

## Gene Note

The 5'-end expressed sequence tags of *Lotus japonicus*Mika Nomura<sup>1,\*</sup>, Dao Van Tan<sup>1</sup>, Hatthaya Arunothayanan<sup>1</sup>, Erika Asamizu<sup>2</sup>, Satoshi Tabata<sup>2</sup>, Hiroshi Kouchi<sup>3</sup>, Shigeyuki Tajima<sup>1</sup><sup>1</sup> Department of Life Science, Faculty of Agriculture, Kagawa University, Miki-cho, Kita-gun, Kagawa 761-0795, Japan;<sup>2</sup> Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan; <sup>3</sup> National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8602, Japan

\*E-mail: nomura@ag.kagawa-u.ac.jp Tel: +81-87-891-3135 Fax: +81-87-891-3021

Received December 10, 2007; accepted January 10, 2008 (Edited by D. Shibata)

**Abstract** For a global analysis of gene expression during nodule tissue formation in a model legume, *Lotus japonicus*, a cDNA array was constructed using 18,144 non-redundant clone set deduced from 3'-end expressed sequence tags (EST, Kouchi et al. 2004). In the current study, to obtain peptide coding information from 5'-ends of clone, we generated 7,320 5' EST sequences. BLAST search against a protein database indicated that 65.5% of the ESTs showed significant similarity to registered sequences. We identified numerous of nodule-enhanced genes, including many involved in diverse cellular processes, suggesting that the EST resources generated in this study will be a useful tool for the identification of genes related to nodulation and symbiotic nitrogen fixation.

**Key words:** ESTs, Legume, *Lotus japonicus*, Nitrogen fixation.

Nitrogen is an essential mineral nutrient for plant growth. Most plant species cannot utilize atmospheric nitrogen directly and thus require nitrogenous compounds, such as nitrate and ammonium (Crawford 1995). One important attribute of legumes, including *Lotus japonicus*, is their ability to form a stable endosymbiotic tissue, the nodule, with soil bacteria, called rhizobia, that use nitrogenase to synthesize ammonia from atmospheric nitrogen and pass it on to the host plant.

Formation of this symbiotic tissue begins with the perception of secreted rhizobial compounds, Nod (nodulation) factors. *Lotus japonicus* has been proposed as a model legume for molecular genetic studies of the complex steps after Nod factor perception in symbiotic nitrogen fixation (Handberg and Stougaard 1992). Many symbiotic mutant lines of *L. japonicus* have been established for elucidating the pathway of nodule formation (Kawaguchi et al. 2002; Schauser et al. 1998; Szczyglowski et al. 1998), and the resources required for molecular genetics, such as expressed sequence tags (ESTs), cDNA arrays and high-density genetic maps, are being prepared (Asamizu et al. 2004, 2000; Hayashi et al. 2001; Kouchi et al. 2004).

In addition to the morphological and functional differentiation corresponding to nodule tissue formation, the acquisition of nitrogen through endosymbiosis with soil bacteria results in further diverse biochemical and

physiological changes in both the host plant and the rhizobia. For the analysis of such complex phenomena, ESTs provide a large amount of information for identifying sequences discovered during genomic research. This approach is useful for understanding gene expression patterns at particular developmental stages of specific tissue types.

Here, we present 5' ESTs from nodule primordia, mature nodules, pods, and roots of *L. japonicus*. The source of the ESTs from *L. japonicus* were normalized and size- selected cDNA libraries of uninfected roots, and developing pods of *L. japonicus* MG20 Miyakojima, nodulated roots of the hyper-nodulating mutant *har1* (Nishimura et al. 2002) 4 days after bacterial infection, and mature nodules (23 days old) of B-129 Gifu. All of the cDNA libraries were sequenced from the 3'-ends, generating a total of 67,916 ESTs (Asamizu et al. 2004). Non-redundant ESTs were selected through sequence comparisons, and 18,144 ESTs were chosen for cDNA array construction (Kouchi et al. 2004). Given that cDNAs from MG-20 seedlings (2 weeks after germination) had already been sequenced from the 5'-end (Asamizu et al. 2000), we selected 9,299 cDNA clones that were not sequenced previously. The EST reads were quality trimmed by the Phred score, at a position where 5 ambiguous bases (Phred score <16) were found within consecutive 15 bases. We obtained 7,320 reads satisfying

Abbreviations: EST, expressed sequence tag

This article can be found at <http://www.jspcmb.jp/>

the condition and used these sequence information for characterization of the Lotus 5' ESTs (Table 1). Each cDNA sequence was translated into its amino acid sequence in six frames and used to search against the UniProt Reference Clusters database, Uniref100 (The UniProt Consortium, 2007), by use of the BLASTX program (Altschul et al. 1990). Similarity between a deduced amino acid sequence and a known sequence was judged as significant when the *E*-value was less than  $1.0^{-10}$ . As a result, 4797 (65.5%) showed significant similarity to known protein sequences. The search result of individual clone is available in our EST Index database (<http://est.kazusa.or.jp/en/plant/lotus/EST/>). In order to obtain information of functional categories, the Lotus 5' ESTs were similarity searched against the Arabidopsis protein sequence dataset provided from the Munich Information Center for Protein Sequencing (MIPS) (<http://mips.gsf.de/proj/thal/db/index.html>). We classified 4797 genes significantly conserved between *L. japonicus* and *A. thaliana* into functional categories according to Gene Ontology (GO), provided by The

Table 1. cDNA libraries for expressed sequence tags (ESTs) in *Lotus japonicus*.

