More Precise QTL Mapping Using STAIRS

S.A.C.N. Perera, T.M. Wilkes[1], M.J.Kearsey[2]

Coconut Research Institute
Lunuwila, Sri Lanka

ABSTRACT. *Many economically important traits of plants and animals are quantitative in nature and are governed by Quantitative Trait Loci (QTL). Mapping of such QTL generates a vast amount of information which helps in genetic improvement of organisms. However, conventional genome mapping methods using segregating populations lack the precision desired in QTL mapping. An alternative method for using segregating populations is to generate populations with special genetic make up by using chromosome engineering methods. The current study uses Chromosome Substitution Strains (CSSs) and STepped Aligned Recombinant Inbred Strains (STAIRS) produced by chromosome engineering, in fine mapping of QTL. Sixteen lines known to be STAIRS of chromosome 3 of Arabidopsis were genotyped with 29 microsatellite markers in order to define the introgressed regions. Subsequently, morphological traits such as flowering time, height at flowering, rosette and cauline leaf numbers at flowering and rosette width at flowering were scored by growing the lines in a controlled environment in a completely randomised design. The genotypic data revealed the genetic make up of the STAIRS. Both genotypic and phenotypic data were combined to map QTL on chromosome 3 of Arabidopsis by least squares model fitting. QTL which affect all the measured traits were located within the top region of chromosome three between 0-14 cM. Two additional QTL affecting plant height at flowering were located between 44-60 cM and 80-100 cM. The experiment provided conclusive evidence to narrow down the interval of QTL affecting flowering time, and rosette leaf number and revealed new QTL affecting cauline leaf number and plant height.*

INTRODUCTION

Genetic improvement of cultivated crops and farm animals for desirable traits is of utmost importance to ensure a sufficient food supply for the ever-increasing population of the world. The availability of genetic information about major and minor traits and their interactions increases the efficiency and probability of success in producing organisms with desired attributes. Genetic maps of organisms generate a vast amount of information which can be used in order to identify, manipulate and complement traits to their maximum advantage (Allen, 1994).

Gene mapping is the positioning of genes at specific locations along chromosomes. The advent of molecular markers facilitated the wide-spread application of Quantitative Trait Loci (QTL) mapping and a better understanding of complex quantitative traits. The use of polymorphic molecular markers in segregating populations such as F_2s, DHLs or

[1] Centre for Cancer Research, University of Birmingham, B15 2TT, UK

[2] School of Biosciences, University of Birmingham, B15 2TT, UK.

RILs enables the construction of the framework map, around which QTL can be located. Yet, there are limitations in this approach, particularly in fine mapping. The main limitation is the difficulty of locating QTL with sufficient accuracy due to wide confidence intervals (normally about 30 cM) of the mapped QTL. Furthermore, 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 also are constraints in using segregating populations for QTL mapping. The precision of mapping with segregating populations may be adequate for marker aided selection, yet far too low for detailed genetic analysis, chromosome walking or for map based cloning of the genes (Hyne and Kearsey, 1995; Kearsey and Farquhar, 1998).

An alternative method for more precise mapping of QTL is 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 this approach is to create two genotypes which are identical apart from a defined region on a particular chromosome (Howell 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 (Burns *et al.*, 2003; Ramsay *et al.*, 1996).

Koumproglou *et al.* (2002) reported the production of whole Chromosome Substitution Strains (CSSs) in the model plant *Arabidopsis thaliana* which has become the most favoured choice as a model organism among plant scientists. They used ecotypes Columbia (Col) as the recurrent parent and Landsberg (L*er*) as the donor parent in producing CSSs. The production of a total number of five CSSs is possible for the five haploid chromosomes in *Arabidopsis* (Fig. 1). Koumproglou *et al.* (2002) further explained the production of a novel resource, STepped Aligned Inbred Recombinant Strains (STAIRS) of *Arabidopsis* and the usage of these resources in mapping QTL.

