*The Plant Journal* (1993) **3**(5), 745–754

TECHNICAL ADVANCE

# An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome

**Brian M. Hauge, Susan M. Hanley, Sam Cartinhour, J. Michael Cherry and Howard M. Goodman***
*Department of Genetics, Harvard Medical School and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA*
**Maarten Koornneef[1] and Piet Stam[1,2]**
*[1]Department of Genetics, Wageningen Agricultural University, Dreyenlaan 2, 6703 HA Wageningen, The Netherlands, and*
*[2]Centre for Plant Breeding and Reproduction Research (CPRO-DLO), PO Box 16, 6700 AA Wageningen, The Netherlands*
**Caren Chang, Sherry Kempin, Leonard Medrano and Elliot M. Meyerowitz**
*Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA*

## Summary

**We have assembled an integrated genetic/restriction fragment length polymorphism (RFLP) linkage map of the nuclear genome of the flowering plant *Arabidopsis thaliana*. The map is based on two independent sets of RFLP data, RFLP data for 123 new markers, and pairwise segregation data of 125 classical genetic markers. Mathematical integration of the independent data sets was performed using the JOINMAP computer package. Sixty-two markers common to two or more data sets were exploited to facilitate integration of the individual maps. The current map, which encompasses a total genetic distance of 520 cM, contains 125 classical genetic markers and 306 RFLP markers. Comparison of the integrated consensus map with the individual maps demonstrates that the overall linear order of the integrated map is in good agreement with the component maps. It must be emphasized, however, that the integrated map represents the 'best fit' which is clearly subject to the statistical limitations of the available data. We present several examples where local differences in map order are observed between the integrated and component maps. It is likely, given the problems associated with statistical integration of mapping data from different populations, that the integrated map will contain additional local inconsistencies and problematic regions. None the less, the unified map provides a framework for building an increasingly accurate and useful map. Subsequent refinements of the map will be available electronically and researchers are invited to submit revised map data to the corresponding author for inclusion in future updates (see Appendix 1).**

Received 23 October 1992; revised 18 December 1992; accepted 22 December 1992.
*For correspondence at Department of Molecular Biology, Wellman 11, Massachusetts General Hospital, Boston, MA 02114, USA (fax +1 617 726 3535).

## Introduction

*Arabidopsis thaliana* is popular as a model system for the study of plant biology. Its small size, short life cycle and large seed output make *Arabidopsis* well suited for classical genetic analysis. Mutations affecting a wide range of developmental and metabolic processes have been described (reviewed in Meyerowitz, 1989) and a genetic linkage map containing 86 morphological markers has been assembled (Koornneef, 1990). For molecular biological studies *Arabidopsis* offers the additional advantage of having a very small genome containing a relatively low content of interspersed repetitive DNA (Leutwiler *et al.*, 1984; Pruitt and Meyerowitz, 1986). The small, relatively simple genome greatly simplifies the cloning of genes which have been identified by mutational analysis.

Many of the tools required for gene isolation and genome mapping in *Arabidopsis* have either been developed or are currently being developed. An increasing number of cloned genes, restriction fragment length polymorphism (RFLP) (Chang *et al.*, 1988; Nam *et al.*, 1989) and random amplified polymorphic DNA (RAPD) (Reiter *et al.*, 1992) markers are available for mapping and cloning studies. An overlapping cosmid map covering 90–95% of the *Arabidopsis* genome has been assembled (Hauge *et al.*, 1991). Ordered arrays of yeast artificial chromosome libraries (YACs) are available (Grill and Somerville, 1991; Ward and Jen, 1990) and are being widely used in the *Arabidopsis* community (Hwang *et al.*, 1991).

Restriction fragment length polymorphisms (RFLPs) are being increasingly employed to assemble linkage maps of complex genomes (Burr and Burr, 1991; Donis-Keller *et al.*, 1987; Helentjaris *et al.*, 1986). Mapped RFLPs provide a collection of DNA probes for mapping Mendelian traits such as genetic diseases and loci of agronomic importance (Helentjaris and Burr, 1989). The RFLP markers also

serve as a source of probes for cloning genes based solely on their mutant phenotype and genetic map position. The general approach for gene isolation based on positional information is to identify linked RFLP probes residing within one to several cM of the gene of interest. The cloning of the gene is then achieved by bridging the intervening gap using techniques such as chromosome walking (Bender *et al.*, 1983) or chromosome jumping (Poustka *et al.*, 1987). Owing to the potential of the RFLP technology for gene mapping and subsequent gene isolation, the development of RFLP maps has proceeded rapidly for many organisms. Accordingly, the large amount of available linkage information has created the need to integrate data which have been independently obtained. For example, two RFLP maps have been published for maize (Burr and Burr, 1991; Coe *et al.*, 1990) and *A. thaliana* (Chang *et al.*, 1988; Nam *et al.*, 1989).

Four independent linkage maps of the *A. thaliana* genome have been published; the classical genetic map (Koornneef, 1990), two independent RFLP maps (Chang *et al.*, 1988; Nam *et al.*, 1989) and more recently a map based on the meiotic segregation of PCR-based RAPD markers (Reiter *et al.*, 1992). While each of these maps contains common markers and the relative order of these markers is maintained in the individual maps, considerable variation in the map distances is observed between markers common to the individual maps. Consequently, it is difficult to align the maps with confidence, and the relative order between markers from the distinct maps is frequently unclear. This is especially true over small intervals and in regions of the genome containing few common markers. This problem is augmented by the fact that the raw data and/or the software used to generate the maps may not be readily available or familiar to the individual researcher. As an initial step towards alleviating these problems we have generated a single common RFLP/genetic linkage map by mathematical integration of the two independent RFLP maps (Chang *et al.*, 1988; Nam *et al.*, 1989), the classic genetic linkage map (Koornneef *et al.*, 1992), and a series of newly mapped RFLP markers (Goodman *et al.*, unpublished data; Meyerowitz *et al.*, unpublished data).

In this communication we present an integrated linkage map of the *A. thaliana* genome. The integrated map is based on two independent sets of RFLP data from eight different crosses and pairwise segregation data of classical genetic markers. Sixty-two markers common to at least two of the independent data sets provide points of contact between the individual maps. Mathematical integration of the data sets was performed using the JoinMap (Stam, 1993) computer program. The integrated map consists of 306 molecular markers and 125 classical genetic markers.

