DNA RESEARCH 5, 163–168 (1998)

A Physical Map of *Arabidopsis thaliana* Chromosome 3 Represented by Two Contigs of CIC YAC, P1, TAC and BAC Clones

Shusei SATO,[1] Hirokazu KOTANI,[1] Reiko HAYASHI,[1,2] Yao-Guang LIU,[3,4] Daisuke SHIBATA,[3] and Satoshi TABATA[1,*]

Kazusa DNA Research Institute, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan,[1] Chiba Prefectural Agricultural Experiment Station, 2-325 Sawara, Sawara, Chiba 287-0822, Japan,[2] Mitusi Plant Biotechnology Research Institute, TCI D-21, 2-1-6 Sengen, Tsukuba, Ibaraki 305-0047, Japan,[3] and Genetic Engineering Laboratory, South China Agricultural University, Guangzhou 510642, People's Republic of China[4]

(Received 30 April 1998)

Abstract

We have constructed a physical map of *Arabidopsis thaliana* chromosome 3 by ordering the clones from CIC YAC, P1, TAC and BAC libraries using the sequences of a variety of genetic and EST markers and terminal sequences of clones. The markers used were 112 DNA markers, 145 YAC end sequences, and 156 end sequences of P1, TAC and BAC clones. The entire genome of chromosome 3, except for the centromeric and telomeric regions, was covered by two large contigs, 13.6 Mb and 9.2 Mb long. This physical map will facilitate map-based cloning experiments as well as genome sequencing of chromosome 3. The map and end sequence information are available on the KAOS (Kazusa *Arabidopsis* data Opening Site) web site at http://www.kazusa.or.jp/arabi/.

Key words: *Arabidopsis thaliana*; chromosome 3; physical map; contig map

1. Introduction

Arabidopsis thaliana is an excellent model organism for the analysis of complex processes in plants using molecular genetic techniques.[1] Many laboratories are currently making intensive efforts to isolate *Arabidopsis* genes of biological importance using a map-based cloning strategy. This plant has also been chosen as a target for sequencing of the entire genome,[2] because the estimated genome size, 130 Mb, is the smallest among known higher plants and the content of repetitive sequences is assumed to be low.[3] Under these circumstances, a complete physical map of the *Arabidopsis* genome should be greatly advantageous for cloning the genetic loci of interest as well as sequencing the entire genome. With this view, yeast artificial chromosome (YAC)-based physical maps of chromosome 2,[4] 4,[5] and 5[6] of *A. thaliana* have been constructed. We also constructed a fine physical map of the entire chromosome 5 by ordering YAC, P1, TAC and BAC clones[7] to support the initial phase of our sequencing project, and based on the map constructed, sequence analysis of chromosome 5 is in progress.[8–12] In the second phase of the project, we focused on chromosome 3 according to the international agreement of the Arabidopsis Genome Initiative. As to the physical map of chromosome 3, a YAC contig map was recently reported [http://genome-www.stanford.edu/Arabidopsis/Chr3-INRA/],[13] although the map consists of nine contigs. Construction of a complete physical map which covers the entire chromosome is essential to proceed with the sequencing project. We present here a physical map of the entire chromosome 3 which was constructed by ordering the clones from YAC, P1, TAC and BAC libraries using the information on the sequences of various DNA markers and the terminal sequences of the clones.

