Structural Analysis of *Arabidopsis thaliana* Chromosome 5.  
X. Sequence Features of the Regions of 3,076,755 bp Covered by Sixty P1 and TAC Clones

Shusei SATO, Yasukazu NAKAMURA, Takakazu KANEKO, Tomohiko KATOII, Erika ASAMIZU, Hirokazu KOTANI, and Satoshi TABATA*

Kazusa DNA Research Institute, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

(Received 24 January 2000)

Abstract

In our ongoing project to deduce the nucleotide sequence of *Arabidopsis thaliana* chromosome 5, non-redundant P1 and TAC clones have been sequenced on the basis of the fine physical map, and as of January, 2000, the sequences of 16.6 Mb representing approximately 60% of chromosome 5 have been accumulated and released at our web site. Along with the sequence determination, structural features of the sequenced regions have been analyzed by applying a variety of computer programs, and we already predicted a total of 2697 potential protein coding genes in the 11,166,130 bp regions, which are covered by 159 P1 and TAC clones. In this paper, we describe the structural features of the 3,076,755 bp regions covered by newly analyzed 60 P1 and TAC clones. A total of 715 potential protein coding genes were identified, giving an average density of the genes identified of 1 gene per 4001 bp. Introns were observed in 80% of the genes, and the average number per gene and the average length of the introns were 4.5 and 147 bp, respectively. These sequence features are nearly identical to those in our latest report in which the data were compiled based on a new standard of gene assignment including the computer-predicted hypothetical genes. The regions also contained 12 tRNA genes when searched by similarity to reported tRNA genes and the tRNA scan-SE program. The sequence data and information on the potential genes are available through the World Wide Web database KAOS (Kazusa *Arabidopsis* data Opening Site) at http://www.kazusa.or.jp/kaos/.

Key words: *Arabidopsis thaliana* chromosome 5; genomic sequence; P1 genomic library; TAC genomic library; gene prediction

In order to investigate the whole genetic system in higher plants, we have been operating a sequencing project of the genome of a dicot model plant *Arabidopsis thaliana*. Of five chromosomes that constitute the *A. thaliana* genome of approximately 120 Mb, we focused our efforts on chromosomes 5 and 3. For precise localization of the clones for DNA sequencing, we constructed the fine physical maps of both chromosomes with clones from YAC, P1, TAC, and BAC libraries.1-2 On the basis of the fine physical map information, P1 and TAC clones were selected and assigned on the map by polymerase chain reaction (PCR), and then subjected to sequence analysis. As of January 2000, the regions of 16.6 Mb representing approximately 60% of chromosome 5 have been sequenced and the data are available at our web site KAOS (Kazusa *Arabidopsis* data Opening Site, http://www.kazusa.or.jp/kaos/). In parallel, potential genes in the sequenced regions have been analyzed using a variety of computer programs for similarity search and gene modeling, and we so far predicted the potential genes in a total of 11,166,130 bp which are represented by 159 P1 and TAC clones.3-11 In this paper, we newly investigated the structural features of the 3,076,755 bp regions covered by an additional 60 P1 and TAC clones.

1. Isolation and Sequencing of P1 and TAC Clones

DNA sources and the method of clone isolation were essentially the same as described in the previous paper.3 The P1 and TAC clones containing the DNA regions which cover a total of 60 DNA markers on chromosome 5 were isolated by screening the Mitsui P112 and TAC13 libraries by means of PCR using primers designed from the sequence information of DNA markers. The DNA markers and selected clones are listed in Table 1. Relative positions of the markers and the sequenced clones...
on chromosome 5 are shown in Fig. 1. The relative orientation of each clone and contig on the chromosome has been confirmed by anchoring both ends of the clone to those at the corresponding positions of the contig map.

The nucleotide sequence of each PI or TAC insert was determined according to the bridging shotgun method described previously. The length of the nucleotide sequence of each PI or TAC insert finally confirmed is listed together with the accession numbers in Table 1.

2. Assignment of Potential Coding Regions

For assignment of the protein coding regions and gene modeling, similarity search and computer prediction were performed as described in the previous paper. Briefly, similarity search against the non-redundant protein sequence database nr (compiled by NCBI) was carried out using the BLASTX program. In parallel, the positions of potential protein coding regions were predicted with the Grail, GENSCAN and NetGene2 computer programs. The transcribed regions were assigned by comparison of the nucleotide sequences with Arabidopsis ESTs in the public databases using the BLASTN program. All the results obtained were compiled with the aid of our new web-based tool, named Arabidopsis Genome Displayer (manuscript in preparation), then assignment of the potential protein coding genes was carried out by taking both similarity to known genes and computer prediction into consideration. Therefore, the regions predicted only by the computer programs with no apparent similarity to known genes were also assigned as genes. This standard of gene assignment has been adopted since the analysis in our last report, while such computer-predicted hypothetical genes were not included in the earlier analyses. To sum up, 715 potential protein-coding genes as well as 54 partial genes located at the terminal regions of the clones and 43 pseudo genes were assigned in the 3,076,755 bp regions, giving an average gene density of 1 gene per 4001 bp. This value is lower than that in our latest report in which the data were compiled based on a new standard of gene assignment described above, and is higher than that observed in regions of chromosomes 2 and 4. The reason for this inconsistency is thought to be the difference in the ratio of heterochromatic regions within the analyzed sequences.

In addition to the protein-coding regions, the RNA coding regions were assigned on the basis of sequence similarity to the reported structural RNAs. For tRNA genes, the prediction by the tRNAscan-SE program was also taken into account. As a result, 12 tRNA genes corresponding to 12 amino acid species and genes for U1, U3 and U4 snRNAs were identified in the 3,076,755 bp regions. Both potential protein and RNA coding genes are denoted by numbers with the clone names followed by sequential numbers from one end to another of the intervals (Mbp).

Figure 1. Relative locations of the sequenced PI and TAC clones and the associated markers on the physical map of chromosome 5. The positions of DNA markers used for PI and TAC isolation and of other major DNA markers were localized on the map on the basis of the YAC tiling path and map information in ref. 1. The vertical open bar represents the entire length of chromosome 5. The names of PI and TAC clones are given at the right side, and those of markers at the left side. The distance (Mbp) from the telomeric site of the top arm is given in the vertical scale.
In this paper, the complete structures of 715 potential protein coding genes were predicted. Structural features of these genes as well as those of 2619 genes including those previously identified are listed in Table 1. They amount for approximately 13.1% of the total gene constituents (2×10^4 genes) assumed for A. thaliana. Approximately 77% of the protein-coding genes contained introns, and the average number per gene and their average length were 4.0 and 167 bp, respectively.

