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

Shusei SATO, Takakazu KANEKO, Hirokazu KOTANI, Yasukazu NAKAMURA, Erika ASAMIZU, Nobuyuki MIYAJIMA, and Satoshi TABATA*

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

(Received 26 January 1998)

**Abstract**

Nineteen P1 and TAC clones, which have been precisely localized to the fine physical map of *Arabidopsis thaliana* chromosome 5, were newly sequenced, and their sequence features were analysed. The total length of the clones sequenced was 1,456,315 bp. Together with the previously reported sequences, the regions of chromosome 5 that have been sequenced to date is now 5,310,105 bp. When the sequences determined in this study were subjected to similarity search against protein and expressed sequence tag (EST) databases and analysis with computer programs for gene modeling, a total of 354 potential protein-coding genes and/or gene segments were identified. The average density of the assigned genes and/or gene segments was one gene per 4,114 bp. Introns were identified in 75% of the potential protein genes, and the average number per gene and the average length of the introns were 3.7 and 194 bp, respectively. These sequence features are essentially identical to those in the previously reported sequences. The numbers of the *Arabidopsis* ESTs matched to each of the predicted genes have been counted to monitor the transcription level. The sequence data and gene information are available on the World Wide Web database KAOS (the Kazusa *Arabidopsis* data Opening Site) at http://www.kazusa.or.jp/arabi/.

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

To understand the whole genetic system in higher plants, we began large-scale structural analysis of the *Arabidopsis thaliana* genome, which is estimated to be approximately 130 Mb long, consisting of five chromosomes. As the initial phase of the project, we focused our effort on DNA sequencing of chromosome 5. The template clones were isolated by screening of the P1 and TAC (Transformation-competent Artificial Chromosome, Mitsui Plant Biotechnology Research Institute, Japan) libraries by means of polymerase chain reaction (PCR) using marker-specific primers. The isolated P1 and TAC clones were then precisely localized by PCR on the yeast artificial chromosome (YAC) contig map and our fine physical map of chromosome 5, prior to shotgun-based sequence analysis. We already reported the sequences of a total of 3.85 Mb which are covered by 50 P1 clones. In this paper, we newly determined the sequences of 19 additional P1 and TAC clones. Gene organization and structural and functional information of the likely genes in the sequenced regions, deduced by computer-aided analysis, are described.

1. **Isolation and Sequencing of P1 and TAC Clones**

DNA sources and the method of clone isolation were the same as described in the previous paper. Two types of genomic clones of *A. thaliana* Columbia, P1 and TAC, which are respectively represented by adding "M" and "K" to the first letters of the clone names, were used as the templates. The average insert length of the P1 and TAC clones was approximately 80 kb. The P1 and TAC clones containing the DNA regions which cover a total of 19 DNA markers on chromosome 5 were isolated by screening the Mitsui P1 and TAC libraries by PCR with the primers designed from the sequence information of DNA markers. The DNA markers and selected clones (in parentheses) are mi97 (MK20), MRO11_Left end (MZF18), CIC5D7L (K18P6), mi323 (K15E6), CIC8D12L (MUL8), CIC12H4R (MYH19), MPO12_Right end (MN13), CIC8D12R (MEE6), MZE20_Left end (MIO24), g4130 (MXC20), CIC6C5R (MCO15), CIC5C7R (MDF20),...
Assignment of the Potential Coding Regions

Assignment of potential protein coding regions and gene modeling were performed by similarity search and computer prediction, as described in the previous papers. In brief, similarity search against the non-redundant protein sequence database, owl (release 29), was carried out using the BLASTP program, and information obtained were integrated into the gene models constructed with the aid of following computer programs: Gene-Finder, GENSCAN and Net-PlantGene programs. The transcribed regions were assigned by comparison of the nucleotide sequences with Arabidopsis ESTs in the non-redundant library of GenBank (release 104) and EMBL (release 52) using the BLASTN program.

