Gautier, M. C. Gomez-Jimenez, P. Duffe, M. Buret, and M. Causse. 2004. Fine mapping of QTLs of chromosome 2 affecting the fruit architecture and composition of tomato. Molecular Breeding **13**:1-14.

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

Lee, I., A. Bleecker, and R. Amasino. 1993. Analysis of Naturally-Occurring Late Flowering in Arabidopsis-Thaliana. Molecular & General Genetics **237**:171-176.

Lindblad-Toh, K., E. Winchester, M. J. Daly, D. G. Wang, J. N. Hirschhorn, J. P. Laviolette, K. Ardlie, D. E. Reich, E. Robinson, P. Sklar, N. Shah, D. Thomas, J. B. Fan, T. Gingeras, J. Warrington, N. Patil, T. J. Hudson, and E. S. Lander. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nature Genetics **24**:381-386.

Lister, C., and C. Dean. 1993. Recombinant Inbred Lines for Mapping Rflp and Phenotypic Markers in Arabidopsis-Thaliana. Plant Journal **4**:745-750.

Loridon, K., B. Cournoyer, C. Goubely, A. Depeiges, and G. Picard. 1998. Length polymorphism and allele structure of trinucleotide microsatellites in natural accessions of Arabidopsis thaliana. Theoretical and Applied Genetics **97**:591-604.

Loudet, O., S. Chaillou, C. Camilleri, D. Bouchez, and F. Daniel-Vedele. 2002. Bay-0 x Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in Arabidopsis. Theoretical and Applied Genetics **104**:1173-1184.

Lukowitz, W., C. S. Gillmor, and W. R. Scheible. 2000. Positional cloning in arabidopsis. Why it feels good to have a genome initiative working for you. Plant Physiology **123**:795-805.

Luo, Z. W., and M. J. Kearsey. 1989. Maximum-Likelihood-Estimation of Linkage between a Marker Gene and a Quantitative Locus. Heredity **63**:401-408.

Luo, Z. W., and J. A. Woolliams. 1993. Estimation of Genetic-Parameters Using Linkage between a Marker Gene and a Locus Underlying a Quantitative Character in F2 Populations. Heredity **70**:245-253.

Luo, Z. W., C. I. Wu, and M. J. Kearsey. 2002. Precision and high-resolution mapping of quantitative trait loci by use of recurrent selection, backcross or intercross schemes. Genetics **161**:915-929.

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

Macaulay, M., L. Ramsay, W. Powell, and R. Waugh. 2001. A representative, highly informative 'genotyping set' of barley SSRs. Theoretical and Applied Genetics **102**:801-809.

Mackay, T. F. C. 2001. The genetic architecture of quantitative traits. Annual Review of Genetics **35**:303-339.

Maguire, T. L. 2001. Producing andexploiting Enriched Microsatellite Libraries *in* R. J. Henry, ed. Plnt Genotyping - The DNA fingerprinting of plants.

Mara, M., T. Kucaba, M. Sekhon, L. Hillier, R. Martienssen, A. Chinwalla, J. Crockett, J. Fedele, H. Grover, C. Gund, W. R. McCombie, K. McDonald, J. McPherson, N. Mudd, L. Parnell, J. Schein, R. Seim, P. Shelby, R. Waterston, and R. Wilson. 1999. zA map for sequence analysis of the Arabidopsis thaliana genome. Nature Genetics **22**:265-270.

Marquez-Cedillo, L. A., P. M. Hayes, A. Kleinhofs, W. G. Legge, B. G. Rossnagel, K. Sato, S. E. Ullrich, and D. M. Wesenberg. 2001. QTL analysis of agronomic traits in barley based on the doubled haploid progeny of two elite North American varieties representing different germplasm groups. Theoretical and Applied Genetics **103**:625-637.

Martinez, O., and R. N. Curnow. 1992. Estimating the Locations and the Sizes of the Effects of Quantitative Trait Loci Using Flanking Markers. Theoretical and Applied Genetics **85**:480-488.

Mather, K., and J. L. Jinks. 1982. Biometrical Genetics.

Matsumoto, F., T. Obayashi, Y. Sasaki-Sekimoto, H. Ohta, K. Takamiya, and T. Masuda. 2004. Gene expression profiling of the tetrapyrrole metabolic pathway in Arabidopsis with a mini-array system. Plant Physiology **135**:2379-2391.

Meyerowitz, E. M., and C. R. Sommerville. 1994. ARABIDOPSIS. Cold Spring Harbour Press.

Mignouna, H. D., R. A. Mank, T. H. N. Ellis, N. van den Bosch, R. Asiedu, M. M. Abang, and J. Peleman. 2002. A genetic linkage map of water yam (Dioscorea alata L.) based on AFLP markers and QTL analysis for anthracnose resistance. Theoretical and Applied Genetics **105**:726-735.

Mitchell-Olds, T., and D. Pedersen. 1998. The molecular basis of quantitative genetic variation in central and secondary metabolism in Arabidopsis. Genetics **149**:739-747.

Mohan, M., S. Nair, A. Bhagwat, T. G. Krishna, M. Yano, C. R. Bhatia, and T. Sasaki. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding **3**:87-103.

Morgante, M., M. Hanafey, and W. Powell. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genetics **30**:194-200.

Mott, R., C. J. Talbot, M. G. Turri, A. C. Collins, and J. Flint. 2000. A method for fine mapping quantitative trait loci in outbred animal stocks. Proceedings of the National Academy of Sciences of the United States of America **97**:12649-12654.

Mozo, T., K. Dewar, P. Dunn, J. R. Ecker, S. Fischer, S. Kloska, H. Lehrach, M. Marra, R. Martienssen, S. Meier-Ewert, and T. Altmann. 1999. A complete BAC-based physical map of the Arabidopsis thaliana genome. Nature Genetics **22**:271-275.

Nadeau, J. H., and W. N. Frankel. 2000. The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs. Nature Genetics **25**:381-384.

Nam, H. G., J. Giraudat, B. Denboer, F. Moonan, W. D. B. Loos, B. M. Hauge, and H. M. Goodman. 1989. Restriction Fragment Length Polymorphism Linkage Map of Arabidopsis-Thaliana. Plant Cell **1**:699-705.

Noh, Y. S., C. M. Bizzell, B. Noh, F. M. Schomburg, and R. M. Amasino. 2004. EARLY FLOWERING 5 acts as a floral repressor in Arabidopsis. Plant Journal **38**:664-672.

Paran, I., and D. Zamir. 2003. Quantitative traits in plants: beyond the QTL. Trends in Genetics **19**:303-306.

Peleman, J. D., and J. R. van der Voort. 2003. Breeding by design. Trends in Plant Science **8**:330-334.

Peng, J. H., Y. Ronin, T. Fahima, M. S. Roder, Y. C. Li, E. Nevo, and A. Korol. 2003. Domestication quantitative trait loci in Triticum dicoccoides, the progenitor of wheat. Proceedings of the National Academy of Sciences of the United States of America **100**:2489-2494.

Pestsova, E., E. Salina, A. Borner, V. Korzun, O. I. Maystrenko, and M. S. Roder. 2000. Microsatellites confirm the authenticity of inter-varietal chromosome substitution lines of wheat (Triticum aestivum L.). Theoretical and Applied Genetics **101**:95-99.

Peters, J. L., Constandt, Hans, Neyt, Pia., Cnops, Gerda., Zethof, Jan., Zabeau, Marc., Gerats, Tom. 2001. A physical amplified fragment- length polymorphism map of Arabidopsis. Plant Physiology **127**:1579-1589.

Pickett, A. A., and N. W. Galwey. 1997. A further evaluation of hybrid wheat. Plant Varieties and Seeds **10**:15-32.

Ponce, M. R., P. Robles, and J. L. Micol. 1999. High-throughput genetic mapping in Arabidopsis thaliana. Molecular and General Genetics **261**:408-415.

Powell, W., G. C. Machray, and J. Provan. 1996. Polymorphism revealed by simple sequence repeats. Trends in Plant Science **1**:215-222.

Putterill, J., R. Laurie, and R. Macknight. 2004. It's time to flower: the genetic control of flowering time. Bioessays **26**:363-373.

Qiagen. 2001. RNeasy Mini Handbook. For total RNA isolation from plants.

Qiagen. 2002. QIAquick Spin Handbook. For PCR purification.

Rae, A. M., E. C. Howell, and M. J. Kearsey. 1999. More QTL for flowering time revealed by substitution lines in Brassica oleracea. Heredity **83**:586-596.

Ramsay, L. D., D. E. Jennings, E. J. R. Bohuon, A. E. Arthur, D. J. Lydiate, M. J. Kearsey, and D. F. Marshall. 1996. The construction of a substitution library of recombinant backcross lines in Brassica oleracea for the precision mapping of quantitative trait loci. Genome **39**:558-567.

Redei, G. P. 1970. *Arabidopsis thaliana*(L.) Heynh. A review of the genetics and biology. Bibliograpfy of Genetics. **20**:1-151.

Redman, J. C., B. J. Hass, G. Tanimoto, and C. D. Town. 2004. Development and evaluation of an Arabidopsis whole genome Affymetrix probe array. (vol 38, pg 545, 2004). Plant Journal **38**:1023-1023.

Reiter, R. S., J. G. K. Williams, K. A. Feldmann, J. A. Rafalski, S. V. Tingey, and P. A. Scolnik. 1992. Global and Local Genome Mapping in Arabidopsis-Thaliana by Using Recombinant Inbred Lines and Random Amplified Polymorphic Dnas. Proceedings of the National Academy of Sciences of the United States of America **89**:1477-1481.

Remington, D. L., and M. D. Purugganan. 2003. Candidate genes, quantitative trait loci, and functional trait evolution in plants. International Journal of Plant Sciences **164**:S7-S20.

Remington, D. L., R. W. Whetten, B. H. Liu, and D. M. O'Malley. 1999. Construction of an AFLP genetic map with nearly complete genome coverage in Pinus taeda. Theoretical and Applied Genetics **98**:1279-1292.

Risch, N. 2001. Implications of multilocus inheritance for gene-disease association studies. Theoretical Population Biology **60**:215-220.

Robert, L. S., F. Robson, A. Sharpe, D. Lydiate, and G. Coupland. 1998. Conserved structure and function of the Arabidopsis flowering time gene CONSTANS in Brassica napus. Plant Molecular Biology **37**:763-772.

Ronin, Y., A. Korol, M. Shtemberg, E. Nevo, and M. Soller. 2003. High-resolution mapping of quantitative trait loci by selective recombinant genotyping. Genetics **164**:1657-1666.

Rousset, M., P. Brabant, R. S. Kota, J. Dubcovsky, and J. Dvorak. 2001. Use of recombinant substitution lines for gene mapping and QTL analysis of bread making quality in wheat. Euphytica **119**:81-87.

Schena, M., and R. W. Davis. 2000. Technology Standards for Microarray Research. Pp. 1-18 *in* M. Schena, ed. Microarray Biochip Technology.

Seki, M., M. Narusaka, J. Ishida, T. Nanjo, M. Fujita, Y. Oono, A. Kamiya, M. Nakajima, A. Enju, T. Sakurai, M. Satou, K. Akiyama, T. Taji, K. Yamaguchi-Shinozaki, P. Carninci, J. Kawai, Y. Hayashizaki, and K. Shinozaki. 2002. Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full- length cDNA microarray. Plant Journal **31**:279-292.

Shen, R., L. R. Kong, and H. Ohm. 2004. Fusarium head blight resistance in hexaploid wheat (Triticum aestivum)-Lophopyrum genetic lines and tagging of the alien chromatin by PCR markers. Theoretical and Applied Genetics **108**:808-813.

Shoemaker, R. C. 1994. Genome Mapping and Agriculture. Pp. 1- 10 *in* P. M. Gresshoff, ed. Plant Genome Analysis. CRC Press.

Simpson, G. G., and C. Dean. 2002. Flowering - Arabidopsis, the rosetta stone of flowering time? Science **296**:285-289.

Song, Q. J., E. W. Fickus, and P. B. Cregan. 2002. Characterization of trinucleotide SSR motifs in wheat. Theoretical and Applied Genetics **104**:286-293.

Soppe, W. J. J., S. E. Jacobsen, C. Alonso-Blanco, J. P. Jackson, T. Kakutani, M. Koornneef, and A. J. M. Peeters. 2000. The late flowering phenotype of fwa mutants is caused by gain- of-function epigenetic alleles of a homeodomain gene. Molecular Cell **6**:791-802.

Sourdille, P., M. Tavaud, G. Charmet, and M. Bernard. 2001. Transferability of wheat microsatellites to diploid Triticeae species carrying the A, B and D genomes. Theoretical and Applied Genetics **103**:346-352.

Sprague, G. F. 1983. Heterosis in maize: theory and ptactice. Monogr. Theoretical and applied Genetics **6**:47-70.

Stajner, N., J. Jakse, P. Kozjak, and B. Javornik. 2005. The isolation and characterization of microsatellites in hop (*Humulus lupulus* L.). Plant Science **168**:213-221.

Stam, P. 1993. Construction of Integrated Genetic-Linkage Maps by Means of a New Computer Package - Joinmap. Plant Journal **3**:739-744.

Stekel, D. 2003. Microarray Bioinformatics. Cambridge University Press.

Stuber, C. W., S. E. Lincoln, D. W. Wolff, T. Helentjaris, and E. S. Lander. 1992. Identification of Genetic-Factors Contributing to Heterosis in a Hybrid from 2 Elite Maize Inbred Lines Using Molecular Markers. Genetics **132**:823-839.

Swarup, K., C. Alonso-Blanco, J. R. Lynn, S. D. Michaels, R. M. Amasino, M. Koornneef, and A. J. Millar. 1999. Natural allelic variation identifies new genes in the Arabidopsis circadian system. Plant Journal **20**:67-77.

Syed, N. H., and Z. J. Chen. 2004. Molecular marker genotypes, heterozygosity and genetic interactions explain heterosis in *Arabidopsis thaliana*. Heredity:1-10.

Taji, T., M. Seki, M. Satou, T. Sakurai, M. Kobayashi, K. Ishiyama, Y. Narusaka, M. Narusaka, J. K. Zhu, and K. Shinozaki. 2004. Comparative genomics in salt tolerance between Arabidopsis and Arabidopsis-related halophyte salt cress using Arabidopsis microarray. Plant Physiology **135**:1697-1709.

Tanksley, S. D. 1993. Mapping Polygenes. Annual Review of Genetics **27**:205-233.

Tasma, I. M., and R. C. Shoemaker. 2003. Mapping flowering time gene homologs in soybean and their association with maturity (E) loci. Crop Science **43**:319-328.

Terryn, N., P. Rouze, and M. Van Montagu. 1999. Plant genomics. Febs Letters **452**:3-6.

Thoday, J. M. 1961. Location of polygenes. Nature **191**:368-370.

Ungerer, M. C., S. S. Halldorsdottir, J. L. Modliszewski, T. F. C. Mackay, and M. D. Purugganan. 2002. Quantitative trait loci for inflorescence development in Arabidopsis thaliana. Genetics **160**:1133-1151.

Ungerer, M. C., S. S. Halldorsdottir, M. A. Purugganan, and T. F. C. Mackay. 2003. Genotype-environment interactions at quantitative trait loci affecting inflorescence development in Arabidopsis thaliana. Genetics **165**:353-365.

van Nocker, S., M. Muszynski, K. Briggs, and R. M. Amasino. 2000. Characterization of a gene from Zea mays related to the Arabidopsis flowering-time gene LUMINIDEPENDENS. Plant Molecular Biology **44**:107-122.

Vanooijen, J. W. 1992. Accuracy of Mapping Quantitative Trait Loci in Autogamous Species. Theoretical and Applied Genetics **84**:803-811.

Virk, P. S., H. S. Pooni, N. H. Syed, and M. J. Kearsey. 1999. Fast and reliable genotype validation using microsatellite markers in Arabidopsis thaliana. Theoretical and Applied Genetics **98**:462-464.

Vision, T. J., D. G. Brown, and S. D. Tanksley. 2000. The origins of genomic duplications in Arabidopsis. Science **290**:2114-2117.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Vandelee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. Aflp - a New Technique for DNA-Fingerprinting. Nucleic Acids Research **23**:4407-4414.

Walsh, B. 2001. Quantitative genetics in the age of genomics. Theoretical Population Biology **59**:175-184.

Weller, J. I. 1986. Maximum-Likelihood Techniques for the Mapping and Analysis of Quantitative Trait Loci with the Aid of Genetic-Markers. Biometrics **42**:627-640.

Wicks, S. R., R. T. Yeh, W. R. Gish, R. H. Waterston, and R. H. A. Plasterk. 2001. Rapid gene mapping in Caenorhabditis elegans using a high density polymorphism map. Nature Genetics **28**:160-164.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey. 1990. Oligo-nucleotide primers of arbitrary sequence amplify DNA polymorphisms which are useful as genetic markers. Nucleic Acids Research **18**:6531-6535.

Wilson, Z. A. 2000. Arabidopsis A Practical Approach. Oxford University Press.

Wolf, N., A. Galecki, R. Lipman, S. Chen, M. Smith-Wheelock, D. Burke, and R. Miller. 2004. Quantitative trait locus mapping for age-related cataract severity and synechia prevalence using four-way cross mice. Investigative Ophthalmology & Visual Science **45**:1922-1929.

Wright, S. 1952. The genetics of quantitative variability *in* E. C. R. Reeve, and C. H. Waddington, eds. Quantitative Inheritance.

www.arabidopsis.org.

www.dur.ac.uk.

Xiao, J. H., J. M. Li, L. P. Yuan, and S. D. Tanksley. 1995. Dominance Is the Major Genetic-Basis of Heterosis in Rice as Revealed by Qtl Analysis Using Molecular Markers. Genetics **140**:745-754.

Yang, Y. H., and T. Speed. 2002. Design issues for cDNA microarray experiments. Nature Reviews Genetics **3**:579-588.

Yates, H. E., A. Frary, S. Doganlar, A. Frampton, N. T. Eannetta, J. Uhlig, and S. D. Tanksley. 2004. Comparative fine mapping of fruit quality QTLs on chromosome 4 introgressions derived from two wild tomato species. Euphytica **135**:283-296.

Zeng, Z. B. 1993. Theoretical Basis for Separation of Multiple Linked Gene Effects in Mapping Quantitative Trait Loci. Proceedings of the National Academy of Sciences of the United States of America **90**:10972-10976.

Zhao, H. Y., and T. P. Speed. 1996. On genetic map functions. Genetics **142**:1369-1377.

APPENDICES

Appendix 01:

DNA Extraction Method

(Based on Gawel and Jarret, Plant Mol. Biol. Rep., 9, 262-266)

1. Take 10 ml of extraction buffer (see below) and add 10 ul of mercaptaethanol in a fume cupboard. Heat this in water bath before starting the homogenisation. This amount is enough for 12 extractions.
2. Take 20 mg of leaf material, place in a 1.5 ml tube and freeze using liquid nitrogen. Homogenise contents to a powder using the hand- held device taking care that the contents do not thaw out.
3. Immdiately, add 700ul of extraction buffer that has been pre heated to 65^{o}. Incubate at 65^{o} for 30 minutes.
4. Add 600ul of chloroform:isoamyl alcohol and mix by inversion for 5 min at room temperature.
5. Centrifuge at 13, 000 rpm for 5 minutes.
6. Take off 500 ul of the supernatant. Place in a fresh tube and add 500 ul of ice cold propanol. Mix by inversion. Centrifuge at 13,000 rpm for 5 min.
7. Pour off isopropanol; stand tubes upside down on a paper towel for a few minutes. Add 300ul of 70% ethanol (room temperature) to the DNA pellet to wash it. Spin at 13,000 rpm for 5 min.
8. Pour off ethanol; stand tubes upside down on a paper towel for a few minutes. With a piece of tissue, ensure that any film of ethanol is removed from the neck of each tube. Place in vacuum desiccator for at least 20 min.
9. Take up DNA pellet in 100ul of sterile distilled water.

Extraction buffer:

4% CTAB, 100mM tris-Hcl, pH8.0,1.4 M NaCl, 20 mM EDTA $(Na)_2$.

Note: 0.1% 2- mercaptaethanol to be added immediately before use.

Chloroform:isoamyl alcohol

24:1 vol:vol

Appendix 02

Tables of ANOVAS in chapter 03.