## Results and discussion

The mapping data used for this analysis are of two types; raw $F_2$ data (coded genotypes) of RFLP and genetic markers (Chang *et al.*, 1988; Nam *et al.*, 1989) and pairwise estimates of recombination frequencies of genetic markers (Koornneef, 1990; Koornneef *et al.*, 1983). Our data set consists of pairwise estimates of recombination frequencies for 125 classical genetic markers (Koornneef, 1990; Koornneef *et al.*, 1983; data available in the literature, see Experimental procedures) and previously published $F_2$ population data (Chang *et al.*, 1988; Nam *et al.*, 1989). The latter consists of 184 RFLP and 19 morphological markers. In addition, data for 123 unpublished RFLP markers are included (Table 1 and Figure 1). To establish points of contact between the individual maps 62 markers

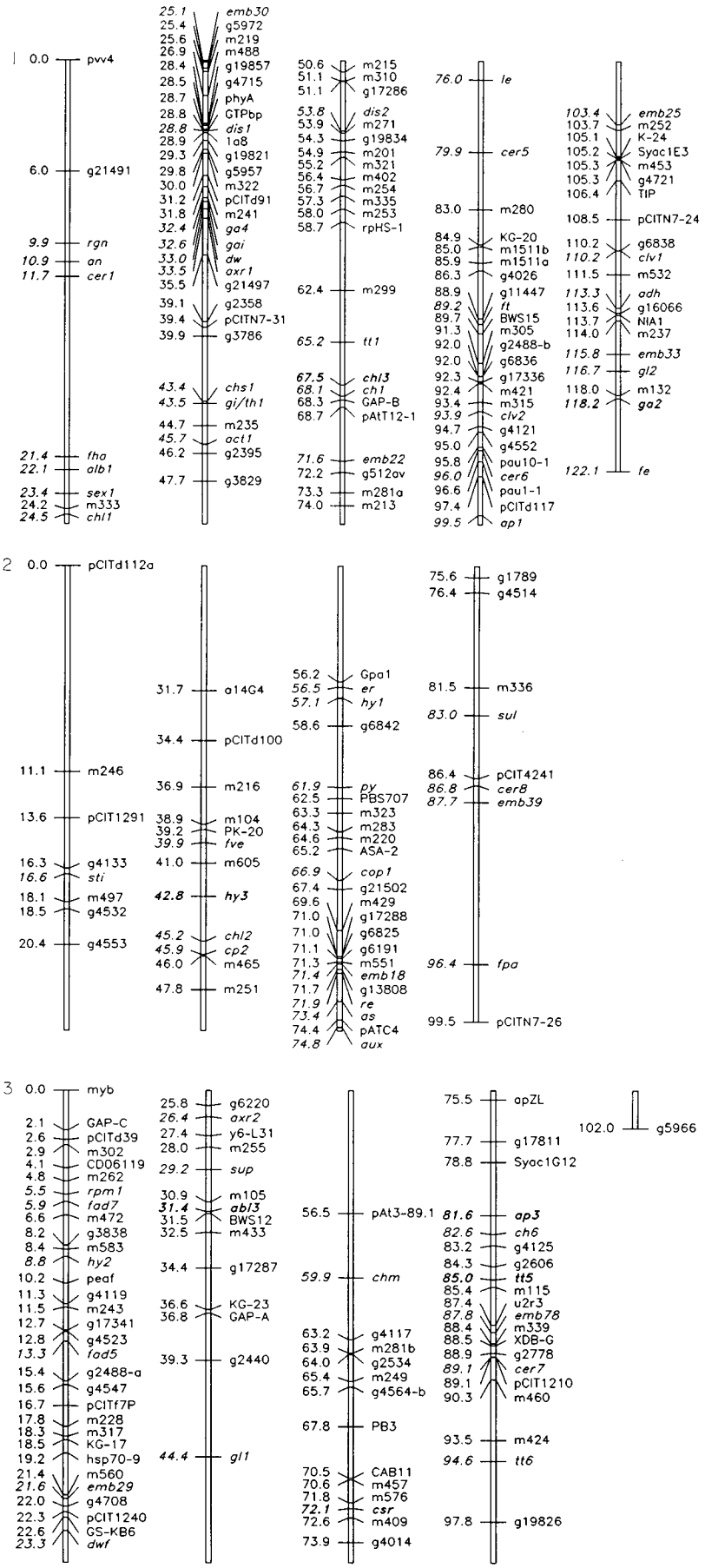
**Table 1.** New markers

| Marker[a] | Chr. no.[b] | Description[c] | Reference[d] |
|---|---|---|---|
| pvv4 | 1 | Acc synthase | Van Der Straeten *et al.* (manuscript in press) |
| GTPbp | 1 | GTP binding protein | Nam (personal communication) |
| 1a8 | 1 | random cosmid clone | Hauge and Goodman (unpublished data) |
| GAP-B | 1 | GAPDH GapB | Shih *et al.* (1991) |
| pAtT12-1 | 1 | telomere related clone | Richards *et al.* (1991) |
| BWS15 | 1 | random clone | Shirley and Goodman (unpublished data) |
| pau10-1 | 1 | urease | Zonia and Polacco (1990) |
| pau1-1 | 1 | urease | Zonia and Polacco (1990) |
| NIA1 | 1 | nitrate reductase | Cheng *et al.* (1988) |
| PHYA | 1 | phytochrome A | Sharrock and Quail (1989) |
| rpHS-1 | 1 | random clone | Sakai and Meyerowitz (unpublished data) |
| KG-20 | 1 | random cosmid clone | Goto and Meyerowitz (unpublished data) |
| Syac1E3 | 1 | YAC 1E3 | Grill and Somerville (1991) |
| K-24 | 1 | random cosmid clone | Goto and Meyerowitz (unpublished data) |
| TIP | 1 | tonoplast intrinsic protein | Höfte *et al.* (1992) |
| STI | 2 | Stichel; unbranched trichomes | Jurgens (personal communication) |
| PBS707 | 2 | myb homolog | Doerner and Lamb (personal communication) |
| ASA-2 | 2 | anthranilate synthase | Niyogi and Fink (personal communication) |
| a14G4 | 2 | random clone | Vijayraghavan (unpublished data) |