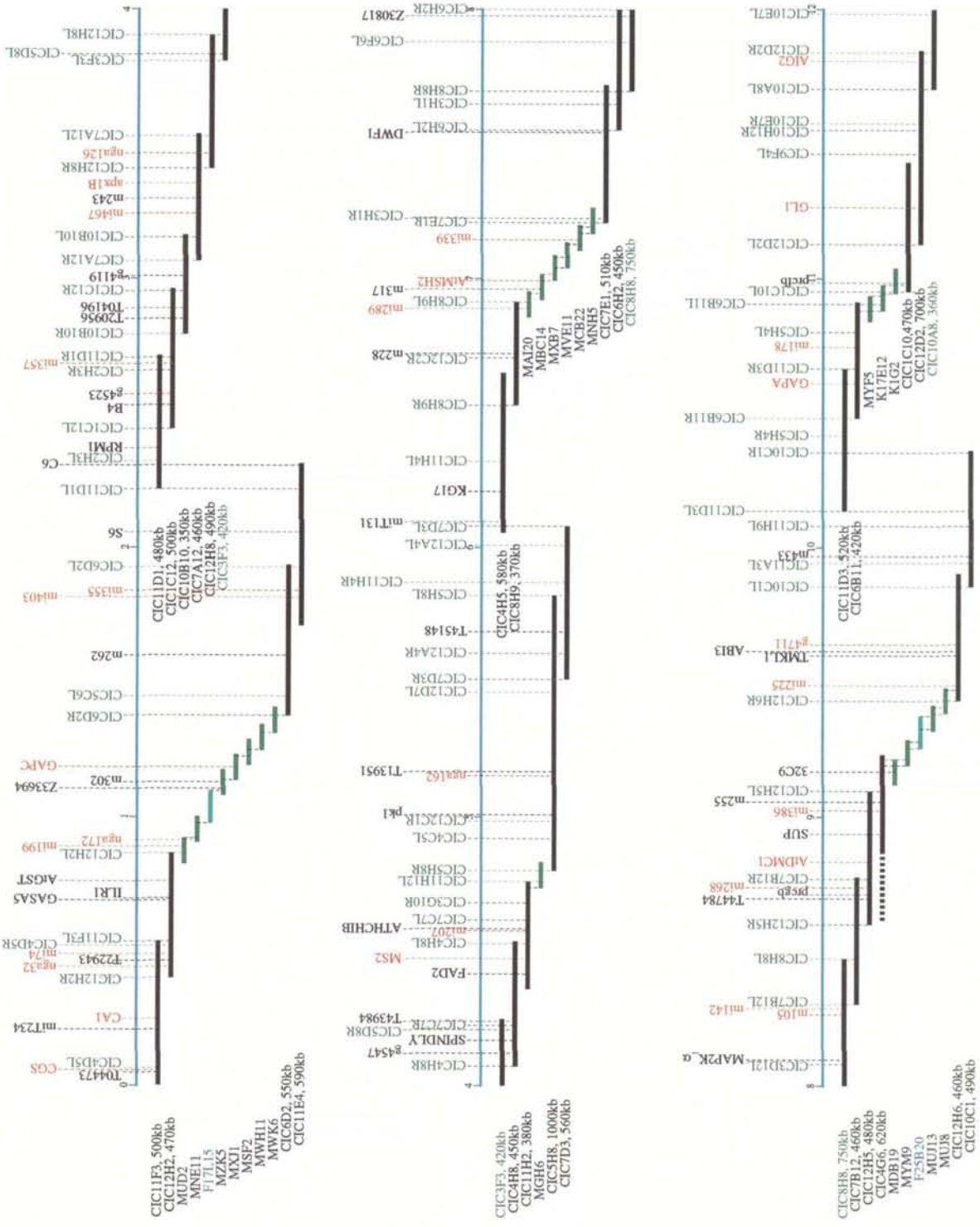
2. Materials and Methods

2.1. Libraries

Five kinds of libraries made from the genome of *A. thaliana* ecotype Columbia were used for the construction of the map: CIC YAC library[14] from Arabidopsis Biological Resource Center at Ohio State University, IGF BAC (Mozo, T. unpublished, http://194.94.225.1/private_workgroups/pg_101/igf_bac_lib.html) and TAMU BAC[15] libraries from Dr. Ian Bancroft at John Innes Center, P1[16] and TAC (Liu, Y.-G.,

Communicated by Mituru Takanami

* To whom correspondence should be addressed. Tel. +81-438-52-3933, Fax. +81-438-52-3934, E-mail: tabata@kazusa.or.jp