### Table 1. Information of the sequenced PI and TAC clones.

| Clone name | DNA markers | Confirmed length (bp) | Accession number |
|------------|-------------|-----------------------|-----------------|
| K1L20      | ends of K2A18&K1P13 | 47665                 | AB022211       |
| K1O13      | ends of MEE&MYC6  | 25275                 | AB019225       |
| K2M18      | CIC1F1L      | 41465                 | AB023651       |
| K2N11      | ends of MFC19&MRA19 | 30340                 | AB022313       |
| K5A21      | MDD2_right end | 13880                 | AB024030       |
| K5F14      | CIC310R      | 31178                 | AB022814       |
| K5U14      | ends of M3C6&MDH9 | 59762                 | AB023032       |
| K6A12      | MXT22_left end | 64136                 | AB024031       |
| K6M13      | mg129        | 77129                 | AB023033       |
| K7M9       | K1P31_right end | 56563                 | AB023034       |
| K7R15      | ends of K1R1&MFC9 | 26052                 | AB027044       |
| K9H21      | ends of MDC2&MLE2 | 15119                 | AB023035       |
| K9P8       | MPF11_right end | 90760                 | AB024032       |
| K111       | MPF12_right end | 55129                 | AB023032       |
| K14B20     | K2A18_left end  | 40521                 | AB018108       |
| K15O15     | CIC1065R     | 23026                 | AB024026       |
| K18F1      | ends of MDF8&MPF20 | 33903                 | AB022270       |
| K18F13     | ends of MCA2A&MDF11 | 19742                 | AB024035       |
| K17N15     | CIC1F10      | 81293                 | AB018109       |
| K17O22     | ends of K2C1&K18C1 | 67720                 | AB019224       |
| K1B18      | ends of K1H9&MNC17 | 35896                 | AB024027       |
| K2D17      | ends of K1P31&K1P13 | 36243                 | AB020228       |
| K2H11      | CIC3811      | 74342                 | AB020742       |
| K2J19      | MCK2_left end  | 41087                 | AB024029       |
| K2J21      | ends of MAC9&MTG10 | 40453                 | AB022212       |
| K2J317     | ends of MPA2A&K1B20 | 12121                 | AB020743       |
| K24C1      | MDA7_left end  | 29498                 | AB023029       |
| K24B7      | CIC1F101     | 73999                 | AB019226       |
| K25F15     | CIC1065L     | 10045                 | AB018121       |
| MBB17      | ends of MGI9&MIH224 | 52717                 | AB019227       |
| MCB7       | mi184        | 87900                 | AB019228       |
| MEF16      | CIC1F10      | 66067                 | AB023037       |
| MFO1       | MMN10_right end | 43570                 | AB019231       |
| MHH17      | CIC1064L     | 76423                 | AB024035       |
| MIP21      | MDF11_right end | 59372                 | AB023039       |
| MHR4       | MSF19_left end | 58589                 | AB019233       |
| MJE7       | K15N18_left end | 74298                 | AB020745       |
| MJM18      | ends of MI024&MSG15 | 16203                 | AB026523       |
| MJT17      | ends of K1W17&K1G13 | 31827                 | AB018115       |
| MKC22      | ends of MCD7&MK19 | 27229                 | AB019234       |
| MNP16      | ends of MTG10&K19B1 | 81736                 | AB019235       |
| MNB6       | MCG0_right end | 46872                 | AB018116       |
| MNF5       | KME11_right end | 70111                 | AB025627       |
| MP110      | mi69         | 29605                 | AB020747       |
| MQ15       | CICI1B8L     | 88398                 | AB018117       |
| MQC1       | ends of K1Y1&MRC11 | 81365                 | AB025633       |
| MRG11      | ends of K1B9&MQ92 | 55125                 | AB017051       |
| MSDL2      | ends of MZA15&MQD22 | 33479                 | AB022221       |
| MSJ10      | CICI118      | 81414                 | AB024037       |
| MSH2       | ends of F13&MUD21 | 69257                 | AB018119       |
| MUD12      | ends of MSNK9&MYH19 | 27601                 | AB022222       |
| MUF8       | ends of MBK23&K1622 | 13776                 | AB025635       |
| MUL3       | MB24_left end  | 82020                 | AB023042       |
| MW222      | K5F12_left end  | 87180                 | AB020144       |
| MWF20      | CIC5F12L     | 91913                 | AB025638       |
| MWJ1       | MPD20_right end | 42356                 | AB018120       |
| MWJ9       | ends of M3H1&MK22 | 13020                 | AB020753       |
| MXK3       | CIC1B4L      | 84194                 | AB019236       |
| MYN8       | ends of K19E1&MCN6 | 54528                 | AB020754       |
| MZN1       | K19M22_left end | 81672                 | AB020755       |

3. Structural Features of Potential Protein Genes

The nucleotide sequence of each of the potential protein coding genes was compared with those in the Arabidopsis EST database, and the number of matched Arabidopsis ESTs was counted to monitor the transcriptional level of each gene. Of 715 complete and 54 partial genes that we have identified in chromosome 5 in this study, 290 carried matched ESTs. The putative products of the genes hit by 10 or more EST files, suggesting to be a class of highly expressed genes, include those showing sequence similarity to multicatalytic endopeptidase complex, proteasome component, alpha subunit in A. thaliana (K2K18.4), yoxilosidase in Aspergillus niger (K7J8.3), hypothetical protein in A. thaliana (K1B18.8), subtilisin-like protease homolog in A. thaliana (K1B18.9), outer membrane lipoprotein Bic precursor in Citrobacter freundii (K2I19.6), 26S protease regulatory subunit 6B homolog in Solanum tuberosum (MCK7.16), unknown protein in A. thaliana (M1F21.5), RNA helicase in A. thaliana (MM19.2), 40S ribosomal protein S20 in A. thaliana (MM19.13), tubulin beta-2/beta-3 chain in A. thaliana (MRG21.11 and MRG21.12), cytoplasmic malate dehydrogenase in A. thaliana (MWF20.2), NO1 protein in A. thaliana (MWJ3.3), and glutamate synthase precursor in Medicago sativa (MYN8.7).

