A Resource of 5,814 Dissociation Transposon-tagged and Sequence-indexed Lines of Arabidopsis Transposed from Start Loci on Chromosome 5

Takuya Ito 1, Reiko Motohashi 2, Takashi Kuromori 2, Yoshiteru Noutoshi 1, Motoaki Seki 1,2, Asako Kamiya 2, Saho Mizukado 1, Tetsuya Sakurai 2,3 and Kazuo Shinozaki 1,2, *  

1 Plant Molecular Biology Laboratory, RIKEN Tsukuba Institute, 3-1-1 Koyadai, Tsukuba, 305-0074 Japan  
2 Plant Mutation Exploration Team, Plant Functional Genomics Research Group, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045 Japan  
3 Genomic Knowledge Base Research Team, Bioinformatics Group, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045 Japan

We report here the generation of an additional collection of Dissociation (Ds) transposon-tagged, sequence-indexed lines of Arabidopsis thaliana. Our RIKEN Ds insertion collection now totals 17,668 lines. Our collection has preferential insertions in chromosomes 1 and 5, because Ds was transposed from start loci on those chromosomes (11,854 and 5,814 lines, respectively). We describe here features of the latter 5,814 lines. The former 11,854 lines have been described previously. We have created a searchable database of the insertion sites and mutated genes (http://rarge.gsc.riken.jp/), and are depositing these lines in the RIKEN BioResource Center (http://www.brc.riken.go.jp/lab/epd/Eng/). Our collection of these mutants will contribute to progress in functional genomics of plants.

Keywords: Ac/Ds — Arabidopsis thaliana — Functional genomics.

Abbreviations: ABRC, Arabidopsis Biological Resource Center; Ac, Activator; Ds, Dissociation.

Arabidopsis thaliana is recognized as a model organism for research in plant biology. Determination of the whole genome sequence revealed that it contains approximately 26,000 genes (Arabidopsis Genome Initiative 2000). Functional genomics, i.e. understanding the functions of each of these genes, is the next challenge. One of the effective approaches for understanding the in vivo functions of these genes is reverse genetics. For this purpose, some groups, including us, have constructed large collections of sequence-indexed insertion mutants by using T-DNA or transposable elements (Parinov et al. 1999, Tissier et al. 1999, Ito et al. 2002, Sessions et al. 2002, Alonso et al. 2003, Kuromori et al. 2004). These collections enable us to create loss-of-function mutations for almost all genes.

We adopted a local transposition system based on a maize Activator (Ac) / Dissociation (Ds) transposon system (Fedoroff and Smith 1993). Using this system, we previously reported a collection of 1,173 Ds transposon-tagged and sequence-indexed Arabidopsis lines (Nossen ecotype) based on three start loci on chromosome 5 (Ito et al. 2002), and 11,854 lines based on five loci on chromosome 1 (Kuromori et al. 2004). Our Ds collection contained far fewer lines bred from transposition from start loci on chromosome 5 than from start loci on chromosome 1. To obtain more disrupted genes and more alleles of the same gene, we report here generation of an additional 4,641 lines from four start loci on chromosome 5.

The additional lines were prepared by outcrossing an Ac transposase donor line and Ds donor lines (Ito et al. 1999). This gave us a total of 5,814 lines bred by transposition from start loci on chromosome 5. The start positions are indicated in Table 1. Most of the transposed lines contained a single copy of the Ds element (data not shown). Plants with a single Ds copy are suitable for functional analysis. We determined partial genomic sequences flanking both 5′ and 3′ ends of the Ds elements by using an adaptor ligation method (Kuromori et al. 2004), and identified the insertion sites by comparing these sequences with the published whole genome sequence (Arabidopsis Genome Initiative 2000). Both genomic sequences were obtained from 79.1% of the lines, and either 5′- or 3′-flanking sequences were obtained from the rest of the lines. (For definition of the 5′ and 3′ ends, see the legend to Fig. 2B.) Including the collection of 1,173 already reported lines (Ito et al. 2002), 44.6% of the 5,814 lines contained Ds on chromosome 5 (Table 2). This collection shows preferential distribution on chromosome 5 because the Ds element tends to transpose near its start position (Fig. 1). In other lines, the Ds elements were intrachromosomally transposed to the other four chromosomes (Table 2 and Fig. 1).