The potential protein-coding regions assigned were divided into three categories. A single exon or a region containing consecutive multiple exons showing similarity to a single reported gene throughout the alignment was assigned as a potential protein gene. They were denoted by numbers with the clone names followed by sequential numbers from one end to another of the insert. A region which matched only to portions of a reported gene and only to Arabidopsis ESTs were assigned as a potential exon(s) and a transcribed region, respectively. These regions were denoted by adding “p” and “t” between the clone names and the sequential numbers in the identifiers, respectively. All the genes and gene segments assigned in each P1 and TAC clone according to the above procedure are schematically represented in Fig. 2, and the assignment data are listed in the table below each figure. To sum up, 281 potential protein genes, 24 potential exons, and 49 transcribed regions were assigned in the 1,456,315 bp regions. The number of genes and gene segments assigned so far in the total of 5,310,105 bp, including the previously reported sequences, is 1,236, and an average density of the genes in the three categories is estimated to be one gene per 4,296 bp. However, it is possible that additional genes may be discovered in the future among the genes and gene segments assigned so far, because our prediction is principally based on similarity to the registered sequences.

In addition to the protein coding regions, RNA coding regions were assigned on the basis of sequence similarity to the reported structural RNAs, and of prediction by the tRNAscan-SE program in the case of tRNA genes. As indicated in Fig. 2, 9 tRNA genes corresponding to 7 amino acid species and 2 snRNA genes (U6-1 and Ula) were identified in the 1,456,315 bp regions. These genes are denoted with the clone names followed by “r” and sequential numbers.
3. Structural Features of the Potential Protein Genes

The complete structures of 281 potential protein genes were predicted in this study. Structural features of these genes as well as those of 744 genes including those previously identified are listed in Table 1. They amount to approximately 3.7% of the total gene constituents (20,000 genes) assumed for A. thaliana. Approximately 78% of the potential genes contained introns, and the average number per gene and their average length were 3.9 and 183 bp, respectively. The GC content of exons (44%) is significantly higher than that of introns (32%).

Table 1. Structural features of potential protein genes in A. thaliana chromosome 5

| Features                          | 281 genes | 744 genes |
|-----------------------------------|-----------|-----------|
| Gene length (bp) including introns| 191-10,099 (2,014) | 164-11,377 (2,029) |
| Product length (amino acids)      | 51-1,530 (134) | 51-1,837 (436) |
| Genes with introns                | 210       | 580       |
| Introns/gene                      | 0-23 (3.7) | 0-42 (3.9) |
| Exon length (bp)                  | 3-3,588 (280) | 2-4,026 (265) |
| Intron length (bp)                | 20-5,405 (194) | 20-5,405 (183) |
| GC content of exon                | 43%       | 44%       |
| GC content of introns             | 32%       | 32%       |

Structural features of the 281 potential protein genes assigned in this study and the 744 genes assigned so far are listed. Average values are shown in parentheses.

4. Expression Level of the Potential Protein Genes and Gene Segments

The nucleotide sequence of each of the potential protein genes and gene segments was compared with those in the Arabidopsis EST database, and the number of matched Arabidopsis ESTs was counted. Of the 354 genes and gene segments that we have identified on chromosome 5 in this study, 155 carried matched ESTs, and 604 out of 1,236 genes and gene segments identified so far matched ESTs, suggesting that the current EST database represents 48.9% of the gene complement in A. thaliana. The putative products of the genes hit by 10 or more EST files include those showing sequence similarity to pyruvate kinase in the plastid of Nicotiana tabacum (nxc20.10), 30s ribosomal protein in the chloroplast precursor in Spinacia oleracea (k18p6.1), and small Ras-like GTP-binding protein in A. thaliana (nxc15.11). These genes are suggested to be a class of highly expressed genes. 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/arabi/.

Acknowledgments: We thank S. Sasamoto and K. Naruo for excellent technical assistance, A. Tanaka for technical advice and the members of DNA Sequencing Laboratory: T. Kimura, T. Hosouchi, K. Kawashima, M. Matsunoto, A. Matsui, E. Mitsui, A. Muraki, N. Nakazaki, S. Okumura, S. Shinpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda and M. Yatabe for their excellent team work. Thanks are also due to the Mitsui Plant Biotechnology Research Institute, the Research Institute of Innovative Research for the Earth and Arabidopsis Biological Resource Center at Ohio State University for providing the DNA libraries and the DNA markers. This work was supported by the Kazusa DNA Research Institute Foundation. We thank M. Takanami for his support and encouragement to perform this project.

References

1. 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.
2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. Solovyev, V. V., Salamov, A. A., and Lawrence, C. B. 1994, Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames, *Nucl. Acids Res.*, **22**, 5156-5163.

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

11. Hebsgaard, S. M., Korning, P. G., Tolstrup, N., Engelbrecht, J., Rouze, P., and Brunak, S. 1996, Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information, *Nucl. Acids Res.*, **24**, 3439-3452.