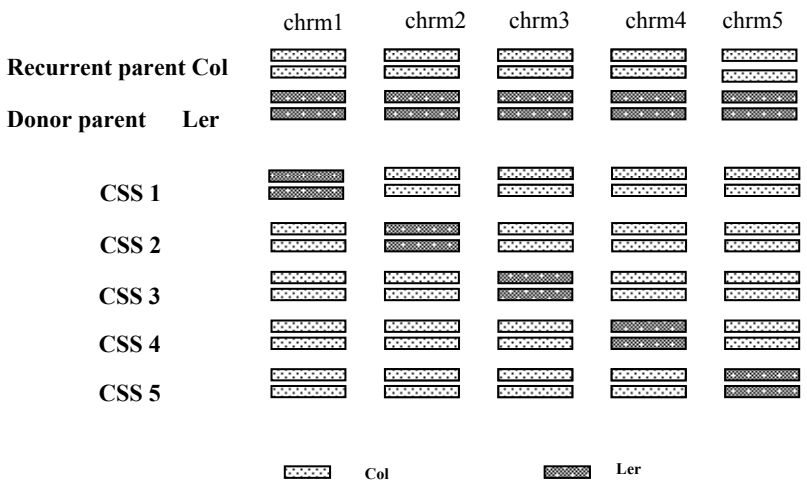
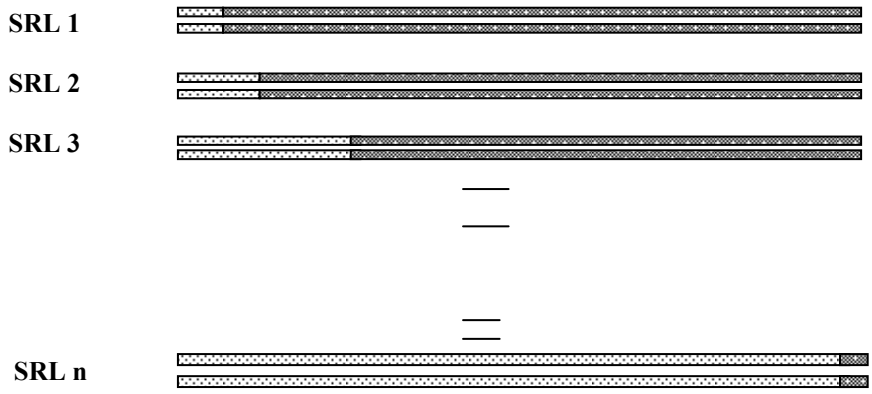


Fig. 1. Chromosome substitution strains of *Arabidopsis thaliana*.

STAIRS are homozygous single recombinant lines (SRL) which are derived from each CSS. They consist of a number of lines, each containing a homozygous chromosome with a single crossover in such a way that the chromosome contains recurrent genotype at one end and donor genotype at the other end. When the SRLs for each chromosome are sequentially stacked, they show a step-like progression with each successive line having a little more donor chromosome (Fig. 2). The term STAIRS thus reflects 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 chromosome.



SRL = Single Recombinant Line

Fig. 2. Structure of STAIRS.

STAIRS are constructed by means of several rounds of back crossing of each CSS to the recurrent parent, using a marker assisted breeding programme. The width of each step in the ladder-like progression when the STAIRS are sequentially stacked, indicates the level of precision in QTL mapping. Determination of the width of each step in STAIRS depends upon the coverage of the chromosome with markers.

As shown in Figure 3, fine mapping of QTL can be achieved in three steps using the resources CSSs and STAIRS by growing these lines in a controlled environment in an appropriate experimental design and scoring for quantitative phenotypes (Koumproglou *et al.*, 2002).

CSSs help locating QTL to particular chromosome/chromosomes. By using wide STAIRS within a chromosome, QTL can be identified to a region of about 10-15 cM along the specific chromosome. This can be followed by the construction of narrow STAIRS within the region resulting from analysis with wide STAIRS for fine mapping of QTL. Theoretically, a map resolution as high as 0.5 cM is possible, given the availability of a sufficient number of narrow STAIRS.

Flowering time is an important trait with regard to initiation of the reproductive period of a plant. Locating genes related to flowering time will be very useful in genetic improvement of many plants. In this study, our objectives were to test the authenticity of

wide STAIRS [some of which were produced by Koumproglou *et al.* (2002) and the rest identified by the authors] by microsatellite genotyping and mapping of QTL for flowering time and related traits within chromosome 3 of *Arabidopsis* using wide STAIRS.