General Linear Model: DOG versus ECOTYPE, tube

```
Factor          Type Levels Values
Line           fixed      2   1 2
rep(line)     random      8   1 2 3 4 5 6 7 8
Analysis of Variance for DOG, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1     0.4737     0.3325     0.3325    0.41  0.547
rep(Line)        6     4.6101     4.6101     0.7683    2.82  0.017
Error           68    18.5478    18.5478     0.2728
Total           75    23.6316
```

General Linear Model: D 20 RL versus ECOTYPE, tube

```
Factor          Type Levels Values
line            fixed   2   1 2
rep(line)       random  8   1 2 3 4 5 6 7 8
Analysis of Variance for D 20 RL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line            1       0.329      0.215      0.215    0.07  0.803
rep(Line)       6      17.918     17.918      2.986    2.86  0.015
Error           68     70.950     70.950      1.043
Total           75     89.197
```

General Linear Model: D20 CL versus ECOTYPE, tube

```
Factor          Type    Levels Values
line           fixed     2     1 2
rep(Line)      random    8     1 2 3 4 5 6 7 8
Analysis of Variance for D20 CL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1      0.118      0.140      0.140    0.16  0.710
rep(Line)        6      5.629      5.629      0.938    0.62  0.715
Error           68    103.240    103.240      1.518
Total           75    108.987
```

General Linear Model: D20 RW versus ECOTYPE, tube

```
Factor          Type   Levels Values
line            fixed    2    1 2
rep(Line)       random   8    1 2 3 4 5 6 7 8
Analysis of Variance for D20 RW, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1     889.47     900.42     900.42   14.56  0.014
rep(Line)        6     388.73     388.73      64.79    0.65  0.687
Error           68    6736.48    6736.48      99.07
Total           75    8014.68
```

General Linear Model: D20 HT versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed      2    1 2
rep(Line)      random     8    1 2 3 4 5 6 7 8
Analysis of Variance for D20 HT, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1      1.316      1.663      1.663    0.33  0.593
rep(Line)        6     31.890     31.890      5.315    0.63  0.709
Error           68    577.478    577.478      8.492
Total           75    610.684
```

General Linear Model: D25 HT versus ECOTYPE, tube

```
Factor          Type     Levels Values
Line           fixed      2     1 2
rep(Line)      random     8      1 2 3 4 5 6 7 8
Analysis of Variance for D25 HT, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1         25         23         23    0.04  0.854
rep(Line)        6       3847       3847        641    0.57  0.751
Error           68      76175      76175       1120
Total           75      80048
```

General Linear Model: D25 RL versus ECOTYPE, tube

```
Factor          Type Levels Values
Line           fixed     2    1 2
rep(Line)      random    8    1 2 3 4 5 6 7 8
Analysis of Variance for D25 RL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1     10.316     10.074     10.074    2.64  0.159
rep(Line)        6     21.814     21.814      3.636    2.28  0.046
Error           68    108.502    108.502      1.596
Total           75    140.632
```

General Linear Model: D25 CL versus ECOTYPE, tube

```
Factor       Type    Levels Values
Line        fixed       2    1 2
rep(Line)   random      8     1 2 3 4 5 6 7 8
Analysis of Variance for D25 CL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1      0.474      0.457      0.457    0.52  0.503
rep(Line)        6      5.327      5.327      0.888    0.87  0.524
Error           68     69.620     69.620      1.024
Total           75     75.421
```

General Linear Model: D25 RW versus ECOTYPE, tube

```
Factor          Type Levels  Values
Line           fixed   2      1 2
rep(Line)      random  8      1 2 3 4 5 6 7 8
Analysis of Variance for D25 RW, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1     2105.3     2243.0     2243.0    6.80  0.045
rep(Line)        6     1926.2     1926.2      321.0    1.49  0.196
Error           68    14679.7    14679.7      215.9
Total           75    18711.2
```

General Linear Model: D35 HT versus ECOTYPE, tube

```
Factor          Type    Levels Values
Line           fixed      2   1 2
rep(Line)      random     8   1 2 3 4 5 6 7 8
Analysis of Variance for D35 HT, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1      13878      13081      13081    4.20  0.094
rep(Line)        6      18468      18468       3078    1.18  0.329
Error           68     177891     177891       2616
Total           75     210237
```

General Linear Model: FD versus ECOTYPE, tube

```
Factor          Type     Levels Values
Line            fixed         2    1 2
rep(Line)       random        8    1 2 3 4 5 6 7 8
Analysis of Variance for FD, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1      4.750      4.331      4.331    0.87  0.391
rep(Line)        6     29.025     29.025      4.838    1.57  0.169
Error           68    209.212    209.212      3.077
Total           75    242.987
```

General Linear Model: HT@F versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed       2    1 2
rep(Line)      random      8    1 2 3 4 5 6 7 8
Analysis of Variance for HT@F, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1      987.8      800.7      800.7    1.29  0.303
rep(Line)        6     3565.3     3565.3      594.2    2.24  0.049
Error           68    18003.5    18003.5      264.8
Total           75    22556.7
```

General Linear Model: RL@F versus ECOTYPE, tube

```
Factor          Type    Levels Values
Line           fixed       2    1 2
rep(Line)      random      8    1 2 3 4 5 6 7 8
Analysis of Variance for RL@F, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1     15.211     14.497     14.497    3.70  0.106
rep(Line)        6     22.328     22.328      3.721    2.69  0.021
Error           68     94.198     94.198      1.385
Total           75    131.737
```

General Linear Model: CL@F versus ECOTYPE, tube

```
Factor          Type    Levels Values
Line            fixed       2    1 2
rep(Line)       random      8    1 2 3 4 5 6 7 8
Analysis of Variance for CL@F, using Adjusted SS for Tests
Source               DF     Seq SS     Adj SS     Adj MS       F      P
Line                  1     1.8947     1.8283     1.8283   22.02  0.011
rep(Line)             6     0.5788     0.5788     0.0965    0.38  0.888
Error           68    17.1581    17.1581     0.2523
Total           75    19.6316
```

General Linear Model: RW@F versus ECOTYPE, tube

```
Factor          Type    Levels Values
Line           fixed       2    1 2
rep(Line)      random      8    1 2 3 4 5 6 7 8
Analysis of Variance for RW@F, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1    1719.61    1733.03    1733.03    8.02  0.032
rep(Line)        6    1239.08    1239.08     206.51    2.20  0.054
Error           67    6301.65    6301.65      94.05
Total           74    9260.35
```

General Linear Model: D30 HT versus ECOTYPE, tube

```
Factor          Type  Levels Values
Line           fixed       2   1 2
rep(Line)     random       8   1 2 3 4 5 6 7 8
Analysis of Variance for D30 HT, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1      13689      13980      13980    2.11  0.203
rep(Line)        6      38915      38915       6486    1.36  0.242
Error           68     323656     323656       4760
Total           75     376261
```

General Linear Model: D30 RL versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed       2    1 2
rep(Line)     random       8    1 2 3 4 5 6 7 8
Analysis of Variance for D30 RL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1      2.224      1.852      1.852    0.57  0.480
rep(Line)        6     18.623     18.623      3.104    1.92  0.090
Error           68    110.035    110.035      1.618
Total           75    130.882
```

General Linear Model: D30 CL versus ECOTYPE, tube

```
Factor          Type    Levels  Values
Line           fixed       2    1 2
rep(Line)      random      8    1 2 3 4 5 6 7 8
Analysis of Variance for D30 CL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1      74.01      71.93      71.93    0.77  0.415
rep(Line)        6     534.12     534.12      89.02    2.00  0.077
Error           68    3024.75    3024.75      44.48
Total           75    3632.88
```

General Linear Model: D30 RW versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed       2   1 2
rep(Line)     random       8   1 2 3 4 5 6 7 8
Analysis of Variance for D30 RW, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1     5889.0     5899.1     5899.1   62.33  0.001
rep(Line)        6      593.4      593.4       98.9    0.67  0.677
Error           68    10086.5    10086.5      148.3
Total           75    16568.9
```

General Linear Model: D40HT versus ECOTYPE, tube

```
Factor          Type   Levels  Values
Line           fixed       2    1 2
rep(Line)     random       8    1 2 3 4 5 6 7 8
Analysis of Variance for D40HT, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             1       6120       5842       5842    4.82  0.086 x
rep(Line)        6       7731       7731       1289    0.59  0.735
Error           68     147732     147732       2173
Total           75     161583
```

General Linear Model: DOG versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed     3     1 2 3
rep(Line)      random    11    1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for DOG, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2     0.5117     0.3775     0.1888    0.25  0.785
rep(Line)        8     5.7497     5.7497     0.7187    3.00  0.005
Error          101    24.1582    24.1582     0.2392
Total          111    30.4196
```

General Linear Model: D 20 RL versus ECOTYPE, tube

```
Factor          Type Levels Values
Line           fixed    3    1 2 3
rep(Line)      random  11   1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D 20 RL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    146.123    141.949     70.974   29.60  0.000
rep(Line)        8     18.340     18.340      2.293    2.62  0.012
Error          101     88.528     88.528      0.877
Total          111    252.991
```

General Linear Model: D20 CL versus ECOTYPE, tube

```
Factor       Type   Levels Values
Line        fixed      3    1 2 3
rep(Line)   random     11   1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D20 CL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2     15.400     16.279      8.140    8.15  0.016
rep(Line)        8      8.124      8.124      1.015    0.81  0.592
Error          101    125.967    125.967      1.247
Total          111    149.491
```

General Linear Model: D20 RW versus ECOTYPE, tube

```
Factor          Type  Levels Values
Line           fixed      3   1 2 3
rep(Line)      random     11  1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D20 RW, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    1730.75    1704.29     852.15   17.27  0.003
rep(Line))       8     413.83     413.83      51.73    0.62  0.760
Error          101    8445.38    8445.38      83.62
Total          111   10589.96
```

General Linear Model: D20 HT versus ECOTYPE, tube

```
Factor          Type  Levels  Values
Line           fixed      3    1 2 3
rep(Line)     random     11    1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D20 HT, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    1408.84    1460.24     730.12   18.12  0.001
rep(Line)        8     311.35     311.35      38.92    1.89  0.069
Error          101    2077.24    2077.24      20.57
Total          111    3797.43
```

General Linear Model: D25 HT versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed       3   1 2 3
rep(Line)     random      11  1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D25 HT, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    24321.0    25245.7    12622.9   11.72  0.006
rep(Line)        8     8562.6     8562.6     1070.3    1.09  0.379
Error          101    99579.7    99579.7      985.9
Total          111   132463.3
```

General Linear Model: D25 RL versus ECOTYPE, tube

```
Factor          Type Levels  Values
Line           fixed      3     1 2 3
rep(Line)     random     11    1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D25 RL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    805.711    797.620    398.810  115.54  0.000
rep(Line)        8     26.441     26.441      3.305    2.48  0.017
Error          101    134.625    134.625      1.333
Total          111    966.777
```

General Linear Model: D25 CL versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed       3   1 2 3
rep(Line)     random      11   1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D25 CL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    180.143    192.613     96.307   18.19  0.001
rep(Line)        8     41.163     41.163      5.145    1.63  0.125
Error          101    318.756    318.756      3.156
Total          111    540.062
```

General Linear Model: D25 RW versus ECOTYPE, tube

```
Factor          Type    Levels Values
Line           fixed       3    1 2 3
rep(Line)      random     11    1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D25 RW, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2     5884.8     5850.1     2925.0   11.79  0.005
rep(Line)        8     1938.6     1938.6      242.3    1.48  0.174
Error          101    16535.3    16535.3      163.7
Total          111    24358.7
```

General Linear Model: D35 HT versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed       3    1 2 3
rep(Line)      random     11   1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D35 HT, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2     332224     325031     162515   63.27  0.000
rep(Line)        8      20330      20330       2541    1.17  0.324
Error          101     219184     219184       2170
Total          111     571737
```

General Linear Model: FD versus ECOTYPE, tube

```
Factor          Type Levels  Values
Line           fixed      3     1 2 3
rep(Line)     random     11     1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for FD, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    169.627    170.933     85.466   20.03  0.001
rep(Line)        8     33.087     33.087      4.136    1.74  0.098
Error          101    239.705    239.705      2.373
Total          111    442.420
```

General Linear Model: HT@F versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed       3   1 2 3
rep(Line)      random     11   1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for HT@F, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    15927.1    14902.0     7451.0   15.12  0.002
rep(Line)        8     3791.5     3791.5      473.9    2.16  0.037
Error          101    22149.7    22149.7      219.3
Total          111    41868.3
```

General Linear Model: RL@F versus ECOTYPE, tube

```
Factor          Type   Levels  Values
Line           fixed       3   1 2 3
rep(Line)      random     11   1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for RL@F, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    663.632    650.168    325.084  104.07  0.000
rep(Line)        8     24.000     24.000      3.000    2.24  0.030
Error          101    135.081    135.081      1.337
Total          111    822.714
```

General Linear Model: CL@F versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed       3   1 2 3
rep(Line))     random     11   1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for CL@F, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2      23.15      14.52       7.26    0.48  0.638
rep(Line)        8     127.01     127.01      15.88    0.58  0.792
Error          101    2762.95    2762.95      27.36
Total          111    2913.11
```

General Linear Model: RW@F versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed       3   1 2 3
rep(Line)      random     11   1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for RW@F, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    18645.6    18607.6     9303.8   54.43  0.000
rep(Line)        8     1321.9     1321.9      165.2    1.88  0.072
Error          100     8805.8     8805.8       88.1
Total          110    28773.3
```

General Linear Model: D30 HT versus ECOTYPE, tube

```
Factor          Type  Levels Values
Line           fixed      3   1 2 3
rep(Line)      random     11  1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D30 HT, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2      47700      45445      22723    4.36  0.057
rep(Line)        8      40711      40711       5089    1.51  0.162
Error          101     339889     339889       3365
Total          111     428300
```

General Linear Model: D30 RL versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed      3   1 2 3
rep(Line)      random     11  1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D30 RL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    1021.83    1012.53     506.27  190.30  0.000
rep(Line)        8      20.68      20.68       2.58    1.65  0.120
Error          101     157.98     157.98       1.56
Total          111    1200.49
```

General Linear Model: D30 CL versus ECOTYPE, tube

```
Factor          Type Levels  Values
Line           fixed      3  1 2 3
rep(Line)      random     11 1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D30 CL, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2     446.52     428.58     214.29    3.09  0.105
rep(Line)        8     535.07     535.07      66.88    1.99  0.055
Error          101    3394.69    3394.69      33.61
Total          111    4376.28
```

General Linear Model: D30 RW versus ECOTYPE, tube

```
Factor          Type   Levels Values
Line           fixed      3   1 2 3
rep(Line)      random     11  1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D30 RW, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2    35455.0    35048.9    17524.4  187.42  0.000
rep(Line)        8      760.1      760.1       95.0    0.82  0.583
Error          101    11636.6    11636.6      115.2
Total          111    47851.7
```

General Linear Model: D40HT versus ECOTYPE, tube

```
Factor          Type    Levels Values
Line           fixed       3    1 2 3
rep(Line)      random     11    1  2  3  4  5  6  7  8  9 10 11
Analysis of Variance for D40HT, using Adjusted SS for Tests
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Line             2     490189     478766     239383  200.43  0.000
rep(Line)        8       9828       9828       1229    0.72  0.675
Error          100     171099     171099       1711
Total          110     671117
```

Appendix 03:

ANOVAS for the quantitative traits in chapter 04.

General Linear Model: d.o.g versus gen. line

```
Factor          Type    Levels Values
gen.           fixed      8    1 2 3 4 5 6 7 8
line(gen.)     random    16   1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for d.o.g, using Adjusted SS for Tests
Source              DF     Seq SS     Adj SS     Adj MS       F      P
gen.                 7    354.627    353.597     50.514    8.31  0.004
line (gen.)          8     48.793     48.793      6.099    4.85  0.000
Error              526    661.585    661.585      1.258
Total              541   1065.006
```

General Linear Model: LN - 20 versus gen., line

```
Factor         Type    Levels Values
gen.           fixed      8   1 2 3 4 5 6 7 8
line (gen.)    random    16  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for LN - 20 using Adjusted SS for Tests
Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1    163.613     64.763     64.763  149.36  0.000
gen.                 7     18.959     19.005      2.715    3.47  0.048
line(gen.)           8      6.357      6.357      0.795    1.83  0.069
Error              496    215.071    215.071      0.434
Total              512    404.000
```

General Linear Model: RW - 20 versus gen., line

```
Factor          Type   Levels Values
gen.           fixed      8   1 2 3 4 5 6 7 8
line(gen.)     random    16   1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for RW - 20, using Adjusted SS for Tests
Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1    3282.80    1143.78    1143.78   39.27  0.000
gen.                 7     814.88     839.18     119.88    5.43  0.012
line(gen.)           8     174.66     174.66      21.83    0.75  0.648
Error              496   14445.49   14445.49      29.12
Total              512   18717.83
```

General Linear Model: H - 30 versus gen., line

```
Factor        Type    Levels Values
gen.          fixed      8   1 2 3 4 5 6 7 8
line(gen.)    random    16  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for H - 30, using Adjusted SS for Tests
Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1     45.251     31.495     31.495   20.76  0.000
gen.                 7    106.389    105.872     15.125    4.60  0.022
line(gen.)           8     26.802     26.802      3.350    2.21  0.026
Error              497    753.992    753.992      1.517
Total              513    932.434
```

General Linear Model: H - 35 versus gen., line

```
Factor         Type    Levels Values
gen.           fixed    8     1 2 3 4 5 6 7 8
line(gen.)     random  16     1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for H - 35, using Adjusted SS for Tests
Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1    2001.74     512.15     512.15    6.90  0.009
gen.                 7    9005.70    8994.92    1284.99   12.79  0.001
line(gen.)           8     810.80     810.80     101.35    1.37  0.209
Error              496   36824.43   36824.43      74.24
Total              512   48642.68
```

General Linear Model: F T versus gen., line

```
Factor           Type Levels Values
gen.          fixed      8    1 2 3 4 5 6 7 8
line(gen.)    random    16    1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for FT, using Adjusted SS for Tests
Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1     4677.5     1193.4     1193.4   50.62  0.000
gen.                 7    15493.2    15322.4     2188.9   21.29  0.000
line(gen.)           8      843.4      843.4      105.4    4.47  0.000
Error              495    11670.4    11670.4       23.6
Total              511    32684.5
```

General Linear Model: H F versus gen., line

```
Factor          Type   Levels Values
gen no.        fixed    8    1 2 3 4 5 6 7 8
line(gen.)     random   16   1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for HF, using Adjusted SS for Tests
Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1     8839.7     1171.2     1171.2    5.97  0.015
gen.                 7    26245.5    26488.0     3784.0    5.71  0.012 line(gen.)
8     5439.2     5439.2      679.9    3.47  0.001
Error              494    96831.3    96831.3      196.0
Total              510   137355.7
```

General Linear Model: LNF versus gen., line

```
Factor         Type    Levels Values
gen.          fixed     8    1 2 3 4 5 6 7 8
line(gen.)    random   16    1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for LNF, using Adjusted SS for Tests
Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1    1116.63     194.47     194.47   10.17  0.002
gen.                 7    8599.82    8729.27    1247.04    9.24  0.003
line(gen.)           8    1112.22    1112.22     139.03    7.27  0.000
Error              496    9480.98    9480.98      19.11
Total              512   20309.65
```

General Linear Model: RLF versus gen., line

```
Factor          Type    Levels Values
gen no.        fixed     8     1 2 3 4 5 6 7 8
line(gen.)     random   16     1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for RLF, using Adjusted SS for Tests
Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1    1047.44     191.58     191.58   11.42  0.001
gen.                 7    7341.50    7479.64    1068.52    8.73  0.003
line(gen.)           8    1007.19    1007.19     125.90    7.50  0.000
Error              495    8306.98    8306.98      16.78
Total              511   17703.11
```

General Linear Model: C LF versus gen., line

```
Factor        Type    Levels Values
gen.         fixed     8    1 2 3 4 5 6 7 8
line(gen.)   random   16    1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for C L @ F, using Adjusted SS for Tests

Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1     1.1083     0.0390     0.0390    0.08  0.773
gen.                 7    54.0972    52.4599     7.4943    5.98  0.010
line(gen.)           8    10.2317    10.2317     1.2790    2.74  0.006
Error              495   231.2816   231.2816     0.4672
Total              511   296.7187
```

General Linear Model: R W@F versus gen., line

```
Factor             Type Levels Values
gen.            fixed      8 1 2 3 4 5 6 7 8
line(gen no.) random     16  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for RWF, using Adjusted SS for Tests
Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1    17008.4       29.9       29.9    0.13  0.714
gen.                 7    93191.3    94721.9    13531.7    9.87  0.002
line(gen.)           8    11293.9    11293.9     1411.7    6.33  0.000
Error              496   110662.6   110662.6      223.1
Total              512   232156.2
```

General Linear Model: H - 45 versus gen., line

```
Factor       Type    Levels Values
gen.         fixed      8   1 2 3 4 5 6 7 8
line(gen.)   random    16   1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16

Analysis of Variance for H - 45, using Adjusted SS for Tests

Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1     353027      65657      65657   34.83  0.000
gen.                 7    1446318    1439581     205654   35.90  0.000
line(gen.)           8      46864      46864       5858    3.11  0.002
Error              495     933081     933081       1885
Total              511    2779290
```

General Linear Model: H - 54 versus gen., line

```
Factor        Type      Levels Values
gen.          fixed      8    1 2 3 4 5 6 7 8
line(gen.)    random     16   1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
Analysis of Variance for H - 54, using Adjusted SS for Tests
Source              DF     Seq SS     Adj SS     Adj MS       F      P
d.o.g                1     967469     166899     166899   28.04  0.000
gen.                 7    3896344    3900244     557178   20.25  0.000
line(gen.)           8     225961     225961      28245    4.74  0.000
Error              495    2946819    2946819       5953
Total              511    8036593
```

Appendix 04:

The introgressed regions, number, observed means, weights and expected means of the wide STAIRS (chapter 04).