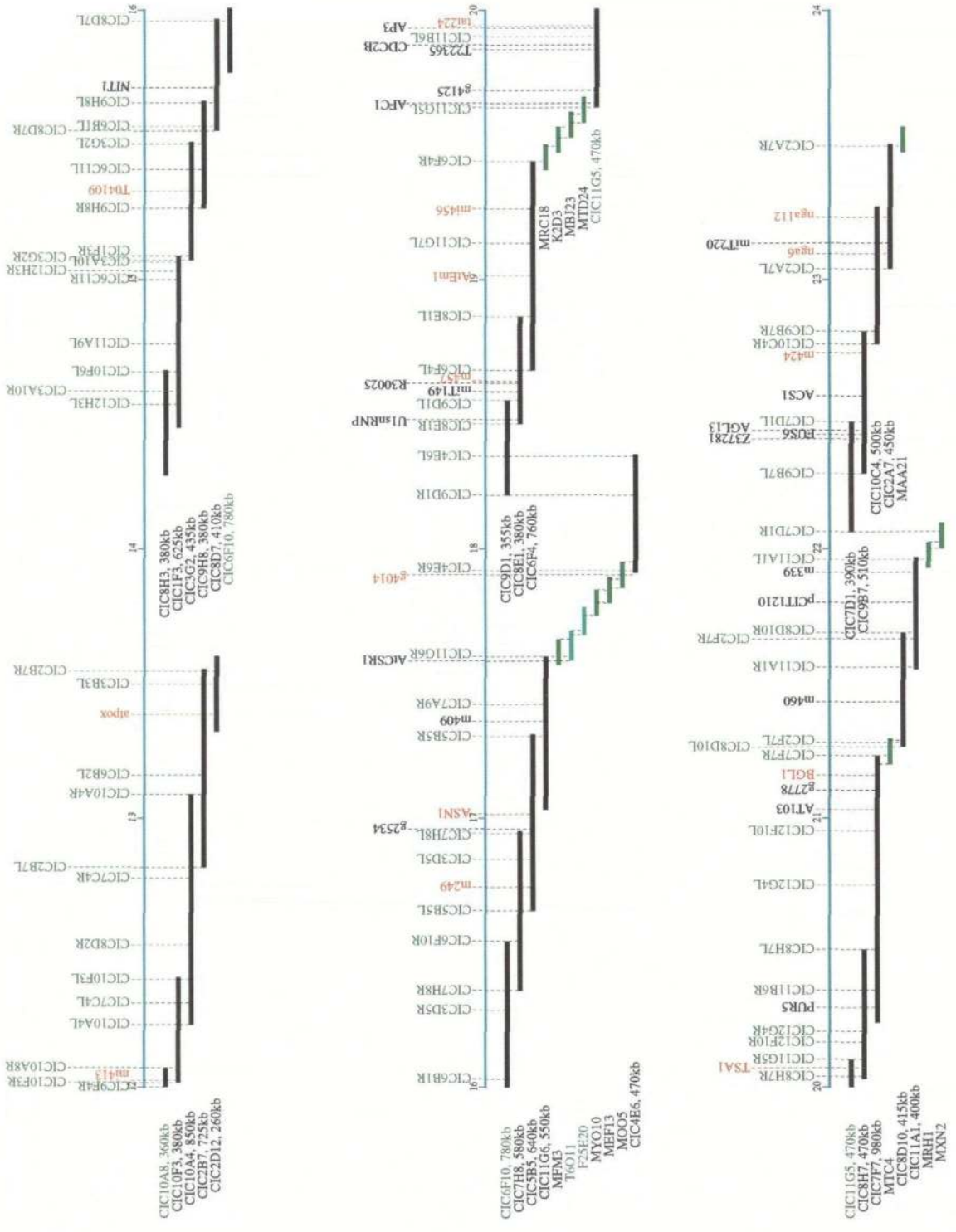


Figure 1. The physical map of chromosome 3. CIC YAC clones are represented by horizontal black bars with lengths reflecting the size of each YAC clone. The name and size of CIC YAC clones are indicated at the left side of the bars. The putative chimeric parts of the YAC clones are indicated by dotted bars. P1, TAC and BAC clones that cover the gaps among YAC contigs are represented by horizontal green bars. The names of bacterial clones are indicated at the left side of the bars as P1 (M####), TAC (K####), IGF BAC (F####), TAMU BAC (T####). RI markers (vertical red letters), other DNA markers (vertical black letters) and YAC end markers (vertical green letters) are shown on the top of each contig, and the approximate location of each marker in the clone is indicated by a vertical line crossing the clone(s).