4. Expression Level of Potential Protein Genes and Gene Segments

The sequence data as well as the gene information shown in this paper are available through the World Wide Web at http://www.kazusa.or.jp/kaos/.

Acknowledgements: We thank S. Sasamoto for excellent technical assistance. Thanks are also due to T. Kimura, T. Hosouchi, K. Ida, and Y. Kamakura, M. Matsunaga, A. Matsuno, A. Murakami, N. Nakazaki, S. Shinpo, T. Takeuchi, T. Wada, A. Watanabe, M. Yamada, and M. Yasuda for their excellent teamwork in sequence analysis. We are grateful to A. Tanaka for technical advice, and Mitsui Plant Biotechnology Research Institute and Arabidopsis Biological Resource Center at the Ohio State University for providing the DNA markers and the DNA libraries. This work was supported by the Kazusa DNA Research Institute Foundation.
Table 2. Structural features of potential protein coding genes in *A. thaliana* chromosome 5.

| Features                               | 715 genes$^a$ | 2619 genes$^b$ |
|----------------------------------------|---------------|---------------|
| Gene length (bp) including introns     | 74-14479 (1993)| 62-14479 (1965)|
| Product length (amino acids)           | 25-2216 (445)  | 19-2756 (433)  |
| Genes with introns                     | 575           | 2012          |
| Number of introns/gene                 | 0-42 (4.5)    | 0-42 (4.0)    |
| Exon length (bp)                       | 3-4473 (245)  | 2-4473 (260)  |
| Intron length (bp)                     | 26-1450 (147) | 8-5405 (167)  |
| GC content of exons                    | 43%           | 43%           |
| GC content of introns                  | 32%           | 32%           |

Structural features of the potential protein-coding genes assigned so far are listed. The 715 genes are assigned based on the new standard in this study$^a$ and the 2619 genes$^b$ include previously assigned 1901 potential protein genes. Average values are shown in parentheses.

References

1. Kotani, H., Sato, S., Liu, Y-G. et al. 1997, A fine physical map of *Arabidopsis thaliana* chromosome 5: Construction of a sequence-ready contig map, *DNA Res.*, 4, 371–378.

2. Sato, S., Kotani, H., Hayashi, R. et al. 1998, A physical map of *Arabidopsis thaliana* chromosome 3 represented by two contigs of CIG YAC, P1, TAC and BAC clones, *DNA Res.*, 5, 163–168.

3. 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.

4. 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, 309–320.

5. 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.

6. Sato, S., Kaneko, T., Kotani, H. 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.

7. Kotani, H., Nakamura, Y., Sato, S. 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, 131–145.

8. Kotani, H., Nakamura, Y., Sato, S. et al. 1998, Structural analysis of *Arabidopsis thaliana* chromosome 5. VI. Sequence features of the regions of 1,367,185 bp covered by 19 physically assigned P1 and TAC clones, *DNA Res.*, 5, 203–216.

9. Nakamura, Y., Sato, S., Asamizu, E. et al. 1998, Structural analysis of *Arabidopsis thaliana* chromosome 5. VII. Sequence features of the regions of 1,013,767 bp covered by sixteen physically assigned P1 and TAC clones, *DNA Res.*, 5, 297–308.

10. Asamizu, E., Sato, S., Kaneko, T. et al. 1998, Structural analysis of *Arabidopsis thaliana* chromosome 5. VIII. Sequence features of the regions of 1,081,958 bp covered by seventeen physically assigned P1 and TAC clones, *DNA Res.*, 5, 379–391.

11. Kaneko, T., Kato, T., Sato, S. et al. 1999, Structural analysis of *Arabidopsis thaliana* chromosome 5. IX. Sequence features of the regions of 1,011,550 bp covered by seventeen P1 and TAC clones, *DNA Res.*, 6, 183–195.

12. 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.

13. Liu, Y.-G., Shirano, Y., Fukaki, H. et al. 1999, Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning, *Proc. Natl. Acad. Sci. USA*, 96, 6535–6540.

14. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990, Basic local alignment search tool, *J. Mol. Biol.*, 215, 403–410.

15. Uberbacher, E. C. and Mural, R. J. 1991, Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach, *Proc. Natl. Acad. Sci. USA*, 88, 11261–11265.

16. Burge, C. and Karlin, S. 1997, Prediction of complete gene structures in human genomic DNA, *J. Mol. Biol.*, 268, 78–94.

17. Hebsgaard, S. M., Korning, P. C., Tolstrup, N. et al. 1996, Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information, *Nucl. Acids Res.*, 24, 3439–3452.

18. Newman, T., Bruijn, F. J., and Green, P. 1994, Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones, *Plant Physiol.*, 106, 1241–1255.

19. Cooke, R., Raynal, M., Laudie, M. et al. 1996, Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5000 non-redundant ESTs, *Plant J.*, 9, 101–124.

20. Lin, X., Kaul, S., Rounsley, S. et al. 1999, Sequence
and analysis of chromosome 2 of the plant Arabidopsis thaliana, Nature, 402, 761-768.

21. Mayer, M., Schüller, C., Wambutt, R. et al. 1999, Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana, Nature, 402, 769-777.

22. Lowe, T. M. and Eddy, S. R. 1997, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucl. Acids Res., 25, 955-964.
Figure 2. Gene organization in the 60 PI and TAC clones. Positions of the identified or predicted genes in each insert of the PI and TAC clones are schematically represented by color-coded boxes above (rightward) and below (leftward) the wide line in the middle which represents the entire insert sequence. The length of sequenced region in each insert was given in parenthesis together with the clone name at the top. The names of the adjacent overlapping clones of which sequences had been reported are shown on the middle bars. Arrowheads indicate the directions of the DNA strands (5' to 3'). Dark and faint blue bars with numbers represent the positions of the assigned potential protein coding genes, and pseudo and partial genes, respectively, and red bars the positions of RNA coding genes. Gray bars indicate the positions of the regions which matched to the Arabidopsis ESTs. The regions which showed similarity to the sequences in the protein database are shown by yellow, orange and red bars, each of which corresponds to BLASTX scores of 70-100, 100-200, and 250 or more, respectively. The green bars indicate the positions of the potential exons predicted by the Grail program. Each of these different colors with increasing depth corresponds to the regions with Grail scores of less than 70, 70-90, and 90 or more, respectively. The potential protein and RNA coding genes assigned as described in the text were listed below each of the figures. In this table, the number of amino acid residues and nucleotide length (in italic) of putative gene products of the respective potential protein and RNA coding genes are indicated.
| No. 1 | S. Sato et al. | 37 |