**Table 1.** Continued

| Marker[a] | Chr. no.[b] | Description[c] | Reference[d] |
|---|---|---|---|
| PK-20 | 2 | protein kinase homology | Chang (unpublished data) |
| Gpa1 | 2 | G protein alpha subunit | Ma *et al.* (1990) |
| pATC4 | 2 | actin gene | Nairn *et al.* (1988) |
| GAP-C | 3 | GAPDH GapC | Shih *et al.* (1991) |
| peaf | 3 | protein kinase | Zhang and Lamb (personal communication) |
| GS-KB6 | 3 | glutamine synthetase | Peterman and Goodman (1991) |
| BWS12 | 3 | random clone | Shirley and Goodman (unpublished data) |
| pAt3-89.1 | 3 | mutational rearrangement at tt5 | Shirley *et al.* (1992) |
| CAB4 | 3 | cab4 | Zhang *et al.* (1991) |
| XDB-G | 3 | glucanase | Dong and Ausubel (personal communication) |
| myb | 3 | myb homolog | Doerner (personal communication) |
| CD06119 | 3 | random clone | Dean (personal communication) |
| hsp70-9 | 3 | heat shock protein | Chen and Vierling (1991) |
| KG-17 | 3 | random cosmid | Goto and Meyerowitz (unpublished data) |
| y6-L31 | 3 | random clone | Sakai and Meyerowitz (unpublished data) |
| KG-23 | 3 | random cosmid clone | Goto and Meyerowitz (unpublished data) |
| CAB11 | 3 | chlorophyll a/b binding protein | McGrath (unpublished data) |
| CSR | 3 | acetolactate synthase | Haughn and Somerville (1986); Sathasivan *et al.* (1988) |
| apZL | 3 | random clone | Liu and Meyerowitz (unpublished data) |
| Syac1G12 | 3 | YAC 1G12 | Grill and Somerville (1991) |
| U2R3 | 3 | U2.3 snRNA gene | Vankan and Filipowicz (1988) |
| U2R9 | 4 | U2.9 snRNA gene | Vankan and Filipowicz (1988) |
| ga1-14 | 4 | random clone | Sun and Ausubel (personal communication) |
| PG11 | 4 | random clone | Gallant and Goodman (unpublished data) |
| 1a5 | 4 | random cosmid clone | Hauge and Goodman (unpublished data) |
| DHS1 & DHS2 | 4 | DAHP synthases | Keith *et al.* (1991) |
| ABP1 | 4 | auxin binding protein | Palme *et al.* (1992) |
| PK-87 | 4 | protein kinase homology | Chang (unpublished data) |
| 06455CD | 4 | random clone | Dean (personal communication) |
| KG-32 | 4 | random cosmid clone | Goto and Meyerowitz (unpublished data) |
| TSB2 | 4 | tryptophan synthase B | Last *et al.* (1991) |
| Wyac23H12 | 4 | YAC 23H12 | Ward and Jen (1990) |
| ASA-1 | 5 | anthranilate synthase | Niyogi and Fink (personal communication) |
| CHP7 | 5 | Chlorina; green yellow | Koornneef *et al.* (1991) |
| PBS811 | 5 | myb homolog | Doerner and Lamb (personal communication) |
| pGATC-11 | 5 | dark-inducible cDNA | Inze and Van Montague (personal communication) |
| GS-L1 | 5 | glutamine synthetase | Peterman and Goodman (1991) |
| GS-R1 | 5 | glutamine synthetase | Peterman and Goodman (1991) |
| pAt5-91.5 | 5 | mutational rearrangement at tt3 | Shirley *et al.* (1992) |
| CRA | 5 | seed storage protein (sAt2105) | Pang *et al.* (1988) |
| pAD1.7 | 5 | protein kinase | Zhang and Lamb (personal communication) |
| PBS813a | 5 | myb homolog | Doerner and Lamb (personal communication) |
| TSB1 | 5 | tryptophan synthase B | Last *et al.* (1991) |
| lox8 | 5 | random clone | Peterman and Goodman (unpublished data) |
| AB5-13 | 5 | ankyrin repeat | Zhang and Goodman (unpublished data) |
| pAtT80 | 5 | telomere clone | Richards *et al.* (1991) |
| ubq6121 | 5 | ubiquitin | Sullivan and Vierstra (1992) |
| CD05629 | 5 | random clone | Dean (personal communication) |
| KG-31 | 5 | random cosmid clone | Goto and Meyerowitz (unpublished data) |
| U2R5 | 5 | U2.5 snRNA gene | Vankan and Filipowicz (1988) |
| AR119 | 5 | anthranilate PR transferase | Rose and Last (personal communication) |
| KG-10 | 5 | random cosmid | Goto and Meyerowitz (unpublished data) |
| CD06455f | 5 | random clone | Dean (personal communication) |
| PHYC | 5 | phytochrome C | Sharrock and Quail (1989) |
| KG-8 | 5 | random cosmid clone | Goto and Meyerowitz (unpublished data) |
| PBS813-b | 5 | myb homolog | Doerner and Lamb (personal communication) |

[a]The name of the new marker (other than new markers with name designations g### or m###; see also Figure 1) or a marker with a name change.
[b]The chromosome to which it has been mapped.
[c]Description of the marker and the name and/or function of the gene to which it corresponds, if it is known.
[d]Literature reference or the name(s) of the person(s) from whom the clone was obtained.


1
0.0 pvv4
6.0 g21491
9.9 rgn
10.9 an
11.7 cer1
21.4 tha
22.1 alb1
23.4 sex1
24.2 m333
24.5 chl1
25.1 emb30
25.4 g5972
25.6 m219
26.9 m488
28.4 g19857
28.5 g4715
28.7 phyA
28.8 GTPbp
28.8 dis1
28.9 1a8
29.3 g19821
29.8 g5957
30.0 m322
31.2 pCITd91
31.8 m241
32.4 ga4
32.6 gai
33.0 dw
33.5 axr1
35.5 g21497
39.1 g2358
39.4 pCITN7-31
39.9 g3786
43.4 chs1
43.5 gi/th1
44.7 m235
45.7 act1
46.2 g2395
47.7 g3829
50.6 m215
51.1 m310
51.1 g17286
53.8 dis2
53.9 m271
54.3 g19834
54.9 m201
55.2 m321
56.4 m402
56.7 m254
57.3 m335
58.0 m253
58.7 rpHS-1
62.4 m299
65.2 tt1
67.5 chl3
68.1 ch1
68.3 GAP-B
68.7 pAtT12-1
71.6 emb22
72.2 g512av
73.3 m281a
74.0 m213
76.0 le
79.9 cer5
83.0 m280
84.9 KG-20
85.0 m1511b
85.9 m1511a
86.3 g4026
88.9 g11447
89.2 ft
89.7 BWS15
91.3 m305
92.0 g2488-b
92.0 g6836
92.3 g17336
92.4 m421
93.4 m315
93.9 clv2
94.7 g4121
95.0 g4552
95.8 pau10-1
96.0 cer6
96.6 pau1-1
97.4 pCITd117
99.5 ap1
103.4 emb25
103.7 m252
105.1 K-24
105.2 Syac1E3
105.3 m453
105.3 g4721
106.4 TIP
108.5 pCITN7-24
110.2 g6838
110.2 clv1
111.5 m532
113.3 adh
113.6 g16066
113.7 NIA1
114.0 m237
115.8 emb33
116.7 gl2
118.0 m132
118.2 ga2
122.1 fe
2
0.0 pCITd112a
11.1 m246
13.6 pCIT1291
16.3 g4133
16.6 sti
18.1 m497
18.5 g4532
20.4 g4553
31.7 a14G4
34.4 pCITd100
36.9 m216
38.9 m104
39.2 PK-20
39.9 fve
41.0 m605
42.8 hy3
45.2 chl2
45.9 cp2
46.0 m465
47.8 m251
56.2 Gpa1
56.5 er
57.1 hy1
58.6 g6842
61.9 py
62.5 PBS707
63.3 m323
64.3 m283
64.6 m220
65.2 ASA-2
66.9 cop1
67.4 g21502
69.6 m429
71.0 g17288
71.0 g6825
71.1 g6191
71.3 m551
71.4 emb18
71.7 g13808
71.9 re
73.4 as
74.4 pATC4
74.8 aux
75.6 g1789
76.4 g4514
81.5 m336
83.0 sul
86.4 pCIT4241
86.8 cer8
87.7 emb39
96.4 fpa
99.5 pCITN7-26
3
0.0 myb
2.1 GAP-C
2.6 pCITd39
2.9 m302
4.1 CD06119
4.8 m262
5.5 rpm1
5.9 fad7
6.6 m472
8.2 g3838
8.4 m583
8.8 hy2
10.2 peaf
11.3 g4119
11.5 m243
12.7 g17341
12.8 g4523
13.3 fad5
15.4 g2488-a
15.6 g4547
16.7 pCITf7P
17.8 m228
18.3 m317
18.5 KG-17
19.2 hsp70-9
21.4 m560
21.6 emb29
22.0 g4708
22.3 pCIT1240
22.6 GS-KB6
23.3 dwf
25.8 g6220
26.4 axr2
27.4 y6-L31
28.0 m255
29.2 sup
30.9 m105
31.4 abl3
31.5 BWS12
32.5 m433
34.4 g17287
36.6 KG-23
36.8 GAP-A
39.3 g2440
44.4 gl1
56.5 pAt3-89.1
59.9 chm
63.2 g4117
63.9 m281b
64.0 g2534
65.4 m249
65.7 g4564-b
67.8 PB3
70.5 CAB11
70.6 m457
71.8 m576
72.1 csr
72.6 m409
73.9 g4014
75.5 apZL
77.7 g17811
78.8 Syac1G12
81.6 ap3
82.6 ch6
83.2 g4125
84.3 g2606
85.0 tt5
85.4 m115
87.4 u2r3
87.8 emb78
88.4 m339
88.5 XDB-G
88.9 g2778
89.1 cer7
89.1 pCIT1210
90.3 m460
93.5 m424
94.6 tt6
97.8 g19826
102.0 g5966