1) Germination time

STAIRS	Introgression cM	N	Observed mean	Weight	Expected mean
1	80-100	70	7.001	11.48	5.61
2	60-100	34	6.92	5.25	5.61
3	44-100	137	7.19	20.82	5.61
4	20-100	35	7.50	5.74	5.61
5	CSS3	62	6.68	8.53	7.51
6	0-20	137	6.87	21.64	5.43
7	0-44	35	6.96	5.74	5.43
8	0-80	32	6.79	5.08	7.51

2) RW – 20

STAIRS	*Introgression cM*	*N*	*Observed mean*	*Weight*	*Expected mean*
1	80-100	70	16.02	2.40	16.12
2	60-100	34	15.67	1.10	16.12
3	44-100	137	16.29	4.36	16.12
4	20-100	35	20.62	1.20	20.62
5	CSS3	62	14.30	1.79	15.35
6	0-20	137	15.52	4.57	15.35
7	0-44	35	16.27	1.20	15.35
8	0-80	32	15.37	1.06	15.35

3) H – 30

STAIRS	*Introgression cM*	*N*	*Observed mean*	*Weight*	*Expected mean*
1	80-100	70	3.74	20.90	3.62
2	60-100	34	3.74	9.55	3.62
3	44-100	137	4.03	38.21	4.13
4	20-100	35	5.23	10.45	5.23
5	CSS3	62	4.30	15.52	4.54
6	0-20	137	3.93	39.70	4.03
7	0-44	35	4.82	10.45	4.54
8	0-80	32	5.05	9.25	4.54

4) H – 35

STAIRS	*Introgression cM*	*N*	*Observed mean*	*Weight*	*Expected mean*
1	80-100	70	9.91	0.94	10.36
2	60-100	34	7.51	0.43	10.36
3	44-100	137	11.31	1.72	10.36
4	20-100	35	24.07	0.47	24.07
5	CSS3	62	8.62	0.70	7.52
6	0-20	137	6.89	1.78	7.52
7	0-44	35	6.54	0.47	7.52
8	0-80	32	9.45	0.42	7.52

5) FT

STAIRS	*Introgression cM*	*N*	*Observed mean*	*Weight*	*Expected mean*
1	80-100	70	40.62	0.664	40.54
2	60-100	34	41.43	0.304	40.54
3	44-100	137	40.64	1.205	40.54
4	20-100	35	39.17	0.332	40.54
5	CSS3	62	49.47	0.493	51.30
6	0-20	137	53.06	1.252	51.30
7	0-44	35	48.76	0.332	51.30
8	0-80	32	49.76	0.294	51.30

6) HF

STAIRS	*Introgression cM*	*N*	*Observed mean*	*Weight*	*Expected mean*
1	80-100	70	38.55	0.103	37.37
2	60-100	34	31.63	0.047	37.37
3	44-100	137	45.40	0.187	46.30
4	20-100	35	52.44	0.051	46.30
5	CSS3	62	26.76	0.076	26.76
6	0-20	137	30.96	0.194	30.19
7	0-44	35	35.70	0.051	39.12
8	0-80	32	39.73	0.044	39.12

7) RLNF

STAIRS	*Introgression cM*	*N*	*Observed mean*	*Weight*	*Expected mean*
1	80-100	70	13.69	0.50	13.84
2	60-100	34	13.98	0.23	13.84
3	44-100	137	13.87	0.91	13.84
4	20-100	35	13.93	0.25	13.84
5	CSS3	62	19.23	0.37	21.06
6	0-20	137	22.85	0.95	21.06
7	0-44	35	17.91	0.25	21.06
8	0-80	32	20.04	0.22	21.06

8) CLNF

STAIRS	*Introgression cM*	*N*	*Observed mean*	*Weight*	*Expected mean*
1	80-100	70	2.85	54.05	2.85
2	60-100	34	2.72	24.71	2.85
3	44-100	137	2.87	98.07	2.85
4	20-100	35	2.92	27.03	2.85
5	CSS3	62	3.23	40.15	3.46
6	0-20	137	3.55	101.93	3.46
7	0-44	35	3.29	27.03	3.46
8	0-80	32	3.62	23.94	3.46

9) RWF

STAIRS	*Introgression cM*	*N*	*Observed mean*	*Weight*	*Expected mean*
1	80-100	70	67.22	0.05	70.14
2	60-100	34	73.02	0.02	70.14
3	44-100	137	70.74	0.09	70.14
4	20-100	35	71.19	0.02	70.14
5	CSS3	62	106.20	0.04	98.00
6	0-20	137	92.74	0.09	98.00
7	0-44	35	101.67	0.02	98.00
8	0-80	32	102.51	0.02	98.00

10) H – 45

STAIRS	*Introgression cM*	*N*	*Observed mean*	*Weight*	*Expected mean*
1	80-100	70	122.77	0.01	118.39
2	60-100	34	92.45	0.01	118.39
3	44-100	137	122.52	0.02	118.39
4	20-100	35	166.43	0.01	166.43
5	CSS3	62	35.59	0.01	21.23
6	0-20	137	13.64	0.02	21.23
7	0-44	35	15.35	0.01	21.23
8	0-80	32	36.54	0.01	21.23

11) H - 54

STAIRS	*Introgression cM*	*N*	*Observed mean*	*Weight*	*Expected mean*
1	80-100	70	296.22	0.002	279.29
2	60-100	34	293.05	0.001	279.29
3	44-100	137	300.28	0.005	312.52
4	20-100	35	310.55	0.001	312.52
5	CSS3	62	136.46	0.002	146.02
6	0-20	137	100.57	0.005	112.79
7	0-44	35	213.88	0.001	146.02
8	0-80	32	139.86	0.001	146.02

Appendix 5a:

An Example of Least square best fit model fitting from chapter 04

(Wide STAIRS flowering time)

INFORMATION MATRIX
4.8767 0.1327
0.1327 4.8767
RHS
223.3058
-22.6188
VARIANCE COVARIANCE MATRIX
0.2052 -0.0056
-0.0056 0.2052
MODEL

Parameter	Theta	Standard Error	c
m	45.95058957 +/-	0.45299961	101.436
a1	-5.88849391 +/-	0.45299961	-12.999

STATISTIC	OBSERVED	EXPECTED
1	39.6100	40.0621
2	42.1300	40.0621
3	40.3200	40.0621
4	38.1400	40.0621
5	52.5000	51.8391
6	52.6500	51.8391
7	47.6300	51.8391
8	52.0300	51.8391

Chisquared = 9.6745 with 6 df

Appendix 5 b: Line means and SEmeans for wide STAIRS

		d o g		lf.no d.20		RW d 20		H d 30		H d 35		F D		H @F	
Line	Ler region	Mean	SEMean	Mean	SEMean	Mean	SEMean	Mean	SEMean	Mean	SEMean	Mean	SEMean	Mean	SEMean
1	80-100 A	5.200	0.069	7.314	0.107	17.014	0.441	3.914	0.155	10.543	1.183	39.429	0.521	40.457	2.540
2	80-100 B	5.171	0.077	7.171	0.145	17.014	0.511	3.900	0.159	10.600	0.989	39.800	0.435	38.629	2.190
3	60-100 A	6.088	0.136	6.750	0.191	14.969	0.648	3.625	0.140	7.063	0.396	42.125	0.733	30.938	2.095
4	44-100 A	6.406	0.291	6.833	0.225	14.400	0.828	3.806	0.188	10.419	1.565	41.807	0.735	45.774	4.569
5	44- 100 B	5.057	0.057	7.686	0.080	19.071	0.315	4.914	0.381	16.114	2.791	38.486	0.302	50.200	2.156
6	44- 100 C	5.657	0.183	7.424	0.107	16.955	0.459	3.848	0.152	10.758	1.272	39.879	0.417	39.273	2.094
7	44- 100 D	6.314	0.366	7.069	0.148	15.759	0.755	3.690	0.199	8.483	0.687	41.464	0.543	47.250	2.780
8	20-100 A	5.171	0.065	7.743	0.085	21.643	2.919	5.400	0.398	24.743	4.072	38.143	0.442	53.457	2.556
9	CSS3	7.774	0.244	5.917	0.158	11.250	0.646	3.875	0.184	6.625	0.247	51.875	1.096	25.625	1.849
10	CSS3	7.677	0.354	6.036	0.150	11.554	0.629	3.714	0.169	6.714	0.329	53.036	1.099	22.250	1.287
11	0-20 A	5.229	0.101	7.118	0.070	16.191	0.452	3.824	0.181	7.412	0.220	48.559	0.829	32.618	2.012
12	0-20 B	5.457	0.095	7.118	0.110	16.779	1.052	4.059	0.126	7.412	0.287	53.765	1.418	39.206	3.039
13	0-20 C	5.514	0.138	6.771	0.124	15.386	0.517	3.971	0.176	6.914	0.233	55.371	1.102	26.743	1.915
14	0-20 D	5.906	0.145	6.867	0.093	15.283	0.515	4.133	0.171	6.931	0.232	52.862	1.638	26.931	2.429
15	0-44 A	5.114	0.055	7.229	0.072	17.400	0.327	5.000	0.192	7.286	0.195	47.629	0.963	36.829	1.979
16	0-80 A	7.156	0.308	6.161	0.237	12.936	0.792	4.726	0.245	8.065	0.781	52.032	1.155	37.433	2.478

Appendix 06

Analysis of narrow STAIRS - ANOVAS for morphological traits

General Linear Model: FT versus genotype, rep. lines

```
Factor          Type    Levels  Values
genotype       fixed        13  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
rep(genotype)  random       35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
                               13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
                               23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
                              33, 34, 35

Analysis of Variance for FT, using Adjusted SS for Tests

Source                   DF    Seq SS    Adj SS  Adj MS      F      P
Germ.                     1    509.69    245.15  245.15  16.19  0.000
genotype                 12   8119.42   8138.63  678.22  16.06  0.000
rep. lines(genotype)     22    927.46    927.46   42.16   2.78  0.000
Error                   992  15017.52  15017.52   15.14
Total                  1027  24574.09
```

General Linear Model: LN-20 versus genotype, rep. lines

```
Factor          Type    Levels  Values
genotype     fixed        13  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
rep(genotype)random       35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
                            13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
                            23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
                            33, 34, 35

Analysis of Variance for LN-20, using Adjusted SS for Tests
Source                DF    Seq SS    Adj SS   Adj MS       F      P
Germ.                  1   229.186   172.995  172.995  120.46  0.000
genotype              12   154.471   155.469   12.956    4.76  0.001
rep   (genotype)      22    60.014    60.014    2.728    1.90  0.007
Error               1004  1441.819  1441.819    1.436
Total               1039  1885.491
```

General Linear Model: RW-20 versus genotype, rep. lines

```
Factor          Type    Levels  Values
genotype     fixed        13  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
rep(genotype)random       35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
                            13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
                            23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
                            33, 34, 35

Analysis of Variance for RW-20, using Adjusted SS for Tests

Source             DF    Seq SS    Adj SS   Adj MS       F      P
Germ.               1  12612.34   9273.69  9273.69  233.08  0.000
genotype           12   8578.54   8632.29   719.36    6.35  0.000
rep.(genotype)     22   2499.45   2499.45   113.61    2.86  0.000
Error            1004  39945.94  39945.94    39.79
Total            1039  63636.28
```

General Linear Model: BT versus genotype, rep. lines

```
Factor       Type    Levels  Values
genotype    fixed       13  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
rep(genotype)  random   35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
                           13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
                           23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
                           33, 34, 35


Analysis of Variance for BT, using Adjusted SS for Tests

Source                  DF    Seq SS    Adj SS  Adj MS      F      P
Germ.                    1    632.40    360.09  360.09  31.40  0.000
genotype                12   7230.09   7243.79  603.65  13.21  0.000
rep. lines(genotype)    22   1008.59   1008.59   45.85   4.00  0.000
Error                 1002  11490.10  11490.10   11.47
Total                 1037  20361.18
```

General Linear Model: HF versus line, rep. lines

```
Factor     Type    Levels  Values
line       fixed       24  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
                           14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24
rep.line)  random      35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
                           14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
                           25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35


Analysis of Variance for HF, using Adjusted SS for Tests

Source              DF    Seq SS    Adj SS  Adj MS     F      P
Germ.                1    2442.2    1662.6  1662.6  3.23  0.073
line                23   39913.7   39628.5  1723.0  2.66  0.047
rep. lines(line)    11    7139.2    7139.2   649.0  1.26  0.243
Error              987  508290.3  508290.3   515.0
Total             1022  557785.4
```

General Linear Model: RWF versus line, rep. lines

```
Factor    Type    Levels  Values
line     fixed       24  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
                        14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24
rep(line)random      35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
                        14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
                        25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35

Analysis of Variance for RWF, using Adjusted SS for Tests

Source              DF    Seq SS    Adj SS  Adj MS     F      P
line                23   22870.9   22783.6   990.6  2.72  0.044
rep. lines(line)    11    4013.5    4013.5   364.9  2.89  0.001
Error              990  125179.7  125179.7   126.4
Total             1024  152064.2
```

General Linear Model: RLNF versus genotype, rep. lines

```
Factor      Type    Levels  Values
genotype    fixed       13  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
rep(genotype)  random      35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
                           13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
                           23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
                           33, 34, 35


Analysis of Variance for RLNF, using Adjusted SS for Tests

Source                  DF    Seq SS    Adj SS  Adj MS     F      P
Germ.                    1    159.36     65.71   65.71  4.13  0.042
genotype                12   3643.35   3656.69  304.72  5.95  0.000
rep. lines(genotype)    22   1127.87   1127.87   51.27  3.22  0.000
Error                  985  15664.22  15664.22   15.90
Total                 1020  20594.81
```

General Linear Model: RLNF versus genotype, rep. lines

```
Factor      Type    Levels  Values
genotype    fixed       13  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
rep.genotype)  random      35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
                           13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
                           23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
                           33, 34, 35


Analysis of Variance for RLNF, using Adjusted SS for Tests

Source                  DF    Seq SS    Adj SS  Adj MS     F      P
genotype                12   3680.30   3685.63  307.14  5.78  0.000
rep. lines(genotype)    22   1172.58   1172.58   53.30  3.34  0.000
Error                  988  15780.44  15780.44   15.97
Total                 1022  20633.32
```

General Linear Model: CLNF versus genotype, rep. lines

```
Factor       Type    Levels  Values
genotype    fixed       13  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
rep.genotype)  random      35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
                           13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
                           23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
                           33, 34, 35


Analysis of Variance for CLNF, using Adjusted SS for Tests

Source                  DF    Seq SS    Adj SS  Adj MS     F      P
genotype                12    998.53   1017.98   84.83  3.34  0.007
rep. lines(genotype)    22    561.60    561.60   25.53  1.87  0.009
Error                  978  13317.82  13317.82   13.62
Total                 1012  14877.96
```

General Linear Model: H-30 versus genotype, rep. lines

```
Factor       Type    Levels  Values
genotype     fixed       13  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
rep.genotype)random      35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
                          13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
                          23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
                          33, 34, 35


Analysis of Variance for H-30, using Adjusted SS for Tests

Source                   DF  Seq SS  Adj SS  Adj MS        F      P
Germ.                     1    9453    4174    4174    13.83  0.000
genotype                 12  598911  598506   49876  2945.37  0.000
rep. lines(genotype)     22     367     367      17     0.06  1.000
Error                   998  301247  301247     302
Total                  1033  909978
```

General Linear Model: H-30 versus genotype, rep. lines

```
Factor       Type    Levels  Values
genotype     fixed       13  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
rep.genotype)random      35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
                          13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
                          23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
                          33, 34, 35


Analysis of Variance for H-30, using Adjusted SS for Tests

Source                   DF  Seq SS  Adj SS  Adj MS        F      P
Germ.                     1    9453    4174    4174    13.83  0.000
genotype                 12  598911  598506   49876  2945.37  0.000
rep. lines(genotype)     22     367     367      17     0.06  1.000
Error                   998  301247  301247     302
Total                  1033  909978
```

General Linear Model: H-36 versus genotype, rep. lines

```
Factor    Type    Levels  Values
genotype fixed       13  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
rep.(genotype)random      35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
                           13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
                           23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
                           33, 34, 35


Analysis of Variance for H-36, using Adjusted SS for Tests

Source                   DF   Seq SS   Adj SS  Adj MS        F      P
Germ.                     1   104708    37456   37456   18.10  0.000
genotype                 12  6295141  6297182  524765  108.49  0.000
rep. lines(genotype)     22   106543   106543    4843    2.34  0.000
Error                   999  2066859  2066859    2069
Total                  1034  8573250
```

General Linear Model: H- final versus genotype, line

```
Factor          Type    Levels  Values
genotype        fixed   13  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
line(genotype)  random  24  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
                            14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24


Analysis of Variance for H- final, using Adjusted SS for Tests

Source              DF   Seq SS   Adj SS  Adj MS     F      P
Germ.                1    42770    34869   34869  9.26  0.002
genotype            12   348396   345164   28764  2.17  0.096
line(genotype)      11   169448   169448   15404  4.09  0.000
Error             1004  3782516  3782516    3767
Total             1028  4343130
```

General Linear Model: H- final versus line, rep. lines

```
Factor  Type    Levels  Values
line    fixed       24  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
                        14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24
rep. lines(line)  random      35  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,
                      12, 13,14, 15, 16, 17, 18, 19, 20, 21, 22,
                      23, 24,25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35


Analysis of Variance for H- final, using Adjusted SS for Tests

Source              DF   Seq SS   Adj SS  Adj MS     F      P
Germ.                1    42770    29460   29460  7.83  0.005
line                23   517844   519297   22578  5.24  0.003
rep. lines(line)    11    47368    47368    4306  1.14  0.322
Error              993  3735148  3735148    3761
Total             1028  4343130
```