### K15O15 (23026 bp)

| Deduced Gene | Position | No. of Exon | No. of EST | Length | Sequence ID | Gene ID | Identity | Definition |
|--------------|----------|-------------|------------|--------|-------------|---------|----------|------------|
| K15O15.1     | 1        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |
| K15O15.2     | 2        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |
| K15O15.3     | 3        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |
| K15O15.4     | 4        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |

### K16F13 (19742 bp)

| Deduced Gene | Position | No. of Exon | No. of EST | Length | Sequence ID | Gene ID | Identity | Definition |
|--------------|----------|-------------|------------|--------|-------------|---------|----------|------------|
| K16F13.1     | 1        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |
| K16F13.2     | 2        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |
| K16F13.3     | 3        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |
| K16F13.4     | 4        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |

### K16E1 (33963 bp)

| Deduced Gene | Position | No. of Exon | No. of EST | Length | Sequence ID | Gene ID | Identity | Definition |
|--------------|----------|-------------|------------|--------|-------------|---------|----------|------------|
| K16E1.1      | 1        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |
| K16E1.2      | 2        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |
| K16E1.3      | 3        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |
| K16E1.4      | 4        | 1           | 5          | 430    | gi|115650| 100.0 | protein A. thaliana |

---

**8.3.2.**
Sequencing of *Arabidopsis thaliana* chromosome 5

**K18B18 (35896 bp)**

- **Gene**
- **EST db hit**
- **Protein db hit**
- **Gralexon**

| Identifier | Direction | S' Exon | A' Exon | No. of EST | Length | Information on the most similar sequence |
|------------|-----------|---------|---------|------------|--------|----------------------------------------|
| K18B18.1   | +         | 2       | 2       | 0          | 700    | c12905446.x (AGADP371) 707 32.3 (AC004516) unknown protein A thaliana |
| K18B18.2   | +         | 1       | 1       | 1          | 170    | c12905446.x (AGADP371) 170 100.0 copper transporter 1 |
| K18B18.3   | +         | 1       | 1       | 1          | 101    | c12905446.x (AGADP371) 101 52.0 copper transporter protein A thaliana |
| K18B18.4   | +         | 3       | 3       | 0          | 325    | c12905446.x (AGADP371) 325 52.0 |
| K18B18.5   | +         | 1       | 1       | 1          | 72     | AB009097.x (AC004522) 72 100.0 LNA Arab(GTC) |
| K18B18.6   | +         | 1       | 1       | 3          | 103    | AB009209.s (AC004523) 103 34.3 LNA Arab(GTC) |
| K18B18.7   | +         | 3       | 3       | 3          | 531    | AB009198.s (AC004522) 531 34.3 LNA Arab(GTC) |
| K18B18.8   | +         | 4       | 4       | 1          | 135    | AB009209.s (AC004522) 135 34.3 LNA Arab(GTC) |
| K18B18.9   | +         | 4       | 4       | 15         | 797    | AB009304.s (AC004522) 797 34.3 LNA Arab(GTC) |
| K18B18.10  | +         | 2       | 2       | 0          | 447    | AB009209.s (AC004522) 447 34.3 LNA Arab(GTC) |

**K22J17 (11211 bp)**

- **Gene**
- **EST db hit**
- **Protein db hit**
- **Gralexon**

| Identifier | Direction | S' Exon | A' Exon | No. of EST | Length | Information on the most similar sequence |
|------------|-----------|---------|---------|------------|--------|----------------------------------------|
| K22J17.1   | +         | 2       | 2       | 0          | 147    | AB009209.s (AC004522) 147 100.0 LNA Arab(GTC) |
| K22J17.2   | +         | 3       | 3       | 0          | 219    | AB009209.s (AC004522) 219 100.0 LNA Arab(GTC) |
| K22J17.3   | +         | 1       | 1       | 1          | 637    | AB009304.s (AC004522) 637 100.0 LNA Arab(GTC) |
| K22J17.4   | +         | 1       | 1       | 1          | 379    | AB009209.s (AC004522) 379 100.0 LNA Arab(GTC) |
### K17N15 (81293 bp)

| Exon | Start | End | Length | Sequence ID | Description |
|------|-------|-----|--------|-------------|-------------|
| 1    | 458   | 604  | 146    | g1594178     | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 2    | 336   | 404  | 68     | g1594179     | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 3    | 1025  | 1121 | 96     | g1594180     | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 4    | 1553  | 1594 | 41     | g1594181     | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 5    | 2101  | 2137 | 37     | g1594182     | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 6    | 2654  | 2655 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 7    | 3205  | 3205 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 8    | 3757  | 3756 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 9    | 4308  | 4307 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 10   | 4859  | 4858 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 11   | 5407  | 5406 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 12   | 5958  | 5957 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |

### K17O22 (67720 bp)

| Exon | Start | End | Length | Sequence ID | Description |
|------|-------|-----|--------|-------------|-------------|
| 1    | 458   | 604  | 146    | g1594178     | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 2    | 336   | 404  | 68     | g1594179     | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 3    | 1025  | 1121| 96     | g1594180     | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 4    | 1553  | 1594 | 41     | g1594181     | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 5    | 2101  | 2137 | 37     | g1594182     | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 6    | 2654  | 2655 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 7    | 3205  | 3205 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 8    | 3757  | 3756 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 9    | 4308  | 4307 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 10   | 4859  | 4858 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 11   | 5407  | 5406 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |
| 12   | 5958  | 5957 | 1      |              | partial cleavage a polyadenylation specificity factor, 154 kDa protein precursor |

Downloaded from https://academic.oup.com/dnaresearch/article-abstract/7/1/31/389236 by guest on 15 March 2020
**Sequencing of Arabidopsis thaliana chromosome 5**

---

### K1L20 (47665 bp)

| Position | No. of Exon | Length | Information on the most similar sequence |
|----------|-------------|--------|----------------------------------------|
| K1L20.1 | 40817       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.2 | 41298       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.3 | 39083       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.4 | 37110       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.5 | 22100       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.6 | 14001       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.7 | 10091       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.8 | 10089       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.9 | 7309        | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.10| 20270       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.11| 20082       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.12| 47427       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.13| 42724       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |
| K1L20.14| 39731       | 1      | identify Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270] |