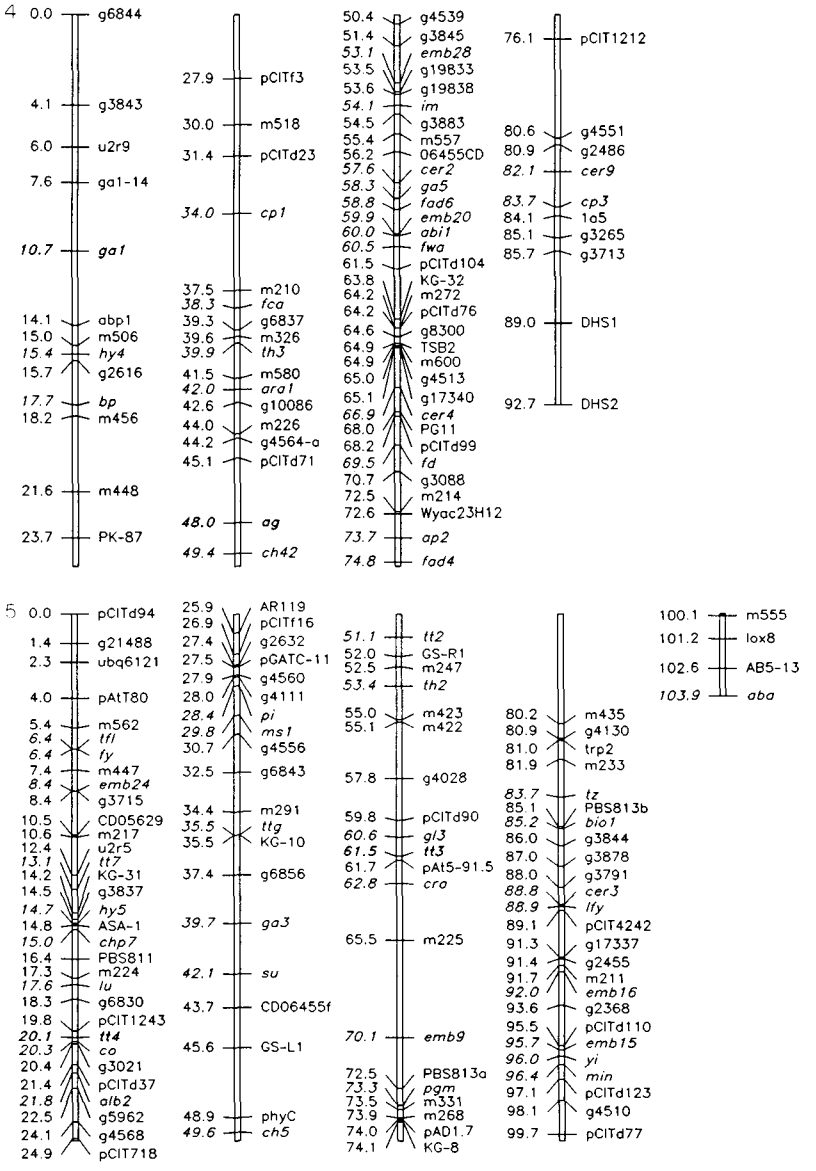



**Figure 1.** The integrated genetic/RFLP map of *A. thaliana*.

Classical genetic markers are indicated in italics (Koornneef, 1987) and classical genetic markers which have also been cloned and mapped by RFLP analysis are indicated in bold italic type. The remaining markers which are listed in plain text correspond to RFLP markers which have not been correlated with a mapped mutation. RFLP markers with names of the form m### are random bacteriophage λ clones from the Meyerowitz laboratory (Chang *et al.*, 1988) previously designated either λbAt### or LEM###; names of the form pCITd###, pCIT### or pCITN7-### are random cosmid clones from the Meyerowitz laboratory; and names of the form g### are random cosmid clones from the Goodman laboratory (Nam *et al.*, 1989). Table 1 indicates the source of the *Arabidopsis* clones used as RFLP markers which are not described above. The numbers on the left indicate the location (in cM) as calculated in JoinMap using the Kosambi mapping function.

Where an RFLP marker from either the Goodman (G) or the Meyerowitz (M) laboratory was deemed to be identical in map position or physically contained within each other the names were changed to the Meyerowitz (m) marker name. Where an RFLP marker from either the Goodman (G) or the Meyerowitz (M) laboratory was deemed to either be identical to, or at least co-segregate with, a classical genetic marker (C) the names were changed to the classical genetic marker name. These 'identities' or very closely linked markers (indicated by *) are as follows:

