Analysis of Expressed Sequence Tags of Flower Buds in *Lotus japonicus*

Makoto Endo, Takao Kokubun, Yoshihito Takahata, Atsushi Higashitani, Satoshi Tabata, and Masao Watanabe

Laboratory of Plant Breeding, Faculty of Agriculture, Iwate University 3-18-8, Ueda, Morioka 020-8550, Japan, Soma Agricultural High School, 1-65, Mishima-machi, Haramachi 975-0012, Japan, Institute of Genetic Ecology, Tohoku University, 2-1-1, Katahira, Aoba-ku, Sendai 980-8577, Japan, and Kazusa DNA Research Institute, 1532-3, Yana, Kisarazu 292-0812, Japan

(Received 28 April 2000)

**Abstract**

In order to study gene expression in a reproductive organ, we constructed a cDNA library of mature flower buds in *Lotus japonicus*, and characterized expressed sequence tags (ESTs) of 842 clones randomly selected. The EST sequences were clustered into 718 non-redundant groups. From BLAST and FASTA search analyses of both protein and DNA databases, 58.5% of the EST groups showed significant sequence similarities to known genes. Several genes encoding these EST clones were identified as pollen-specific genes, such as pectin methylesterase, ascorbate oxidase, and polygalacturonase, and as homologous genes involved in pollen-pistil interaction. Comparison of these EST sequences with those derived from the whole plant of *L. japonicus*, revealed that 64.8% of EST sequences from the flower buds were not found in EST sequences of the whole plant. Taken together, the EST data from flower buds generated in this study is useful in dissecting gene expression in floral organ of *L. japonicus*.

**Key words:** Legume; *Lotus japonicus*; ESTs; flower bud; pollen-specific gene

The family of Fabaceae is highly divergent at the species level, and contains several important species that are used for food and forage. Additionally, in root hairs of legumes, symbiotic responses to *Rhizobium* and related bacteria for nitrogen fixation were observed. *Lotus japonicus* has emerged as a model legume, because it has good characteristics (small genome size, short generation time, and self-compatibility) for genome analysis. The size of flower of *L. japonicus* is larger than those of other model plant species, such as *Arabidopsis* and rice; thus, this characteristic may be useful in dissecting the gene expression of floral reproductive organs. Anther and pistil are highly differentiated sexual reproductive organs. To date, several male and female organ-specific genes have been identified and characterized in several other plants. Some of these specific genes are involved in cell-cell communication between male (pollen) and female (pistil) in plants.

Expressed sequence tags (ESTs) have been providing a large amount of information for gene identification during genome research. EST analysis in reproductive organ has already been performed in Chinese cabbage and liverwort. In *L. japonicus*, EST analyses have been performed with two cDNA libraries derived from vegetative organs and roots (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html), but not from floral organs.

In this report, we determined partial sequences of 842 randomly selected cDNA clones from flower buds of *L. japonicus*, identified the reproductive organ-specific genes by comparison with public databases, and estimated the floral organ-specific ESTs by comparison with the ESTs derived from whole plants at the vegetative growth stage.

1. Construction and Characterization of cDNA Library from Mature Flower Buds

Plants of *L. japonicus* B-129, Gifu were grown in a greenhouse, and flower buds of approximately 7-10 mm without sepal, which were just before the anther dehisced, were harvested. Total RNA was isolated from the collected flower buds as described previously. Isolation of poly(A)^+ RNA was performed from the total RNA using a FastTrack 2.0 mRNA isolation kit (Invitrogen, San Diego, CA, USA), and double-stranded cDNA was syn-
thesized using a AZAPII cDNA synthesis kit (Stratagene, La Jolla, CA, USA). The cDNA was ligated to AZAPII arms (Stratagene), and packaged in vitro using a Gigapack III extract (Stratagene). The phage library was converted to the plasmid form by mass excision according to the procedure described by Stratagene. A total of 898 clones containing inserts of more than 500 bp were selected from the library by using PCR with M13 universal primer set for sequence analysis. The insert sizes of the analyzed cDNA clones ranged from 0.5 to 2.5 kb, with the majority (63.8%) falling between 0.5 and 1.0 kb. The 5' ends of 898 clones were sequenced and registered to the public DNA databases (Accession nos. AU088751 to AU089648). The ESTs of low quality were excluded and the remaining 842 ESTs were subjected to further analyses. The average length of the EST sequences was 380 bp.