---

### K1O13 (25275 bp)

| Position | No. of Exon | Length | Information on the most similar sequence |
|----------|-------------|--------|----------------------------------------|
| K1O13.1 | 3062       | 2      | [Vol. 7, 6002270] |
| K1O13.2 | 3075       | 2      | [Vol. 7, 6002270] |
| K1O13.3 | 3088       | 2      | [Vol. 7, 6002270] |
| K1O13.4 | 31016      | 2      | [Vol. 7, 6002270] |
| K1O13.5 | 31029      | 2      | [Vol. 7, 6002270] |
| K1O13.6 | 31032      | 2      | [Vol. 7, 6002270] |
| K1O13.7 | 31033      | 2      | [Vol. 7, 6002270] |
| K1O13.8 | 31034      | 2      | [Vol. 7, 6002270] |
| K1O13.9 | 31035      | 2      | [Vol. 7, 6002270] |
| K1O13.10| 31036      | 2      | [Vol. 7, 6002270] |

---

**Legend**

- **Protein db hit**: Matches to a protein database.
- **EST db hit**: Matches to an EST database.
- **Gene**: Identifies a gene.
- **Grail exon**: Identified as an exon in the Grail database.
- **Definition**: Provides a description or function of the sequence.

**Similarity Measures**

- **Identity**: Percentage of identical nucleotides.
- **Overlap**: Length of the overlap between sequences.
- **Percentage**: Additional metrics related to sequence comparison.

**Sequence IDs**

- **K1L20.1**: Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270]
- **K1O13.1**: Arabidopsis thaliana chromosome 5 [Vol. 7, 6002270]
Xo. 1

K20J1 (36243 bp)

| Exon ID | Position | No. of EST | No. of EST | Length | Information on the most similar sequence |
|---------|----------|------------|------------|--------|-----------------------------------------|
| K20J1.1 | 1        | 4          | 1          | 1      | K20J1.3 | X33192 |
| K20J1.2 | 784      | 449        | 1          | 0      | 100 | 100.0 | K20J1.3 |
| K20J1.3 | 1118     | 2153       | 1          | 0      | 540 | 54.0 | K20J1.3 |

K21H1 (74342 bp)

| Exon ID | Position | No. of EST | No. of EST | Length | Information on the most similar sequence |
|---------|----------|------------|------------|--------|-----------------------------------------|
| K21H1.1 | 1        | 4          | 1          | 1      | K20J1.3 | X33192 |
| K21H1.2 | 784      | 449        | 1          | 0      | 100 | 100.0 | K20J1.3 |
| K21H1.3 | 1118     | 2153       | 1          | 0      | 540 | 54.0 | K20J1.3 |

S. Sato et al.
### Sequencing of Arabidopsis thaliana chromosome 5

#### K21L19 (41087 bp)

| Identifier | Direction | Orientation | Position | No. of Exon | No. of EST | Length | Sequence ID | Start | Identity | Definition |
|------------|------------|-------------|----------|-------------|------------|--------|-------------|-------|-----------|------------|
| K21L19.1   | +          | 5' to 3'    | 1        | 1           | 1          | 387    | gi|43337216/gi|APX17312.1| 322 41.6| partial (APX17312) protein x 0001 Homo sapiens |
| K21L19.3   | +          | 5' to 3'    | 1        | 3           | 3          | 192    | gi|43337216/gi|APX17312.1| 199 91.3| partial (APX17312) protein x 0001 Homo sapiens |
| K21L19.5   | +          | 5' to 3'    | 1        | 1           | 139       | 1597   | gi|43337216/gi|APX17312.1| 100 90.4| partial (APX17312) protein x 0001 Homo sapiens |
| K21L19.6   | +          | 5' to 3'    | 1        | 2           | 2          | 189    | gi|43337216/gi|APX17312.1| 198 79.0| partial (APX17312) protein x 0001 Homo sapiens |
| K21L19.7   | +          | 5' to 3'    | 1        | 1           | 136       | 1553   | gi|43337216/gi|APX17312.1| 101 90.4| partial (APX17312) protein x 0001 Homo sapiens |
| K21L19.8   | +          | 5' to 3'    | 1        | 1           | 136       | 1553   | gi|43337216/gi|APX17312.1| 101 90.4| partial (APX17312) protein x 0001 Homo sapiens |
| K21L19.9   | +          | 5' to 3'    | 1        | 1           | 136       | 1553   | gi|43337216/gi|APX17312.1| 101 90.4| partial (APX17312) protein x 0001 Homo sapiens |
| K21L19.10  | +          | 5' to 3'    | 1        | 1           | 136       | 1553   | gi|43337216/gi|APX17312.1| 101 90.4| partial (APX17312) protein x 0001 Homo sapiens |
| K21L19.11  | +          | 5' to 3'    | 1        | 1           | 136       | 1553   | gi|43337216/gi|APX17312.1| 101 90.4| partial (APX17312) protein x 0001 Homo sapiens |
| K21L19.12  | +          | 5' to 3'    | 1        | 1           | 136       | 1553   | gi|43337216/gi|APX17312.1| 101 90.4| partial (APX17312) protein x 0001 Homo sapiens |
| K21L19.13  | +          | 5' to 3'    | 1        | 1           | 136       | 1553   | gi|43337216/gi|APX17312.1| 101 90.4| partial (APX17312) protein x 0001 Homo sapiens |
| K21L19.14  | +          | 5' to 3'    | 1        | 1           | 136       | 1553   | gi|43337216/gi|APX17312.1| 101 90.4| partial (APX17312) protein x 0001 Homo sapiens |