Table 1. DNA markers used in this work.

Categoly	Markers	Resources
RFLP markers	m302, m262, m243, m228, m317, m105, m255, m433, m249, m457, m460, m339, m424, g4523, g4119, g4547, KG17, g4711, g2534 g4014, g4125, AP3, g2778, pCIT1210, 560B1	Arabidopsis Biological Resource Centor (Columbus, Ohaio)
	miT234, mi74, mi199, mi355, mi403, mi357, mi467, mi207, miT131, mi289, mi339, mi142, mi268, mi386 mi225, mi178, mi413, miT149, mi456, miT136, miT220	Mitsui Plant Biotechnology Research Institute (Tsukuba, Japan)[23]
CAPS markers	CA1, S6, C6, B4, 32C9, CDC2B, PUR5, BGL1, FUS6, TSA1	http://genome-www.stanford.edu /Arabidopsis/maps/CAPS_Chr3.html
SSLP markers	nga32, nga172, nga126, ATHCHIB, nga162, nga6, nga112	Bell and Ecker[19]
EST markers	T04473, T22943, Z33694, T04196, T23365, Z37281, T04109, T20956, T43984, T13951, T45148, T44784, R30025, Z30817, pk1	http://genome-www.stanford.edu /Arabidopsis/EST2CIC.html
Cloned genes	CGS(U83500), GASA5(U53221), ILR1(U23794), AtGST(U70672), GAPC(M64119), RPM1(X87851), apx1B(X80036), SPINDLY(U62135), FAD2(L26296), MS2(X73652), AtMSH2(AF003005), DWF1(U12400), MAP2K_α(Y07694), prcgb(Y13692), AtDMC1(U76670), SUP(U38964), TMKL1(X72863), ABI3(AJ002473), GAPA(M64117), prcfb(Y13174), GL1(L22786), AIG2(U40857), atpox(X97075), NIT1(S45384), ASN1(L29083), AtCSR1(X51514), U1 snRNP(U52909), AtEm1(AF049236), AFC1(U16176), tai224(Z26409), AT103(U38232), AGL13(U20183), ACS1(U26542)	DDBJ/EMBL/GenBank DNA databases

manuscript in preparation) libraries from Mitsui Plant Biotechnology Research Institute.

2.2. DNA markers

Restriction fragment length polymorphism (RFLP) markers, cleaved amplified polymorphic sequence (CAPS) markers, simple sequence length polymorphism (SSLP) markers, expression sequence tags (EST) markers and cloned genes used for designation of PCR primers and their sources are listed in Table 1.

2.3. Screening and clone analysis

For physical map construction, clones harboring DNA markers were selected from P1, TAC and BAC libraries by PCR using DNA pools produced from each library as templates. The method of end sequence analysis of selected clones was described in the previous paper.[7] The sizes of CIC YAC clones, for which multiple size information was described[14] or whose size information was inconsistent with their allocation in the physical map, were confirmed by PFG electrophoresis (CHEF Mapper system; Bio-Rad), followed by Southern hybridization with PCR products which were amplified using the sequence information of the DNA markers on the clones.

