

TECHNICAL ADVANCE

Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*

Clare Lister and Caroline Dean*

AFRC IPSR Cambridge Laboratory, John Innes Centre,
Colney Lane, Norwich NR4 7UH, UK

Summary

Recombinant inbred (RI) lines of *Arabidopsis thaliana* (*Arabidopsis*) have been generated from a cross between the ecotypes Landsberg *erecta* and Columbia. Progeny of 300 individual F₂ seedlings were taken by single seed descent to the F₈ generation. Sixty-seven loci, scored using 64 RFLP probes and one phenotypic marker, chosen at approximately 20 cM intervals from the two previously published RFLP maps, were mapped using 100 of these RI lines. More than 500 other new loci are currently being mapped using these RI lines by several other groups. These 100 RI lines thus provide the material to map new probes or phenotypic traits polymorphic between Landsberg *erecta* and Columbia, relative to an increasing number of molecular markers. Higher resolution mapping of distinct chromosomal regions can be achieved by analysing the segregation of particular markers on the additional 200 RI lines.

Introduction

The small crucifer *Arabidopsis thaliana* (*Arabidopsis*) has many advantages as a model plant system for molecular genetic analysis. In particular, its small genome size (approximately 100 Mb; Meyerowitz, 1992) and low abundance of repetitive sequences (Leutwiler *et al.*, 1984; Pruitt and Meyerowitz, 1986) mean that gene isolation by chromosome walking is a viable prospect. Well-defined RFLP maps are the starting point for these walks. They also provide the framework for mapping new probes and for the generation of physical maps (Hwang *et al.*, 1991).

The two published RFLP maps for the *Arabidopsis* genome analysed the segregation of markers in F₃ families (Chang *et al.*, 1988; Nam *et al.*, 1989). The disadvantage of mapping with these populations is that the seed stocks were eventually exhausted. Recombinant inbred (RI) populations provide an alternative population for

mapping. Although they take much longer to produce, once they are made they are nearly homozygous and can be replicated indefinitely whilst maintaining the genetic fidelity of each line. They can be widely distributed enabling many different laboratories to map on to the same population and thus to directly compare mapping data. Data can be collected and held in a central database resulting in the continual addition and refinement of the map. The accuracy of this map should be significantly higher than a map derived from the statistical integration of mapping data derived from similarly sized but different mapping populations.

RI populations have been used for mapping in mouse (Bailey, 1981), wheat (Snape *et al.*, 1985), pea (Ellis *et al.*, 1992), maize (Burr *et al.*, 1988; Burr and Burr, 1991) and *Arabidopsis* (Reiter *et al.* 1992). Reiter *et al.* (1992) produced 150 RI lines from a cross between the *Arabidopsis* ecotype WS and the line W100, multiply marked with phenotypic markers (Koornneef *et al.*, 1987). Nine phenotypic markers were mapped on all 150 lines, 252 RAPD markers were mapped on to 46 of the lines and 60 published RFLP markers (44 from Chang *et al.*, (1988) and 16 from Nam *et al.*, (1989)) were mapped on to 115 of the lines.

We have generated a second, larger family of *Arabidopsis* RI lines using the ecotypes Landsberg *erecta* (La-er) and Columbia (Col). Having a second population of RI lines to draw on will significantly increase the chance of finding a polymorphism with a particular probe and thus increase the number of markers which can be mapped. We describe here the construction of the lines and the mapping of 64 previously described RFLP markers. These were selected with the aim of providing the best integration of the two published RFLP maps.

Results

Generation of the recombinant inbred lines

The ecotypes La-er and Col were crossed, using La-er as the female. Seed from one F₁ plant was grown and allowed to self-fertilize. F₂ seed were sown and self-fertilized. This procedure was repeated to the F₈ generation. In order to minimize any bias in the selection of plants taken through to the next generation 10–15 seeds were sown and a single plant nearest one corner of the

pot was selected for transplanting. F_9 seed were bulk harvested from 330 single F_8 plants. Three hundred of these were selected for distribution and of these, 100 have been characterized with respect to the segregation of 64 previously published RFLP markers. For subsequent generations, it will be important to sow at least 20 seeds and bulk harvest seed from all the plants. This should minimize any deviation in populations, due to any remnant heterozygosity (described below), from the initial population used for mapping.