Appendix 6 b: Line means and SEmeans for narrow STAIRS

			BT		FT		LN20		RW20		RLNF		CLNF		HT-30	
line	genotype	Line name	mean	SEMean	mean	SEMean	mean	SEMean	mean	SEMean	mean	SEMean	mean	SEMean	mean	SEMean
Col	1	Col	25.23	0.29	28.27	0.30	11.27	0.18	11.27	0.18	14.23	0.28	3.50	0.10	135.13	12.46
1	2	814/11	28.70	0.81	32.07	0.83	10.96	0.25	10.96	0.25	16.00	0.51	4.12	0.16	58.60	10.23
2	3	628/7	33.97	0.46	37.57	0.55	10.20	0.21	10.20	0.21	20.80	0.42	4.87	0.32	6.67	0.18
3	4	982/4	27.87	0.70	31.33	0.74	10.13	0.21	10.13	0.21	15.93	0.65	4.10	0.37	66.97	10.84
4	5	982/6	37.00	0.77	40.00	0.94	9.23	0.36	9.23	0.36	21.68	0.78	8.70	1.50	6.45	0.14
5	6	982/8	35.77	0.62	40.17	0.75	9.67	0.15	9.67	0.15	22.30	0.49	7.53	0.99	6.67	0.17
6	7	58/8	36.10	0.43	39.87	0.51	9.77	0.19	9.77	0.19	21.43	0.38	6.23	0.27	6.63	0.12
7	8	1008/6	35.47	0.51	39.43	0.59	9.80	0.14	9.80	0.14	21.38	0.56	6.07	0.35	6.63	0.12
8	9	6/5	34.17	0.47	37.59	0.55	9.70	0.27	9.70	0.27	20.62	0.39	5.86	0.67	7.10	0.21
9	10	255/1	35.53	0.64	39.77	0.69	9.60	0.20	9.60	0.20	21.43	0.42	7.90	0.98	6.50	0.16
10	11	255/2	37.14	0.73	40.11	0.61	8.59	0.44	8.59	0.44	22.35	2.43	5.92	0.22	6.46	0.17
11	12	432/3	35.17	0.36	39.23	0.52	9.80	0.19	9.80	0.19	20.50	0.37	5.90	0.44	6.87	0.22
12	13	432/4	33.43	0.67	36.66	0.59	10.00	0.19	10.00	0.19	19.07	0.37	5.77	0.36	7.57	0.30
13	14	879/1	37.10	0.88	38.70	1.57	10.20	0.14	10.20	0.14	21.03	0.53	8.77	1.20	6.57	0.12
14	15	879/5	36.63	0.92	39.82	0.87	9.37	0.19	9.37	0.19	20.57	0.63	6.46	0.72	9.23	2.75
15	16	879/7	36.30	0.72	39.93	0.77	9.70	0.19	9.70	0.19	20.60	0.51	6.57	0.57	6.80	0.25
16	17	947/5	35.97	0.65	40.18	0.87	9.80	0.16	9.80	0.16	20.21	0.41	8.21	1.74	6.53	0.13
17	18	947/6	34.93	0.65	38.86	0.72	9.67	0.27	9.67	0.27	23.10	2.26	7.24	1.00	6.80	0.21
19	20	1006/6	38.97	0.56	42.69	0.42	8.97	0.14	8.97	0.14	24.97	0.46	7.03	0.17	6.17	0.12
20	21	1006/8	35.36	0.49	39.68	0.55	9.82	0.29	9.82	0.29	21.89	0.88	6.50	0.53	6.78	0.14
21	22	1155/7	35.53	0.61	39.21	0.72	9.93	0.17	9.93	0.17	21.13	0.75	6.61	0.46	6.67	0.13
22	23	1270/1	36.43	0.62	40.33	0.67	9.60	0.16	9.60	0.16	21.00	0.33	6.67	0.42	6.63	0.17
23	24	35/4	34.97	0.59	39.03	0.72	10.30	0.14	10.30	0.14	20.93	0.47	6.60	0.62	6.93	0.14
24	25	749/3	37.83	0.83	40.65	1.39	9.87	0.18	9.87	0.18	19.89	0.92	9.04	1.01	6.43	0.12
25	26	1101/5	34.67	0.56	38.07	0.52	9.53	0.18	9.53	0.18	19.37	0.43	5.83	0.21	6.83	0.15
26	27	855/2	33.77	0.59	37.60	0.67	9.40	0.28	9.40	0.28	20.37	0.74	6.00	0.53	7.57	0.57
27	28	855/9	34.70	0.67	38.37	0.69	9.17	0.36	9.17	0.36	20.52	0.51	6.14	0.52	6.90	0.26
28	29	354/7	33.67	0.51	37.23	0.55	9.63	0.22	9.63	0.22	18.60	0.39	5.90	0.58	7.00	0.21
29	30	308.00	34.07	0.63	37.48	0.54	9.90	0.17	9.90	0.17	19.59	0.38	5.55	0.20	7.13	0.18
30	31	854/4	32.77	0.46	37.03	0.82	9.70	0.17	9.70	0.17	20.21	0.39	5.14	0.30	7.27	0.19
31	32	854/5	35.17	0.68	39.38	0.81	9.23	0.38	9.23	0.38	21.86	0.59	6.52	0.72	7.55	0.75
32	33	1171/3	36.57	0.51	40.03	0.84	9.70	0.23	9.70	0.23	22.77	0.72	6.90	0.41	6.79	0.10
33	34	1276/2	34.57	0.45	38.93	0.61	9.10	0.23	9.10	0.23	19.23	0.46	6.13	0.58	6.73	0.18
34	35	410/3	32.40	0.55	35.90	0.57	10.10	0.22	10.10	0.22	18.21	0.58	5.14	0.32	8.03	0.45
35	36	0-21C	31.13	0.59	34.90	0.70	10.30	0.21	10.30	0.21	16.97	0.44	4.47	0.33	11.17	1.60
36	37	0-21E	33.27	0.62	36.70	0.67	9.40	0.27	9.40	0.27	17.47	0.49	5.83	0.81	7.47	0.48
37	38	CSS3	27.73	0.64	31.17	0.64	10.67	0.26	10.67	0.26	15.43	0.46	4.10	0.26	51.57	10.22

Appendix 07

Total RNA Extraction Procedure- Gene expression Analysis

(Follows the procedure for Isolation of total RNA from Plant Cells and tissues described in Qiagen RNeasy Mini Handbook

(All the buffers and columns were from the Qiagen RNA extraction kit.

1) ß Mercaptoethanol was added to buffer RLT (10 µl-ME per 1 ml buffer RLT).
2) Weighed 80-100 mg of plant material and added 500 µl of the RLT buffer in a 1.5 µl eppendorf tube to which a 2mm ball-bearing was added.
3) Placed the tubes quickly in a mill box (without letting the sample thaw at any time) and milled for 8 minutes until the material was homogenized.
4) Pooling of lysates from 5 samples was done to make a single sample at this stage. 100 µl of each of 5 samples from the replicate was pippetted out onto a QIAshredder spin column (lilac) placed in a 2 ml collection tube, and centrifuged for 2 min at maximum speed. The supernatant of the flow-through was transferred to a new centrifuge tube without disrupting the cell-debris pellet in the collection tube.
5) 0.5 volume of ethanol was added to the lysate, mixed immediately by pipetting.
6) The sample was applied to an RNeasy mini column (pink) and centrifuged for 15 seconds at 10,000 rpm (approximately 8000 x g).
7) The flow through was then discarded and 700 µl Buffer RW1 added to the column and centrifuged for a further 15 seconds at 10,000 rpm (approximately 8000 x g). Both the flow through and the collection tube were then discarded.
8) The RNeasy column was transferred to a new 2 ml collection tube, 500 µl Buffer RPE added and centrifuged for 15 seconds at 10,000 rpm (approximately 8000 x g).
9) The flow through was discarded and this step was repeated by adding another 500 µl Buffer RPE and centrifuged for 2 minutes at 10,000 rpm (approximately 8000 x g), the flow through discarded and centrifuged again for a further 1 minute.
10) To elute the RNA the RNeasy column was transferred to a new 2 ml collection tube, 30 µl RNase free water pipetted onto the RNeasy silica-gel.
11) The elution was repeated with another 30 µl RNase free water.

Appendix 08: Significant, differentially expressed gene lists from expression analysis (GeneSpring) experiment.

Experiment 01

Gene Name	t-test p-value	TAIR Accession	Description
mwgarabidopsis#08161	0.000793	At5g43440	2-oxoglutarate-dependent dioxygenase, putative
mwgarabidopsis#13973	0.000942	At3g44080	F-box family protein
mwgarabidopsis#03050	0.000996	At2g14750	adenylylsulfate kinase 1 (AKN1)
mwgarabidopsis#20563	0.00241	At2g01740	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#17895	0.00292	0	
mwgarabidopsis#11639	0.00365	At1g31000	F-box family protein
mwgarabidopsis#06431	0.00413	At4g21760	glycosyl hydrolase family 1 protein
mwgarabidopsis#09218	0.00435	At3g02080	40S ribosomal protein S19 (RPS19A)
mwgarabidopsis#09067	0.00494	At5g66750	SNF2 domain-containing protein / helicase domain-containing protein
mwgarabidopsis#07366	0.00559	At5g13180	no apical meristem (NAM) family protein
mwgarabidopsis#20420	0.00563	At4g16490	armadillo/beta-catenin repeat family protein
mwgarabidopsis#22456	0.00654	At3g05670	PHD finger family protein
mwgarabidopsis#23851	0.00691	At3g47980	integral membrane HPP family protein
mwgarabidopsis#11884	0.00929	0	
mwgarabidopsis#17111	0.011	At5g11610	exostosin family protein
mwgarabidopsis#23336	0.0116	At3g22160	VQ motif-containing protein
mwgarabidopsis#10499	0.0123	At1g06890	transporter-related
mwgarabidopsis#07969	0.0148	At5g38120	4-coumarate--CoA ligase protein / 4-coumaroyl-CoA synthase family protein
mwgarabidopsis#02196	0.0151	At1g69370	chorismate mutase, putative
mwgarabidopsis#18857	0.0156	At5g56490	FAD-binding domain-containing protein
mwgarabidopsis#04899	0.0158	At3g11010	disease resistance family protein / LRR family protein
mwgarabidopsis#15394	0.0162	At4g11180	disease resistance-responsive family protein / dirigent family protein
mwgarabidopsis#10356	0.0178	At1g03910	expressed protein
mwgarabidopsis#00245	0.018	At1g01460	phosphatidylinositol-4-phosphate 5-kinase family protein
mwgarabidopsis#14060	0.0189	At3g45740	hydrolase family protein / HAD-superfamily protein
mwgarabidopsis#08751	0.0191	At5g59120	subtilase family protein
mwgarabidopsis#00262	0.0194	At1g01910	anion-transporting ATPase, putative
mwgarabidopsis#08095	0.0201	At5g41550	disease resistance protein (TIR-NBS-LRR class), putative
mwgarabidopsis#02892	0.021	0	
mwgarabidopsis#17001	0.0218	At5g09460	expressed protein
mwgarabidopsis#00928	0.0223	At1g20950	pyrophosphate--fructose-6-phosphate 1-phosphotransferase-related
mwgarabidopsis#10763	0.0224	At1g12460	leucine-rich repeat transmembrane protein kinase, putative
mwgarabidopsis#04603	0.0226	At3g03250	UTP--glucose-1-phosphate uridylyltransferase, putative / UDP-glucose
mwgarabidopsis#19468	0.0228	At4g01610	cathepsin B-like cysteine protease, putative
mwgarabidopsis#16916	0.0229	At5g06790	expressed protein
mwgarabidopsis#15981	0.0231	At4g26800	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#09925	0.0238	At4g18370	protease HhoA, chloroplast (SPPA) (HHOA)
mwgarabidopsis#00356	0.025	At1g05030	hexose transporter, putative
mwgarabidopsis#11804	0.0258	At1g35030	hypothetical protein
mwgarabidopsis#19046	0.0267	At5g61120	expressed protein
mwgarabidopsis#00703	0.0269	At1g14190	glucose-methanol-choline (GMC) oxidoreductase family protein
mwgarabidopsis#19951	0.027	At1g54310	expressed protein
mwgarabidopsis#05040	0.0271	At3g14470	disease resistance protein (NBS-LRR class), putative
mwgarabidopsis#08941	0.0273	At5g63810	beta-galactosidase, putative / lactase, putative
mwgarabidopsis#15362	0.0274	At4g10730	protein kinase family protein
mwgarabidopsis#01068	0.0275	At1g26150	protein kinase family protein
mwgarabidopsis#00820	0.0292	At1g17550	protein phosphatase 2C-related / PP2C-related
mwgarabidopsis#15221	0.0294	0	
mwgarabidopsis#14684	0.0298	At3g58890	syntaxin-related family protein
mwgarabidopsis#03190	0.0301	At2g18130	purple acid phosphatase (PAP11)
mwgarabidopsis#13996	0.0309	At3g44520	esterase/lipase/thioesterase family protein
mwgarabidopsis#00162	0.031	At1g51740	syntaxin 81 (SYP81)
mwgarabidopsis#20938	0.0311	At2g18000	YEATS family protein

mwgarabidopsis#21889	0.0322	At2g40850	phosphatidylinositol 3- and 4-kinase family protein
mwgarabidopsis#04450	0.0329	At2g46860	inorganic pyrophosphatase, putative (soluble)/
mwgarabidopsis#21704	0.0334	At2g37070	expressed protein
mwgarabidopsis#15295	0.0335	At4g09420	disease resistance protein (TIR-NBS class), putative
mwgarabidopsis#07214	0.0341	At5g07990	flavonoid 3'-monooxygenase / flavonoid 3'-hydroxylase (F3'H)
mwgarabidopsis#20642	0.0344	At2g03800	expressed protein
mwgarabidopsis#11248	0.0347	At1g21670	expressed protein
mwgarabidopsis#21989	0.0351	At2g43060	expressed protein
mwgarabidopsis#06855	0.0353	At4g36670	mannitol transporter, putative
mwgarabidopsis#12554	0.0354	At1g61030	expressed protein
mwgarabidopsis#01440	0.036	At1g44170	aldehyde dehydrogenase, putative (ALDH)
mwgarabidopsis#02450	0.036	At1g75350	ribosomal protein L31 family protein
mwgarabidopsis#22621	0.0372	At3g09320	zinc finger (DHHC type) family protein
mwgarabidopsis#07227	0.0378	At5g08300	succinyl-CoA ligase (GDP-forming) alpha-chain, mitochondrial, putative
mwgarabidopsis#14075	0.0399	At3g46180	UDP-galactose/UDP-glucose transporter-related
mwgarabidopsis#21888	0.04	At2g40830	zinc finger (C3HC4-type RING finger) family protein
mwgarabidopsis#10623	0.0404	At1g09610	expressed protein
mwgarabidopsis#18232	0.0404	At5g41470	hypothetical protein
mwgarabidopsis#10426	0.0408	At1g05350	thiF family protein
mwgarabidopsis#11151	0.0419	At1g19950	abscisic acid-responsive HVA22 family protein
mwgarabidopsis#04580	0.042	At3g02740	aspartyl protease family protein
mwgarabidopsis#19880	0.0427	0	
mwgarabidopsis#18259	0.0429	At5g42070	expressed protein
mwgarabidopsis#05516	0.043	At3g52160	beta-ketoacyl-CoA synthase family protein
mwgarabidopsis#04141	0.0436	At2g39980	transferase family protein
mwgarabidopsis#06077	0.0436	At4g11320	cysteine proteinase, putative
mwgarabidopsis#09812	0.0436	At4g30280	xyloglucan:xyloglucosyl transferase, putative
mwgarabidopsis#08195	0.0438	At5g44300	dormancy/auxin associated family protein
mwgarabidopsis#21317	0.0439	At2g27490	dephospho-CoA kinase family
mwgarabidopsis#22732	0.0439	At3g11770	expressed protein
mwgarabidopsis#00860	0.0459	At1g18500	2-isopropylmalate synthase, putative
mwgarabidopsis#12080	0.0461	At1g49120	AP2 domain-containing transcription factor, putative
mwgarabidopsis#19033	0.0461	At5g60830	bZIP transcription factor family protein
mwgarabidopsis#11960	0.0465	At2g10350	Ulp1 protease family protein
mwgarabidopsis#08938	0.0469	At5g63750	IBR domain-containing protein
mwgarabidopsis#01019	0.0487	At1g23800	aldehyde dehydrogenase, mitochondrial (ALDH3)
mwgarabidopsis#22599	0.0487	At3g08850	transducin family protein / WD-40 repeat family protein
mwgarabidopsis#02440	0.0491	At1g75030	pathogenesis-related thaumatin family protein
mwgarabidopsis#06662	0.0512	At4g29820	expressed protein
mwgarabidopsis#12904	0.0524	At1g68180	zinc finger (C3HC4-type RING finger) family protein
mwgarabidopsis#07393	0.0528	At5g13790	floral homeotic protein AGL-15 (AGL15)
mwgarabidopsis#01065	0.0545	0	
mwgarabidopsis#13381	0.0554	At1g78940	protein kinase family protein
mwgarabidopsis#21632	0.0561	At2g35310	transcriptional factor B3 family protein
mwgarabidopsis#14973	0.0562	At4g00240	phospholipase D beta 2 / PLD beta 2 (PLDBETA2) / PLDdelta1
mwgarabidopsis#22939	0.0564	At3g15480	expressed protein
mwgarabidopsis#13326	0.0565	At1g77710	expressed protein
mwgarabidopsis#05529	0.0572	At3g52880	monodehydroascorbate reductase, putative
mwgarabidopsis#07619	0.0577	At5g19920	transducin family protein / WD-40 repeat family protein
mwgarabidopsis#00365	0.0585	At1g05200	glutamate receptor family protein (GLR3.4)
mwgarabidopsis#23915	0.0585	At3g52500	aspartyl protease family protein
mwgarabidopsis#19856	0.0588	At1g03490	no apical meristem (NAM) family protein
mwgarabidopsis#00453	0.0588	At1g07590	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#11709	0.0593	At1g32460	expressed protein
mwgarabidopsis#24129	0.0593	At4g05030	heavy-metal-associated domain-containing protein
mwgarabidopsis#16044	0.0593	At4g27980	expressed protein
mwgarabidopsis#10875	0.0602	At1g14530	tobamovirus multiplication protein 3, putative / TOM3
mwgarabidopsis#16357	0.0617	At4g34530	basic helix-loop-helix (bHLH) family protein
mwgarabidopsis#03784	0.0623	0	