#### K22G18 (45453 bp)

| Identifier | Direction | Orientation | Position | No. of Exon | No. of EST | Length | Sequence ID | Start | Identity | Definition |
|------------|------------|-------------|----------|-------------|------------|--------|-------------|-------|-----------|------------|
| K22G18.1   | +          | 5' to 3'    | 1        | 1           | 1          | 348    | gi|4159264/gi|AAP51116.1| 502 98.5| partial (AAP51116) porphobilinogen deaminase-like A. thaliana |
| K22G18.2   | +          | 5' to 3'    | 1        | 1           | 1          | 348    | gi|4159264/gi|AAP51116.1| 502 98.5| partial (AAP51116) porphobilinogen deaminase-like A. thaliana |
| K22G18.3   | +          | 5' to 3'    | 1        | 1           | 1          | 348    | gi|4159264/gi|AAP51116.1| 502 98.5| partial (AAP51116) porphobilinogen deaminase-like A. thaliana |
| K22G18.4   | +          | 5' to 3'    | 1        | 1           | 1          | 348    | gi|4159264/gi|AAP51116.1| 502 98.5| partial (AAP51116) porphobilinogen deaminase-like A. thaliana |
| K22G18.5   | +          | 5' to 3'    | 1        | 1           | 1          | 348    | gi|4159264/gi|AAP51116.1| 502 98.5| partial (AAP51116) porphobilinogen deaminase-like A. thaliana |
| K22G18.6   | +          | 5' to 3'    | 1        | 1           | 1          | 348    | gi|4159264/gi|AAP51116.1| 502 98.5| partial (AAP51116) porphobilinogen deaminase-like A. thaliana |
| K22G18.7   | +          | 5' to 3'    | 1        | 1           | 1          | 348    | gi|4159264/gi|AAP51116.1| 502 98.5| partial (AAP51116) porphobilinogen deaminase-like A. thaliana |
| K22G18.8   | +          | 5' to 3'    | 1        | 1           | 1          | 348    | gi|4159264/gi|AAP51116.1| 502 98.5| partial (AAP51116) porphobilinogen deaminase-like A. thaliana |
| K22G18.9   | +          | 5' to 3'    | 1        | 1           | 1          | 348    | gi|4159264/gi|AAP51116.1| 502 98.5| partial (AAP51116) porphobilinogen deaminase-like A. thaliana |
| K22G18.10  | +          | 5' to 3'    | 1        | 1           | 1          | 348    | gi|4159264/gi|AAP51116.1| 502 98.5| partial (AAP51116) porphobilinogen deaminase-like A. thaliana |
| K22G18.11  | +          | 5' to 3'    | 1        | 1           | 1          | 348    | gi|4159264/gi|AAP51116.1| 502 98.5| partial (AAP51116) porphobilinogen deaminase-like A. thaliana |

#### K21L19 (41087 bp)

- **Graft exon**
- **EST db hit**
- **Gene**
- **EST db hit**
- **Protein db hit**
- **Graft exon**

#### K22G18 (45453 bp)

- **Graft exon**
- **Protein db hit**
- **EST db hit**
- **Gene**
- **EST db hit**
- **Protein db hit**
- **Graft exon**
Sequencing of *Arabidopsis thaliana* chromosome 5

**K24C1 (29498 bp)**

- **Deduced genes**
  - **Gene**
  - **EST db hit**
  - **Protein db hit**

**K2N11 (30340 bp)**

- **Deduced genes**
  - **Gene**
  - **EST db hit**
  - **Protein db hit**

**K5A21 (13874 bp)**

- **Deduced genes**
  - **Gene**
  - **EST db hit**
  - **Protein db hit**
### MAB16 (70475 bp)

| Deduced Gene | Position  | No. of Exons  | No. of EST | Length | Information on the Most Similar Sequence | Sequence ID | Quality | Identity | Definition |
|--------------|-----------|---------------|------------|--------|------------------------------------------|-------------|--------|----------|------------|
| MAB16.1      | -         | 3334          | 3          | 0      | 875                                      | gi|21324095|6| AA | 94.5 | (A1715555) contains similarity to homeobox protein MDRK (MAB17/FPA A. thaliana) |
| MAB16.2      | +         | 9152          | 2          | 0      | 112                                      | gi|2897707|6| -CA | 33.8 | (A032045) hypothetical protein A. thaliana |
| MAB16.3      | +         | 14490         | 3          | 0      | 141                                      | gi|6687682|6| -AF | 73.3 | (AC006011) hypothetical protein A. thaliana |
| MAB16.4      | -         | 17990         | 3          | 0      | 336                                      | gi|4666496|6| AG | 40.1 | (AC007320) Y2R1.1 A. thaliana |
| MAB16.5      | +         | 26369         | 4          | 0      | 477                                      | gi|6219278|6| -AA | 21.4 | (A006621) putative cytochrome P450 A. thaliana |
| MAB16.6      | -         | 35152         | 3          | 0      | 512                                      | gi|1902884|6| -AA | 42.2 | (D04020) hypothetical protein. Hypothetical protein. |
| MAB16.7      | +         | 35941         | 3          | 0      | 1146                                     | gi|9977121|6| -AF | 62.4 | putative nuclear polyprotein - A. thaliana retention sequence Tes-3 |
| MAB16.8      | -         | 36433         | 4          | 0      | 140                                      | gi|4881278|6| -AA | 49.4 | (AC006851) putative cytochrome P450 A. thaliana |
| MAB16.9      | +         | 37364         | 14         | 0      | 216                                      | gi|6801693|6| -AF | 29.9 | CYTOCHROME. Predicted |
| MAB16.10     | -         | 39809         | 12         | 0      | 178                                      | gi|4688688|6| -AG | 76.0 | (AC007360) hypothetical. No definition line based A. thaliana |
| MAB16.11     | +         | 40990         | 5          | 0      | 456                                      | gi|6132593|6| -AA | 84.9 | (A001098) hypothetical tyrosine aminotransferase. A. thaliana |
| MAB16.12     | -         | 42714         | 4          | 0      | 488                                      | gi|1671770|6| -AF | 77.9 | (AC006279) hypothetical tRNA nucleotidyltransferase. A. thaliana |
| MAB16.13     | +         | 64430         | 14         | 0      | 515                                      | gi|6508853|6| -AF | 90.7 | (AC006564) putative nucleoside pyrophosphatase. A. thaliana |
| MAB16.14     | -         | 68944         | 8          | 1      | 145                                      | gi|4031818|6| -AG | 47.7 | (AC012854) hypothetical protein A. thaliana |