Segregation pattern of known loci in the RI lines

RFLP markers from the Meyerowitz and Goodman laboratories, 31 and 33 respectively (Chang *et al.*, 1988; Hauge *et al.*, 1993; Nam *et al.*, 1989), were hybridized to Southern blots carrying DNA from the 100 RI lines picked

to become the basic mapping population. The markers chosen were selected as being evenly distributed (approximately every 20 cM from each map), while minimizing as much as possible the number of different enzyme digests required to analyse all the markers. This was to allow a more accurate integration of the two published RFLP maps than had previously been obtained. The *erecta* mutation was also scored. Analysis of the segregation patterns using the MAPMAKER program gave the expected five linkage groups. The most likely orders for the markers (except for chromosome 4, see later) are shown in Figure 1. All the markers mapped to the appropriate chromosome in the same order as in the original maps. There was a good integration of the RFLP markers from the two different maps indicating that there were no significant differences in the distribution of markers from the two previous maps.

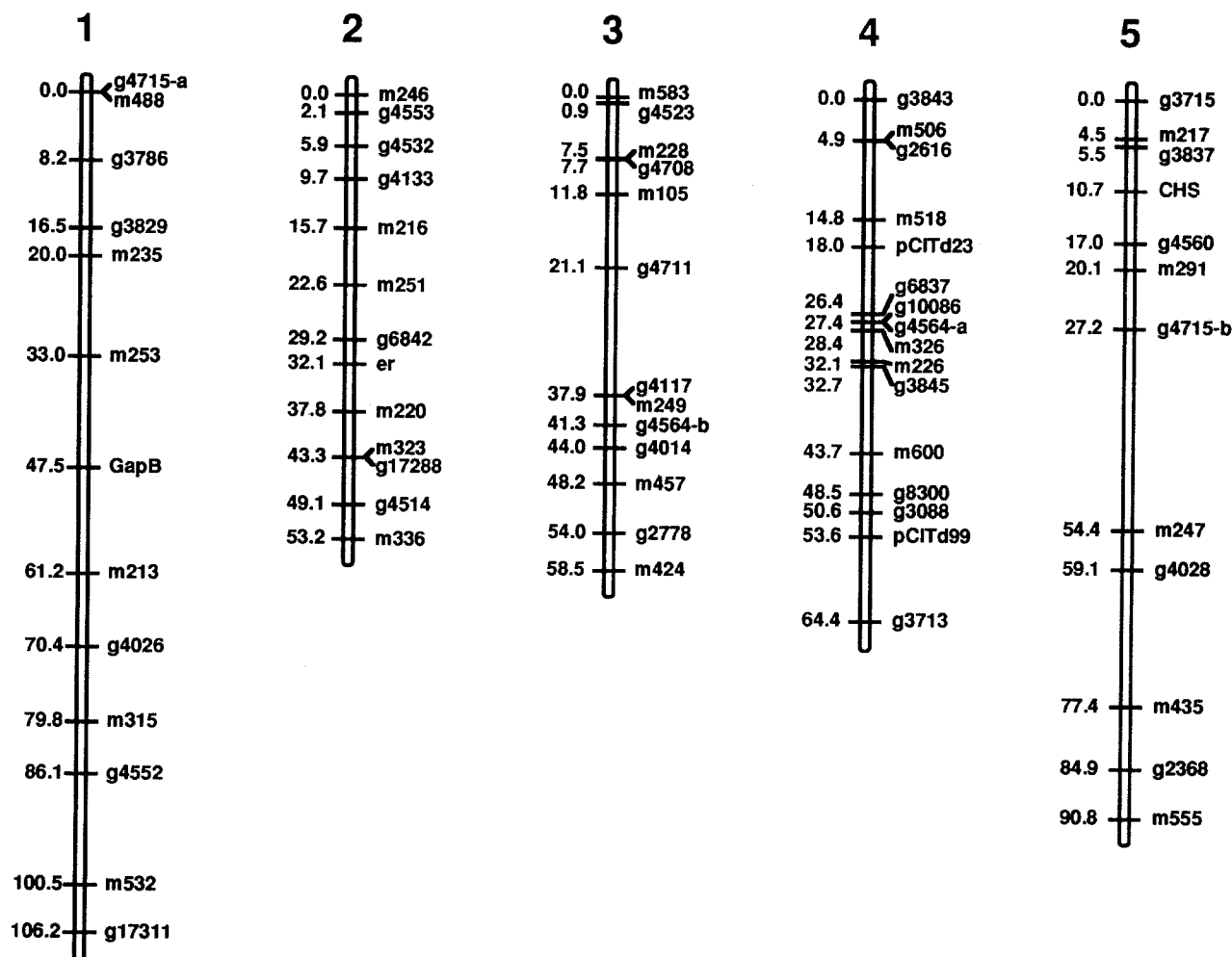


Figure 1. Linkage map of Arabidopsis generated using these RI lines.

Map distances are shown in centimorgans. The first marker at the top of each of the five chromosomes has been given the arbitrary position 0. All loci were linked and ordered using the MAPMAKER program with a LOD score of 6.0. Recombination frequencies and cM values were calculated as described in the Experimental procedures. Two markers at the same position have an equally likely order. The letter a or b after the marker name indicates that more than one locus (from cross-hybridizing sequences) was mapped with the marker.

Segregation frequencies

Statistically, by the F_8 generation, the probability that a particular locus will still be heterozygous is 0.78%. When the frequency of heterozygosity was analysed at the 67 loci, in each of the lines, the averaged frequency was found to be 0.42%, with no locus having significantly more than the expected value. Thus, there does not seem to have been any selection for maintaining heterozygosity in the different parts of the *Arabidopsis* genome.