3. Results and Discussion

3.1. Construction of physical map

Using the PCR primers designed from the sequences of 98 genetically mapped markers, 14 EST markers and 145 YAC end sequences, we isolated 115 CIC YAC clones. By ordering the clones on the basis of overlap analysis of the end sequences of each clone, a physical map which consists of 11 contigs and covers 20 Mb was first constructed. To close the remaining gaps, chromosome walking by using P1, TAC and BAC clones was carried out. A total of 128 P1, TAC and BAC clones were anchored on the regions encompassing the gaps by the use of 10 molecular markers, 20 YAC end sequences and 156 P1, TAC and BAC end sequences. Nine out of 10 gaps were closed by arrangement of the clones identified. Consequently, chromosome 3 could be covered by two contigs of 13.6 Mb (contig 1) and 9.2 Mb (contig 2) leaving a single gap in the middle. The sizes of two contigs were estimated by using the size information of individual clones which were placed on the mid-point of its allocated position. By comparison of the sizes of P1, TAC and BAC clones allocated to the overlapped regions of YAC clones, we assume the estimation error for the contig size to be less than 10%. A minimum tiling path of chromosome 3 composed of 55 CIC YAC and 37 P1, TAC and BAC clones is shown in Fig. 1. As discussed in section 3.3, we tentatively postulated the size between two contigs to be 0.7 Mb, and the physical scale of the chromosome was indicated by the number in Mb from the end of contig 1 containing the top marker. Forty-five chromosome-specific RI markers, whose genetic distances have been derived from up to 100 lines of a recombinant inbred mapping population,[17] were positioned on the physical map as summarized in Fig. 2: 32 markers from CGS (0.0 cM) to atpox (46.75 cM) fell in contig 1, and 13 markers from T04109 (53.72 cM) to

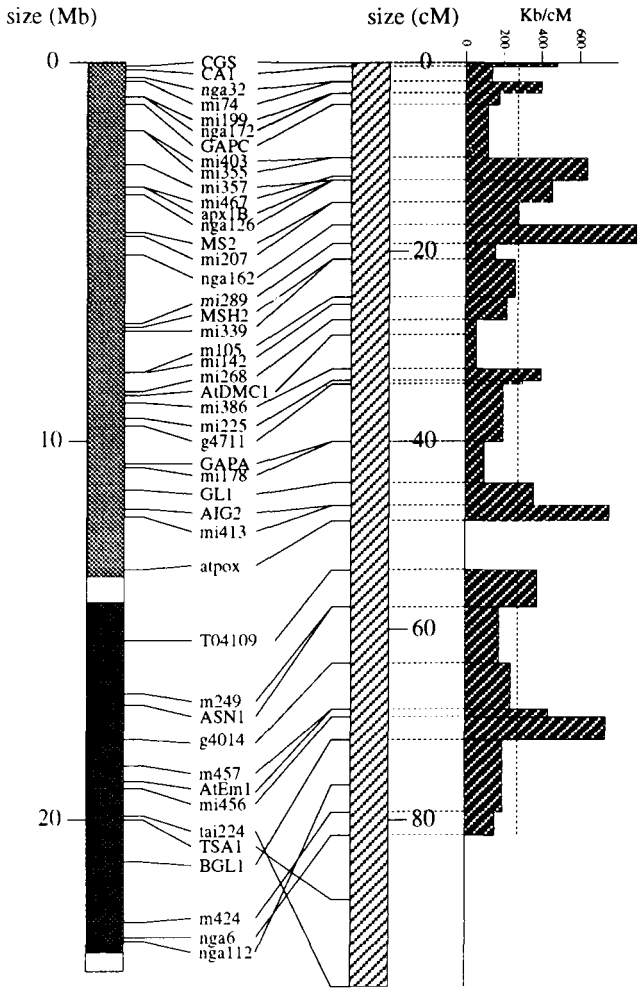


Figure 2. The alignment of the physical and RI maps of chromosome 3 and the ratio of the physical (Kb) and genetic (cM) scales along the RI map. Forty-five markers which have been mapped on the RI map were assigned on the physical map according to Fig. 1. The gray boxes separated by an open box at the left represent the two contigs. The size of the physical map (Mb) is shown on the left side. The shaded box at the middle indicates the RI map with the size (cM) on the right. The ratio of the physical scale (Kb) to the genetic scale (cM) was calculated from the physical and genetic distances, and the average ratios are shown by vertical broken lines.

tai224 (97.68 cM) fell in contig 2.