mwgarabidopsis#13392	0.0634	At1g79130	auxin-responsive family protein
mwgarabidopsis#16062	0.0638	At4g28260	expressed protein
mwgarabidopsis#12710	0.0639	At1g64185	lactoylglutathione lyase family protein / glyoxalase I family
mwgarabidopsis#02074	0.0639	At1g66390	myb family transcription factor, putative
mwgarabidopsis#13001	0.0639	At1g70200	RNA recognition motif (RRM)-containing protein
mwgarabidopsis#08981	0.064	At5g64740	cellulose synthase, catalytic subunit, putative
mwgarabidopsis#13209	0.065	At1g75060	expressed protein
mwgarabidopsis#19996	0.0654	At4g15090	far-red impaired response protein (FAR1)
mwgarabidopsis#00532	0.0659	At1g57860	60S ribosomal protein L21
mwgarabidopsis#14141	0.0659	At3g47810	calcineurin-like phosphoesterase family protein
mwgarabidopsis#20909	0.0662	At2g17300	expressed protein
mwgarabidopsis#11572	0.067	At1g29560	expressed protein
mwgarabidopsis#12157	0.0681	At1g50530	hypothetical protein
mwgarabidopsis#20696	0.069	At2g05400	meprin and TRAF homology domain-containing protein
mwgarabidopsis#15742	0.0691	At4g22120	early-responsive to dehydration protein-related / ERD protein-related
mwgarabidopsis#14176	0.0695	At3g48660	hypothetical protein
mwgarabidopsis#21691	0.0697	0	
mwgarabidopsis#14526	0.0704	At3g56370	leucine-rich repeat transmembrane protein kinase, putative
mwgarabidopsis#11699	0.0706	At1g32240	myb family transcription factor (KAN2)
mwgarabidopsis#07102	0.0709	At5g05390	laccase, putative / diphenol oxidase, putative
mwgarabidopsis#02111	0.071	At1g67430	60S ribosomal protein L17 (RPL17B)
mwgarabidopsis#14433	0.0711	At3g54200	expressed protein
mwgarabidopsis#15427	0.0715	At4g11830	phospholipase D gamma 2 / PLD gamma 2 (PLDGAMMA2)
mwgarabidopsis#12598	0.0722	At1g61900	expressed protein
mwgarabidopsis#03243	0.0722	At2g19130	S-locus lectin protein kinase family protein
mwgarabidopsis#14675	0.0732	At3g58770	expressed protein
mwgarabidopsis#08392	0.0739	At5g49360	glycosyl hydrolase family 3 protein
mwgarabidopsis#06771	0.074	At4g33790	acyl CoA reductase, putative
mwgarabidopsis#04079	0.0751	At2g38470	WRKY family transcription factor
mwgarabidopsis#10973	0.0754	At1g16290	expressed protein
mwgarabidopsis#13203	0.0766	At1g74880	expressed protein
mwgarabidopsis#16732	0.0768	At5g02980	kelch repeat-containing F-box family protein
mwgarabidopsis#15992	0.0783	At4g26970	aconitate hydratase, cytoplasmic, putative / citrate hydro-lyase
mwgarabidopsis#23919	0.0791	At3g52610	expressed protein
mwgarabidopsis#05395	0.0793	At3g48100	two-component responsive regulator / response regulator 5 (ARR5)
mwgarabidopsis#16192	0.0794	At4g31530	expressed protein
mwgarabidopsis#24256	0.0795	At4g14850	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#23719	0.0803	At3g29760	NLI interacting factor (NIF) family protein
mwgarabidopsis#07019	0.0804	At5g02900	cytochrome P450, putative
mwgarabidopsis#20184	0.0818	At2g01270	thioredoxin family protein
mwgarabidopsis#13927	0.0819	At3g43290	hypothetical protein
mwgarabidopsis#04134	0.0821	At2g39850	subtilase family protein
mwgarabidopsis#08794	0.0841	At5g60080	protein kinase family protein
mwgarabidopsis#14674	0.085	At3g58760	ankyrin protein kinase, putative
mwgarabidopsis#01730	0.0865	At1g55010	plant defensin-fusion protein, putative (PDF1.5)
mwgarabidopsis#10449	0.0885	At1g05740	hypothetical protein
mwgarabidopsis#10272	0.0886	At1g02530	multidrug resistance P-glycoprotein, putative
mwgarabidopsis#22745	0.089	At3g11964	S1 RNA-binding domain-containing protein
mwgarabidopsis#07328	0.0899	At5g11590	AP2 domain-containing transcription factor, putative
mwgarabidopsis#02202	0.09	At1g69570	Dof-type zinc finger domain-containing protein
mwgarabidopsis#05042	0.0906	At3g14510	geranylgeranyl pyrophosphate synthase, putative / GGPP synthetase
mwgarabidopsis#06401	0.0906	At4g20990	carbonic anhydrase family protein
mwgarabidopsis#20196	0.0909	At1g05805	basic helix-loop-helix (bHLH) family protein
mwgarabidopsis#13111	0.0912	At1g72590	3-oxo-5-alpha-steroid 4-dehydrogenase family protein /
mwgarabidopsis#05898	0.0913	At4g01010	cyclic nucleotide-regulated ion channel, putative (CNGC13)
mwgarabidopsis#13681	0.0924	At3g24480	leucine-rich repeat family protein / extensin family protein
mwgarabidopsis#03427	0.0937	At2g23380	curly leaf protein (CURLY LEAF) / polycomb-group protein
mwgarabidopsis#19188	0.0937	At5g64430	octicosapeptide/Phox/Bem1p (PB1) domain-containing protein
mwgarabidopsis#06292	0.0943	At4g17510	ubiquitin carboxyl-terminal hydrolase, putative /

mwgarabidopsis#09559	0.0949	At1g21140	nodulin, putative
mwgarabidopsis#11879	0.095	0	
mwgarabidopsis#00494	0.0954	At1g08750	GPI-anchor transamidase, putative
mwgarabidopsis#18869	0.0954	At5g56770	hypothetical protein
mwgarabidopsis#01839	0.0955	At1g59780	disease resistance protein (CC-NBS-LRR class), putative
mwgarabidopsis#16085	0.0958	At4g28840	expressed protein
mwgarabidopsis#15079	0.0961	At4g03140	short-chain dehydrogenase/reductase (SDR) family protein
mwgarabidopsis#02310	0.0964	At1g72280	endoplasmic reticulum oxidoreductin 1 (ERO1) family protein
mwgarabidopsis#21671	0.0967	0	
mwgarabidopsis#18707	0.0968	At5g52930	expressed protein
mwgarabidopsis#16738	0.0969	At5g03060	expressed protein
mwgarabidopsis#17038	0.0971	At5g10340	F-box protein-related / SLF-related
mwgarabidopsis#21506	0.0974	0	
mwgarabidopsis#20865	0.0976	At2g16030	expressed protein
mwgarabidopsis#05692	0.0976	At3g57050	cystathionine beta-lyase, chloroplast / beta-cystathionase /
mwgarabidopsis#14777	0.0979	At3g60360	expressed protein
mwgarabidopsis#07140	0.0985	At5g06460	ubiquitin activating enzyme 2 (UBA2)
mwgarabidopsis#07970	0.0991	At5g38130	transferase family protein
mwgarabidopsis#12541	0.0993	At1g60610	expressed protein

Experiment 02

Gene Name	t-test p-value	TAIR Accession	Description
mwgarabidopsis#21338	7.52E-05	At2g27950	expressed protein
mwgarabidopsis#01360	0.00085	0	
mwgarabidopsis#13973	0.000942	At3g44080	F-box family protein
mwgarabidopsis#03050	0.000996	At2g14750	adenylylsulfate kinase 1 (AKN1)
mwgarabidopsis#03625	0.00103	At2g28050	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#15447	0.00127	At4g12110	sterol desaturase family protein
mwgarabidopsis#19673	0.00176	At5g28990	hypothetical protein
mwgarabidopsis#07791	0.00229	At5g25420	xanthine/uracil permease family protein
mwgarabidopsis#22736	0.00235	At3g11810	expressed protein
mwgarabidopsis#14496	0.00237	At3g55900	F-box family protein
mwgarabidopsis#18698	0.0025	At5g52720	heavy-metal-associated domain-containing protein
mwgarabidopsis#04688	0.00294	At5g27840	serine/threonine protein phosphatase PP1 isozyme 8
mwgarabidopsis#18207	0.00313	At5g40960	expressed protein
mwgarabidopsis#02392	0.00349	At1g74040	2-isopropylmalate synthase 1 (IMS1)
mwgarabidopsis#11639	0.00365	At1g31000	F-box family protein
mwgarabidopsis#06431	0.00413	At4g21760	glycosyl hydrolase family 1 protein
mwgarabidopsis#09218	0.00435	At3g02080	40S ribosomal protein S19 (RPS19A)
mwgarabidopsis#09067	0.00494	At5g66750	SNF2 domain-containing protein / helicase domain
mwgarabidopsis#16774	0.00495	At5g03540	exocyst subunit EXO70 family protein
mwgarabidopsis#07366	0.00559	At5g13180	no apical meristem (NAM) family protein
mwgarabidopsis#20420	0.00563	At4g16490	armadillo/beta-catenin repeat family protein
mwgarabidopsis#23458	0.00586	0	
mwgarabidopsis#24200	0.00592	0	
mwgarabidopsis#09224	0.00608	At4g30800	40S ribosomal protein S11 (RPS11B)
mwgarabidopsis#24564	0.00642	At4g34060	expressed protein
mwgarabidopsis#08003	0.00647	At5g38960	germin-like protein, putative
mwgarabidopsis#23076	0.00653	At3g17980	C2 domain-containing protein
mwgarabidopsis#22456	0.00654	At3g05670	PHD finger family protein
mwgarabidopsis#23851	0.00691	At3g47980	integral membrane HPP family protein
mwgarabidopsis#04279	0.00713	At2g43000	no apical meristem (NAM) family protein
mwgarabidopsis#18467	0.0074	At5g47440	expressed protein
mwgarabidopsis#06950	0.0076	At4g39210	glucose-1-phosphate adenylyltransferase (APL3)

mwgarabidopsis#17693	0.00764	At5g25470	expressed protein
mwgarabidopsis#02122	0.00857	At1g67690	peptidase M3 family protein / thimet oligopeptidase
mwgarabidopsis#14273	0.00898	At3g50750	brassinosteroid signalling positive regulator-related
mwgarabidopsis#09469	0.00905	At2g30540	glutaredoxin family protein
mwgarabidopsis#11884	0.00929	0	
mwgarabidopsis#12564	0.00936	At1g61160	expressed protein
mwgarabidopsis#21948	0.00954	At2g42210	mitochondrial import inner membrane translocase t
mwgarabidopsis#23562	0.00974	At3g26910	hydroxyproline-rich glycoprotein family protein
mwgarabidopsis#13616	0.0101	At4g20680	receptor-like protein kinase-related
mwgarabidopsis#04340	0.0102	At5g44640	glycosyl hydrolase family 1 protein
mwgarabidopsis#15485	0.0103	At4g12820	F-box family protein
mwgarabidopsis#20719	0.0105	At2g06520	membrane protein, putative
mwgarabidopsis#17111	0.011	At5g11610	exostosin family protein
mwgarabidopsis#00309	0.0112	At1g03520	glycosyltransferase family 14 protein / core-2/I- enzyme
mwgarabidopsis#09788	0.0115	At1g35670	calcium-dependent protein kinase 2 (CDPK2)
mwgarabidopsis#06109	0.0115	At4g12720	MutT/nudix family protein
mwgarabidopsis#10336	0.0116	At1g07050	CONSTANS-like protein-related
mwgarabidopsis#23336	0.0116	At3g22160	VQ motif-containing protein
mwgarabidopsis#18734	0.0117	At5g53440	expressed protein
mwgarabidopsis#24025	0.0118	At4g01290	expressed protein
mwgarabidopsis#24215	0.0118	At4g13800	permease-related
mwgarabidopsis#12435	0.0119	At1g56270	hypothetical protein
mwgarabidopsis#16197	0.012	At4g31600	UDP-glucuronic acid/UDP-N-acetylgalactosamine
mwgarabidopsis#10499	0.0123	At1g06890	transporter-related
mwgarabidopsis#10138	0.0126	At5g14370	expressed protein
mwgarabidopsis#21680	0.0128	At2g36400	expressed protein
mwgarabidopsis#17905	0.0129	0	
mwgarabidopsis#14245	0.0129	At3g50090	exonuclease family protein
mwgarabidopsis#12738	0.0137	At1g64700	expressed protein
mwgarabidopsis#07969	0.0148	At5g38120	4-coumarate--CoA ligase family protein
mwgarabidopsis#19581	0.0151	At1g24250	paired amphipathic helix repeat-containing protein
mwgarabidopsis#02196	0.0151	At1g69370	chorismate mutase, putative
mwgarabidopsis#05366	0.0151	At3g47170	transferase family protein
mwgarabidopsis#18857	0.0156	At5g56490	FAD-binding domain-containing protein
mwgarabidopsis#04899	0.0158	At3g11010	disease resistance family protein / LRR
mwgarabidopsis#04596	0.0161	At3g03000	calmodulin, putative
mwgarabidopsis#10123	0.0162	At2g39190	ABC1 family protein
mwgarabidopsis#15394	0.0162	At4g11180	disease resistance-responsive family protein /
mwgarabidopsis#00468	0.0164	At1g07910	expressed protein
mwgarabidopsis#21779	0.0166	At2g38590	F-box family protein
mwgarabidopsis#08593	0.0166	At4g26720	serine/threonine protein phosphatase PP-X isozyme 1
mwgarabidopsis#05708	0.0172	At3g57520	alkaline alpha galactosidase, putative
mwgarabidopsis#24160	0.0172	At4g08880	Ulp1 protease family protein
mwgarabidopsis#02165	0.0175	At1g68570	proton-dependent oligopeptide transport (POT)
mwgarabidopsis#20128	0.0177	At2g33735	DNAJ heat shock N-terminal domain-containing protein
mwgarabidopsis#10356	0.0178	At1g03910	expressed protein
mwgarabidopsis#14479	0.0182	At3g55510	expressed protein
mwgarabidopsis#21383	0.0184	At2g29240	Ulp1 protease family protein
mwgarabidopsis#07257	0.0188	At5g09590	heat shock protein 70 / HSP70 (HSC70-5)
mwgarabidopsis#14060	0.0189	At3g45740	hydrolase family protein / HAD-superfamily protein
mwgarabidopsis#10976	0.0191	At1g16360	LEM3 (ligand-effect modulator 3) / CDC50 family
mwgarabidopsis#21146	0.0191	At2g23160	F-box family protein
mwgarabidopsis#08008	0.0191	At5g39030	protein kinase family protein
mwgarabidopsis#08751	0.0191	At5g59120	subtilase family protein
mwgarabidopsis#11037	0.0194	At1g17760	suppressor of forked / SUF family protein
mwgarabidopsis#01061	0.0195	At1g25490	serine/threonine protein phosphatase 2A (PP2A
mwgarabidopsis#21437	0.0196	At2g30820	expressed protein
mwgarabidopsis#04909	0.0197	At3g11250	60S acidic ribosomal protein P0 (RPP0C)
mwgarabidopsis#05191	0.0198	At3g26280	cytochrome P450 family protein

mwgarabidopsis#14926	0.0201	At3g62900	expressed protein
mwgarabidopsis#08095	0.0201	At5g41550	disease resistance protein (TIR-NBS-LRR class), putative
mwgarabidopsis#02892	0.021	0	
mwgarabidopsis#08701	0.021	At5g58210	hydroxyproline-rich glycoprotein family protein
mwgarabidopsis#19948	0.0212	0	
mwgarabidopsis#17105	0.0212	At5g11540	FAD-binding domain-containing protein
mwgarabidopsis#19274	0.0212	At5g66830	F-box family protein
mwgarabidopsis#11247	0.0214	At1g21640	ATP-NAD kinase family protein
mwgarabidopsis#04054	0.0216	At2g37770	aldo/keto reductase family protein
mwgarabidopsis#10582	0.0217	0	
mwgarabidopsis#24385	0.0222	At4g19550	expressed protein
mwgarabidopsis#00928	0.0223	At1g20950	pyrophosphate--fructose-6-phosphate 1
mwgarabidopsis#10763	0.0224	At1g12460	leucine-rich repeat transmembrane protein kinase,
mwgarabidopsis#21377	0.0224	At2g28890	protein phosphatase 2C family protein /
mwgarabidopsis#19468	0.0228	At4g01610	cathepsin B-like cysteine protease, putative
mwgarabidopsis#03595	0.0229	At2g27220	homeodomain-containing protein
mwgarabidopsis#16916	0.0229	At5g06790	expressed protein
mwgarabidopsis#15981	0.0231	At4g26800	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#04498	0.0233	At2g47750	auxin-responsive GH3 family protein
mwgarabidopsis#15920	0.0233	At4g25660	expressed protein
mwgarabidopsis#07793	0.0234	At5g25450	ubiquinol-cytochrome C reductase complex 14 kDa protein,
mwgarabidopsis#10998	0.0236	At1g16840	expressed protein
mwgarabidopsis#09925	0.0238	At4g18370	protease HhoA, chloroplast (SPPA) (HHOA)
mwgarabidopsis#23072	0.0245	At3g17900	expressed protein
mwgarabidopsis#11344	0.0246	At1g23810	paired amphipathic helix repeat-containing protein
mwgarabidopsis#00356	0.025	At1g05030	hexose transporter, putative
mwgarabidopsis#12929	0.0253	At1g68630	expressed protein
mwgarabidopsis#21221	0.0255	At2g25310	expressed protein
mwgarabidopsis#17786	0.0257	At5g27370	hypothetical protein
mwgarabidopsis#07789	0.0258	At5g25380	cyclin 3a (CYC3a)
mwgarabidopsis#14577	0.026	At3g57120	protein kinase family protein
mwgarabidopsis#13634	0.0265	At3g22980	elongation factor Tu family protein
mwgarabidopsis#11033	0.0266	At1g17690	expressed protein
mwgarabidopsis#24296	0.0266	At4g16070	lipase class 3 family protein
mwgarabidopsis#07068	0.0267	At5g04370	S-adenosyl-L-methionine:carboxyl methyltransferase protein
mwgarabidopsis#08208	0.0267	At5g44510	disease resistance protein (TIR-NBS-LRR class), putative
mwgarabidopsis#19046	0.0267	At5g61120	expressed protein
mwgarabidopsis#19951	0.027	At1g54310	expressed protein
mwgarabidopsis#09764	0.0271	At1g26730	EXS family protein / ERD1/XPR1/SYG1 family protein
mwgarabidopsis#15504	0.0273	At4g13130	DC1 domain-containing protein
mwgarabidopsis#08941	0.0273	At5g63810	beta-galactosidase, putative / lactase, putative
mwgarabidopsis#22153	0.0274	At2g46790	pseudo-responser 9 (APRR9) / timing of CAB expression
mwgarabidopsis#15362	0.0274	At4g10730	protein kinase family protein
mwgarabidopsis#06303	0.0274	At4g17750	heat shock factor protein 1 (HSF1) /
mwgarabidopsis#01068	0.0275	At1g26150	protein kinase family protein
mwgarabidopsis#01560	0.0276	At1g50490	ubiquitin-conjugating enzyme 20 (UBC20)
mwgarabidopsis#17272	0.0277	At5g14990	hypothetical protein
mwgarabidopsis#16295	0.0281	At4g33360	terpene cyclase/mutase-related
mwgarabidopsis#22244	0.0283	At3g01590	aldose 1-epimerase family protein
mwgarabidopsis#13700	0.029	At3g25160	ER lumen protein retaining receptor family protein
mwgarabidopsis#00820	0.0292	At1g17550	protein phosphatase 2C-related / PP2C-related
mwgarabidopsis#15221	0.0294	0	
mwgarabidopsis#17687	0.0294	At5g25330	hypothetical protein
mwgarabidopsis#18118	0.0294	At5g38750	asparaginyl-tRNA synthetase family
mwgarabidopsis#14684	0.0298	At3g58890	syntaxin-related family protein
mwgarabidopsis#16688	0.0299	At5g02280	synbindin, putative
mwgarabidopsis#12040	0.0301	At1g48310	SNF2 domain-containing protein / helicase domain- protein
mwgarabidopsis#03190	0.0301	At2g18130	purple acid phosphatase (PAP11)
mwgarabidopsis#16518	0.0304	At4g38280	expressed protein