If there had been no bias in the selection of individual plants taken from generation to generation then each marker would be expected to segregate 1:1 for each parental allele. When the ratio of La-er:Col alleles for each line was determined, it was clear that some were significantly different (at the 5% level) from the expected 1:1 ratio. The deviation was not huge as most ratios were below 2:1. Markers showing significant deviation also tended to be clustered in certain regions of the genome. This is illustrated schematically in Figure 2. A region encompassing five markers (m213, g4026, m315, g4552, m532) on the lower part of chromosome 1 showed ratios of 1.7:1, 2:1, 2.3:1, 2.4:1 and 1.9:1 in favour of the Col allele. Eleven of the 13 markers on chromosome 2 showed ratios between 2.1:1 to 1.7:1, in favour of the La-er allele. On chromosome 4, the markers g2616 and m506 showed ratios in favour of La-er (1.5:1, 1.6:1) and on chromosome 5 one marker (m555) showed a 1.6:1 ratio in favour of Col and six markers (g3715, m217, g3837, CHS, g4560, m291), covering the top arm showed ratios on favour of the La-er allele (between 2.7:1 to 1.5:1). There has clearly been selection for the pres-

ence of La-er or Col DNA in distinct regions of the genome.

Comparison of the map with the statistically integrated *Arabidopsis* map

A map integrating the phenotypic marker map (Koornneef, 1990), and the two published RFLP maps (Chang *et al.*, 1988; Nam *et al.*, 1989) has recently been produced (Cherry *et al.*, 1992; Hauge *et al.*, 1993, available on AAtDB version 1.3). The integration relied on the use of a new program called JOINMAP (Stam *et al.*, 1993), that statistically integrates recombination data measured in different populations. A comparison of the marker order and interval distances between the map generated from the RI lines and the integrated map showed good correspondence over a large proportion of the genome. However, there were some regions where the marker order was significantly different. These most likely reflect inaccuracies caused by combining data from several populations. The most significant of these was a region on chromosome 4, where the RI map gave the order: g6837; g10086; g4564a; m326; m226; g3845; m600; g8300; g3088; pCITd99 instead of an order: g6837; m326; g10086; m226; g4564a; g3845; g8300; m600; pCITd99; g3088. The most likely order from the RI data placed g3845 between m226 and m326 (10^4 more likely), however, physical mapping data in this region conclusively placed m226 between m326 and g3845 (Schmidt and Dean, unpublished data), so this order was chosen. Physical mapping also confirmed the order of the markers g6837 to g3845 (Schmidt and Dean unpublished data). On chromosome 1, the most likely order from the RI map switched the two markers g3829 and m235. On chromosome 2, the RI map gave the order: g4553; g4532 and g4133 (as in the original map, Nam *et al.*, 1989), rather than: g4133; g4532; g4553. Lastly, the order of the two markers g4014 and m457 chromosome 3 was reversed relative to the integrated map.