We recognized characteristic transposition hot spots near the ends of every chromosome (arrows in Fig. 1). Such hot spots were also observed after transposition from start loci on
A resource of Ds-tagged Arabidopsis lines

Table 1  Physical map positions of the Ds–GUS–T-DNA constructs in the donor lines used in this study, and the number of transposed lines generated from each of the donor line

<table>
<thead>
<tr>
<th>Ds donor line</th>
<th>Line name</th>
<th>Start position</th>
<th>No. of constructed transposed lines from each Ds donor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ds388–30</td>
<td>Ds51</td>
<td>F12E4 (7378607)</td>
<td>45,681</td>
</tr>
<tr>
<td>Ds391–20</td>
<td>Ds52</td>
<td>MQM1 (10178221)</td>
<td>52,533</td>
</tr>
<tr>
<td>Ds392–13</td>
<td>Ds53</td>
<td>MGO3 (3869070)</td>
<td>36,740</td>
</tr>
<tr>
<td>Ds389–13</td>
<td>Ds54</td>
<td>K2A18 (2924651)</td>
<td>75,249</td>
</tr>
</tbody>
</table>

Total: 5,814 lines

Table 2  Distribution of the Ds insertions among five chromosomes transposed from four start loci on chromosome 5

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.0</td>
</tr>
<tr>
<td>2</td>
<td>13.7</td>
</tr>
<tr>
<td>3</td>
<td>14.3</td>
</tr>
<tr>
<td>4</td>
<td>14.5</td>
</tr>
<tr>
<td>5</td>
<td>44.6</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 3  Distribution of Ds insertions in predicted genes transposed from four start loci on chromosome 5

<table>
<thead>
<tr>
<th>Insertion</th>
<th>Line number</th>
<th>Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coding region a</td>
<td>2,878</td>
<td>49.5</td>
</tr>
<tr>
<td>5’-flanking region c</td>
<td>1,127</td>
<td>19.4</td>
</tr>
<tr>
<td>3’-flanking region d</td>
<td>294</td>
<td>5.1</td>
</tr>
<tr>
<td>Others</td>
<td>1,515</td>
<td>26.1</td>
</tr>
<tr>
<td>Total</td>
<td>5,814</td>
<td>100.0</td>
</tr>
</tbody>
</table>

*These donor lines are described by Smith et al. (1996). The four lines are aligned from top to bottom of chromosome 5. For positions, see Fig. 1. These lines are available from ABRC at Ohio State University, USA. Stock numbers are CS8512 (Ds388–30), CS8522 (Ds391–20), CS8525 (Ds392–13) and CS8518 (Ds389–13). The Ac transposase donor line used in this study is also available from ABRC; the stock number is CS8538.

a  Line name used in our collection.

b  Mapped and sequenced P1 or BAC clone in which Ds–GUS–T-DNA is inserted. Numbers on parentheses indicate DDBJ/EMBL/GenBank GI number of the P1/BAC clone sequence.

c  Coordinates of the P1/BAC sequence (DDBJ/EMBL/GenBank entry) flanking the left border of the Ds–GUS–T-DNA. Construction of the Ds–GUS–T-DNA is described by Smith et al. (1996).

d  Including already published 1,173 lines described by Ito et al. (2002).

e  Ds–GUS–T-DNA of this start line is inserted into a protein coding region (protein entry code At5g66210). We found that the homozygous line showed a semi-dwarf phenotype (data not shown). Therefore, pay attention to this background phenotype when using Ds insertion mutants derived from the Ds54 start line.

chromosome 1 (Kuromori et al. 2004). Our observation suggests that these regions may form a transposon-accessible chromatin structure, although, at the moment, a correlation between the insertion mechanism and the chromatin structure is not clear. On the other hand, the Ds element was inserted somewhat less frequently into pericentromeric regions (Fig. 1), where the density of active genes is lower and the density of transcriptionally silent transposons and pseudogenes is higher (Arabidopsis Genome Initiative 2000). These regions constitute heterochromatin structure (Haupt et al. 2001), suggesting that transposons are less accessible to these regions.

We next examined the distribution of Ds insertions in predicted genes (Table 3). Among the 5,814 lines, 49.5% of the Ds elements were inserted into known or predicted protein-coding regions, 19.4% within 500 bp upstream of the start codon of protein-coding regions, and 5.1% within 200 bp downstream of the stop codon. In total, 74.0% could have the potential to affect gene expression. Our analysis of these lines revealed mutations in 2,924 of the 26,000 predicted genes. The other 26.0% of the insertions were in intergenic regions, and probably would not affect gene expression. Our results indicate that about half of the Ds elements are inserted into protein-coding regions (Kuromori et al. 2004, this study). This ratio is higher than that of T-DNA elements. In T-DNA-tagged insertion mutants, <35% of the T-DNA elements were inserted into protein-coding regions, including introns (Sessions et al. 2002, Szabados et al. 2002, Alonso et al. 2003). These results suggest that transposable elements are more accessible to transcribed regions than are T-DNA elements.
In our previous report, 1,173 lines had contained mutations in 728 of the 26,000 predicted genes (Ito et al. 2002). Among these, two or more alleles had been identified for 59 genes, which corresponded to 8.1% of the disrupted genes. On the other hand, in this study, 5,814 lines, including the previous 1,173 lines, have caused mutations in 2,924 of the 26,000 genes. Among these, two or more alleles have been identified for 599 genes (20.5% of the disrupted genes), indicating that we could obtain more alleles for functional analysis.