3.2. Comparison of genetic and physical maps

A comparison of genetic and physical maps is shown in Fig. 2. The order of genetic markers in the contig 1 region well coincides with that of the physical map except for the markers mi357 and mi467 that are located in close proximity. In the region of contig 2, however, the positions of markers TSA1[18] and tai224 were significantly inconsistent. In the RI map, tai224 and TSA1 have been assigned as the bottom markers of chromosome 3, but these markers were mapped to the 20-Mb position of the physical map. To verify the position of tai224 and TSA1, we constructed a detailed contig map of P1 and TAC clones in the region between YAC end markers CIC6F4R (19.4 Mb) and CIC11G5R (20.1 Mb). According to the fine contig map constructed, tai224 was mapped to the 19.9 Mb position closely linked to RFLP marker AP3, and TSA1 was mapped to 100–150 kb below tai224 which is linked to YAC end marker CIC8H7R. In addition, all the P1 and TAC clones isolated by screening with 3 sets of primers designed from the sequence information of tai224 and TSA1 were integrated into that contig (data not shown). Thus, it is unlikely that some other sequences identical or similar to tai224 and TSA1 were located in the bottom region of chromosome 3. The order of markers m424, nga6 and nga112 was also different between the physical and RI maps. The order of these markers on the physical map rather corresponds to the original genetic map of nga markers.[19]

The ratio of the physical to genetic distance between markers varied significantly along the chromosome (Fig. 2). Within the range of this resolution, relatively cold spots of recombination are seen in the middle region of each contig (mi403 to mi289 and m457 to BGL1) in addition to the region flanking the gap between the two contigs. This pattern seems to resemble those of chromosomes 4[20] and 5[7] of *A. thaliana*.

3.3. Comparison of cytogenetic and physical maps

Due to the lack of suitable probes, the centromere was not mapped on the physical map in the present study. It seems most likely that the centromeric region of chromosome 3 is positioned in the gap between contigs 1 and 2, because the gap position also roughly coincides with the prediction by genetic recombination analysis.[21] In addition, several YAC clones allocated in this region contained the 180-bp repeats[14,22] and other repetitive sequences (unpublished observation). The genetic distance between atpox and T04109 containing the gap is 5.16 cM. Using a conversion factor of 0.37 Mb/cM estimated from the neighboring T04109-m249 region, we estimated the physical distance between atpox and T04109 to be 1.9 Mb. Since 1.2 Mb of the 1.9-Mb region has already been occupied by the assigned clones, we assume that the gap size is 0.7 Mb, unless recombination has strongly been suppressed in this region. The physical distances of the telomeres to the end of each contig have not been estimated.

The physical map of *A. thaliana* chromosome 3 presented here will facilitate the isolation and characterization of genetically mapped genes. The map will also support the genome sequencing project.

Acknowledgments: We are grateful to Drs. S. Choi and R. Wing of Texas A & M University, T. Altman of Max-Plank Institute and I. Bancroft of John Innes

Center for providing the TAMU and IGF BAC libraries, and to the Mitsui Plant Biotechnology Research Institute, the Research Institute of Innovative Research for the Earth and Arabidopsis Biological Resource Center at the Ohio State University for providing the DNA libraries and the DNA markers. Thanks are also due to T. Wada, K. Naruo, T. Hosouchi, and N. Nakazaki for their excellent technical assistance. This work was supported by the Kazusa DNA Research Institute Foundations.