12. Newman, T., Bruijn, F. J., 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.

13. 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**, 101-121.

14. 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 19 PI and TAC clones. Positions of the assigned or predicted genes and gene segments in each insert of the PI and TAC clones are schematically presented by color-coded boxes above (rightward) and below (leftward) the wide line in the middle which represents the entire sequenced region. The insert length is given in parenthesis together with the clone name at the top. Arrowheads indicate the directions of the DNA strands (5' to 3'). Dark and faint blue boxes with numbers represent the positions of the assigned potential protein genes and potential exons, respectively, and red bars represent the positions of structural RNA genes. Gray boxes with numbers indicate the positions of the transcribed regions. 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 BLASTP scores of 70-100, 100-250, and 250 or more, respectively. The green bars indicate the positions of the potential exons predicted by the Grail program. Each of the three different colors with increasing depth corresponds to the region with the Grail scores of less than 70, 70-90, and 90 or more, respectively. The potential protein genes, the gene segments and the potential RNA genes assigned as described in the text are listed below each of the figures. The accession numbers are as follows: AB009048 (K15E6), AB010068 (K18P6), AB010069 (MAC9), AB010070 (MBK20), AB009049 (MCD7), AB010071 (MCO15), AB009050 (MDF20), AB010072 (MFE6), AB010073 (MFB13), AB010074 (MIO24), AB009051 (MJH22), AB009052 (MNF13), AB010075 (MPA24), AB009053 (MQB2), AB010076 (MUL8), AB009055 (MUC20), AB010077 (MYH19) and AB009056 (MZF18).
### Sequencing of Arabidopsis Thaliana Chromosome 5

#### MCO15 (82918 bp)

**Potential protein genes**

| Location | Accession | Start (bp) | End (bp) | Length (bp) | Number | Coverage | Identity (of gene) | Definiton | Species |
|----------|-----------|------------|----------|-------------|--------|----------|-------------------|-----------|---------|
| 446-574  | X64838    | 469        | 829      | 360         | 2      | 90.1     | 100.0             | Arabidopsis thaliana | Protein db hit |
| 592-737  | AF120691  | 592        | 737      | 145         | 1      | 98.9     | 100.0             | Arabidopsis thaliana | EST db hit |
| 737-829  | AF076260  | 737        | 829      | 92          | 2      | 95.5     | 100.0             | Arabidopsis thaliana | EST db hit |
| 829-931  | AF076260  | 829        | 931      | 102         | 3      | 95.5     | 100.0             | Arabidopsis thaliana | EST db hit |

**TRANSCRIPTION ENGAGES**

| Location | Accession | Start (bp) | End (bp) | Length (bp) | Number | Coverage | Identity (of gene) | Definiton | Species |
|----------|-----------|------------|----------|-------------|--------|----------|-------------------|-----------|---------|
| 829-931  | AF076260  | 829        | 931      | 102         | 3      | 95.5     | 100.0             | Arabidopsis thaliana | Protein db hit |
| 737-829  | AF076260  | 737        | 829      | 92          | 2      | 95.5     | 100.0             | Arabidopsis thaliana | EST db hit |
| 592-737  | AF120691  | 592        | 737      | 145         | 1      | 98.9     | 100.0             | Arabidopsis thaliana | EST db hit |
| 446-574  | X64838    | 469        | 829      | 360         | 2      | 90.1     | 100.0             | Arabidopsis thaliana | Protein db hit |

#### MYH19 (77380 bp)

**Potential protein genes**

| Location | Accession | Start (bp) | End (bp) | Length (bp) | Number | Coverage | Identity (of gene) | Definiton | Species |
|----------|-----------|------------|----------|-------------|--------|----------|-------------------|-----------|---------|
| 829-931  | AF076260  | 829        | 931      | 102         | 3      | 95.5     | 100.0             | Arabidopsis thaliana | Protein db hit |
| 446-574  | X64838    | 469        | 829      | 360         | 2      | 90.1     | 100.0             | Arabidopsis thaliana | Protein db hit |
| 737-829  | AF076260  | 737        | 829      | 92          | 2      | 95.5     | 100.0             | Arabidopsis thaliana | EST db hit |
| 592-737  | AF120691  | 592        | 737      | 145         | 1      | 98.9     | 100.0             | Arabidopsis thaliana | EST db hit |
### MIO24 (86212 bp)

| Gene   | EST dbhit | Protein dbhit | EST dbhit | Gene | EST dbhit | Protein dbhit |
|--------|-----------|---------------|-----------|------|-----------|---------------|
|        |           |               |           |      |           |               |