### MBM17 (52717 bp)

| Deduced Gene | Position  | No. of Exons  | No. of EST | Length | Information on the Most Similar Sequence | Sequence ID | Quality | Identity | Definition |
|--------------|-----------|---------------|------------|--------|------------------------------------------|-------------|--------|----------|------------|
| MBM17.1      | +         | 656           | 96         | 1      | 605                                      | gi|15681349|6| -AF | 66.9 | DNA topoisomerase III |
| MBM17.2      | +         | 9159          | 2          | 1      | 1103                                     | gi|2993773|6| -AG | 56.6 | (AC003236) putative receptor-like protein kinase A. thaliana |
| MBM17.3      | +         | 14377         | 10         | 1      | 701                                      | gi|5449087|6| -AG | 60.0 | (AC004329) Potentially protein A. thaliana |
| MBM17.4      | -         | 18966         | 30         | 0      | 1013                                     | gi|4816343|6| -AG | 34.6 | (AC005234) putative SNP/RAD54 family DNA repair and recombination protein A. thaliana |
| MBM17.5      | +         | 22778         | 10         | 0      | 1081                                     | gi|3913232|6| -AG | 81.3 | DNA polymerase-Delta catalytic chain |
| MBM17.6      | +         | 33115         | 10         | 0      | 387                                      | gi|9016704|6| -AF | 80.3 | (AC009925) unknown protein A. thaliana |
| MBM17.7      | +         | 35026         | 3           | 1      | 353                                      | gi|3913168|6| -AF | 68.0 | 3'A.1'-DNA-topoisomerase like: 3'2'-DNA-topoisomeraselike (3'Top2)|
| MBM17.8      | +         | 39195         | 8           | 0      | 312                                      | gi|2705667|6| -CA | 64.2 | (TA2015) 3'2'-DNA-topoisomerase like: A. thaliana |
| MBM17.9      | +         | 40639         | 8           | 0      | 347                                      | gi|2705667|6| -CA | 50.4 | (TA2015) 3'2'-DNA-topoisomerase like: A. thaliana |
| MBM17.10     | +         | 40699         | 2           | 1      | 396                                      | gi|3336254|6| -AF | 75.9 | (AC006094) unknown protein A. thaliana |
| MBM17.11     | -         | 41986         | 1           | 0      | 400                                      | gi|3327518|6| -AG | 44.8 | (partial) (AC009481) unknown protein A. thaliana |
| Protein | Description | Accession |
|---------|-------------|-----------|
| MGO3.3  | Deduced Gunw | |
Sequencing of Arabidopsis thaliana chromosome 5

**MIF21 (59372 bp)**

| Position | No. of Exon | Length | Information on the most similar sequence |
|----------|-------------|--------|------------------------------------------|
| MIF21.1  | 1           | 1115   | (U76219) DNA repair/transcription factor  |
| MIF21.9  | 1           | 1115   | (U76219) DNA repair/transcription factor  |

**MJB24 (58589 bp)**

| Position | No. of Exon | Length | Information on the most similar sequence |
|----------|-------------|--------|------------------------------------------|
| MJB24.1  | 1           | 19000  | (U76219) DNA repair/transcription factor  |
| MJB24.10 | 1           | 19000  | (U76219) DNA repair/transcription factor  |

**Gene**

- MIF21
- MJB24

**EST db hit**

- MIF21
- MJB24

**Protein db hit**

- MIF21: predicted protein phosphorylation factor A. thaliana
- MJB24: predicted protein phosphorylation factor A. thaliana
| Gene ID | Description | Length (bp) | Identity (%) | Definition |
|--------|-------------|-------------|--------------|------------|
| MJP23.1 | partial (AC001756) hypothetical protein A. thaliana | 323 | 45.4 | Hypothetical protein A. thaliana |
| MJP23.2 | partial (AC001756) hypothetical protein A. thaliana | 402 | 41.7 | Hypothetical protein A. thaliana |
| MJP23.3 | partial (AC001756) hypothetical protein A. thaliana | 131 | 60.6 | Hypothetical protein A. thaliana |
| MJP23.4 | partial (AC001756) hypothetical protein A. thaliana | 118 | 44.5 | Hypothetical protein A. thaliana |

**Gene Expression Analysis**

| Gene ID | EST Bit Hit | Protein DB Hit | Gene Expression |
|--------|-------------|----------------|-----------------|
| MJP23.1 | | | |
| MJP23.2 | | | |
| MJP23.3 | | | |
| MJP23.4 | | | |

**RNA-Seq Analysis**

| Gene ID | Protein DB Hit | Gene Expression |
|--------|----------------|-----------------|
| MJP23.1 | | |
| MJP23.2 | | |
| MJP23.3 | | |
| MJP23.4 | | |

**Transcriptome Analysis**

| Gene ID | Protein DB Hit | Gene Expression |
|--------|----------------|-----------------|
| MJP23.1 | | |
| MJP23.2 | | |
| MJP23.3 | | |
| MJP23.4 | | |

**Protein Domain Analysis**

| Gene ID | Protein DB Hit | Gene Expression |
|--------|----------------|-----------------|
| MJP23.1 | | |
| MJP23.2 | | |
| MJP23.3 | | |
| MJP23.4 | | |

**Phylogenetic Analysis**

| Gene ID | Protein DB Hit | Gene Expression |
|--------|----------------|-----------------|
| MJP23.1 | | |
| MJP23.2 | | |
| MJP23.3 | | |
| MJP23.4 | | |

**Functional Annotation**

| Gene ID | Protein DB Hit | Gene Expression |
|--------|----------------|-----------------|
| MJP23.1 | | |
| MJP23.2 | | |
| MJP23.3 | | |
| MJP23.4 | | |

**Conclusion**

The analysis of the transcriptome data from the MJP23 gene indicates that this gene is expressed in all analyzed samples. The gene is also present in the protein databases, suggesting its importance in the biological processes of A. thaliana. Further studies are needed to understand the exact function and regulation of this gene in the plant's development and stress responses.
### MQM1 (81365 bp)

| Identifier | Position | Exon | EST | Gene | Protein db hit | EST db hit | Gene db hit | Protein db hit |
|------------|----------|------|-----|------|---------------|-----------|-------------|---------------|
| MQM1.1     | 596      | 1    | 600 |      |               |           |             |               |
| MQM1.2     | 1095     | 1    | 1094|      |               |           |             |               |
| MQM1.3     | 5898     | 13   | 5644|      |               |           |             |               |
| MQM1.4     | 11812    | 13   | 11801|      |               |           |             |               |
| MQM1.5     | 11578    | 3    | 11573|      |               |           |             |               |
| MQM1.6     | 17802    | 12  | 17781|      |               |           |             |               |
| MQM1.7     | 20946    | 1    | 20945|      |               |           |             |               |
| MQM1.8     | 35619    | 6    | 35612|      |               |           |             |               |
| MQM1.9     | 41919    | 9    | 41910|      |               |           |             |               |
| MQM1.10    | 35813    | 9    | 35804|      |               |           |             |               |
| MQM1.11    | 41471    | 8    | 41462|      |               |           |             |               |