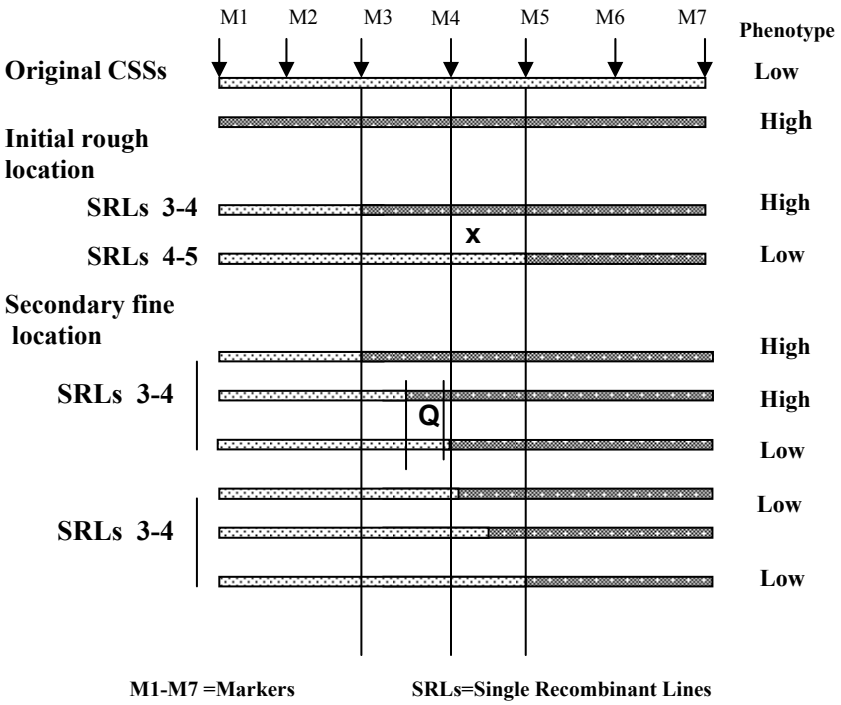


Fig. 3. Stages of fine mapping using CSSs and STAIRS.

MATERIALS AND METHODS

Genotyping wide STAIRS

Sixteen lines known to be STAIRS of chromosome 3 of *Arabidopsis* [some lines from Koumproglou *et al.* (2002) and some identified in the current study] were genotyped to identify the region of donor introgression using nine microsatellite markers positioned along chromosome 3. Five out of the nine markers used were newly developed microsatellite markers in the current research [F16B3 (2 cM), T22K18 (14 cM), MAG2 (21cM), T16K5a (88 cM) and T17J13b 99 cM)] while four markers [nga172 (6.9 cM), nga162 (21 cM), AthGapAB (44 cM), Th620B (59 cM)] were published earlier (Bell and Ecker, 1994). The lines were further genotyped with 20 more markers located on the remaining four chromosomes in order to verify the authenticity of STAIRS. These 20 markers are scattered approximately at equal distances along chromosomes 1, 2, 4 and 5 and they were also published markers by Bell and Ecker (1994).

Extraction of DNA, PCR and gel electrophoresis

DNA of the sixteen lines was extracted from two-week old plants grown in a growth chamber (16 h day length, 24^0C) using a modified CTAB DNA extraction protocol. DNA quantification was carried out using spectrophotometer (Eppendorf Bio Photometer) and diluted in TE buffer pH 8.0, to a final concentration of 50 ng/µl for use in PCR. Ten µl PCR reactions were set up in 0.5 ml PCR tubes with the following ingredients and quantities.

Template DNA 50 ng/µl, 2 µl, 10x PCR buffer (Moltaq) 1 µl, 50 mM $MgCl_2$ 0.2 µl, mM dNTPs, 0.4 µl,Forward primer (10 pM/µl), 0.4 µl, Reverse primer (10 pM/µl) 0.4 µl, Taq DNA polymerase (Moltaq) (5 u/µl), 0.08 µl, Sterile distilled water to volumerize to 10 µl. The reactions were run in a thermal cycler (HYBAID Omnigene) using the following amplification cycles.

1 x (94^0C, 5 min; yy^{0}C, 30 s"; 72^0C, 1 min)

30 x (94^0C, 30 s"; yy^{0}C, 30 s"; 72^0C, 1 min)

1 x (94^0C, 30 s"; yy^{0}C, 30 s"; 72^0C, 5 min)

min = minutes; s''= seconds ; yy = annealing temperature for each primer.

PCR products were electrophoresed on 8% non-denaturing polyacrylamide AccuGel 29:1 (National Diagnostics) and stained in 2 µl/100ml Ethidium bromide solution for 30 minutes for detection and scoring of bands under UV light.

Scoring of quantitative trait phenotypes

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. The environment in the growth chamber was controlled, and maintained at 16 h photoperiod and 24^0C temperature.