#### Potential protein genes

| Gene   | EST dbhit | Protein dbhit | EST dbhit | Gene | EST dbhit | Protein dbhit |
|--------|-----------|---------------|-----------|------|-----------|---------------|
|        |           |               |           |      |           |               |

#### Potential RNA genes

| Gene   | EST dbhit | Protein dbhit | EST dbhit | Gene | EST dbhit | Protein dbhit |
|--------|-----------|---------------|-----------|------|-----------|---------------|
|        |           |               |           |      |           |               |

### MPA24 (84440 bp)

| Gene   | EST dbhit | Protein dbhit | EST dbhit | Gene | EST dbhit | Protein dbhit |
|--------|-----------|---------------|-----------|------|-----------|---------------|
|        |           |               |           |      |           |               |

#### Potential protein genes

| Gene   | EST dbhit | Protein dbhit | EST dbhit | Gene | EST dbhit | Protein dbhit |
|--------|-----------|---------------|-----------|------|-----------|---------------|
|        |           |               |           |      |           |               |

#### Potential RNA genes

| Gene   | EST dbhit | Protein dbhit | EST dbhit | Gene | EST dbhit | Protein dbhit |
|--------|-----------|---------------|-----------|------|-----------|---------------|
|        |           |               |           |      |           |               |
Sequencing of Arabidopsis Thaliana Chromosome 5

MEE6 (83698 bp)

| Identifier | Accession | Length | Overlap | Identity | Description |
|------------|-----------|--------|---------|----------|-------------|
| mst I      | AF007269  |        |         |          |             |
| mst II     | AC000132  |        |         |          |             |
| mst III    | S63758    |        |         |          |             |
| mst IV     | U53154    |        |         |          |             |
| mst V      | X95343    |        |         |          |             |
| mst VI     | S59392    |        |         |          |             |

Grail exon
Protein db hit
EST db hit
Gene

potential protein genes

MFB13 (80376 bp)

| Identifier | Accession | Length | Overlap | Identity | Description |
|------------|-----------|--------|---------|----------|-------------|
| mft 1      | P40631    |        |         |          |             |
| mft 2      | T50902    |        |         |          |             |
| mft 3      | Z71258    |        |         |          |             |
| mft 4      | ACO02337  |        |         |          |             |
| mft 5      | E22845    |        |         |          |             |
| mft 6      | Z73295    |        |         |          |             |
| mft 7      | 57322     |        |         |          |             |

Grail exon
Protein db hit
EST db hit
Gene

potential RNA genes
MBK20 (78172 bp)

| Exon | Gene | EST db hit | Protein db hit | EST db hit | Gene | EST db hit | Protein db hit | Gene | EST db hit | Protein db hit |
|------|------|------------|---------------|------------|------|------------|---------------|------|------------|---------------|
| 1    |      |            |               |            |      |            |               |      |            |               |
| 2    |      |            |               |            |      |            |               |      |            |               |
| 3    |      |            |               |            |      |            |               |      |            |               |

MUB3 (82188 bp)

| Exon | Gene | EST db hit | Protein db hit | EST db hit | Gene | EST db hit | Protein db hit | Gene | EST db hit | Protein db hit |
|------|------|------------|---------------|------------|------|------------|---------------|------|------------|---------------|
| 1    |      |            |               |            |      |            |               |      |            |               |
| 2    |      |            |               |            |      |            |               |      |            |               |
| 3    |      |            |               |            |      |            |               |      |            |               |
Sequencing of Arabidopsis Thaliana Chromosome 5

### Potential Proteins

| Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID |
|---------|----------|--------------------|-----------------|------------|---------|----------|--------------------|-----------------|------------|---------|----------|--------------------|----------------|------------|---------|----------|--------------------|----------------|------------|---------|
| K18P6   | 74589 bp  |                    |                 |            |         |          |                    |                 |            |         |          |                    |                |            |         |          |                    |                |            |         |
| MAC9    | 57246 bp  |                    |                 |            |         |          |                    |                 |            |         |          |                    |                |            |         |          |                    |                |            |         |