mwgarabidopsis#05458	0.0307	At3g50300	transferase family protein
mwgarabidopsis#13996	0.0309	At3g44520	esterase/lipase/thioesterase protein
mwgarabidopsis#00162	0.031	At1g51740	syntaxin 81 (SYP81)
mwgarabidopsis#10853	0.0311	At1g14150	oxygen evolving enhancer 3 (PsbQ) family protein
mwgarabidopsis#20938	0.0311	At2g18000	YEATS family protein
mwgarabidopsis#19367	0.0314	At1g54530	calcium-binding EF hand family protein
mwgarabidopsis#07098	0.0315	At5g05280	zinc finger (C3HC4-type RING finger) family protein
mwgarabidopsis#11898	0.0316	0	
mwgarabidopsis#07949	0.0318	At5g37500	guard cell outward rectifying K+ channel (GORK)
mwgarabidopsis#19278	0.0319	At5g66950	expressed protein
mwgarabidopsis#05580	0.032	At3g54140	proton-dependent oligopeptide transport (POT) protein
mwgarabidopsis#07644	0.0322	At5g20500	glutaredoxin, putative
mwgarabidopsis#04450	0.0329	At2g46860	inorganic pyrophosphatase, putative (soluble) /
mwgarabidopsis#12728	0.0334	At1g64540	F-box family protein
mwgarabidopsis#21704	0.0334	At2g37070	expressed protein
mwgarabidopsis#15295	0.0335	At4g09420	disease resistance protein (TIR-NBS class), putative
mwgarabidopsis#23833	0.0337	At3g46310	expressed protein
mwgarabidopsis#07670	0.0337	At5g21090	leucine-rich repeat protein, putative
mwgarabidopsis#07214	0.0341	At5g07990	flavonoid 3'-monooxygenase / 3'-hydroxylase (F3'H)
mwgarabidopsis#13390	0.0342	At1g79110	expressed protein
mwgarabidopsis#17252	0.0342	At5g14530	transducin family protein / WD-40 repeat protein
mwgarabidopsis#12463	0.0343	At1g57770	amine oxidase family
mwgarabidopsis#06206	0.0343	At4g15475	F-box family protein (FBL4)
mwgarabidopsis#20642	0.0344	At2g03800	expressed protein
mwgarabidopsis#18529	0.0344	At5g48990	kelch repeat-containing F-box family protein
mwgarabidopsis#01334	0.0346	At1g34140	polyadenylate-binding protein,/ PABP, putative
mwgarabidopsis#11248	0.0347	At1g21670	expressed protein
mwgarabidopsis#04282	0.0347	At2g43050	pectinesterase family protein
mwgarabidopsis#21989	0.0351	At2g43060	expressed protein
mwgarabidopsis#06855	0.0353	At4g36670	mannitol transporter, putative
mwgarabidopsis#07804	0.0353	At5g25880	malate oxidoreductase, putative
mwgarabidopsis#12554	0.0354	At1g61030	expressed protein
mwgarabidopsis#01440	0.036	At1g44170	aldehyde dehydrogenase, putative (ALDH)
mwgarabidopsis#02450	0.036	At1g75350	ribosomal protein L31 family protein
mwgarabidopsis#24203	0.036	At4g13270	expressed protein
mwgarabidopsis#13462	0.0361	At3g16460	jacalin lectin family protein
mwgarabidopsis#23001	0.0364	At3g16870	zinc finger (GATA type) family protein
mwgarabidopsis#09072	0.0364	At5g66900	disease resistance protein (CC-NBS-LRR class), putative
mwgarabidopsis#01122	0.0366	At1g28040	zinc finger (C3HC4-type RING finger) family protein
mwgarabidopsis#06977	0.0366	At5g01180	proton-dependent oligopeptide transport (POT) protein
mwgarabidopsis#01564	0.0367	At1g50590	pirin, putative
mwgarabidopsis#05115	0.0369	0	
mwgarabidopsis#05193	0.0369	At3g26650	glyceraldehyde 3-phosphate dehydrogenase A,
mwgarabidopsis#22621	0.0372	At3g09320	zinc finger (DHHC type) family protein
mwgarabidopsis#07227	0.0378	At5g08300	succinyl-CoA ligase (GDP-forming) alpha-chain,
mwgarabidopsis#12755	0.0379	At1g65010	expressed protein
mwgarabidopsis#17663	0.0386	At5g24820	aspartyl protease family protein
mwgarabidopsis#13369	0.0387	At1g78750	F-box family protein
mwgarabidopsis#13854	0.0389	0	
mwgarabidopsis#22034	0.0389	At2g44190	expressed protein
mwgarabidopsis#17843	0.0389	At5g28380	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#05452	0.0393	At3g50080	F-box family protein (FBL16)
mwgarabidopsis#00007	0.0394	At4g17785	myb family transcription factor (MYB39)
mwgarabidopsis#05035	0.0397	At3g14415	(S)-2-hydroxy-acid oxidase, peroxisomal,/ glycolate oxidase,
mwgarabidopsis#14075	0.0399	At3g46180	UDP-galactose/UDP-glucose transporter-related
mwgarabidopsis#21888	0.04	At2g40830	zinc finger (C3HC4-type RING finger) family protein
mwgarabidopsis#04783	0.04	At3g07920	eukaryotic translation initiation factor 2 subunit 2,
mwgarabidopsis#09488	0.0401	At5g09810	actin 7 (ACT7) / actin 2
mwgarabidopsis#15824	0.0402	At4g23810	WRKY family transcription factor

mwgarabidopsis#16356	0.0403	At4g34510	fatty acid elongase, putative
mwgarabidopsis#16783	0.0403	At5g03800	exostosin family protein / pentatricopeptide (PPR)
mwgarabidopsis#18777	0.0403	At5g54580	RNA recognition motif (RRM)-containing protein
mwgarabidopsis#02868	0.0404	0	
mwgarabidopsis#18232	0.0404	At5g41470	hypothetical protein
mwgarabidopsis#09546	0.0406	At1g07770	40S ribosomal protein S15A (RPS15aA)
mwgarabidopsis#15838	0.0406	At4g24050	short-chain dehydrogenase/reductase (SDR) family protein
mwgarabidopsis#15185	0.0407	At1g34740	Ulp1 protease family protein
mwgarabidopsis#20543	0.041	At2g01240	reticulon family protein (RTNLB15)
mwgarabidopsis#06233	0.0414	At4g16160	mitochondrial import inner membrane translocase subunit
mwgarabidopsis#18672	0.0414	At5g52060	BAG domain-containing protein
mwgarabidopsis#04009	0.0415	At2g36870	xyloglucan:xyloglucosyl transferase, putative /
mwgarabidopsis#06533	0.0415	At4g25140	glycine-rich protein / oleosin
mwgarabidopsis#23123	0.0419	At3g18670	ankyrin repeat family protein
mwgarabidopsis#04580	0.042	At3g02740	aspartyl protease family protein
mwgarabidopsis#12366	0.0424	At1g55080	expressed protein
mwgarabidopsis#24517	0.0424	At4g30050	expressed protein
mwgarabidopsis#05527	0.0425	At3g52820	purple acid phosphatase (PAP22)
mwgarabidopsis#08400	0.0426	At5g49630	amino acid permease 6 (AAP6)
mwgarabidopsis#19880	0.0427	0	
mwgarabidopsis#24559	0.0427	At4g33690	expressed protein
mwgarabidopsis#20746	0.0428	At2g10340	hypothetical protein
mwgarabidopsis#23207	0.0428	At3g19920	expressed protein
mwgarabidopsis#14546	0.0428	At3g56670	hypothetical protein
mwgarabidopsis#11043	0.0429	At1g17870	expressed protein
mwgarabidopsis#13153	0.043	At1g73650	expressed protein
mwgarabidopsis#06152	0.043	At4g14190	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#04389	0.0432	At2g45340	leucine-rich repeat transmembrane protein kinase, putative
mwgarabidopsis#22430	0.0433	At3g05130	expressed protein
mwgarabidopsis#21171	0.0434	At2g23920	hypothetical protein
mwgarabidopsis#06077	0.0436	At4g11320	cysteine proteinase, putative
mwgarabidopsis#09812	0.0436	At4g30280	xyloglucan:xyloglucosyl transferase, putative /
mwgarabidopsis#08195	0.0438	At5g44300	dormancy/auxin associated family protein
mwgarabidopsis#00633	0.0439	At1g11930	alanine racemase family protein
mwgarabidopsis#22732	0.0439	At3g11770	expressed protein
mwgarabidopsis#10922	0.0443	At1g15490	hydrolase, alpha/beta fold family protein
mwgarabidopsis#22063	0.0445	0	
mwgarabidopsis#15178	0.0447	At4g05210	bacterial transferase hexapeptide repeat-containing protein
mwgarabidopsis#23134	0.0448	At3g18850	phospholipid/glycerol acyltransferase family protein
mwgarabidopsis#19286	0.0448	At5g67110	basic helix-loop-helix (bHLH) family protein
mwgarabidopsis#19429	0.0449	At5g36080	hypothetical protein
mwgarabidopsis#16534	0.045	0	
mwgarabidopsis#12483	0.0452	At1g58330	transcription factor-related
mwgarabidopsis#19582	0.0455	At5g03210	expressed protein
mwgarabidopsis#17364	0.0457	At5g16660	expressed protein
mwgarabidopsis#00860	0.0459	At1g18500	2-isopropylmalate synthase, putative
mwgarabidopsis#19033	0.0461	At5g60830	bZIP transcription factor family protein
mwgarabidopsis#21805	0.0464	At2g39050	hydroxyproline-rich glycoprotein family protein
mwgarabidopsis#11960	0.0465	At2g10350	Ulp1 protease family protein
mwgarabidopsis#15109	0.0468	0	
mwgarabidopsis#08938	0.0469	At5g63750	IBR domain-containing protein
mwgarabidopsis#22353	0.047	At3g03620	MATE efflux family protein
mwgarabidopsis#08951	0.047	At5g64000	3'(2'),5'-bisphosphate nucleotidase, putative /
mwgarabidopsis#07400	0.0475	At5g13980	glycosyl hydrolase family 38 protein
mwgarabidopsis#03586	0.0478	At2g27050	ethylene-insensitive3-like1 (EIL1)
mwgarabidopsis#06403	0.0478	At4g21050	Dof-type zinc finger domain-containing protein
mwgarabidopsis#01418	0.048	At1g43170	60S ribosomal protein L3 (RPL3A)
mwgarabidopsis#09668	0.0486	At1g22440	alcohol dehydrogenase, putative
mwgarabidopsis#01019	0.0487	At1g23800	aldehyde dehydrogenase, mitochondrial (ALDH3)

mwgarabidopsis#02440	0.0491	At1g75030	pathogenesis-related thaumatin family protein
mwgarabidopsis#20748	0.0494	0	
mwgarabidopsis#14435	0.0494	At3g54240	hydrolase, alpha/beta fold family protein
mwgarabidopsis#20624	0.0495	At2g03390	uvrB/uvrC motif-containing protein
mwgarabidopsis#07648	0.0497	At5g20570	ring-box protein-related
mwgarabidopsis#17115	0.0499	At5g11660	hypothetical protein
mwgarabidopsis#02676	0.0506	At1g80770	expressed protein
mwgarabidopsis#21683	0.0506	At2g36430	expressed protein
mwgarabidopsis#00477	0.0507	At1g08170	histone H2B family protein
mwgarabidopsis#03773	0.0509	0	
mwgarabidopsis#06676	0.0509	At4g30270	MERI-5 protein (MERI-5) (MERI5B) / endo-xyloglucan transferase
mwgarabidopsis#14910	0.0511	At3g62590	lipase class 3 family protein
mwgarabidopsis#19245	0.0511	At5g66100	La domain-containing protein
mwgarabidopsis#06662	0.0512	At4g29820	expressed protein
mwgarabidopsis#16944	0.0512	At5g07640	zinc finger (C3HC4-type RING finger) family protein
mwgarabidopsis#20336	0.0514	At4g17890	human Rev interacting-like family protein / hRIP protein
mwgarabidopsis#08729	0.0515	At5g58750	wound-responsive protein-related
mwgarabidopsis#00986	0.0516	At1g22770	gigantea protein (GI)
mwgarabidopsis#17085	0.0517	At5g11270	expressed protein
mwgarabidopsis#21326	0.052	At2g27680	aldo/keto reductase family protein
mwgarabidopsis#20722	0.0522	At2g06860	Ulp1 protease family protein
mwgarabidopsis#12904	0.0524	At1g68180	zinc finger (C3HC4-type RING finger) family protein
mwgarabidopsis#13312	0.0524	At1g77350	expressed protein
mwgarabidopsis#02013	0.0525	At1g64560	hypothetical protein
mwgarabidopsis#21514	0.0525	At2g32590	barren family protein
mwgarabidopsis#11768	0.0526	At1g33980	Smg-4/UPF3 family protein
mwgarabidopsis#07393	0.0528	At5g13790	floral homeotic protein AGL-15 (AGL15)
mwgarabidopsis#18064	0.0528	At5g37570	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#17472	0.0531	At5g19090	heavy-metal-associated domain-containing protein
mwgarabidopsis#07668	0.0532	At5g21030	PAZ domain-containing protein / piwi domain-protein
mwgarabidopsis#04812	0.0535	At3g08950	electron transport SCO1/SenC family protein
mwgarabidopsis#09536	0.0542	At2g27970	cyclin-dependent kinase, putative / CDK, putative
mwgarabidopsis#12888	0.0544	At1g67860	expressed protein
mwgarabidopsis#01065	0.0545	0	
mwgarabidopsis#06989	0.0547	At5g01540	lectin protein kinase, putative
mwgarabidopsis#15006	0.0549	At4g01280	myb family transcription factor
mwgarabidopsis#10011	0.0555	At4g24290	expressed protein
mwgarabidopsis#10509	0.0558	At1g07060	expressed protein
mwgarabidopsis#10267	0.0561	At1g02420	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#21632	0.0561	At2g35310	transcriptional factor B3 family protein
mwgarabidopsis#22939	0.0564	At3g15480	expressed protein
mwgarabidopsis#08298	0.0564	At5g46880	homeobox-leucine zipper protein / Lipid-binding START protein
mwgarabidopsis#09738	0.0565	At5g38950	germin-like protein-related
mwgarabidopsis#18590	0.0566	At5g50200	expressed protein
mwgarabidopsis#10524	0.0568	At1g07310	C2 domain-containing protein
mwgarabidopsis#04904	0.0569	At3g11130	clathrin heavy chain, putative
mwgarabidopsis#05321	0.057	At3g46020	RNA-binding protein, putative
mwgarabidopsis#05655	0.057	At3g55780	glycosyl hydrolase family 17 protein
mwgarabidopsis#21155	0.0572	At2g23420	nicotinate phosphoribosyltransferase family protein / NAPRTase protein
mwgarabidopsis#05529	0.0572	At3g52880	monodehydroascorbate reductase, putative
mwgarabidopsis#11329	0.0574	At1g23440	pyrrolidone-carboxylate peptidase family protein
mwgarabidopsis#09152	0.0574	At5g01320	pyruvate decarboxylase, putative
mwgarabidopsis#07619	0.0577	At5g19920	transducin family protein / WD-40 repeat family protein
mwgarabidopsis#14818	0.058	At3g61070	peroxisomal biogenesis factor 11 protein / PEX11 protein
mwgarabidopsis#15930	0.0581	At4g25835	AAA-type ATPase family protein
mwgarabidopsis#00365	0.0585	At1g05200	glutamate receptor family protein (GLR3.4)
mwgarabidopsis#23915	0.0585	At3g52500	aspartyl protease family protein
mwgarabidopsis#17328	0.0585	At5g15890	expressed protein
mwgarabidopsis#03978	0.0586	At2g35980	harpin-induced family protein (YLS9) / HIN1 family protein /

mwgarabidopsis#23146	0.0587	At3g18990	transcriptional factor B3 family protein
mwgarabidopsis#17203	0.0587	At5g13540	expressed protein
mwgarabidopsis#19856	0.0588	At1g03490	no apical meristem (NAM) family protein
mwgarabidopsis#20710	0.0593	At2g06030	expressed protein
mwgarabidopsis#16044	0.0593	At4g27980	expressed protein
mwgarabidopsis#12138	0.0597	At1g50150	hypothetical protein
mwgarabidopsis#18125	0.0598	At5g38920	hypothetical protein
mwgarabidopsis#20712	0.06	At2g06120	hypothetical protein
mwgarabidopsis#11775	0.0601	At1g34110	leucine-rich repeat transmembrane protein kinase,
mwgarabidopsis#19534	0.0601	At1g56530	hydroxyproline-rich glycoprotein family protein
mwgarabidopsis#21732	0.0601	At2g37750	expressed protein
mwgarabidopsis#10875	0.0602	At1g14530	tobamovirus multiplication protein 3, / TOM3, (THH1)
mwgarabidopsis#10345	0.0603	At1g03770	zinc finger (C3HC4-type RING finger) protein
mwgarabidopsis#01554	0.0603	At1g50360	myosin family protein
mwgarabidopsis#13998	0.0606	At3g44580	hypothetical protein
mwgarabidopsis#12756	0.0607	At1g65020	expressed protein
mwgarabidopsis#20446	0.0612	At5g23405	high mobility group (HMG1/2) family protein
mwgarabidopsis#07201	0.0615	At5g07590	WD-40 repeat protein family
mwgarabidopsis#05932	0.0616	At4g02550	expressed protein
mwgarabidopsis#16357	0.0617	At4g34530	basic helix-loop-helix (bHLH) family protein
mwgarabidopsis#18112	0.0619	At5g38680	kelch repeat-containing F-box family protein
mwgarabidopsis#09068	0.0621	At5g66760	succinate dehydrogenase (ubiquinone) flavoprotein, mitochondrial /
mwgarabidopsis#00972	0.0622	At1g22550	proton-dependent oligopeptide transport (POT) protein
mwgarabidopsis#03784	0.0623	0	
mwgarabidopsis#21709	0.0626	At2g37160	transducin family protein / WD-40 repeat family protein
mwgarabidopsis#14859	0.0626	At3g61720	C2 domain-containing protein
mwgarabidopsis#13140	0.0627	At1g73380	expressed protein
mwgarabidopsis#10061	0.0631	At1g56145	leucine-rich repeat family protein / protein kinase protein
mwgarabidopsis#08136	0.0632	0	
mwgarabidopsis#03160	0.0633	At2g17520	protein kinase family protein / Ire1 homolog-2 (IRE1-2)
mwgarabidopsis#22217	0.0635	At3g01020	iron-sulfur cluster assembly complex protein,
mwgarabidopsis#03245	0.0638	At2g19170	subtilase family protein
mwgarabidopsis#02964	0.0639	0	
mwgarabidopsis#12710	0.0639	At1g64185	lactoylglutathione lyase family protein / glyoxalase I protein
mwgarabidopsis#02074	0.0639	At1g66390	myb family transcription factor, / anthocyanin pigment 2 protein (PAP2)
mwgarabidopsis#15043	0.064	At4g02180	DC1 domain-containing protein
mwgarabidopsis#08589	0.0641	At5g55200	co-chaperone grpE protein, putative
mwgarabidopsis#22994	0.0649	At3g16750	expressed protein
mwgarabidopsis#13209	0.065	At1g75060	expressed protein
mwgarabidopsis#21734	0.065	At2g37800	DC1 domain-containing protein
mwgarabidopsis#22703	0.0651	At3g11210	GDSL-motif lipase/hydrolase family protein
mwgarabidopsis#21980	0.0653	At2g42780	expressed protein
mwgarabidopsis#18085	0.0653	At5g37970	S-adenosyl-L-methionine:carboxyl methyltransferase protein
mwgarabidopsis#19996	0.0654	At4g15090	far-red impaired response protein (FAR1) /
mwgarabidopsis#18513	0.0656	At5g48590	expressed protein
mwgarabidopsis#04530	0.0657	At3g01480	peptidyl-prolyl cis-trans isomerase, / cyclophilin, / rotamase, putative
mwgarabidopsis#11284	0.0658	At1g22420	hydroxyproline-rich glycoprotein family protein
mwgarabidopsis#13920	0.0659	0	
mwgarabidopsis#00532	0.0659	At1g57860	60S ribosomal protein L21
mwgarabidopsis#14141	0.0659	At3g47810	calcineurin-like phosphoesterase family protein
mwgarabidopsis#09797	0.0659	At5g60390	elongation factor 1-alpha / EF-1-alpha
mwgarabidopsis#02479	0.066	At1g75960	AMP-binding protein, putative
mwgarabidopsis#06470	0.066	At4g22850	expressed protein
mwgarabidopsis#20909	0.0662	At2g17300	expressed protein
mwgarabidopsis#13601	0.0663	At3g21650	serine/threonine protein phosphatase 2A (PP2A) regulatoryt B',
mwgarabidopsis#13964	0.0663	At3g43950	hypothetical protein
mwgarabidopsis#00902	0.0664	At1g20030	pathogenesis-related thaumatin family protein
mwgarabidopsis#15248	0.0665	At4g08530	protease inhibitor/seed storage/lipid transfer protein (LTP) protein
mwgarabidopsis#14552	0.0669	At3g56750	expressed protein