For evaluation of the quality of the library, we surveyed whether or not the translation initiation codon (ATG) was contained in each EST clone encoding for a ubiquitous protein. Among the 12 clones encoding proteins whose molecular size ranged from 100 to 300 amino acid residues (6 clones for ribosomal proteins, 2 clones for ATP synthase, and 4 clones for GTP-binding proteins), only 2 were found to lack the 5' translation initiation codon. In the case of 22 clones encoding proteins whose molecular size ranged from 300 to 500 amino acid residues (6 clones for actin, 3 clones for omega-6 desaturase, 2 clones for elongation factor, 6 clones for GDSL-motif lipase/acylhydrolase, and 5 clones for pyruvate kinase), 12 were found to contain the 5' translation initiation codon. The result indicated that the library constructed in this study has a high proportion of full-length cDNA clones encoding proteins up to 300 amino acid residues, at least.

In order to identify the number of independent EST species, clustering of the EST sequences was performed. The 5' end sequences were compared with a dataset of itself using the BLASTN program, and the sequences that showed over 95% identity for more than 100 bp were included in the same group. As a result, 718 non-redundant groups were generated. However, it is possible that the number of non-redundant groups is overestimated since they could be non-overlapping sequences derived from the same genes.

2. Identification of Gene Function by Database Search

In order to identify the putative function of 718 non-redundant EST groups, they were subjected to similarity search against the non-redundant protein database, nr, provided by NCBI using the BLAST algorithm. Similarity was considered to be significant when the optimized similarity BLAST score was higher than 100 at the amino acid level. When the optimized similarity score was lower than 100, the EST sequences were analyzed using the DDBJ database using FASTA algorithm at the nucleotide level. In both cases, we estimated the putative function of each EST clone from the highest-scored target gene.

Combining the data of BLAST and FASTA analyses, a total of 420 (58.5%) of the 718 non-redundant EST groups were found to be similar in sequence to genes registered in the databases. Two hundred sixteen EST groups (30.1%) showed similarity to protein and/or DNA sequences of unknown function. The remaining 82 EST groups (11.4%) showed no significant similarity to the sequences of registered genes, and were classified as novel sequences (Table 1).

The relative abundance of the mRNA would be approximately reflected in the abundance of its corresponding cDNA, because we analyzed a non-normalized library. Therefore, random sequencing of cDNA yields information about the relative expression levels of the genes represented by the ESTs. Out of 718 non-redundant groups, 84 contained at least two EST sequences. Groups that contained three or more ESTs are listed in Table 2.

We identified four potential pollen-specific EST groups. These EST groups showed similarity to genes encoding GBP169-12 of Petunia hybrida; NTP303 of Nicotiana tabacum, which was homologous to ascorbate oxidase; pectin methylesterase of Salix gilgiana; and polygalacturonase of Medicago sativa. It was suggested that these pollen-specific proteins were required for pollen tube growth and cell expansion because these proteins provide wall precursors for rapidly growing pollen tube.

The gene encoding pectin methylesterase, which was characterized as pollen-specific gene and is related to pollen tube growth, was tagged four times (Table 2). Multiple clones of other genes (expansin, extensin, β-galactosidase, β-glucanase, xylosidase), which were required for pollen tube growth and cell expansion, were also isolated.

Several EST clones whose genes could encode calcium-binding pollen allergen, receptor kinase, and pollen coat protein were identified. These genes are involved in

### Table 1. The results of similarity search against the public database. The number of EST groups and clones that showed similarity to genes of known function and to hypothetical genes that no definition of known function are given.

| Similarity                | Number of groups | Number of clones |
|---------------------------|------------------|-----------------|
| Genes of known function   | 420              | 500             |
| Hypothetical genes        | 216              | 252             |
| No similarity             | 82               | 90              |
| Total                     | 718              | 842             |
This suggests that the genes corresponding to

By guest

http://www.kazusa.or.jp/en/plant/lotus/EST/. The search results are provided through World Wide Web at

Table 2. Groups containing more than two ESTs.