### Exons

| Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID |
|---------|----------|--------------------|-----------------|------------|---------|----------|--------------------|-----------------|------------|---------|----------|--------------------|----------------|------------|---------|----------|--------------------|----------------|------------|---------|
| K18P6   | 74589 bp  |                    |                 |            |         |          |                    |                 |            |         |          |                    |                |            |         |          |                    |                |            |         |
| MAC9    | 57246 bp  |                    |                 |            |         |          |                    |                 |            |         |          |                    |                |            |         |          |                    |                |            |         |

### Protein DB Hits

| Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID |
|---------|----------|--------------------|-----------------|------------|---------|----------|--------------------|-----------------|------------|---------|----------|--------------------|----------------|------------|---------|----------|--------------------|----------------|------------|---------|
| K18P6   | 74589 bp  |                    |                 |            |         |          |                    |                 |            |         |          |                    |                |            |         |          |                    |                |            |         |
| MAC9    | 57246 bp  |                    |                 |            |         |          |                    |                 |            |         |          |                    |                |            |         |          |                    |                |            |         |

### EST DB Hits

| Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID | Position | Number of EST hits | Number of Genes | Gene Name | Gene ID |
|---------|----------|--------------------|-----------------|------------|---------|----------|--------------------|-----------------|------------|---------|----------|--------------------|----------------|------------|---------|----------|--------------------|----------------|------------|---------|
| K18P6   | 74589 bp  |                    |                 |            |         |          |                    |                 |            |         |          |                    |                |            |         |          |                    |                |            |         |
| MAC9    | 57246 bp  |                    |                 |            |         |          |                    |                 |            |         |          |                    |                |            |         |          |                    |                |            |         |
**MDF20 (86699 bp)**

| Gene | EST db hit | Protein db hit | Grail exon |
|------|------------|----------------|------------|
| nufiRni | | | |
| rndf2 | | | |
| mdRN | | | |
| niJfl | | | |
| mdf20 | | | |

**Potential protein genes**

| Gene | No. of EST | Length (bp) | Protein db hit | EST db hit | Grail exon |
|------|------------|-------------|----------------|------------|------------|
| niJfl | | | | | |
| mdf20 | | | | | |

**Gene regions**

| Gene | No. of EST | Length (bp) | Protein db hit | EST db hit | Grail exon |
|------|------------|-------------|----------------|------------|------------|
| niJfl | | | | | |
| mdf20 | | | | | |

---

**MJH22 (27856 bp)**

| Gene | EST db hit | Protein db hit | Grail exon |
|------|------------|----------------|------------|
| hjil | | | |
| poten | | | |
| HNA | | | |

**Potential protein genes**

| Gene | No. of EST | Length (bp) | Protein db hit | EST db hit | Grail exon |
|------|------------|-------------|----------------|------------|------------|
| hjil | | | | | |
| poten | | | | | |
| HNA | | | | | |

**Gene regions**

| Gene | No. of EST | Length (bp) | Protein db hit | EST db hit | Grail exon |
|------|------------|-------------|----------------|------------|------------|
| hjil | | | | | |
| poten | | | | | |
| HNA | | | | | |
### Sequencing of *Arabidopsis Thaliana* Chromosome 5

**K15E6 (71736 bp)**

| Potential protein genes | Position | Length | Accession | Coding | Identity | Definition |
|-------------------------|----------|--------|-----------|--------|----------|------------|
| kis2-3                   | 1        | 3021   | K15E6     | 3      | 0.52     | 78         |
| kis2-4                   | 1        | 3955   | K15E6     | 4      | 0.56     | 74         |
| kis2-5                   | 1        | 20122  | K15E6     | 10     | 0.85     | 151        |
| kis2-6                   | 1        | 2956   | K15E6     | 11     | 1.00     | 128        |
| kis2-7                   | 1        | 28289  | K15E6     | 16     | 0.99     | 133        |

**Potential genes**

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

**Potential regions**

- **Accession**
- **Identity**
- **Definition**

**Potential RNA genes**

- **Accession**
- **Identity**
- **Definition**

**MCID7 (87665 bp)**

| Potential protein genes | Position | Length | Accession | Coding | Identity | Definition |
|-------------------------|----------|--------|-----------|--------|----------|------------|
| kis1-1                   | 1        | 33307  | MCD7     | 3      | 0.56     | 74         |

**Potential regions**

- **Accession**
- **Identity**
- **Definition**

**Potential RNA genes**

- **Accession**
- **Identity**
- **Definition**

---

*Note: The diagram and tables represent the sequencing data for the *Arabidopsis Thaliana* chromosome 5, including potential protein, RNA, and gene regions.*