### MRG21 (55151 bp)

| Identifier | Position | Exon | EST | Gene | Protein db hit | EST db hit | Gene db hit | Protein db hit |
|------------|----------|------|-----|------|---------------|-----------|-------------|---------------|
| MRG21.1    | 34514    | 7    | 34511|      |               |           |             |               |
| MRG21.2    | 41686    | 3    | 41673|      |               |           |             |               |
| MRG21.3    | 47379    | 2    | 47367|      |               |           |             |               |
| MRG21.4    | 49882    | 6    | 49868|      |               |           |             |               |
| MRG21.5    | 50577    | 11   | 50562|      |               |           |             |               |
| MRG21.6    | 56446    | 5    | 56436|      |               |           |             |               |
| MRG21.7    | 58960    | 6    | 58952|      |               |           |             |               |
| MRG21.8    | 72879    | 5    | 72874|      |               |           |             |               |
| MRG21.9    | 79829    | 9    | 79820|      |               |           |             |               |
| MRG21.10   | 77688    | 9    | 77680|      |               |           |             |               |

---

**Gene**
- **Protein db hit**
- **EST db hit**
- **Gene**
- **EST db hit**
- **Protein db hit**
- **Gene db hit**
| Gene | EST Database | Genbank ID | Length (bp) |
|------|-------------|------------|-------------|
| MSK10.21 | GB:10457 | gi|45678901| 10457 |
| MSK10.17 | GB:25678 | gi|45678902| 12345 |
| MSK10.15 | GB:45678 | gi|45678903| 67890 |
| MSK10.13 | GB:25678 | gi|45678904| 56789 |
| MSK10.8 | GB:45678 | gi|45678905| 45678 |
| MSK10.C | GB:45678 | gi|45678906| 78901 |
| MSK10.5 | GB:45678 | gi|45678907| 89012 |
| MSK10.4 | GB:45678 | gi|45678908| 45678 |
| MSK10.3 | GB:45678 | gi|45678909| 12345 |
| MSK10-1 | GB:45678 | gi|45678910| 56789 |

Sequencing of Arabidopsis thaliana chromosome 5
No. 1

S. Sato et al.

MSN2 (62927 bp)

Grail exon
Protein db hit
EST db hit
Gene
Gene
EST db hit
Protein db hit
Grail exon

MUD12 (22601 bp)

Grail exon
Protein db hit
EST db hit
Gene
Gene
EST db hit
Protein db hit
Grail exon
Sequencing of Arabidopsis thaliana chromosome 5

**MUL3 (82010 bp)**

| Position | Sequence ID | No. of Genes | Length |
|----------|-------------|--------------|--------|
| 1        | MUL3.1      | 3            | 657    |
| 2        | MUL3.2      | 62           |        |
| 3        | MUL3.3      | 613          | 4      |
| 4        | MUL3.4      | 2514         | 32     |
| 5        | MUL3.5      | 2541         | 8      |
| 6        | MUL3.6      | 1259         | 32     |
| 7        | MUL3.7      | 1549         | 32     |
| 8        | MUL3.8      | 3841         | 8      |
| 9        | MUL3.9      | 4173         | 5      |

- **Overlap Identity Definition**
  - Deduced genes

**MWD22 (87180 bp)**

| Position | Sequence ID | No. of Genes | Length |
|----------|-------------|--------------|--------|
| 1        | MWD22.1     | 3            | 5732   |
| 2        | MWD22.2     | 6993         | 12     |
| 3        | MWD22.3     | 7123         | 1      |
| 4        | MWD22.4     | 12533        | 8      |
| 5        | MWD22.5     | 29699        | 72     |
| 6        | MWD22.6     | 24579        | 4      |
| 7        | MWD22.7     | 13209        | 8      |
| 8        | MWD22.8     | 34441        | 9      |
| 9        | MWD22.9     | 22717        | 5      |
| 10       | MWD22.10    | 29693        | 7      |
| 11       | MWD22.11    | 26431        | 7      |
| 12       | MWD22.12    | 17459        | 6      |
| 13       | MWD22.13    | 3133         | 9      |
| 14       | MWD22.14    | 13209        | 8      |
| 15       | MWD22.15    | 26431        | 7      |
| 16       | MWD22.16    | 29699        | 7      |
| 17       | MWD22.17    | 24579        | 4      |
| 18       | MWD22.18    | 17459        | 6      |
| 19       | MWD22.19    | 3133         | 9      |
| 20       | MWD22.20    | 13209        | 8      |
| 21       | MWD22.21    | 26431        | 7      |
| 22       | MWD22.22    | 29699        | 7      |
| 23       | MWD22.23    | 24579        | 4      |
| 24       | MWD22.24    | 17459        | 6      |
| 25       | MWD22.25    | 3133         | 9      |

- **Overlap Identity Definition**
  - Proteins
### MWP19 (11026 bp)

| Identifier | Position (bp) | Length (bp) | Information on the most similar sequence |
|------------|---------------|-------------|------------------------------------------|
| MWP19.1    | 1544          | 1515        | (AC014302) putative sensor protein in A. thaliana |
| MWP19.2    | 7984          | 412         | (AC013690) putative sensor protein in A. thaliana |

### MXK3 (8149 bp)

| Identifier | Position (bp) | Length (bp) | Information on the most similar sequence |
|------------|---------------|-------------|------------------------------------------|
| MXK3.1     | 434           | 195         | (AC007301) putative protein in A. thaliana |
| MXK3.2     | 1131          | 131         | (AC014302) putative sensor protein in A. thaliana |

### Exon Positions

| Position | Gene |
|----------|------|
| 7984     |      |
| 1244     |      |
| 31981    |      |
| 29088    |      |
| 27114    |      |
| 73035    |      |
| 71250    |      |
| 64998    |      |
| 60900    |      |
| 58285    |      |
| 56144    |      |
| 52832    |      |
| 51222    |      |

### EST and Protein Database Hits

- **EST hit**: (AC008298) putative protein in A. thaliana
- **Protein hit**: (AC009243) F28K19.24 A. thaliana

---

*Note: The above text and tables contain information on gene sequences, positions, lengths, and corresponding databases and identifiers.*