Discussion

We have generated RI lines between the ecotypes *Landsberg erecta* and *Columbia*. Three hundred F_8 lines are now currently available. Sixty-seven loci have been mapped on to 100 of these lines. Sixty-six were scored using 64 RFLP markers from the two previously published maps and the 67th was the phenotypic *erecta* mutation. In addition, a large number of markers are currently being mapped on to the RI lines by a number of different laboratories: R. Whittier, Mitsui Corp. Japan; J. Ecker, University of Pennsylvania; M. Zabeau, Keygene, Holland; B. Osborne and B. Baker, USDA, Albany; GDR-Arabidopsis, France. Our aim is to generate a very

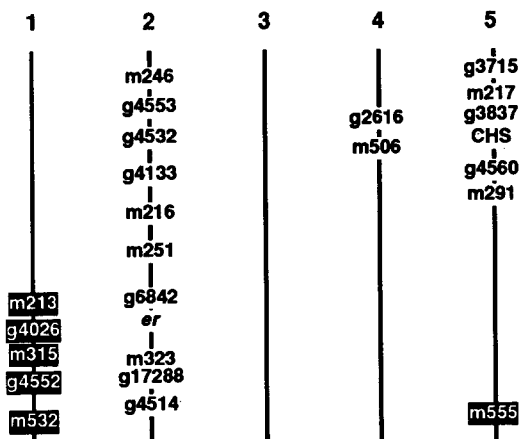


Figure 2. Markers showing a significant deviation from a 1:1 segregation ratio.

Markers showing a significant deviation from a 1:1 ratio for the *Landsberg erecta*: *Columbia* allele (at the 5% level) are indicated. Black letters on a white background represents markers where significantly higher numbers of lines showed the *Columbia* allele. White letters on a black background represents markers where significantly higher numbers of lines showed the *Landsberg erecta* allele.

high density RFLP map. At a certain point, ordering of closely linked markers will no longer be possible with the 100 RI lines. Further ordering of these markers could be achieved using all 300 of the RI lines. Indeed, we have utilized all 300 lines to facilitate ordering 30 markers covering a 15 cM section of chromosome 4 (Stammers and Dean unpublished data).

A high density RFLP map will greatly facilitate the fine mapping of new mutations prior to chromosome walking. Recombinants closely linked to the locus would be used to order the markers mapping nearby. It may be possible to identify two markers flanking the locus of interest that are so close that they hybridize to the same YAC clone, thereby eliminating all the chromosome walking steps needed for gene isolation. A high density RFLP map will also be valuable in the construction of an overlapping genomic library of the Arabidopsis genome (Hwang *et al.*, 1991, Schmidt *et al.*, 1992).

In addition to mapping new markers, these RI lines can be used to map (to a high resolution if all 300 lines are used) any phenotypic trait that is polymorphic between the two parents. Now that sufficient molecular markers have been mapped on to the lines, all that is required to map a new trait is to score that trait in the 100 RI lines and then compare the segregation data with that for the known markers. The RI lines have been actively exploited to map disease resistance genes, for example resistance to the fungus *Peronospora parasitica* (Parker *et al.*, 1993).

The RI lines also provide excellent material to map quantitative trait loci (QTLs; Knapp and Bridges, 1990; Lander and Botstein, 1989; Paterson *et al.*, 1988). A wide range of phenotypic traits, including stature, rosette size, leaf shape and flowering time are segregating in the lines. We are currently analysing flowering time in the 100 RI lines which have been scored for RFLP marker segregation. Even though both parents show early flowering characteristics, there is considerable variation in flowering time in the RI lines. Thus, for quantitative traits even when both parents appear to be phenotypically similar for a given trait, there may be considerable variation between the RI lines suggesting that modifiers affecting that trait can be mapped.