| | | |
|---|---|---|
| **chr1** | | |
| 1 | NIA2 (G) = m24 (M) = *chl*3 (C) | Change all to *chl*3 |
| 2 | g17311 (G) = *ga*2 (C) | Change to *ga*2 |
| 3 | *gi* (C) = *th*1* (C) | Change to *gi*/*th*1 |
| 4 | m3012 (M) = *adh* (C) | Change to *adh* |
| **chr2** | | |
| 1 | phyB (M) = *hy*3 (C) | Change to *hy*3 |
| **chr3** | | |
| 1 | ch1-14 (G) = *tt*5 (C) | Change to *tt*5 |
| 2 | g4711 (G, M)) = *abi*3 (C) | Change all to *abi*3 |
| 3 | DEF-N7-12 (M) = *ap*3 (C) | Change to *ap*3 |
| 4 | g5970 (G) = m243 (M) | Change to m243 |
| 5 | pGH1 (G, M) = *csr* (C) | Change all to *csr* |
| **chr4** | | |
| 1 | g4108 (G) = m210 (M) | Change to m210 |
| 2 | pCIT505 (M) = *ag* (C) | Change to *ag* |
| 3 | m455 (M) = m326 (M) | Change to m326 |
| 4 | λGA1-3 (G) = *ga*1 (C) | Change to *ga*1 |
| **chr5** | | |
| 1 | g6833* (G) = ICHS2* (M) = *tt*4 (C) | Change all to *tt*4 |
| 2 | m558* (M) = *bio*1* (C) | Change to *bio*1 |
| 3 | dfr-1 (G) = *tt*3 (C) | Change to *tt*3 |
| 4 | TSB1(G, M) = *trp*2 (C) | Change to *trp*2. |

**Table 2.** Summary of the markers in common between data sets

| Chromosome no. | Markers in common[a] | | | |
|---|---|---|---|---|
| | All three (G, M, & C) | RFLP (G & M) | Classical (G & C) | Classical (M & C) |
| 1 | 2<br>*AP1*<br>***CHL3*** | 8<br>m322<br>g4715<br>m241<br>m299<br>m305<br>m315<br>m252<br>m132 | 5<br>***CHL3***<br>*CH1*<br>*AP1*<br>*GL2*<br>***GA2*** | 4<br>*AN*<br>***CHL3***<br>*AP1*<br>*CLV1* |
| 2 | 1<br>*ER* | 3<br>m251<br>g4514<br>m336 | 4<br>*CP2*<br>*ER*<br>*AS*<br>*CER8* | 3<br>***HY3***<br>*ER*<br>*PY* |
| 3 | 3<br>***ABI3***<br>*GL1*<br>***CSR*** | 7<br>m302<br>m243<br>g4119<br>m433<br>g2440<br>m249<br>g4014 | 5<br>*HY2*<br>***ABI3***<br>*GL1*<br>***TT5***<br>***CSR*** | 4<br>***ABI3***<br>*GL1*<br>***AP3***<br>***CSR*** |
| 4 | 3<br>*AP2*<br>*CER2*<br>*BP* | 5<br>U2R9<br>m210<br>m557<br>m272<br>g3088 | 4<br>***GA1***<br>*AP2*<br>*CER2*<br>*BP* | 4<br>***AG***<br>*AP2*<br>*CER2*<br>*BP* |
| 5 | 2<br>***TT4***<br>***TT3*** | 9<br>PBS811<br>g6843<br>m247<br>CRA<br>m331<br>m435<br>TRP2<br>m211<br>m555 | 3<br>*TTG*<br>***TT4***<br>***TT3*** | 5<br>*PI*<br>*TZ*<br>*BIO1*<br>***TT4***<br>***TT3*** |

[a]The number and the names of the markers in common between all three data sets, Goodman (G), Meyerowitz (M), and classical (C); the $F_2$ data sets only (RFLP: G & M); classical markers in the Goodman data set (classical: G & C); classical markers in the Meyerowitz data set (classical: M & C). In these columns the name of the marker is repeated if it applies to more than one column, classical markers are in italics, and classical mutations which have also been mapped by RFLP analysis are in bold face.

common to at least two of the independent data sets were exploited (Table 2). Integration of the data sets was performed using the JoinMap computer program (Stam, 1993).

In Figure 1 we present an integrated linkage map of the *A. thaliana* genome. The map is based on 306 molecular markers and 125 classical genetic markers which together encompass 520 map units. The total number of markers on the map is 413; 306 RFLPs + 125 classical −18 renamed RFLPs (see Figure 1 legend).

To assess the fidelity of the map comparisons were made between the integrated map and the published maps (Chang *et al.*, 1988; Koornneef, 1990; Nam *et al.*, 1989). For additional comparison, new RFLP maps containing the unpublished RFLP markers were assembled using MAPMAKER (Lander *et al.*, 1987). Comparison of the integrated map with the independent maps reveals that the overall linear order of markers is in good agreement and that the integrated map is consistent with the component maps. It should be noted, however, that in several regions of the map a clear reordering of markers has occurred. This is not unexpected since in various regions several orders are equally probable.

It must be emphasized that maps are assembled based on probabilities and represent the 'best fit'. Clearly any calculated map is only as good as the data will allow. The integration of distinct maps is strongly dependent on the number of markers which are common to the individual maps. The map presented in Figure 1 is built around 62 common markers (Table 2). Therefore, 14.4% (62/431) of the markers are common to at least two of the independent data sets. While each chromosome contains at least nine common markers (Table 2), these markers are not uniformly distributed along the lengths of the chromosomes. Clearly, the order of markers in regions of the genome containing a low density of common markers will be less likely to be correct than those containing a uniform distribution of common markers.

A second limitation is the mapping of those markers for which the total amount of linkage information in the data set is poor (small numbers, misclassifications, significant linkage with less than four other markers). In general, markers of this class tend to get placed at the extremities of the linkage groups. This may explain in part why for several of the chromosomes apparent map expansion is observed near the ends.

The map presented in Figure 1 represents the best fit based largely on the raw data. For several markers however, the map position initially generated by JoinMap was in conflict with the known map position. For example, based on three point analysis, *fwa* has been shown to map between *ga*5 and *ap*2 (Koornneef *et al.*, 1991), while JoinMap produces the order *fwa-cer*2-*ap*2 instead of *cer*2-*fwa-ap*2. The discrepancy occurs because the *ap*2-*fwa* (and *cer*2-*ap*2) distance is somewhat larger than the previous estimate which was based on 6556 $F_2$ coupling phase plants (Koornneef *et al.*, 1991). To correct for known discrepancies in map position, JoinMap allows the user to specify a list of 'fixed orders'. By imposing a 'fixed order' JoinMap will produce a map which does not contradict any of the known data (for discussion see Stam, 1993). To

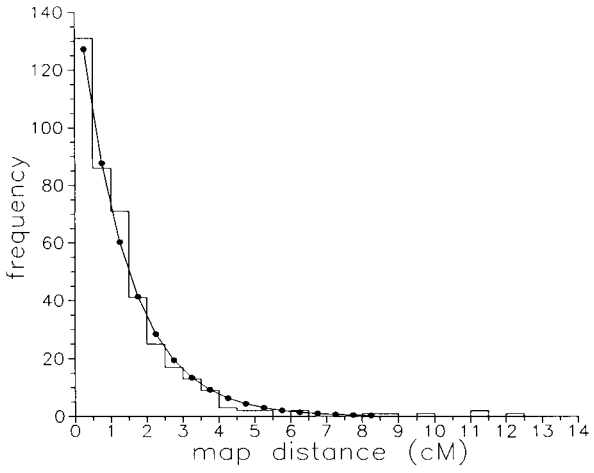



**Figure 2.** Comparison of observed and theoretical distribution of map distances between adjacent markers.
The observed distribution (histogram) and expected distribution (curve) of map distances between adjacent markers is shown. The expected curve is the (scaled) distribution when marker positions are randomly distributed. This is an exponential distribution with the same mean as the observed mean distance, i.e. $f(x) = 0.76\ e^{-0.76x}$.