Source	Library	5'-EST
MG-20 pods	N <sup>a</sup> (MPD)	1487
	L <sup>b</sup> (MPDL)	1104
MG-20 roots	N (MR)	1231
	L (MRL)	267
Gifu nodule primordia (har1; 4d after inoculation)	N (GN)	1474
	L (GNL)	120
Gifu mature nodules (23 d)	N (GEN)	801
	L (GENL)	836
Total		7320

<sup>a</sup> N, normalized library; <sup>b</sup> L, size selected library.

Arabidopsis Information Resource (TAIR, <http://www.arabidopsis.org/>) (Figure 1). The EST sequences reported here can be found in the GenBank / EMBL / DDBJ database under accession numbers DC592852 through DC600171.

Because the cDNA inserts in libraries can be less than full length or represent the fusion of multiple sequences, knowing the sequence of cDNA clones from both the 3'- and 5'ends is very helpful. Further, ESTs (including those we obtained in the present study), cDNA arrays and clustering of expression patterns, the availability of molecular markers, mutant strains, and the mapping of robust quantitative trait loci all provide valuable tools for assessing the function of many genes in plant-rhizobium interactions.

### Acknowledgements

We thank Dr. Peter Lutes for critical reading the document. We also thank Ms. M. Kohara for excellent technical assistance. This work was supported by the Special Coordination Funds for Promoting Science and Technology from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

### References

- Asamizu E, Nakamura Y, Sato S, Tabata S (2004) Characteristics of the *Lotus japonicus* gene repertoire deduced from large-scale expressed sequence tag (EST) analysis. *Plant Mol Biol* 54: 405–414
- Asamizu E, Nakamura Y, Sato S, Tabata S (2000) Generation of 7137 non-redundant expressed sequence tags from *Lotus japonicus*. *DNA Res* 7: 127–130
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410
- Crawford NM (1995) Nitrate: nutrient and signal for plant growth.

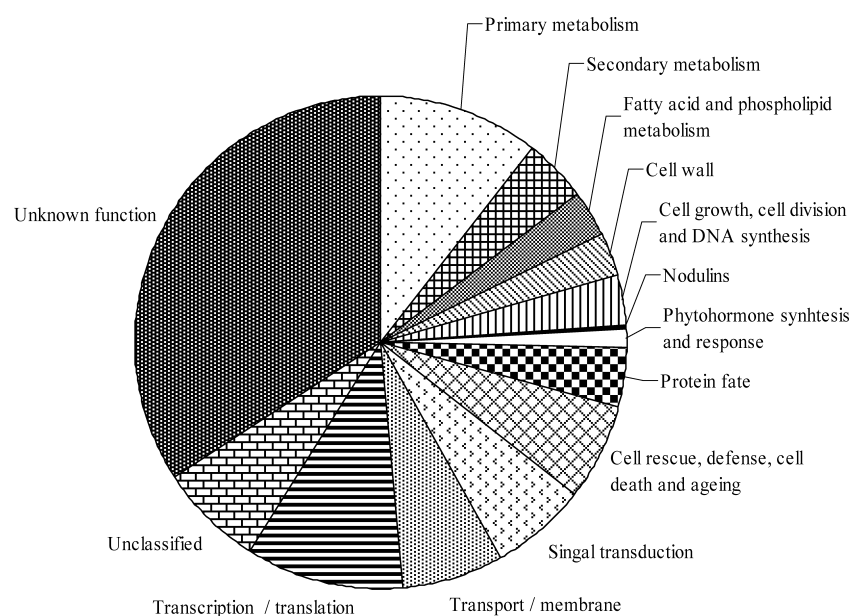


Figure 1. Chart of ESTs categorized by function. A total of 7320 ESTs from various libraries are included.

- Plant Cell* 7: 859–868
- Handberg K, Stougaard J (1992) *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *Plant J* 2: 487–496
- Hayashi M, Miyahara M, Sato S, Kato T, Yoshikawa M, Taketa M, Hayashi M, Pedrosa M, Onda R, Imaizumi-Anraku H, Bachmair A, Sanda N, Stougaard J, Murooka Y, Tabata S, Kawasaki S, Kawaguchi M, Harada K (2001) Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F<sub>2</sub> population. *DNA Res* 8: 301–310
- Kawaguchi M, Imaizumi-Anraku H, Koiwa H, Niwa S, Ikuta A, Syono K, Akao S. (2002) Root, root hair, and symbiotic mutants of the model legume *Lotus japonicus*. *Mol Plant Microbe Interac* 15: 17–26
- Kouchi H, Shimomura K, Hata S, Hirota A, Wu GJ, Kumagai H, Tajima S, Suganuma N, Suzuki A, Aoki T, Hayashi M, Yokoyama T, Ohyama T, Asamizu E, Kuwata C, Shibata D, Tabata S (2004) Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, *Lotus japonicus*. *DNA Res* 411: 263–274
- Nishimura R, Hayashi M, Wu GJ, Kouchi H, Imaizumi-Anraku H, Murakami Y, Kawasaki S, Akao S, Ohmori M, Nagasawa M, Harada K, Kawaguchi M (2002) HAR1 mediates systemic regulation of symbiotic organ development. *Nature* 420: 426–429
- Schauser L, Handberg K, Sandal N, Stiller J, Thykjær T, Pajuelo E, Nielsen A, Stougaard J (1998) Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of *Lotus japonicus*. *Mol Gen Geneti* 259: 414–423
- Szczyglowski K, Shaw RS, Wopereis J, Copeland S, Hamburger D, Kasiborski B, Dazzo FB, de Bruijn FJ (1998) Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. *Mol Plant Microbe Interac* 11: 684–697
- The UniProt Consortium (2007) The universal protein resource (UniProt). *Nucleic Acids Res* 35: D193–197