The 67 loci mapped on to these RI lines were selected so as to give a good integration of the two published RFLP maps. Previous integrations relied on 17 common markers (Nam *et al.*, 1989) or 16 common markers (Reiter *et al.*, 1992). Statistical integration of different maps using the JOINMAP program (Stam, 1993) has given an integrated map showing good correspondence with this RI map. However, a map generated statistically, from data from similarly sized but different populations, is likely to be less reliable than a map generated on one population. Now that the RI lines have been characterized,

future mapping with loci polymorphic between La-er and Col, can all been done on these lines and so mapping data can be compared directly.

Parents were chosen for the generation of the RI lines that did not carry the phenotypic markers present in the W100 line (except for the *ERECTA* locus). This was to avoid, as much as possible, any bias for parental type in different parts of the genome due to under-representation of certain mutant phenotypes. Despite this, there was a slight bias for certain parental types indicating that there had been selection for the presence of La-er or Col DNA in distinct regions of the genome. It would be interesting to compare if this had also happened in the other Arabidopsis RI lines (Reiter *et al.*, 1992) and if similar genomic regions were involved. The slight bias in parental types in certain parts of the genome does not seem to impair the resolution of mapping with these lines.

The characterization of the segregation profile of the different markers on each of the RI lines enables lines to be pooled to facilitate the identification of new markers in certain regions of the genome. This has been previously described by Reiter *et al.* (1992) using RI lines and Michelmore *et al.* (1991) using a technique they termed 'bulked segregant analysis' for use with segregating populations. Targeting markers will be an important facility, especially as new PCR-based markers are developed, for both chromosome walking experiments and for global physical mapping (Arratia *et al.*, 1991; Ewens *et al.*, 1991).

Our intent is to distribute these RI lines as widely as possible. The seed is being bulked up on a very large scale (to avoid individual laboratories having to do this) and will be housed at the Nottingham Arabidopsis Stock Centre. Seed will also be available in the long term from the Arabidopsis Biological Resource Center at Ohio State University. A mini-database carrying the segregation data of the 67 loci described above is available through AAtBD. We intend to have only one database housing all the segregation data on these RI lines which will initially be at Norwich. Users will be encouraged to submit new mapping information to the central RI database in order to map their new markers relative to all other markers mapped using this RI population. Updates of the RI map will be sent to AAtDB at regular intervals.

Experimental procedures

Plant handling

Landsberg *erecta* and Columbia seed (obtained from M. Koornneef and C. Somerville, respectively) were sown in a soil mix consisting of John Innes No.1, vermiculite and grit in a ratio of 1.5:1:1. Imbibed seed were cold treated for 3 days and then grown in the glasshouse or controlled environment cabinets under long or continuous day conditions. Plants from which

genomic DNA was to be prepared were grown under short-day conditions to increase the amount of leaf material.

Preparation of genomic DNA and Southern blot analysis

Genomic DNA was prepared from leaf material from 4–5-week-old plants. The material was frozen in liquid N₂ prior to extraction using a miniprep CTAB method (Dean *et al.*, 1992). Genomic DNA (1–2 µg) was digested with restriction enzymes in a Cutsall buffer (20 mM Tris-HCl pH 7.5, 7 mM MgCl₂, 100 mM KCl, 2 mM β-mercaptoethanol), subjected to electrophoresis in 0.8% agarose gels, and transferred to Hybond-N (Amersham) according to the manufacturer's recommendations. Hybridizations were carried out in 5 × SSC, 0.5% SDS, 5 × Denhardt's solution. Filters were washed in 2 × SSC, 1% SDS, at 65°C, followed by 0.1 × SSC, 1% SDS at 65°C. Filters were exposed on a Kodak X-Omat XAR X-ray film for 1–14 days. Between hybridizations, filters were stripped using 0.4 M NaOH, at 45°C for 30 min followed by 0.1 × SSC, 0.1% SDS, 0.2 M Tris-HCl, pH 7.5 at 45°C for 15 min. Filters could be reused at least 10 times.