Transposable elements are widely spread among many species, including bacteria, plants and animals. The insertion site specificities of some transposable elements have been determined. For example, bacterial Tn10 targets a particular 6 bp symmetrical sequence (GCTNAGC) (Halling and Kleckner 1982), and Tc1 of nematode recognizes dinucleotide TA (Mori et al. 1988). The Tourist family in plants also recognizes the target site consensus sequence TAA (Bureau and Wessier 1992). In contrast, the target site specificity of Ac/Ds family transposable elements has not been reported. We previously collected 2,799 perfect 8 bp target site sequences of the Ds element from the earlier 11,854 lines (Kuromori et al. 2004). By adding data obtained from the 5,814 lines reported herein, we now have 4,672 perfect target site sequences. By analysing huge numbers of these sample sequences, we can discuss the target site specificity of the Ac/Ds elements from a statistical viewpoint. Fig. 2 indicates nucleotide frequency in the 8 bp target site sequence. Among eight positions of nucleotide pairs, the frequencies of complementary nucleotides at positions 1–8, 2–7, 3–6 and 4–5 were similar, although the target sequence was not palindromic. We also observed more or less frequent nucleotides at some positions. For example, the frequency of a T at positions 2 and 3, and a complementary A at positions 7 and 6, was much higher than the average 32.5% calculated from the AT content (65%) in the whole genome. On the other hand, the frequency of an A at positions 2 and 3, and a complementary T at positions 7 and 6, was much lower than 32.5%. However, we could not recognize exclusive nucleotide frequencies at any position. From these results, we conclude that the Ac/Ds element has no target site specificity, although some biases are present.

Our RIKEN group has prepared a total of 17,668 Ds insertion lines. Among them, 11,854 lines were bred by transposition from start loci on chromosome 1, and about half of them have insertions on chromosome 1 (Kuromori et al. 2004). The other 5,814 lines were bred by transposition from chromosome 5, about half of them with insertions on chromosome 5 (Table 2.
A resource of Ds-tagged Arabidopsis lines

Therefore, our collection has a high rate of disruption of genes on chromosomes 1 and 5 (Table 4). We estimate that these 17,668 Ds insertion sites cause mutations in 7,338 of the 26,054 predicted genes, or 28.1% of all Arabidopsis genes. Among the 7,338 genes, 2,532 are on chromosome 1 and 1,748 are on chromosomes 5. In addition, two or more alleles have been identified for 2,338 genes. These multiple alleles will be useful for analyses of the gene function, since mutations are not always tagged by the Ds element (Ito et al. 2002). We have created a searchable database of the insertion sites and mutated genes (http://rarge.gsc.riken.jp/) for the benefit of Arabidopsis researchers.

We are depositing these lines in the RIKEN BioResource Center (http://www.brc.riken.go.jp/lab/epd/Eng/) for academic use. Our transposon insertion lines were generated from the Nossen ecotype (Fedoroff and Smith 1993). However, we have found that some genetic and physiological traits of the Ds donor lines differ from those of available Nossen plants (e.g. ABRC stock no. CS3081) (Noutoshi et al. unpublished data). Therefore, we recommend using the start lines (the Ds donor and the Ac transposase donor) as control lines for functional analysis.

Materials and Methods

The Ds-transposed lines were constructed as described by Ito et al. (1999) and Kuromori et al. (2004). High-throughput analysis of Ds-flanking sequences was carried out as described by Kuromori et al. (2004). Ds insertion sites were determined as described by Ito et al. (2002) and Kuromori et al. (2004).

Acknowledgments

We thank J. Ishida and M. Nakajima for sequencing, T. Hirayama and K. Akiyama for constructing the database, M. Kobayashi for providing us with laboratory space, F. Myouga for discussion, and S. Kawamura and H. Kobayashi for maintaining the greenhouse. This work was supported by a grant for Genome Research from RIKEN, and in part by the Special Coordination Fund of the Science and Technology Agency, and a grant-in-aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan to K.S.

References


A resource of Ds-tagged Arabidopsis lines


(Received January 13, 2005; Accepted April 15, 2005)