correct for known errors in the map, the maps of chromosomes 1 and 4 (Figure 1) were re-calculated after imposing the following 'fixed orders': chromosome 1, *an-act*1-*ch*1, *an-act*1-*ap*1, *fha-ga*4-*dis*2, and *tt*1-*chl*1-*ch*1; chromosome 4, *bp-ara*1-*cer*2 and *ap*2-*fwa-ga*5-*cer*2. Since it is known that the current map (Figure 1) sometimes differs in local regions from that determined for example by chromosome walking, the use of additional 'fixed orders' will be increasingly employed for future refinement of the map.

The relatively high density of markers enables us to examine the observed distribution of map distances between adjacent markers and compare it to the theoretical distribution (Figure 2). The theoretical distribution is based on the assumption that the markers are randomly distributed over the total length of the map. As indicated in Figure 2, the observed distribution closely fits the theoretical model, suggesting that the markers in Figure 1 are randomly distributed. None the less, the close fit between the observed and theoretical distributions does not constitute proof that the markers are randomly distributed over the physical length of the genome. To demonstrate definitively that the markers are randomly distributed it will be necessary to establish the relationship between physical and genetic distance. In a similar analysis of the tomato linkage map (Tanksley *et al.*, 1992) a clear non-random distribution of markers was observed along individual chromosomes. The authors suggest that the markers in tomato are randomly distributed along the physical length of the chromosomes and the observed clustering reflects suppression of recombination especially in the centromeric region. By contrast, in *Arabidopsis* there is no indication of clustering of markers in known centromeric regions (chromosomes 1 and 5; Koornneef *et al.*, 1983). The significance of this apparent difference between *Arabidopsis* and tomato remains to be determined.

The map presented in Figure 1 represents the status as of September 1992. As additional information is added, the map will continue to change and subsequently be refined. Both the raw mapping data as well as a version of the integrated map that is continuously updated with new data (see Appendix 1) are readily available in AAtDB (Cherry *et al.*, 1992).

## Experimental procedures

### *Crosses*

The crosses and the resultant $F_2$ populations for RFLP analysis have previously been described (Chang *et al.*, 1988; Nam *et al.*, 1989), except that one new cross of the W100 multiple marker strain in a Landsberg *erecta* background to Col-0 wild-type was used by the Meyerowitz laboratory for additional $F_2$ analysis. Two factor cross data were obtained mainly by analyzing $F_2$ populations segregating for several morphological markers (in the same population) in the Landsberg *erecta* background. Data from the literature have been included in previously published maps (Koornneef, 1987, 1990; Koornneef *et al.*, 1983). Included in this paper are published linkage data for *ara*1 (Dolezal and Cobbett, 1991); *act*1, *fad*4, *fad*5, *fad*6, and *fad*7 (Hugly *et al.*, 1991); *axr*1 (Lincoln *et al.*, 1990); *axr*2 (Wilson *et al.*, 1990); *bio* and *emb* genes (Patton *et al.*, 1991); *cop*1 (Deng and Quail, 1992); *chm* (Redei, 1973); *chs*1 (Hugly *et al.*, 1990); *lfy* (Schultz and Haughn, 1991); *rpm*1 (Debener *et al.*, 1991); *sex*1 (Caspar *et al.*, 1991); *sup=flo*10 (Schultz *et al.*, 1991); and *tfl*1 (Alvarez *et al.*, 1992; Shannon and Meeks-Wagner, 1991).

### *Cloned genes*

The cloned *Arabidopsis* genes used as RFLP markers in this study and two previously unpublished morphological markers (*sti* on chromosome 2 and *chp*7 on chromosome 5) are listed in Table 1.

### *Nucleic acid extractions*

Procedures for the isolation of total *Arabidopsis* genomic DNA, cosmid, plasmid and bacteriophage λ DNA have been described previously (Chang *et al.*, 1988; Nam *et al.*, 1989).

### *Genomic blots*

Genomic blots were performed either as previously published (Chang *et al.*, 1988; Nam *et al.*, 1989) or as follows. Total *Arabidopsis* genomic DNA (1–2.5 μg) was incubated for 16 h with a fivefold excess of a restriction enzyme. The enzymes used for RFLP determinations were *Hin*dIII, *Eco*RI, *Bcl*I, *Cla*I, *Dra*I, *Xba*I and *Bgl*II. The digested DNA was then size-fractionated by electrophoresis in either 0.5 or 0.8% agarose (endomosis 0.1) gels for 640 volt hours. Following electrophoresis, the gels were depurinated with 0.25 M HCl for 30 min, denatured with 0.5 M NaOH, 1.5 M NaCl for 20 min and neutralized with three washes of 25 mM $NaPO_4$ (pH 6.5) for 10 min each. DNA was transferred by capillary

action to 0.2 micron neutral nylon filters (Biotrans from ICN Biomedicals) with 25 mM $NaPO_4$ (pH 6.5). Following the overnight DNA transfer, the DNA was UV cross-linked to the membrane. [$^{32}$P]DNA probes were generated by random primer extension (Feinberg and Vogelstein, 1983) and filters were hybridized ($10^6$–$10^8$ c.p.m. $ml^{-1}$) for 16–20 h. Filters were washed twice at 55°C in 40 mM $NaPO_4$ (pH 7.2), 5% SDS, 0.5% BSA Fraction 5, 1 mM EDTA (pH 8.0) and subsequently washed three times at 55°C in 40 mM $NaPO_4$ pH 7.2, 1% SDS, 1 mM EDTA (pH 8.0) (Church and Gilbert, 1984). Filters were stripped for re-hybridization by submerging filters for 30 min at 70°C in 2 mM Tris (pH 8.2), 2 mM EDTA (pH 8.0), 0.1% SDS. The filters were stored in fresh deprobe buffer.

### *Mapping*

Integration of the data sets was accomplished by using a new computer mapping program 'JoinMap' which is described in detail in the accompanying paper (Stam, 1993). Briefly, JoinMap is able to process two basic types of data, raw population data (i.e. backcross, $F_2$, recombinant inbred lines) and pairwise estimates of recombination frequencies coupled with their respective standard error. Following data input, the pairwise recombination frequencies and the corresponding LOD scores are calculated. Based on the pairwise estimates and the LOD scores the markers are placed into linkage groups. Following assignment of linkage groups, JoinMap sequentially builds the map starting with marker pairs containing the highest LOD score. The calculations for ordering markers and estimating map distances are described in detail in the accompanying paper (Stam, 1993). Map distances were calculated using Kosambi's (Kosambi, 1944) mapping function (for discussion see Stam, 1993).