mwgarabidopsis#10646	0.0673	At1g09980	expressed protein
mwgarabidopsis#15423	0.0674	At4g11770	kelch repeat-containing F-box family protein
mwgarabidopsis#17281	0.068	At5g15130	WRKY family transcription factor
mwgarabidopsis#12157	0.0681	At1g50530	hypothetical protein
mwgarabidopsis#04045	0.0686	At2g37630	myb family transcription factor (MYB91)
mwgarabidopsis#13143	0.0687	At1g73430	sec34-like family protein
mwgarabidopsis#20696	0.069	At2g05400	meprin and TRAF homology domain-protein / MATH domain- protein
mwgarabidopsis#17775	0.0692	At5g27210	expressed protein
mwgarabidopsis#21056	0.0694	0	
mwgarabidopsis#09251	0.0694	At4g39150	DNAJ heat shock N-terminal domain-containing protein
mwgarabidopsis#14176	0.0695	At3g48660	hypothetical protein
mwgarabidopsis#21691	0.0697	0	
mwgarabidopsis#08730	0.0697	At5g58770	dehydrodolichyl diphosphate synthase, / DEDOL-PP synthase,
mwgarabidopsis#05380	0.0698	At3g47570	leucine-rich repeat transmembrane protein kinase, putative
mwgarabidopsis#16020	0.07	At4g27480	glycosyltransferase family 14 protein / core-2/I-branching enzyme protein
mwgarabidopsis#10708	0.0702	At1g11360	universal stress protein (USP) family protein
mwgarabidopsis#18385	0.0702	At5g45450	iron transporter-related
mwgarabidopsis#10601	0.0703	At1g09150	pseudouridine synthase and archaeosine transglycosylase (PUA)
mwgarabidopsis#03406	0.0703	At2g23030	protein kinase, putative
mwgarabidopsis#14526	0.0704	At3g56370	leucine-rich repeat transmembrane protein kinase, putative
mwgarabidopsis#03629	0.0706	0	
mwgarabidopsis#02323	0.0707	At1g72550	tRNA synthetase beta subunit family protein
mwgarabidopsis#19725	0.0708	At2g17030	F-box family protein
mwgarabidopsis#13706	0.0709	At3g25420	serine carboxypeptidase S10 family protein
mwgarabidopsis#15440	0.0709	At4g12020	protein kinase family protein
mwgarabidopsis#07102	0.0709	At5g05390	laccase, putative / diphenol oxidase, putative
mwgarabidopsis#02111	0.071	At1g67430	60S ribosomal protein L17 (RPL17B)
mwgarabidopsis#18794	0.071	At5g54980	integral membrane family protein
mwgarabidopsis#14433	0.0711	At3g54200	expressed protein
mwgarabidopsis#15427	0.0715	At4g11830	phospholipase D gamma 2 / PLD gamma 2 (PLDGAMMA2)
mwgarabidopsis#16344	0.0717	At4g34330	expressed protein
mwgarabidopsis#12598	0.0722	At1g61900	expressed protein
mwgarabidopsis#03243	0.0722	At2g19130	S-locus lectin protein kinase family protein
mwgarabidopsis#13900	0.0723	At3g42920	hypothetical protein
mwgarabidopsis#14832	0.0723	At3g61290	hypothetical protein
mwgarabidopsis#17526	0.0724	At5g20260	exostosin family protein
mwgarabidopsis#20847	0.0725	At2g15670	hypothetical protein
mwgarabidopsis#08312	0.0725	At5g47240	MutT/nudix family protein
mwgarabidopsis#24412	0.0727	At4g21890	expressed protein
mwgarabidopsis#16137	0.0727	At4g30130	expressed protein
mwgarabidopsis#04248	0.0728	At2g42220	rhodanese-like domain-containing protein
mwgarabidopsis#15071	0.0729	At4g03010	leucine-rich repeat family protein
mwgarabidopsis#14675	0.0732	At3g58770	expressed protein
mwgarabidopsis#11425	0.0736	At1g26370	RNA helicase, putative
mwgarabidopsis#19678	0.0738	At1g32290	hypothetical protein
mwgarabidopsis#08721	0.0743	At5g58480	glycosyl hydrolase family 17 protein
mwgarabidopsis#17497	0.0746	At5g19580	glyoxal oxidase-related
mwgarabidopsis#10919	0.0749	At1g15460	anion exchange family protein
mwgarabidopsis#13348	0.075	At1g78280	transcription factor jumonji (jmjC) domain-protein
mwgarabidopsis#11787	0.0751	At1g34400	hypothetical protein
mwgarabidopsis#12350	0.0751	At1g54770	expressed protein
mwgarabidopsis#04079	0.0751	At2g38470	WRKY family transcription factor
mwgarabidopsis#12883	0.0753	At1g67570	expressed protein
mwgarabidopsis#06587	0.0755	At4g27150	2S seed storage protein 2 / 2S albumin storage protein /
mwgarabidopsis#08845	0.0757	At5g61350	protein kinase family protein
mwgarabidopsis#01942	0.0758	At1g62680	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#17893	0.0759	0	
mwgarabidopsis#11365	0.0759	At1g24220	paired amphipathic helix repeat-containing protein
mwgarabidopsis#07243	0.076	At5g09230	transcriptional regulator Sir2 family protein

mwgarabidopsis#02317	0.0764	At1g72360	ethylene-responsive element-binding protein, putative
mwgarabidopsis#18080	0.0764	At5g37890	seven in absentia (SINA) protein, putative
mwgarabidopsis#18291	0.0765	At5g43050	expressed protein
mwgarabidopsis#13203	0.0766	At1g74880	expressed protein
mwgarabidopsis#01117	0.0768	At1g27940	multidrug resistance P-glycoprotein, putative
mwgarabidopsis#16732	0.0768	At5g02980	kelch repeat-containing F-box family protein
mwgarabidopsis#20381	0.0769	At4g01895	systemic acquired resistance (SAR) regulator protein
mwgarabidopsis#08524	0.0769	At5g53460	glutamate synthase (NADH), chloroplast, putative
mwgarabidopsis#21672	0.0773	At2g36210	auxin-responsive family protein
mwgarabidopsis#05160	0.0775	At3g23490	cyanate lyase family
mwgarabidopsis#17102	0.0778	At5g11480	expressed protein
mwgarabidopsis#11946	0.0779	At1g44222	hypothetical protein
mwgarabidopsis#15326	0.078	At4g10030	hydrolase, alpha/beta fold family protein
mwgarabidopsis#07357	0.078	At5g12910	histone H3, putative
mwgarabidopsis#05641	0.0781	At3g55480	adaptin family protein
mwgarabidopsis#04560	0.0782	At3g02340	zinc finger (C3HC4-type RING finger) family protein
mwgarabidopsis#15992	0.0783	At4g26970	aconitate hydratase, cytoplasmic, / citrate hydro-lyase/aconitase,
mwgarabidopsis#14815	0.0784	At3g61010	glycosyl hydrolase family protein 85
mwgarabidopsis#17891	0.0789	0	
mwgarabidopsis#23919	0.0791	At3g52610	expressed protein
mwgarabidopsis#24013	0.0791	At4g01050	hydroxyproline-rich glycoprotein family protein
mwgarabidopsis#08976	0.0794	At5g64610	histone acetyltransferase, putative
mwgarabidopsis#23204	0.0795	At3g19880	F-box family protein
mwgarabidopsis#24256	0.0795	At4g14850	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#06863	0.0797	At4g36760	aminopeptidase P
mwgarabidopsis#17008	0.0797	At5g09620	octicosapeptide/Phox/Bem1p (PB1) domain-protein
mwgarabidopsis#20362	0.0799	At1g15040	glutamine amidotransferase-related
mwgarabidopsis#04437	0.08	0	
mwgarabidopsis#12966	0.08	At1g69480	EXS family protein / ERD1/XPR1/SYG1 protein
mwgarabidopsis#01969	0.0805	At1g63430	leucine-rich repeat transmembrane protein kinase
mwgarabidopsis#04442	0.0809	At2g46700	calcium-dependent protein kinase, putative / CDPK,
mwgarabidopsis#23006	0.081	At3g16960	hypothetical protein
mwgarabidopsis#07288	0.0812	At5g10430	arabinogalactan-protein (AGP4)
mwgarabidopsis#01175	0.0816	At1g29940	DNA-directed RNA polymerase family protein
mwgarabidopsis#04134	0.0821	At2g39850	subtilase family protein
mwgarabidopsis#05607	0.0821	At3g54860	vacuolar protein sorting protein, putative
mwgarabidopsis#11257	0.0822	At1g21880	peptidoglycan-binding LysM domain- protein
mwgarabidopsis#21251	0.0824	At2g25920	expressed protein
mwgarabidopsis#24264	0.0826	At4g15080	zinc finger (DHHC type) family protein
mwgarabidopsis#14797	0.0832	At3g60710	F-box family protein
mwgarabidopsis#23904	0.0839	At3g52230	expressed protein
mwgarabidopsis#06800	0.0841	At4g34910	DEAD/DEAH box helicase, putative (RH16)
mwgarabidopsis#08794	0.0841	At5g60080	protein kinase family protein
mwgarabidopsis#15235	0.0846	At4g08250	scarecrow transcription factor family protein
mwgarabidopsis#12172	0.085	At1g50800	hypothetical protein
mwgarabidopsis#20611	0.085	At2g03180	hypothetical protein
mwgarabidopsis#20747	0.0853	At2g10560	hypothetical protein
mwgarabidopsis#01376	0.0854	At1g36050	expressed protein
mwgarabidopsis#10290	0.0862	At1g02870	expressed protein
mwgarabidopsis#20663	0.0862	At2g04560	glycotransferase family protein 19
mwgarabidopsis#06444	0.0863	At4g22130	protein kinase family protein
mwgarabidopsis#15985	0.0863	At4g26860	alanine racemase family protein
mwgarabidopsis#01730	0.0865	At1g55010	plant defensin-fusion protein, putative (PDF1.5)
mwgarabidopsis#04980	0.0866	At3g13220	ABC transporter family protein
mwgarabidopsis#09960	0.0867	At2g22090	UBP1 interacting protein 1a (UBA1a)
mwgarabidopsis#02505	0.0869	At1g76570	chlorophyll A-B binding family protein
mwgarabidopsis#01901	0.087	At1g61500	S-locus protein kinase, putative
mwgarabidopsis#04109	0.087	At2g39200	seven transmembrane MLO family protein /(MLO12)
mwgarabidopsis#15370	0.0872	0	

mwgarabidopsis#06012	0.0872	At4g08140	hypothetical protein
mwgarabidopsis#20353	0.0873	At1g17455	expressed protein
mwgarabidopsis#21263	0.0877	At2g26130	zinc finger (C3HC4-type RING finger) family protein
mwgarabidopsis#22221	0.0877	At3g01100	early-responsive to dehydration protein-related /
mwgarabidopsis#02669	0.0878	At1g80640	protein kinase family protein
mwgarabidopsis#19768	0.0879	At1g65410	ABC transporter family protein
mwgarabidopsis#04400	0.0883	At2g45650	MADS-box protein (AGL6)
mwgarabidopsis#04644	0.0885	At3g04280	two-component responsive regulator family protein /
mwgarabidopsis#13564	0.0885	At3g20210	vacuolar processing enzyme, putative / asparaginyl endopeptidase
mwgarabidopsis#10272	0.0886	At1g02530	multidrug resistance P-glycoprotein, putative
mwgarabidopsis#17316	0.0886	At5g15620	F-box family protein
mwgarabidopsis#02859	0.0889	At2g05630	autophagy 8d (APG8d)
mwgarabidopsis#14429	0.0889	At3g54100	expressed protein
mwgarabidopsis#08030	0.0891	At5g39660	Dof-type zinc finger domain-containing protein
mwgarabidopsis#10158	0.0894	At1g73670	mitogen-activated protein kinase, putative / MAPK, (MPK15)
mwgarabidopsis#13677	0.0894	At3g24330	glycosyl hydrolase family 17 protein
mwgarabidopsis#04562	0.0897	At3g02380	zinc finger protein CONSTANS-LIKE 2 (COL2)
mwgarabidopsis#02048	0.0899	At1g65860	flavin-containing monooxygenase family protein / FMO family protein
mwgarabidopsis#13046	0.0899	At1g71090	auxin efflux carrier family protein
mwgarabidopsis#07328	0.0899	At5g11590	AP2 domain-containing transcription factor, putative
mwgarabidopsis#02202	0.09	At1g69570	Dof-type zinc finger domain-containing protein
mwgarabidopsis#05042	0.0906	At3g14510	geranylgeranyl pyrophosphate synthase, putative / GGPP synthetase,
mwgarabidopsis#10735	0.0907	At1g12020	expressed protein
mwgarabidopsis#16195	0.0907	At4g31570	expressed protein
mwgarabidopsis#13111	0.0912	At1g72590	3-oxo-5-alpha-steroid 4-dehydrogenase family protein /
mwgarabidopsis#05898	0.0913	At4g01010	cyclic nucleotide-regulated ion channel, putative (CNGC13)
mwgarabidopsis#15385	0.0919	At4g11060	single-strand-binding family protein
mwgarabidopsis#23090	0.0922	At3g18250	expressed protein
mwgarabidopsis#13681	0.0924	At3g24480	leucine-rich repeat family protein / extensin family protein
mwgarabidopsis#10014	0.0925	At4g28250	beta-expansin, putative (EXPB3)
mwgarabidopsis#21421	0.0929	At2g30430	hypothetical protein
mwgarabidopsis#07700	0.0933	At5g22980	serine carboxypeptidase III, putative
mwgarabidopsis#00271	0.0935	At1g02300	cathepsin B-like cysteine protease, putative
mwgarabidopsis#12766	0.0935	At1g65050	meprin and TRAF homology domain-Protein / MATH domain-protein
mwgarabidopsis#01108	0.0936	At1g27640	expressed protein
mwgarabidopsis#18997	0.0936	At5g60060	F-box family protein
mwgarabidopsis#03427	0.0937	At2g23380	curly leaf protein (CURLY LEAF) / polycomb-group protein
mwgarabidopsis#19188	0.0937	At5g64430	octicosapeptide/Phox/Bem1p (PB1) domain-containing protein
mwgarabidopsis#20660	0.0938	At2g04480	hypothetical protein
mwgarabidopsis#05766	0.0939	At3g60120	glycosyl hydrolase family 1 protein
mwgarabidopsis#06292	0.0943	At4g17510	ubiquitin carboxyl-terminal hydrolase, putative / ubiquitin thiolesterase
mwgarabidopsis#18144	0.0946	At5g39650	expressed protein
mwgarabidopsis#02974	0.0947	0	
mwgarabidopsis#20518	0.0947	At1g80610	expressed protein
mwgarabidopsis#14119	0.0947	At3g47350	short-chain dehydrogenase/reductase (SDR) family protein
mwgarabidopsis#09559	0.0949	At1g21140	nodulin, putative
mwgarabidopsis#21460	0.095	At2g31310	LOB domain protein 14 / lateral organ boundaries protein
mwgarabidopsis#22676	0.095	At3g10580	myb family transcription factor
mwgarabidopsis#16911	0.0951	At5g06660	expressed protein
mwgarabidopsis#23169	0.0953	At3g19340	expressed protein
mwgarabidopsis#07465	0.0953	At5g15950	adenosylmethionine decarboxylase family protein
mwgarabidopsis#18869	0.0954	At5g56770	hypothetical protein
mwgarabidopsis#01839	0.0955	At1g59780	disease resistance protein (CC-NBS-LRR class), putative
mwgarabidopsis#22445	0.0955	At3g05410	expressed protein
mwgarabidopsis#19640	0.0957	At1g20470	auxin-responsive family protein
mwgarabidopsis#16085	0.0958	At4g28840	expressed protein
mwgarabidopsis#19742	0.0959	At3g26085	CAAX amino terminal protease family protein
mwgarabidopsis#00416	0.0962	At1g06670	DEIH-box RNA/DNA helicase
mwgarabidopsis#24432	0.0962	At4g23110	hypothetical protein

mwgarabidopsis#15987	0.0963	At4g26890	protein kinase family protein
mwgarabidopsis#21115	0.0964	At2g22340	hypothetical protein
mwgarabidopsis#05440	0.0964	At3g49690	myb family transcription factor
mwgarabidopsis#09062	0.0964	At5g66680	dolichyl-diphosphooligosaccharide-protein glycosyltransferase 48kDa
mwgarabidopsis#21671	0.0967	0	
mwgarabidopsis#15420	0.0967	At4g11720	hypothetical protein
mwgarabidopsis#19763	0.0968	At1g38950	hypothetical protein
mwgarabidopsis#18707	0.0968	At5g52930	expressed protein
mwgarabidopsis#15255	0.0969	At4g08690	SEC14 cytosolic factor family protein / phosphoglyceride transfer protein
mwgarabidopsis#16738	0.0969	At5g03060	expressed protein
mwgarabidopsis#06007	0.0971	0	
mwgarabidopsis#02511	0.0971	At1g76800	nodulin, putative
mwgarabidopsis#17038	0.0971	At5g10340	F-box protein-related / SLF-related
mwgarabidopsis#12499	0.0972	At1g59720	pentatricopeptide (PPR) repeat-containing protein
mwgarabidopsis#20865	0.0976	At2g16030	expressed protein
mwgarabidopsis#05692	0.0976	At3g57050	cystathionine beta-lyase, chloroplast / beta-cystathionase / (CBL)
mwgarabidopsis#09225	0.0984	At1g48630	guanine nucleotide-binding protein / activated protein kinase C receptor,
mwgarabidopsis#02356	0.0985	At1g73250	GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase (GER1)
mwgarabidopsis#07970	0.0991	At5g38130	transferase family protein
mwgarabidopsis#17650	0.0997	At5g24380	transporter, putative
mwgarabidopsis#00910	0.0998	At1g20260	vacuolar ATP synthase subunit B, / V-ATPase B / vacuolar proton pump B
mwgarabidopsis#09250	0.0998	At4g35100	plasma membrane intrinsic protein (SIMIP)
mwgarabidopsis#19307	0.1	At5g67590	NADH-ubiquinone oxidoreductase-related

Appendix 09

Chapter 08: Genes list within the polymorphic region (2-3 cM on top of chromosome 3) for the 2 STAIRS used in gene expression analysis

TAIR Accession	Locus	Gene Model	Gene Type	Description
Gene:3698200	AT3G02370	AT3G02370.1	protein_coding	expressed protein,
Gene:4370595	AT3G02380	COL2	proein_coding	homologous to the flowering-time gene CONSTANS (CO) encoding zinc-finger proteins transcription factor activity, flower, leaf, stem, regulation of flower development true
Gene:3698228	AT3G02380	AT3G02380.1	protein_coding	Zinc finger protein CONSTANS-LIKE 2 (COL2), identical to putative flowering-time gene CONSTANS (COL2) GB:, zinc ion binding true
Gene:3698191	AT3G02390	AT3G02390.1	protein_coding	expressed protein F11A12.8; At3g02390; F11A12_8
Gene:2074556	AT3G02400	AT3G02400.1	protein_coding	forkhead-associated domain-containing protein /
Gene:2076978	AT3G02410	AT3G02410.1	protein_coding	hypothetical protein, weak similarity to kynurenine formamidase (Mus musculus)
Gene:2076873	AT3G02420	AT3G02420.1	protein_coding	expressed protein F16B3_5; F16B3.5; At3g02420
Gene:2076893	AT3G02430	AT3G02430.1	protein_coding	hypothetical protein, contains Pfam profile PF05078: Protein of unknown function (DUF679)
Gene:2076908	AT3G02440	AT3G02440.1	protein_coding	expressed protein F16B3.7; F16B3_7; At3g02440
Gene:2076928	AT3G02450	AT3G02450.1	protein_coding	cell division protein ftsH, putative, similar to SWISS-PROT:P46469 cell division protein ftsH homolog
Gene:2076948	AT3G02460	AT3G02460.1	protein_coding	plant adhesion molecule, putative, strong similarity to plant adhesion molecule 1 (Arabidopsis thaliana) GI:3511223; contains Pfam profile PF00566: TBC domain
Gene:1005714688	AT3G02460	AT3G02460.2	protein_coding	plant adhesion molecule, putative, strong similarity to plant adhesion molecule 1
Gene:2076948	AT3G02460	AT3G02460.1	protein_coding	plant adhesion molecule, putative, strong similarity to plant adhesion molecule 1 (Arabidopsis thaliana)
Gene:1005714688	AT3G02460	AT3G02460.2	protein_coding	plant adhesion molecule, putative, strong similarity to plant adhesion molecule 1 (Arabidopsis thaliana)
Gene:2076833	AT3G02470	AT3G02470.1	protein_coding	adenosylmethionine decarboxylase family protein, contains Pfam profile: PF01536 adenosylmethionine decarboxylase At3g02470; F16B3.10; F16B3_10 AGI-
Gene:3354458	AT3G02470	SAMDC	protein_coding	S-adenosylmethionine decarboxylase (SAMdc) mRNA, complete
Gene:2076843	AT3G02480	AT3G02480.1	protein_coding	ABA-responsive protein-related, similar to ABA-inducible protein (Fagus sylvatica) GI:3901016,
Gene:2076853	AT3G02490	AT3G02490.1	protein_coding	pentatricopeptide (PPR) repeat-containing protein, contains Pfam profile PF01535: PPR repeat
Gene:2076868	AT3G02500	AT3G02500.1	protein_coding	expressed protein, F16B3.13; At3g02500; F16B3_13 AGI-TIGR
Gene:2076888	AT3G02510	AT3G02510.1	protein_coding	regulator of chromosome condensation (RCC1) family protein, similar to UVB-resistance protein UVR8 (Arabidopsis thaliana)
Gene:504952382	AT3G02520	GRF7	protein_coding	14-3-3 gene
Gene:2076903	AT3G02520	AT3G02520.1	protein_coding	14-3-3 protein GF14 nu (grf7), identical to 14-3-3 protein GF14 nu GI:1531631 from (Arabidopsis thaliana)
Gene:3354460	AT3G02520	GF14 NU NU	protein_coding	14-3-3 protein isoform GF14 nu mRNA, complete cds
Gene:2076923	AT3G02530	AT3G02530.1	protein_coding	chaperonin, putative, similar to SWISS-PROT:P80317-T-complex protein 1, zeta subunit (TCP-1-zeta)
Gene:2076943	AT3G02540	AT3G02540.1	protein_coding	ubiquitin family protein, contains Pfam profiles PF00240: Ubiquitin family, PF00627
Gene:1006228439	AT3G02540	AT3G02540.2	protein_coding	ubiquitin family protein, contains Pfam profiles
Gene:2076943	AT3G02540	AT3G02540.1	protein_coding	ubiquitin family protein, contains Pfam profiles PF00240: Ubiquitin family, PF00627: UBA/TS-N domain;
Gene:1006228439	AT3G02540	AT3G02540.2	protein_coding	ubiquitin family protein, contains Pfam profiles
Gene:2076828	AT3G02550	AT3G02550.1	protein_coding	LOB domain protein 41 / lateral organ boundaries domain protein 41 (LBD41), identical to LOB
Gene:3698210	AT3G02555	AT3G02555.1	protein_coding	Expressed protein At3g02555 AGI-
Gene:2076838	AT3G02560	AT3G02560.1	protein_coding	40S ribosomal protein S7 (RPS7B), similar to ribosomal protein S7 GB:AAD26256 from (Secale cereale)
Gene:1005714689	AT3G02560	AT3G02560.2	protein_coding	40S ribosomal protein S7 (RPS7B), similar to ribosomal protein S7 GB:AAD26256 from (Secale cereale) true