Details of which enzymes would reveal polymorphisms between the Col and La-er ecotypes for most of the markers were provided by the Meyerowitz and Goodman laboratories. Most of this information is available on the database AAtDB version 1.3 (Cherry *et al.*, 1992) and so has not been given here. These were confirmed on parental blots before they were used on the 100 RI lines. Polymorphisms for markers not listed on AAtDB were obtained with the following enzymes: m583-BglI; m600-XbaI; pCITd99-XbaI; CHS-XbaI; m220-EcoRV; m424-EcoRV; pCITd23-EcoRI.

Mapping the GAP-B locus.

A co-dominant ecotype-specific PCR based marker had been designed for this locus by Konieczny and Ausubel (1993). Primers hybridizing to the GAP-B locus were used to PCR amplify a fragment which was then digested with DdeI. The products were resolved on a 2.5% agarose gel. The presence of a 605 bp fragment represented the Col allele, whilst the presence of a 350 bp fragment represented the La-er allele.

Linkage analysis

The markers were scored as L (La-er homozygote), C (Col homozygote), H for definite heterozygotes or U for unclassified. The data was entered into an Mac Excel spreadsheet and the linkage analysis done using a Macintosh version no. 1.0 of the MAPMAKER mapping program of Lander *et al.* (1987), supplied by S. Tingey (DuPont Co.). This program was designed for mapping with F₂ families resulting in a doubling of the LOD scores when RI data is analysed; thus a LOD score of 6.0 was used for the linkage threshold. Recombination frequency (r) was calculated from the fraction of recombinants (R; estimated using the MAPMAKER F₂ algorithm) using the equation $r=R/2(1-R)$ (Haldane and Waddington, 1931). Recombination frequencies (r) were converted to map distances in cM using the Kosambi mapping function (Kosambi, 1944).

Acknowledgements

We thank Melanie Stammers and Veronique Szabo for unpublished segregation data and Kristina Bergstrom

for technical help with the seed harvesting and DNA preparation. The authors gratefully acknowledge polymorphism information and RFLP markers supplied by the Meyerowitz and Goodman laboratories. We also acknowledge financial support from the AFRC PMB II programme to fund the propagation of large amounts of seed for distribution. We thank Drs John Snape, Noel Ellis and David Flanders for critically reading the manuscript.

References

- Arratia, R., Lander, E.S., Tavare, S and Waterman, M.S. (1991) Genomic mapping by anchoring random clones: A mathematical analysis. *Genomics*, **11**, 806–827.
- Bailey, D.W. (1981) Recombinant inbred strains and bilineal congenic strains. In *The Mouse in Biomedical Research*, 1 (Foster, H.J., Small, J.D. and Fox, J.G. eds.). New York: Academic Press, pp. 223–239.
- Burr, B., Burr, F.A., Thompson, K.H., Albertson, M.C. and Stuber, C.W. (1988) Gene mapping with recombinant inbreds in maize. *Genetics*, **118**, 519–526.
- Burr, B. and Burr, F.A. (1991) Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. *Trends Genet.* **7**, 55–60.
- Chang, C., Bowman, J.L., DeJohn, A.W., Lander, E.S. and Meyerowitz, E.M. (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **85**, 6856–6860.
- Cherry, J.M., Cartinhour, S.W. and Goodman, H.M. (1992) AAtDB, an *Arabidopsis thaliana* database. *Plant Mol. Biol. Rep.* **10**, 308.
- Dean, C., Sjodin, C., Page, T., Jones, J. and Lister, C. (1992) Behaviour of the maize transposable element Ac in *Arabidopsis thaliana*. *Plant J.* **1**, 69–81.
- Ellis, T.H.N., Turner, L., Hellens, R.P., Lee, D., Harker, C.L., Enard, C., Domoney, C. and Davies, D.R. (1992) Linkage maps in pea. *Genetics*, **130**, 649–663.
- Ewens, W.J., Bell, C.J., Donnelly, P.J., Dunn, P., Matallana, E. and Ecker, J.R. (1991) Genome mapping with anchored clones: Theoretical aspects. *Genomics*, **11**, 799–805.
- Haldane, J.B.S. and Waddington, C.H. (1931) Inbreeding and linkage. *Genetics*, **16**, 357–374.
- Hauge, B., Hanley, S., Cartinhour, S., *et al.* (1993) An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. *Plant J.* **3**, 745–754.
- Hwang, I., Kohchi, T., Hauge, B.M., *et al.* (1991) Identification and map position of YAC clones comprising one third of the *Arabidopsis* genome. *Plant J.* **1**, 367–374.
- Konieczny, A. and Ausubel, F.M. (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* **4**, 403–410.
- Koornneef, M., Hanhart, C.J., van Loenen Martinet, E.P., Peeters, A.J.M. and van der Veen, J.H. (1987) A marker line, that allows the detection of linkage on all *Arabidopsis* chromosomes. *Arabidopsis Inf. Serv. Frankfurt*. **23**, 46–50.
- Koornneef, M. (1990) Linkage map of *Arabidopsis thaliana* (2n=10). In *Genetic Maps* (O'Brien, S.J., ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 6.95–6.97.
- Kosambi, D.D. (1944) The estimation of map distance from recombination values. *Ann. Eugen.* **12**, 172–175.
- Knapp, S.J. and Bridges, W.C. (1990) Using molecular markers to estimate quantitative trait locus parameters: power and