| Group Number | Putative identification of ESTs |
|--------------|---------------------------------|
| 1            | 7 (unknown)                     |
| 2            | 5 (unknown)                     |
| 3            | 5 abscisic stress ripening protein |
| 4            | 5 metallothionein                |
| 5            | 4 (unknown)                     |
| 6            | 4 (unknown)                     |
| 7            | 4 (unknown)                     |
| 8            | 4 (unknown)                     |
| 9            | 4 argonaute protein             |
| 10           | 4 GAST1                         |
| 11           | 4 pectin methylesterase         |
| 12           | 3 (unknown)                     |
| 13           | 3 (unknown)                     |
| 14           | 3 60S ribosomal protein L5      |
| 15           | 3 extragenic suppressor         |
| 16           | 3 glyceraldehyde 3-phosphate dehydrogenase |
| 17           | 3 lipid transfer protein precursor |
| 18           | 3 methionine synthase           |
| 19           | 3 microsomal omega-6 desaturase |
| 20           | 3 mucin-2                       |
| 21           | 3 phosphoglycerate kinase 1     |
| 22           | 3 polyubiquitin                 |
| 23           | 3 surface protein               |

therefore, could be potential floral organ-specific genes, though there remains the possibility of overestimation of the number of floral organ-specific EST groups.

As described earlier, we have identified 82 novel EST groups. Of these, 76 EST groups (92.7%) were classified into floral organ-specific EST groups, which strongly suggests that a large number of novel genes were specifically expressed in flower bud in *L. japonicus*. The result indicates that the cDNA library generated from mature flower bud is a good source to discover reproduction-related genes. The ESTs obtained in this study will also be useful for analysis of the spatial and temporal expression patterns using DNA microarray technology.

Acknowledgements: This work was supported in part by a grant from the Ministry of Agriculture, Forestry and Fisheries of Japan (Rice Genome Project MA-2211), and by the Joint Research Program of the Institute of Genetic Ecology, Tohoku University. The authors are grateful to Masayoshi Kawaguchi, University of Tokyo, for providing seeds of *L. japonicus* B-129 Gifu. We also thank Manabu Yamada (Kazusa DNA Research Institute) and Akiko Ito (Iwate University) for their excellent technical assistance with analysis of ESTs data, and Erika Asamizu (Kazusa DNA Research Institute), Tohru Tsuchiya (Mie University), Yukihiro Ito (National Institute of Genetics), and Katsuyuki Yamato (Kyoto University), for their many helpful discussions.

References

1. Brewin, N. J. 1991, Development of the legume root nodule, *Annu. Rev. Cell Biol.*, 7, 191–226.
2. Handberg, K. and Stougaard, J. 1992, *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics, *Plant J.*, 2, 487–496.
3. McCormick, S. 1993, Male gametophyte development, *Plant Cell*, 5, 1265–1275.
4. Tsuchiya, T., Toriyama, K., Nasrallah, M. E., and Ejiri, S. 1992, Isolation of genes abundantly expressed in rice anthers at the microspore stage, *Plant Mol. Biol.*, 20, 1189–1193.
5. Taylor, L. P. and Hepler, P. K. 1997, Pollen germination and tube growth, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 48, 461–491.
6. Doughty, J., Dixon, S., Hiscock, S. J., Willis, A. C., Parkin, I. A. P., and Dickinson, H. G. 1998, PGP-A1, a defensin-like *Brassica* pollen coat protein that binds the *S* locus glycoprotein, is the product of gametophytic gene expression, *Plant Cell*, 10, 1333–1347.
7. Suzuki, G., Kai, N., Hirose, T. et al. 1999, Genomic organization of the *S* locus: identification and characterization of genes in SLC/SLR region of *S* haplotype of *Brassica campestris* (syn. *Brassica*), *Genetics*, 153, 391–400.
8. McCubbin, A. G. and Kao, T.-h. 1999, The emerging complexity of self-incompatibility of self-incompatibility (*S*) loci, *Sex. Plant Reprod.*, 12, 1–5.
9. Watanabe, M., Suzuki, G., Hatakeyama, K., Isogai, A., and Hinata, K. 1999, Molecular biology of self-
incompatibility in *Brassica* species, *Plant Biotech.*, 16, 263-272.

10. Okada, T., Zhang, Z., Russell, S. D., and Toriyama, K. 1999, Localization of the Ca$^{2+}$-binding protein, *Bra r 1*, in anthers and pollen tubes, *Plant Cell Physiol.*, 40, 1243-1252.

11. Takayama, S., Shibata, H., Iwano, M. et al. 2000, The pollen determinant of self-incompatibility in *Brassica campestris*, *Proc. Natl. Acad. Sci. USA*, 97, 1920-1925.

10. Okada, T., Zhang, Z., Russell, S. D., and Toriyama, K. 1999, Localization of the Ca$^{2+}$-binding protein, *Bra r 1*, in anthers and pollen tubes, *Plant Cell Physiol.*, 40, 1243-1252.