The phenotypic traits such as Germination Time (GT), Flowering Time (FT) (days from germination to flowering), Height at Flowering (HF), Rosette Width at Flowering (RWF), Rosette Leaf Number at Flowering (RLNF) and Cauline Leaf Number at Flowering (CLNF) were scored in each plant.

Analysis of phenotypic data

Phenotypic data were analysed in MINITAB version 16 using General Linear Models procedure. A hierarchical ANOVA was performed considering genotypes and the lines within respective genotypes as the sources of variation. All the lines having the same genotype for all the marker bins were considered as common STAIR genotype. Because the time for germination was revealed to be significantly different among the genotypes it was used as a covariate in the ANOVA for all the other traits to account for the variation this impose on the other traits scored.

Least squares model fitting

Results of the ANOVA were combined with genotypic data to construct a model in locating QTL by least squares model fitting for significantly different traits among STAIR genotypes. Weights were calculated (weight = n/variance) for the means of the STAIR genotypes for each of the significantly different trait and for fitting the model. Models which were 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 (recurrent genotype) alleles by Landsberg (donor) alleles were calculated for each QTL region.

RESULTS AND DISCUSSION

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 genotypes of each line for 9 markers within chromosome 3 are given in Table 1. The results of the 20 markers along the rest of the 4 chromosomes verified the homozygous recurrent genotype for all the 20 markers in all the lines.

Table 1. Genotype scores of wide STAIRS for the nine microsatellite markers on chromosome 3 (11 = Col homozygous; 22 = L*er* homozygous).

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

Based on the genotypes sixteen lines were grouped into eight STAIR genotypes (including CSS3) demarcated by different marker bins. 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.

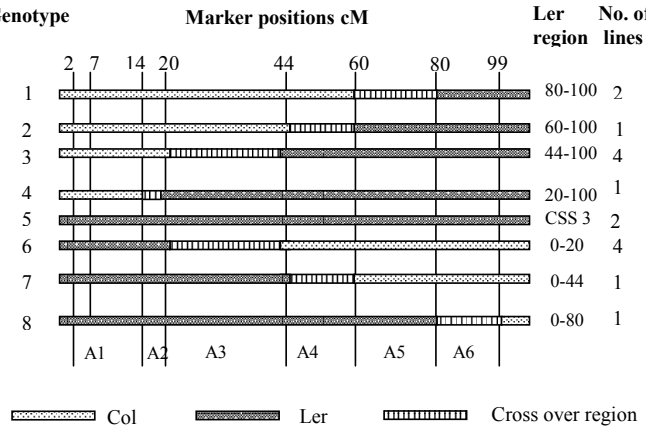


Fig. 4. Wide STAIRS: marker bins, genotypes, lines within genotypes and introgressed and crossover regions.

Analysis of quantitative trait variation by Least Squares Model fitting

ANOVA revealed significant differences between genotypes for all the traits scored. The model used for detecting and estimating genetic differences among STAIRS is given in Table 2. According to this model the marker demarcated regions in genetic distances are as follows.

a1 = 0-14 cM, a2 = 14-20 cM, a3 = 20-44 cM,
a4 = 44-60 cM, a5 = 60-80 cM, a6 = 80-100 cM

Table 2: Model for detecting and estimating genetic differences among STAIRS. Column m accounts for the mean and a1 to a6 accounts for the effect of each marker demarcated bin in the model. (-) denotes regions of introgressed L*er* donor alleles and (+) denotes regions of recurrent Col alleles.

STAIRS	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

Table 3. Significance levels in ANOVA (P value), chi-squared value of the best-fit model and regions of QTL.

Trait	P value	χ^2 of the best-fit model	QTL regions	No.of QTL regions
- FT	>0.0001	9.25, 6 d.f	a1	1
- HF	0.0010	4.52, 4 d.f	a1, a3, a6	3
- RLNF	>0.0001	7.04, 6 d.f	a1	1
- CLNF	0.0100	5.01, 6 d.f	a1	1
- RWF	0.0020	6.31, 6 d.f	a1	1
Total				7

Effects of substitution of Col alleles by L*er* Alleles

The substitution of Col alleles in region a1 by L*er* alleles causes delaying in flowering by 10.767 days under these experimental conditions. 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. The substitution of Col in region a1 by L*er* has an increasing effect on rosette leaf number at flowering by 7.214 mm. When Landsberg substitutes Columbia in region a1, the cauline leaf number at flowering increases by 0.606. Finally, 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.