References

1. Meyerowitz, E. M. and Somervill, C. R. (eds) 1994, *Arabidopsis*, Cold Spring Harbor Laboratory Press.
2. Kaiser, J. 1996, First global sequencing effort begins, *Science*, **274**, 30.
3. Luetwiler, L. S., Hough-Evans, B. R., and Meyerowitz, E. M. 1984, The DNA of *Arabidopsis thaliana*, *Mol. Gen. Genet.*, **194**, 15–23.
4. Zachgo, E. A., Wang, M. L., Dewdney, J. et al. 1996, A physical map of Chromosome 2 of *Arabidopsis thaliana*, *Genome Res.* **6**, 19–25.
5. Schmidt, R., West, J., Cnops, G. et al. 1996, Detailed description of four YAC contigs representing 17 Mb of chromosome 4 of *Arabidopsis thaliana* ecotype Columbia, *Plant J.*, **9**, 755–765.
6. Schmidt, R., Love, K., West, J. et al. 1997, Description of 31 YAC contigs spanning the majority of *Arabidopsis thaliana* chromosome 5, *Plant J.* **11**, 563–572.
7. Kotani, H., Sato, S., and Fukami, M. 1997, A fine physical map of *Arabidopsis thaliana* chromosome 5: Construction of a sequence-ready contig map, *DNA Res.*, **4**, 371–378.
8. Sato, S., Kotani, H., Nakamura, Y. et al. 1997, Structural analysis of *Arabidopsis thaliana* chromosome 5. I. Sequence features of the 1.6 Mb regions covered by twenty physically assigned P1 clones, *DNA Res.*, **4**, 215–230.
9. Kotani, H., Nakamura, Y., Sato, S. et al. 1997, Structural analysis of *Arabidopsis thaliana* chromosome 5. II. Sequence features of the regions of 1,044,062 bp covered by thirteen physically assigned P1 clones, *DNA Res.*, **4**, 291–300.
10. Nakamura, Y., Sato, S., Kaneko, T. et al. 1997, Structural analysis of *Arabidopsis thaliana* chromosome 5. III. Sequence features of the regions of 1,191,918 bp covered by seventeen physically assigned P1 clones, *DNA Res.*, **4**, 401–414.
11. Sato, S., Kaneko, T., Nakamura, Y. et al. 1998, Structural analysis of *Arabidopsis thaliana* chromosome 5. IV. Sequence features of the regions of 1,456,315 bp covered by nineteen physically assigned P1 and TAC clones, *DNA Res.*, **5**, 41–54.
12. Kaneko, T., Kotani, H., Nakamura, Y. et al. 1998, Structural analysis of *Arabidopsis thaliana* chromosome 5. V. Sequence features of the regions of 1,381,565 bp covered by twenty one physically assigned P1 and TAC clones, *DNA Res.*, **5**, 41–54.
13. Camilleri, C., Lafleuriel, J., Macadre, C. et al. 1998, A YAC contig map of *Arabidopsis thaliana* chromosome 3, *Plant J.* in press.
14. Creusot, F., Fouilloux, E., Dron, M. et al. 1995, The CIC library: A large insert YAC library for genome mapping in *Arabidopsis thaliana*, *Plant J.*, **8**, 763–770.
15. Choi, S., Creelman, R. A., Mullet, J. E., and Wing, R. A. 1995., Construction and characterization of a bacterial artificial chromosome library of *Arabidopsis thaliana*, *Weeds World*, **2**, 17–20.
16. Liu, Y.-G., Mitsukawa, N., Vazquez-Tello, A., and Whittier, R. F. 1995, Generation of a high-quality P1 library of *Arabidopsis* suitable for chromosome walking, *Plant J.*, **7**, 351–358.
17. Lister, C. and Dean, C. 1993, Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J.* **10**, 733–736.
18. Radwanski, E. R., Zhao, J., and Last, R. L. 1995, *Arabidopsis thaliana* tryptophan synthase alpha: gene cloning, expression, and subunit interaction, *Mol. Gen. Genet.*, **248**, 657–667.
19. Bell, C. J. and Ecker, J. R. 1994, Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*, *Genomics*, **19**, 137–144.
20. Schmidt, R., West, J., Love, K. et al. 1995, Physical map and organization of *Arabidopsis thaliana* chromosome 4, *Science*, **270**, 480–483.
21. Copenhaver, G. P., Browne, W. E., and Preuss, D. 1998, Assaying genome-wide recombination and centromere functions with *Arabidopsis* tetrads, *Proc. Natl. Acad. Sci. USA*, **95**, 247–252.
22. Maluszynska, J. and Heslop-Harrison, J. S. 1991, Localization of tandemly repeated sequences in *Arabidopsis thaliana*, *Plant J.*, **1**, 159–166.
23. Liu, Y.-G., Mitsukawa, N., Lister, C. et al. 1996, Isolation and mapping of a new set of 129 RFLP markers in *Arabidopsis thaliana* using recombinant inbred lines, *Plant J.*, **10**, 733–736.