11. Takayama, S., Shibata, H., Iwano, M. et al. 2000, The pollen determinant of self-incompatibility in *Brassica campestris*, *Proc. Natl. Acad. Sci. USA*, 97, 1920-1925.

12. Luu, D.-T., Heizmann, P, Dumas, C, Trick, M., and Cappadocia, M. 1997, Involvement of *SLR1* genes in pollen adhesion to the stigmatic surface in Brassicaceae, *Sex. Plant Reprod.*, 10, 227-235.

13. Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A., and Hinata, K., 2000, The S receptor kinase determines self-incompatibility in *Brassica* stigma, *Nature*, 403, 913-916.

14. Cheung, A. Y., Wang, H., and Wu, H.-M. 1995, A floral transmitting tissue-specific glycoprotein attracts pollen tubes and stimulates their growth, *Cell*, 82, 383-393.

15. Uchimiya, H., Kidou, S. I., Shimazaki, T. et al. 1992, Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.), *Plant J.*, 2, 1005-1009.

16. Delseny, M., Cooke, R., Raynal, M., and Grellet, F. 1997, The *Arabidopsis thaliana* cDNA sequencing project, *FEBS Lett.*, 403, 221-224.

17. Lim, C. O., Kim, H. Y., Kim, M. G. et al. 1996, Expressed sequence tags of Chinese cabbage flower bud cDNA, *Plant Physiol.*, 111, 577-588.

18. Covita, P. A., Smith, L. S., and Long, S. R. 1998, Expressed sequence tags from a root-hair-enriched *Medicago truncatula* cDNA library, *Plant Physiol.*, 117, 1325-1332.

19. Nagai, J., Yamato, K. T., Sakaïda, M., Yoda, H., Fukuzawa, H., and Ohyama, K. 1999, Expressed sequence tags from immature female sexual organ of a liverwort, *Marchantia polymorpha*, *DNA Res.*, 6, 1-11.

20. Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S. 2000, Generation of 7137 non-redundant expressed sequence tags from legume, *Lotus japonicus*, *DNA Res.*, 7, 127-130.

21. Chang, S., Puryear, J., and Cairney, J. 1993, A simple and efficient method for isolation RNA from pine trees, *Plant Mol. Biol. Rep.*, 11, 113-116.

22. 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.

23. Pearson, W. R. and Lipman, D. J. 1988, Improved tools for biological sequence comparison, *Proc. Natl. Acad. Sci. USA*, 85, 2444-2448.

24. Brown, S. M. and Crouch, M. L. 1990, Characterization of a gene family abundantly expressed in *Oenothera organensis* pollen that shows sequence similarity to polygalacturonase, *Plant Cell*, 2, 263-274.

25. Lin, L. S. and Varner, J. E. 1991, Expression of ascorbate oxidase in zucchini squash (*Cucurbita pepo* L.), *Plant Physiol.*, 96, 159-165.

26. Wakeley, P. R., Rogers, H. J., Rozycka, M., Greenland, A. J., and Hussey, P. J. 1998, A maize pectin methylesterase-like gene, *ZmC5*, specifically expressed in pollen, *Plant Mol. Biol.*, 37, 187-192.

27. Futamura, N., Mori, H., Kouchi, H., and Shinohara, K. 2000, Male flower-specific expression of genes for polygalacturonase, pectin methylesterase and $\beta$-1, 3-glucanase in a dioecious willow, *Plant Cell Physiol.*, 41, 16-26.

28. Cosgrove, D. J. 1998, Cell wall loosening by expansins, *Plant Physiol.*, 118, 333-339.

29. Ruiter, R. K., van Eldik, G. J., van Herpen, M. M. A., Schrauwen, J. A. M., and Wullems, G. J. 1999, Hydration-dependent gene expression in *Brassica oleracea* anthers, *Sex. Plant Reprod.*, 12, 135-143.

30. Toriyama, K., Okada, T., Watanabe, M. et al. 1995, A cDNA clone encoding an IgE-binding protein from *Brassica* anther has significant sequence similarity to Ca$^{2+}$-binding proteins, *Plant Mol. Biol.*, 29, 1157-1165.

31. Kehoe, D. M., Villand, P., and Somerville, S. 1999, DNA microarrays for studies of higher plants and other photosynthetic organisms, *Trends Plant Sci.*, 4, 38-41.