- genetic variances for unreplicated and replicated progeny. *Genetics*, **126**, 769–777.
- Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Day, M.J., Lincoln, S.E. and Newberg, L.** (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, **1**, 174–181.
- Lander, E.S. and Botstein, D.** (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.
- Leutwiler, L.S., Hough-Evans, B.R. and Meyerowitz, E.M.** (1984) The DNA of *Arabidopsis thaliana*. *Mol. Gen. Genet.* **194**, 15–23.
- Meyerowitz, E.M.** (1992) Introduction to the *Arabidopsis* genome. In *Methods in Arabidopsis Research* (Koncz, C., Chua, N-H., Schell, J., eds). Singapore: World Scientific Publishing Co. Pte. Ltd., pp. 101–118.
- Michelmore, R.W., Paran, I. and Kessell, R.V.** (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. *Proc. Natl. Acad. Sci. USA*, **88**, 9828–9832.
- Nam, H-G., Giraudat, J., den Boer, B., Moonan, F., Loos, W.D.B., Hauge, B.M. and Goodman, H.** (1989) Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. *Plant Cell*, **1**, 699–705.
- Parker, J., Szabo, V., Staskawicz, B., Lister, C., Dean, C., Daniels, M. and Jones J.** (1993) Phenotypic characterization and molecular mapping of the *Arabidopsis thaliana* locus *RPP5* determining resistance to *Peronospora parasitica*. *Plant J.* **4**, in press
- Paterson, A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E. and Tanksley, S.** (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature*, **335**, 721–726.
- Pruitt, R.E. and Meyerowitz, E.M.** (1986) Characterization of the genome of *Arabidopsis thaliana*. *J. Mol. Biol.* **187**, 169–183.
- Reiter, R.S., Williams, J.G.K., Feldmann, K.A., Rafalski, J.A., Tingey, S.V. and Scolnik, P.A.** (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. *Proc. Natl. Acad. Sci. USA*, **89**, 1477–1481.
- Schmidt, R., Cnops, G., Bancroft, I. and Dean, C.** (1992) Construction of an overlapping YAC library of the *Arabidopsis thaliana* genome. *Aust. J. Plant Physiol.* **19**, 341–51.
- Snape J.W., Flavell, R.B., O'Dell, M., Hughes, W.G. and Payne, P.P.** (1985) Intrachromosomal mapping of the nucleolar organiser region relative to three marker loci on chromosome 1B of wheat (*Triticum aestivum*). *Theor. Appl. Genet.* **69**, 263–270.
- Stam, P.** (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. *Plant J.* **3**, 739–744.