Location of QTL for traits scored

A total of 7 QTL affecting the 5 traits were located along chromosome 3. Results indicated the presence of QTL in region a1 (0-14 cM) affecting all the traits scored. Two additional QTL were located for height at flowering in regions a3 (20-44 cM) and a6 (80-100 cM). The presence of QTL in region a1 for all the measured traits may be due to tight linkage of different genes in this region. But it is more likely that all or most of the traits being affected by a single gene showing pleiotropy because all the traits scored were related to flowering.

No QTL were located within the regions a2, a4 and a5. This may simply be due to the absence of QTL in these regions affecting the traits under study. Or else, if in fact there are QTL 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 but a much wider region (Jansen, 1996). Furthermore, Koornneef *et al.* (1998) reported QTL on top of chromosome 3 using a L*er* x Col RIL population. Koumproglou *et al.* (2002) using some of the same STAIRS, located QTL for flowering time in a more wider region *i.e.* 0-21 cM on top of chromosome 3 for flowering time and rosette leaf number. In the current research we were able to narrow down the

previously reported QTL region for flowering time, height and rosette leaf number, locate new QTL for cauline leaf number and rosette width and also locate two additional QTL for plant height along chromosome 3.

STAIRS in QTL location

Koumproglou *et al.* (2002) reported the first demonstration of QTL mapping using STAIRS and during the current study we were able to map QTL affecting more traits to a finer region. 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 or else can be used to map different QTL which show polymorphism between Col and L*er*. With the invention of new polymorphic markers within marker bins the same STAIRS can be re-genotyped to achieve better precision in QTL mapping. The STAIRS have been introduced as a tool for fine mapping of QTL. The region of QTL location in the current study is still 10-15 cM and in order to achieve finer mapping, it is needed to proceed to the third stage of this procedure, which is the use of narrow STAIRS (Perera, 2005).

CONCLUSIONS

STAIRS are recently produced powerful resources in locating QTL. The top region of chromosome 3 houses QTL for flowering time and flowering related traits, height, rosette leaf number, cauline leaf number and rosette width at flowering. Two additional loci controlling plant height are located in the middle and towards the end of chromosome 3.

In order to fine map the QTL related to flowering time, it is necessary to proceed to the construction of narrow STAIRS within the top 14 cM of chromosome 3 of *Arabidopsis*.

ACKNOWLEDGEMENTS

The CSSs and some of the STAIRS used in the current research were produced at The University of Birmingham by Koumproglou *et al*., The first author wishes to thank the group for making these lines available for the current study.

The first author gratefully acknowledges the postgraduate research scholarship by the ADB Science and Technology Personnel Development Project in Sri Lanka. The research was supported by a grant from BBSRC, U.K. Valuable guidance from the local supervisor Mr. Everard Jayamanne/CRI and the technical assistance from Sue Bradshaw, Karen Stapler, Sandya Fernando, Nissanka Herath and Ayodya Fernando is gratefully acknowledged.

REFERENCES

Allen, F.L. (1994). Usefulness of Plant Genome Mapping to Plant Breeding. *In*: Gresshoff, P.M. (Ed). Plant Genome Analysis. CRC Press.

Bell, C.J. and Ecker, J.R. (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics. 19: 137-144.

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

Hyne, V. and Kearsey, M.J. (1995). QTL Analysis - further uses of marker regression. Theoreti. and Appl. Geneti. 91: 471-476.

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

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

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

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

Koornneef, M., Alonso-Blanco, C., Peeters, A.J.M. and Soppe, W. (1998). Genetic control of flowering time in *Arabidopsis*. Annual Rev. of Plant Physio. and Plant Molecu. Biolo. 49: 345-370.

Koumproglou, R., Wilkes, T.M., Townson, P., Wang, X.Y., Beynon, J., Pooni, H.S., Newbury, H.J. and Kearsey, M.J. (2002). STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. Plant Journal 31: 355-364.

Perera, S.A.C.N. (2005). Fine mapping of QTL and microarray gene expression studies in *Arabidopsis* using STAIRS. Ph.D. thesis, The University of Birmingham, UK.

Ramsay, L.D., Jennings, D.E., Bohuon, E.J.R., Arthur, A.E., Lydiate, D.J., Kearsey, M.J. and Marshall, D.F. (1996). The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. Genome. 39: 558-567.