## Acknowledgements

This work was supported by grants from Hoechst AG, the National Science Foundation (DIR-9113935), and the US Department of Agriculture Plant Genome Research Program through the National Agricultural Library (58-0520-1-150) to HMG and the National Science Foundation (MCB-9204839) to EMM.

## References

**Alvarez, J., Guli, C.L., Yu, X.-H. and Smyth, D.R.** (1992) *Terminal flower*: a gene affecting inflorescence development in *Arabidopsis thaliana*. *Plant J.* **2**, 103–116.

**Bender, W., Spierer, P. and Hogness, D.S.** (1983) Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in *Drosophila melanogaster*. *J. Mol. Biol.* **168**, 17–33.

**Burr, B. and Burr, F.A.** (1991) Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. *Trends Genet.* **7**, 55–60.

**Caspar, T., Lin, T.P., Kakefuda, G., Benbow, L., Preiss, J. and Somerville, C.** (1991) Mutants of *Arabidopsis* with altered regulation of starch degradation. *Plant Physiol.* **95**, 1181–1188.

**Chang, C., Bowman, J.L., De John, 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.

**Chen, Q. and Vierling, E.** (1991) Analysis of conserved domains identifies a unique structural feature of a chloroplast heat shock protein. *Mol. Gen. Genet.* **226**, 425–431.

**Cheng, C.L., Dewdney, J., Nam, H.G., Den Boer, B.G.W. and Goodman, H.M.** (1988) A new locus (NIA1) in *Arabidopsis thaliana* encoding nitrate reductase. *EMBO J.* **7**, 3309–3314.

**Cherry, J.M., Cartinhour, S.W. and Goodman, H.M.** (1992) AAtDB, an *Arabidopsis thaliana* database. *Plant Mol. Biol. Rep.* **10**, 308–309, 409–411.

**Church, G.M. and Gilbert, W.** (1984) Genomic sequencing. *Proc. Natl Acad. Sci. USA*, **83**, 1991–1995.

**Coe, E., Hoisington, D. and Chao, S.** (1990) Gene list and working maps. *Maize Genet. Coop. Newslett.* **64**, 154–163.

**Debener, T., Lehnackers, H., Arnold, M. and Dangl, J.L.** (1991) Identification and molecular mapping of a single *Arabidopsis thaliana* locus determining resistance to a phytopathogenic *Pseudomonas syringae* isolate. *Plant J.* **1**, 289–302.

**Deng, X.-W. and Quail, P.H.** (1992) Genetic and phenotypic characterization of cop1 mutants of *Arabidopsis thaliana*. *Plant J.* **2**, 83–95.

**Dolezal, O. and Cobbett, C.S.** (1991) Arabinose kinase-deficient mutant of *Arabidopsis thaliana*. *Plant Physiol.* **96**, 1255–1260.

**Donis-Keller, H., Green, P., Helms, C., *et al.*** (1987) A genetic linkage map of the human genome. *Cell*, **51**, 319–337.

**Feinberg, A.P. and Vogelstein, B.** (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6–13.

**Grill, E. and Somerville, C.** (1991) Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. *Mol. Gen. Genet.* **226**, 484–490.

**Hauge, B.M., Hanley, S., Giraudat, J. and Goodman, H.M.** (1991) Mapping the *Arabidopsis* genome. In *Molecular Biology of Plant Development* (Jenkins, G. and Schurch, W., eds). Cambridge: Cambridge University Press, pp. 45–56.

**Haughn, G.W. and Somerville, C.** (1986) Sulfonylurea-resistant mutants of *Arabidopsis thaliana*. *Mol. Gen. Genet.* **204**, 430–434.

**Helentjaris, T. and Burr, B.** (1989) *Development and Application of Molecular Markers to Problems in Plant Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

**Helentjaris, T., Wright, S. and Weber, D.F.** (1986) Construction of a genetic linkage map in maize using restriction fragment polymorphisms. *Maize Genet. Coop. Newslett.* **60**, 118–120.

**Höfte, H., Hubbard, L., Reizer, J., Ludevid, D., Herman, E.M. and Crispeels, M.J.** (1992) Vegetative and seed-specific forms of tonoplast intrinsic protein in the vacuolar membrane of *Arabidopsis thaliana*. *Plant Physiol.* **99**, 561–570.

**Hugly, S., McCourt, P., Browse, J., Patterson, G.W. and Somerville, C.** (1990) A chilling sensitive mutant of *Arabidopsis* with altered steryl-ester metabolism. *Plant Physiol.* **93**, 1053–1062.

**Hugly, S., Kunst, L. and Somerville, C.** (1991) Linkage relationships of mutations that affect fatty acid composition in *Arabidopsis*. *J. Heredity*, **82**, 484–488.

**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.

**Keith, B., Dong, X., Ausubel, F.M. and Fink, G.R.** (1991) Differential induction of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase genes in *Arabidopsis thaliana* by wounding and pathogenic attack. *Proc. Natl Acad. Sci. USA*, **88**, 8821–8825.

**Koornneef, M.** (1987) Linkage map of *Arabidopsis thaliana* (2n = 10). In *Genetic Maps*, Fourth Edn. (O'Brien, S.J., ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 742–745.

**Koornneef, M.** (1990) Linkage map of *Arabidopsis thaliana* (2n = 10). In *Genetic Maps*, Fifth Edn. (O'Brien, S.J., ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 6.95–6.97.

**Koornneef, M., Van Eden, J., Hanhart, C.J., Stam, P., Braaksma, F.J. and Feenstra, W.J.** (1983) Linkage map of *Arabidopsis thaliana*. *J. Heredity*, **74**, 265–272.

**Koornneef, M., Hanhart, C.J. and Van Der Veen, J.H.** (1991) A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **229**, 57–66.

**Koornneef, M., Chang, C., Goodman, H.M., *et al.*** (1992) Linkage map of *Arabidopsis thaliana* 2N = 10. In *Genetic Maps*, Sixth Edn. (O'Brien, S.J., ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. in press.

**Kosambi, D.D.** (1944) The estimation of map distance from recombination values. *Ann. Eugen.* **12**, 172–175.

**Krol, E.** (1992) *The Whole Internet User's Guide & Catalog.* Sebastopol, CA: O'Reilly & Associates, Inc.

**Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, 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.

**Last, R.L., Bissinger, P.H., Mahoney, D.J., Radwanski, E.R. and Fink, G.R.** (1991) Tryptophan mutants in *Arabidopsis* the consequences of duplicated tryptophan synthase beta genes. *Plant Cell*, **3**, 345–358.

**Leutwiler, L.S., Hough-Evans, B.R. and Meyerowitz, E.M.** (1984) The DNA of *Arabidopsis thaliana*. *Mol. Gen. Genet.* **194**, 15–23.

**Lincoln, C., Britton, J.H. and Estelle, M.** (1990) Growth and development of the axr1 mutants of *Arabidopsis*. *Plant Cell*, **2**, 1071–1080.