Gene:2076838	AT3G02560	AT3G02560.1	protein_coding 40S ribosomal protein S7 (RPS7B), similar to ribosomal protein S7 GB:AAD26256 from (Secale cereale)
Gene:1005714689	AT3G02560	AT3G02560.2	protein_coding 40S ribosomal protein S7 (RPS7B), similar to ribosomal protein S7 GB:AAD26256 from (Secale cereale) true
Gene:2076863	AT3G02570	AT3G02570.1	protein_coding phosphomannose isomerase type I family protein, contains Pfam profile: PF01238 phosphomannose isomerase type
Gene:2076883	AT3G02580	AT3G02580.1	protein_coding delta 7-sterol-C5-desaturase (STE1), identical to sterol-C5-desaturase GB:AAD12944 GI:4234768 from (Arabidopsis thaliana)
Gene:1944897	AT3G02580	DWF7	protein_coding Brassinosteroid biosynthetic enzyme, catalyzes delta7 sterol C-5 desaturation step. Mutant has dwarf phenotype
Gene:3685810	AT3G02580	STE1	protein_coding Brassinosteroid biosynthetic enzyme, catalyzes delta7 sterol C-5 desaturation step.
Gene:2076898	AT3G02590	AT3G02590.1	protein_coding delta 7-sterol-C5-desaturase, putative, similar to delta7 sterol C-5 desaturase GI:5031219 from (Arabidopsis thaliana
Gene:2076918	AT3G02600	AT3G02600.1	protein_coding phosphatidic acid phosphatase family protein / PAP2 family protein, similar to phosphatidic acid phosphatase (Mus musculus) GI:1487873;
Gene:2076938	AT3G02610	AT3G02610.1	protein_coding acyl-(acyl-carrier-protein) desaturase, putative / stearoyl-ACP desaturase, putative, similar to Acyl-(acyl-carrier protein)
Gene:2076958	AT3G02620	AT3G02620.1	protein_coding acyl-(acyl-carrier-protein) desaturase, putative / stearoyl-ACP desaturase, putative, similar to Acyl-(acyl-carrier protein)
Gene:2076968	AT3G02630	AT3G02630.1	protein_coding acyl-(acyl-carrier-protein) desaturase, putative / stearoyl-ACP desaturase, putative, similar to Acyl-(acyl-carrier protein)
Gene:2076848	AT3G02640	AT3G02640.1	protein_coding expressed protein F16B3_27; F16B3.27; At3g02640 AGI-TIGR integral to membrane,
Gene:2076858	AT3G02650	AT3G02650.1	protein_coding pentatricopeptide (PPR) repeat-containing protein, contains Pfam profile PF01535:
Gene:2076878	AT3G02660	AT3G02660.1	protein_coding tRNA synthetase class I (W and Y) family protein, similar to SP|P00952 Tyrosyl-tRNA synthetase (EC
Gene:2076913	AT3G02670	AT3G02670.1	protein_coding proline-rich family protein, contains proline rich extensin domains
Gene:2076933	AT3G02680	AT3G02680.1	protein_coding hypothetical protein At3g02680; F16B3_31; F16B3.31 AGI-
Gene:2076953	AT3G02690	AT3G02690.1	protein_coding integral membrane family protein, similar to PecM protein At3g02690; F16B3_32 AGI-TIGR chloroplast,
Gene:2076963	AT3G02700	AT3G02700.1	protein_coding NC domain-containing protein, contains Pfam domain, PF04970: NC domain At3g02700; F16B3.33;
Gene:2076973	AT3G02710	AT3G02710.1	protein_coding nuclear associated protein-related / NAP-related, similar to Nuclear associated protein (NAP) (NYD-SP19) (Swiss-Prot:Q8WYA6) (Homo sapiens)
Gene:1005714687	AT3G02715	AT3G02715.1	pre_trna tRNA-Ser (anticodon: AGA) 60284.tRNA-Ser-1; 60277.tRNA-Ser-1
Gene:3698204	AT3G02720	AT3G02720.1	protein_coding DJ-1 family protein / protease-related, similar to Intracellular Protease (Pyrococcus horikoshii) GI:11513902; Contains Pfam profile: PF01965 ThiJ/PfpI family At3g02720;
Gene:2075521	AT3G02730	AT3G02730.1	protein_coding thioredoxin, putative, similar to SP|P29450 Thioredoxin F-type, chloroplast precursor (TRX-F) {Pisum sativum};
Gene:3354462	AT3G02730	AT3G02730	protein_coding thioredoxin f1 mRNA, complete cds F13E7.33; F13E7_
Gene:2075511	AT3G02740	AT3G02740.1	protein_coding aspartyl protease family protein, contains Pfam profile: PF00026 eukaryotic aspartyl protease
Gene:2075486	AT3G02750	AT3G02750.1	protein_coding protein phosphatase 2C family protein / PP2C family protein, similar to protein phosphatase-2C; PP2C
Gene:2075471	AT3G02760	AT3G02760.1	protein_coding histidyl-tRNA synthetase, putative / histidine—tRNA ligase, putative, similar to SP|P12081 Histidyl-tRNA synthetase (EC 6.1.1.21)
Gene:2075431	AT3G02770	AT3G02770.1	protein_coding dimethylmenaquinone methyltransferase family protein, similar to bacterial S-adenosylmethionine:
Gene:2075411	AT3G02780	AT3G02780.1	protein_coding isopentenyl-diphosphate delta-isomerase II / isopentenyl diphosphate:dimethylallyl diphosphate isomerase II (IPP2), identical to isopentenyl diphosphate:dimethylallyl diphosphate isomerase
Gene:3354464	AT3G02780	IPIAT1	protein_coding clone Atdyi isopentenyl pyrophosphate:dimethyllallyl GenBank true
Gene:3354466	AT3G02780	IPP2	protein_coding isopentenyl diphosphate:dimethylallyl diphosphate isomerase GenBank true
Gene:2075396	AT3G02790	AT3G02790.1	protein_coding zinc finger (C2H2 type) family protein, contains Pfam profile: PF00096 zinc finger,
Gene:2075526	AT3G02800	AT3G02800.1	protein_coding tyrosine specific protein phosphatase family protein, contains tyrosine specific protein phosphatases active
Gene:2075516	AT3G02810	AT3G02810.1	protein_coding protein kinase family protein, contains protein kinase domain, Pfam:PF00069 F13E7_25; F13E7.25;

Gene:2075506	AT3G02820	AT3G02820.1	protein_coding zinc knuckle (CCHC-type) family protein, contains Pfam domain, PF00098: Zinc knuckle F13E7_24;
Gene:2075476	AT3G02830	AT3G02830.1	protein_coding zinc finger (CCCH-type) family protein, contains Pfam domain, PF00642: Zinc finger C-x8-C-x5-C-x3-
Gene:3354468	AT3G02830	ZFN1	protein_coding zinc finger protein 1 (zfn1) mRNA, complete cds GenBank true
Gene:2075456	AT3G02840	AT3G02840.1	protein_coding immediate-early fungal elicitor family protein, similar to immediate-early fungal elicitor protein CMPG1
Gene:3714742	AT3G02850	SKOR	protein_coding member of Stelar K+ outward rectifying channel (SKOR)
Gene:2075441	AT3G02850	AT3G02850.1	protein_coding stelar K+ outward rectifier (SKOR) / potassium channel protein, identical to SKOR (Arabidopsis thaliana)
Gene:2075421	AT3G02860	AT3G02860.1	protein_coding expressed protein F13E7.20; At3g02860; F13E7_20 AGI-TIGR nucleus, metalloendopeptidase activity,
Gene:1005714712	AT3G02860	AT3G02860.2	protein_coding expressed protein nucleus, metalloendopeptidase activity, nucleic acid binding
Gene:2075421	AT3G02860	AT3G02860.1	protein_coding expressed protein F13E7.20; At3g02860; F13E7_20 AGI-TIGR nucleus, metalloendopeptidase activity, nucleic acid binding,
Gene:1005714712	AT3G02860	AT3G02860.2	protein_coding expressed protein nucleus, metalloendopeptidase activity, nucleic acid binding,
Gene:2075391	AT3G02870	AT3G02870.1	protein_coding inositol-1(or 4)-monophosphatase, putative / inositol monophosphatase,
Gene:1944568	AT3G02875	ILR1	protein_coding Hydrolyzes amino acid conjugates of the plant growth regulator indole-3-acetic acid (IAA), including IAA-Leu and IAA-Phe.
Gene:2075381	AT3G02875	AT3G02875.1	protein_coding IAA-amino acid hydrolase 1 (ILR1), identical to IAA-amino acid hydrolase 1 (ILR1) (Arabidopsis thaliana)
Gene:2075501	AT3G02880	AT3G02880.1	protein_coding leucine-rich repeat transmembrane protein kinase, putative, contains Pfam profiles: PF00069 Eukaryotic protein kinase domain, PF00560 Leucine Rich
Gene:3697927	AT3G02885	AT3G02885.1	protein_coding gibberellin-regulated protein 5 (GASA5) / gibberellin-responsive protein 5, identical to GASA5 (Arabidopsis thaliana)
Gene:1944696	AT3G02885	GASA5	protein_coding true
Gene:2075491	AT3G02890	AT3G02890.1	protein_coding PHD finger protein-related, contains low similarity to PHD-finger domain proteins
Gene:2075461	AT3G02900	AT3G02900.1	protein_coding expressed protein, F13E7.15; At3g02900; F13E7_15 AGI-TIGR
Gene:2075436	AT3G02910	AT3G02910.1	protein_coding expressed protein, contains Pfam domain PF03674: Uncharacterised protein family (UPF0131)
Gene:2075416	AT3G02920	AT3G02920.1	protein_coding replication protein-related, similar to replication protein A 30kDa (Oryza sativa (japonica cultivar-group))
Gene:2075401	AT3G02930	AT3G02930.1	protein_coding expressed protein, ; expression supported by MPSS F13E7.12; At3g02930; F13E7_12 AGI-TIGR chloroplast, delta-DNA polymerase cofactor complex,
Gene:2075386	AT3G02940	AT3G02940.1	protein_coding myb family transcription factor (MYB107), contains Pfam profile: PF00249 Myb-like DNA-binding domain F13E7_11; F13E7.11; At3g02940 AGI-
Gene:3354470	AT3G02940	MYB107	protein_coding putative transcription factor (MYB107) mRNA, complete cds GenBank true
Gene:2075376	AT3G02950	AT3G02950.1	protein_coding expressed protein F13E7.10; At3g02950; F13E7_10 AGI-TIGR
Gene:2075496	AT3G02960	AT3G02960.1	protein_coding copper-binding protein-related, low similarity to copper homeostasis factor gi:3168840 from Arabidopsis thaliana;
Gene:2075481	AT3G02970	AT3G02970.1	protein_coding phosphate-responsive 1 family protein, similar to phi-1 (phosphate-induced gene) (Nicotiana tabacum) GI:3759184; contains Pfam profile PF04674:
Gene:2075466	AT3G02980	AT3G02980.1	protein_coding GCN5-related N-acetyltransferase (GNAT) family protein, contains Pfam profile: PF00583 acetyltransferase (GNAT) family At3g02980; F13E7.7;
Gene:3714172	AT3G02990	ATHSFA1E	protein_coding member of Heat Stress Transcription Factor (Hsf) family
Gene:2075446	AT3G02990	AT3G02990.1	protein_coding heat shock factor protein 2 (HSF2) / heat shock transcription factor 2 (HSTF2), identical to heat shock transcription factor 2 (HSF2) SP:Q96320 from (Arabidopsis thaliana);
Gene:2075426	AT3G03000	AT3G03000.1	protein_coding calmodulin, putative, similar to calmodulin SP:P04352 from (Chlamydomonas reinhardtii); contains Pfam profile:
Gene:2075536	AT3G03010	AT3G03010.1	protein_coding expressed protein F13E7.4; At3g03010; F13E7_4
Gene:2075451	AT3G03020	AT3G03020.1	protein_coding expressed protein F13E7.3; At3g03020; F13E7_3
Gene:2075406	AT3G03030	AT3G03030.1	protein_coding F-box family protein, contains F-box domain Pfam: PF00646

Gene:3701969	AT3G03040	AT3G03040.1	protein_coding F-box family protein, contains F-box domain Pfam:PF00646
Gene:2097699	AT3G03050	AT3G03050.1	protein_coding cellulose synthase family protein (CslD3), similar to cellulose synthase catalytic subunit gi:2827143 from
Gene:3713624	AT3G03050	CSLD3	protein_coding encodes a cellulose synthase like protein. Mutations initiate root hairs that rupture at their tip soon after initiation.
Gene:2097689	AT3G03060	AT3G03060.1	protein_coding AAA-type ATPase family protein, contains a ATP/GTP-binding site motif A (P-loop), PROSITE:PS00017 T17B22_25; At3g03060; T17B22.25 AGI-TIGR
Gene:2097809	AT3G03070	AT3G03070.1	protein_coding NADH-ubiquinone oxidoreductase-related, contains weak similarity to NADH-ubiquinone oxidoreductase 13 kDa-A subunit, mitochondrial precursor (EC 1.6.5.3)
Gene:3354472	AT3G03070	AT3G03070	protein_coding AT3g03070 (AT3g03070/T17B22_24) mRNA, complete cds T17B22_24; T17B22.24 GenBank true
Gene:2097794	AT3G03080	AT3G03080.1	protein_coding NADP-dependent oxidoreductase, putative, similar to probable NADP-dependent oxidoreductase (zeta-crystallin homolog)
Gene:2097779	AT3G03090	AT3G03090.1	protein_coding sugar transporter family protein, similar to xylose permease (Bacillus megaterium) GI:1924928; contains Pfam profile PF00083: major facilitator superfamily protein
Gene:2097764	AT3G03100	AT3G03100.1	protein_coding NADH:ubiquinone oxidoreductase family protein, contains Pfam PF05071: NADH:ubiquinone oxidoreductase 17.2 kD subunit; similar to ethylene-regulated
Gene:2097749	AT3G03110	AT3G03110.1	protein_coding exportin 1, putative, strong similarity to Exportin1 (XPO1) protein (Arabidopsis thaliana)
Gene:1006153946	AT3G03110	XPO1B	protein_coding true
Gene:2097724	AT3G03120	AT3G03120.1	protein_coding ADP-ribosylation factor, putative, similar to ADP-ribosylation factor 1; ARF 1 (GP:385340) {Drosophila melanogaster}
Gene:2097714	AT3G03130	AT3G03130.1	protein_coding expressed protein, ; expression supported by MPSS T17B22_18; At3g03130; T17B22.18
Gene:2097804	AT3G03140	AT3G03140.1	protein_coding expressed protein T17B22_17; At3g03140; T17B22.17 AGI-TIGR
Gene:2097789	AT3G03150	AT3G03150.1	protein_coding expressed protein T17B22.16; T17B22_16; At3g03150 AGI-TIGR
Gene:2097774	AT3G03160	AT3G03160.1	protein_coding expressed protein T17B22_15; At3g03160; T17B22.15 AGI-
Gene:2097759	AT3G03170	AT3G03170.1	protein_coding expressed protein T17B22.14; T17B22_14; At3g03170 AGI-TIGR
Gene:2097744	AT3G03180	AT3G03180.1	protein_coding Got1-like family protein, contains Pfam profile: PF04178 Got1-like
Gene:1944639	AT3G03190	ATGSTF11	protein_coding Encodes glutathione transferase belonging to the phi class of GSTs. Naming convention according to Wagner et al.
Gene:2097729	AT3G03190	AT3G03190.1	protein_coding glutathione S-transferase, putative, identical to glutathione S-transferase GB:AAB09584 from
Gene:2097719	AT3G03200	AT3G03200.1	protein_coding no apical meristem (NAM) family protein, similar to NAC2 (GI:6456751) {Arabidopsis thaliana}
Gene:2097709	AT3G03210	AT3G03210.1	protein_coding expressed protein, T17B22.10; At3g03210; T17B22_10 AGI-TIGR
Gene:3714105	AT3G03220	ATEXPA13	protein_coding member of Alpha-Expansin Gene Family. true
Gene:2097694	AT3G03220	AT3G03220.1	protein_coding expansin, putative (EXP13), similar to expansin precursor GB:AAD13631 from (Lycopersicon esculentum);
Gene:2097684	AT3G03230	AT3G03230.1	protein_coding esterase/lipase/thioesterase family protein, contains Interpro entry IPR000379 T17B22_8; At3g03230; T17B22.8
Gene:2097799	AT3G03240	AT3G03240.1	protein_coding esterase/lipase/thioesterase family protein, contains Interpro entry IPR000379 T17B22_7; T17B22.7;
Gene:2097784	AT3G03250	AT3G03250.1	protein_coding UTP--glucose-1-phosphate uridylyltransferase, putative / UDP-glucose
Gene:2097769	AT3G03260	AT3G03260.1	protein_coding homeobox-leucine zipper family protein / lipid-binding START domain-containing protein, similar to L1 specific homeobox gene
Gene:1005027795	AT3G03270	AT3G03270.2	protein_coding universal stress protein (USP) family protein / early nodulin ENOD18 family protein,
Gene:2097754	AT3G03270	AT3G03270.1	protein_coding universal stress protein (USP) family protein / early nodulin ENOD18 family protein, contains Pfam profile
Gene:1005027795	AT3G03270	AT3G03270.2	protein_coding universal stress protein (USP) family protein / early nodulin ENOD18 family protein,
Gene:2097754	AT3G03270	AT3G03270.1	protein_coding universal stress protein (USP) family protein / early nodulin ENOD18 family protein, contains Pfam profile PF00582: universal stress protein family;
Gene:2097739	AT3G03280	AT3G03280.1	protein_coding expressed protein, At3g03280; T17B22_3; T17B22.3 AGI-

Gene:2097734	AT3G03290	AT3G03290.1	protein_coding universal stress protein (USP) family protein, contains Pfam profile: PF00582 universal stress protein family
Gene:2097704	AT3G03300	AT3G03300.1	protein_coding DEAD/DEAH box helicase carpel factory-related, similar to RNA helicase GB:AAF03534 At3g03300; T17B22.28
Gene:1006240211	AT3G03300	DCL2	protein_coding Arabidopsis thaliana DEAD/DEAH box helicase carpel factory-related mrna
Gene:3701949	AT3G03305	AT3G03305.1	protein_coding calcineurin-like phosphoesterase family protein, contains Pfam profile: PF00149 calcineurin-like phosphoesterase
Gene:2099608	AT3G03310	AT3G03310.1	protein_coding lecithin:cholesterol acyltransferase family protein / LACT family protein, weak similarity to LCAT-like lysophospholipase (LLPL)