**Ma, H., Yanofsky, M.F. and Meyerowitz, E.M.** (1990) Molecular cloning and characterization of GPA1, a G protein alpha subunit gene from *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **87**, 3821–3825.

**Meyerowitz, E.M.** (1989) *Arabidopsis*, a useful weed. *Cell*, **56**, 263–270.

**Nairn, C.J., Winesett, L. and Ferl, R.J.** (1988) Nucleotide sequence of an actin gene from *Arabidopsis thaliana*. *Gene*, **65**, 247–258.

**Nam, H.G., Giraudat, J., Den Boer, B., Moonan, F., Loos, W.D.B., Hauge, B.M. and Goodman, H.M.** (1989) Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. *Plant Cell*, **1**, 699–705.

**Palme, K., Hesse, T., Campos, N., Garbers, C., Yanofsky, M.F. and Schell, J.** (1992) Molecular analysis of an auxin binding protein gene located on chromosome 4 of *Arabidopsis*. *Plant Cell*, **4**, 193–201.

**Pang, P.P., Pruitt, R.E. and Meyerowitz, E.M.** (1988) Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of *Arabidopsis thaliana*. *Plant Mol. Biol.* **11**, 805–820.

**Patton, D.A., Franzmann, L.H. and Meinke, D.W.** (1991) Mapping genes essential for embryo development in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **227**, 337–347.

**Peterman, T.K. and Goodman, H.M.** (1991) The glutamine synthetase gene family of *Arabidopsis thaliana* light-regulation and differential expression in leaves roots and seeds. *Mol. Gen. Genet.* **230**, 145–154.

**Poustka, A., Pohl, T.M., Barlow, D.P., Frischauf, A.-M. and Lehrach, H.** (1987) Construction and use of human chromosome jumping libraries from *Not*I-digested DNA. *Nature*, **325**, 353–355.

**Pruitt, R.E. and Meyerowitz, E.M.** (1986) Characterization of the genome of *Arabidopsis thaliana*. *J. Mol. Biol.* **187**, 169–184.

**Redei, G.P.** (1973) Extrachromosomal mutability determined by a nuclear gene locus in *Arabidopsis*. *Mutation Res.* **18**, 149–162.

**Reiter, R.S., Williams, J.G.K., Feldmann, K.A., Rafalski, A., Tingey, S.V. and Scolnick, 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.

**Richards, E.J., Goodman, H.M. and Ausubel, F.M.** (1991) The centromere region of *Arabidopsis thaliana* chromosome-1 contains telomere-similar sequences. *Nucl. Acids Res.* **19**, 3351–3358.

**Sathasivan, K., Haughn, G.W. and Murai, N.** (1988) Isolation and sequencing of acetolactate synthase gene from imazapyr-resistant *Arabidopsis thaliana* (L.) *Plant Physiol.* **86**, Suppl. 4, 136.

**Schultz, E.A. and Haughn, G.W.** (1991) Leafy, a homeotic gene that regulates inflorescence development in *Arabidopsis*. *Plant Cell*, **3**, 771–781.

**Schultz, E.A., Pickett, F.B. and Haughn, G.W.** (1991) The FLO10 gene product regulates the expression domain of homeotic gene AP3 and PI in *Arabidopsis* flowers. *Plant Cell*, **3**, 1221–1237.

**Shannon, S. and Meeks-Wagner, D.R.** (1991) A mutation in the *Arabidopsis* TFL1 gene affects inflorescence meristem development. *Plant Cell*, **3**, 877–892.

**Sharrock, R.A. and Quail, P.H.** (1989) Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution, and differential expression of a plant regulatory photoreceptor family. *Genes Devel.* **3**, 1745–1757.

**Shih, M.C., Heinrich, P. and Goodman, H.M.** (1991) Cloning and chromosomal mapping of nuclear genes encoding chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenase from *Arabidopsis thaliana*. *Gene*, **104**, 133–138.

**Shirley, B.W., Hanley, S. and Goodman, H.M.** (1992) Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis transparent testa* mutations. *Plant Cell*, **4**, 333–347.

**Stam, P.** (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J.* **3**, 739–744.

**Sullivan, M.L. and Vierstra, R.D.** (1992) Cloning of a 16-kDa ubiquitin carrier protein from wheat and *Arabidopsis thaliana*: Identification of functional domains by the vitro mutagenesis. *J. Biol. Chem.* **266**, 23 878–23 885.

**Tanksley, S.D., Ganal, M.W., Prince, J.P., *et al.*** (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics*, **132**, 1141–1160.

**Vankan, P. and Filipowicz, W.** (1988) Structure of U2 small nuclear RNA genes of *Arabidopsis thaliana* and their expression in electroporated plant protoplasts. *EMBO J.* **7**, 791–800.

**Ward, E.R. and Jen, G.C.** (1990) Isolation of single-copy-sequence clones from a yeast artificial chromosome library of randomly-sheared *Arabidopsis thaliana* DNA. *Plant Mol. Biol.* **14**, 561–568.

**Wilson, A.K., Pickett, F.B., Turner, J.C. and Estelle, M.** (1990) A dominant mutation in *Arabidopsis* confers resistance to auxin, ethylene and abscisic acid. *Mol. Gen. Genet.* **222**, 377–383.

**Zhang, H., Hanley, S. and Goodman, H.M.** (1991) Isolation characterization and chromosomal location of a new cab gene from *Arabidopsis thaliana*. *Plant Physiol.* **96**, 1387–1388.

**Zonia, L.E. and Polacco, J.C.** (1990) A molecular and biochemical study of urease in *Arabidopsis*. *Plant Physiol.* **93**, Suppl. 1, 15.

**Appendix 1**

A continuously updated collection of *Arabidopsis* genetic maps and access to a text only version of the AAtDB (Cherry *et al.*, 1992) database is available through the Internet worldwide computer network using Gopher and WAIS. This service is called the Arabidopsis Research Companion and is provided on the computer called weeds.mgh.harvard.edu (IP 132.183.190.21)

Gopher and WAIS (reviewed in Krol, 1992) use specialized software on an Internet connected computer to connect with a remote information server. These specialized client software programs are available for most types of personal and workstation computers, as well as for mainframes. The WAIS clients are available via Anonymous FTP from think.com and samba.oit.unc.edu. Gopher clients are available from boombox.micro.umn.edu.

A public account with access to the Arabidopsis Research Companion Gopher server is provided by the Department of Molecular Biology at Massachusetts General Hospital. If the appropriate WAIS or Gopher software cannot be obtained for your networked computer this special account on ochre.mgh.harvard.edu (IP 132.183.190.25) provides access. The command "telnet ochre.mgh.harvard.edu" from many networked computers will establish a connection with the MGH computer. Enter the username "gopher" and the password "thaliana" to complete the login procedure.

For more information on using the Arabidopsis Research Companion contact the AAtDB Project via electronic mail at curator@frodo.mgh.harvard.edu or via fax at +1- 617-726-6893.