A.J. Márquez (Editorial Director). 2005. *Lotus japonicus* Handbook. pp. 135-138.
http://www.springer.com/life+sci/plant+sciences/book/978-1-4020-3734-4

Chapter 4.1

GENOME STRUCTURE ANALYSIS

Satoshi Tabata

Kazusa DNA Research Institute, Department of Plant Gene Research, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, JAPAN

Email: tabata@kazusa.or.jp

Phone: +81 438 52 3933   Fax: +81 438 52 3934

Keywords: *Agrobacterium rhizogenes*, *Rhizobium*, nodulation

A large-scale structural analysis of the genome of Lotus japonicus *is in progress. This project involves the interfacing of three approaches: EST (Expressed sequence tags) analysis, genome sequencing, and generation of DNA markers. Each approach is described here.*

EST ANALYSIS

EST analysis is performed to make a catalogue of expressed portions of the genome by developing anonymous partial cDNA sequences. So far, 110,000 Lotus ESTs have been deposited in the public EST database. These include ESTs from roots, nodule primordial, immature nodules, mature nodules, flower buds, pods, and whole plants (Asamizu et al. 2000, Endo et al. 2000, Asamizu et al. 2004). The sequence information and search results are available at the web site at http://www.kazusa.or.jp/en/plant/lotus/EST/. These ESTs as well as the corresponding cDNA clones are serving as valuable information and material resources for the functional analysis of individual genes and of the genome. The EST clones corresponding to ESTs can be requested at www.kazusa.or.jp/clonereq/.

GENOME SEQUENCING

Genome sequences, which complement the ESTs, provide a whole picture of the genetic information carried by living organisms. To understand the genetic systems carried by legume plants, large scale genome sequencing of *L. japonicus* accession Miyakojima MG-20 is in progress.

Genomic libraries of *L. japonicus* were constructed as a source for sequencing using a transformation competent artificial chromosome (TAC) and a standard BAC

A.J. Márquez (Editorial Director). 2005. *Lotus japonicus* Handbook. pp. 135-138.
http://www.springer.com/life+sci/plant+sciences/book/978-1-4020-3734-4

pBeloBAC11 as vectors. Genome sequencing is performed from the multiple seed points of the genome, which are selected based on the EST information (Sato et al 2001). TAC clones are selected by screening the three-dimensional DNA pools of the TAC genomic libraries by PCR with oligonucleotide primers designed based on nucleotide sequences of the public ESTs and cDNAs from *L. japonicus*. The nucleotide sequence of each TAC insert is determined according to the shotgun method. Assignment of the protein-encoding regions and gene annotation were performed by combination of similarity searches and computer predictions. As of February 28, 2004, a total of 1,568 seed clones have been selected, and 320 clones covering 32,537,698 bp genomic regions have been sequenced, annotated, and released to public databases (Kato et al 2003).

The structural features of the 3,136 potential protein-encoding genes in *L. japonicus* were compared to those of *Arabidopsis thaliana* (Table 1) (Kato et al 2003). The genes of two plants share several common features: average length of the coding exons (296 bp vs 256 bp) and the average number of introns per gene (3.6 vs 4.0). However, the average length of genes including introns, (2,706 bp), was longer in Lotus than in Arabidopsis (1,918 bp), which is due to a longer average intron length in Lotus (378 bp vs 157 bp). The average gene density in Lotus was one gene in every 10.4 kb, half the density of Arabidopsis (4.5 kb).

	Lotus	Arabidopsis
Number of assigned genes	3,136[a]	6,451[b]
Gene length (bp) including introns	165-21,418 (2,706)	78-17,203 (1,918)
Product length (amino acids)	16-2,185 (450)	25-4,706 (427)
Genes with introns	2283 (73%)	4,906 (76%)
Number of introns/gene	0-37 (3.6)	0-48 (4.0)
Exon length (bp)	2-5,604 (296)	2-5,966 (256)
Intron length (bp)	30-8,862 (378)	23-2,989 (157)
GC content of exons	45%	44%
GC content of introns	33%	32%

Table 1. Structural Features of potential protein-encoding genes in L. japonicus *and* A. thaliana*. The 3,136 genes assigned in* L. japonicus[a] *and the 6,451 genes previously assigned in our* A. thaliana *genome sequencing project*[b] *were compared. Average values are shown in parentheses.*

LINKAGE MAPPING

The sequenced TAC clones are mapped onto a genetic linkage map by sequence repeat length polymorphism (SSLP) markers generated utilizing the sequence information of each clone. In the Lotus genome, Simple sequence repeats such as (AT)n, (GT)n, (AAT)n of equal to or longer than 15 bp occurred every 12 kb on average (Sato et al 2001). Genetic mapping of the generated DNA markers was performed using the 127 F2 mapping population derived from a cross between two Lotus accessions, Miyakojima MG-20 and Gifu B-129 (Figure 1). At the time of

A.J. Márquez (Editorial Director). 2005. *Lotus japonicus* Handbook. pp. 135-138.
http://www.springer.com/life+sci/plant+sciences/book/978-1-4020-3734-4

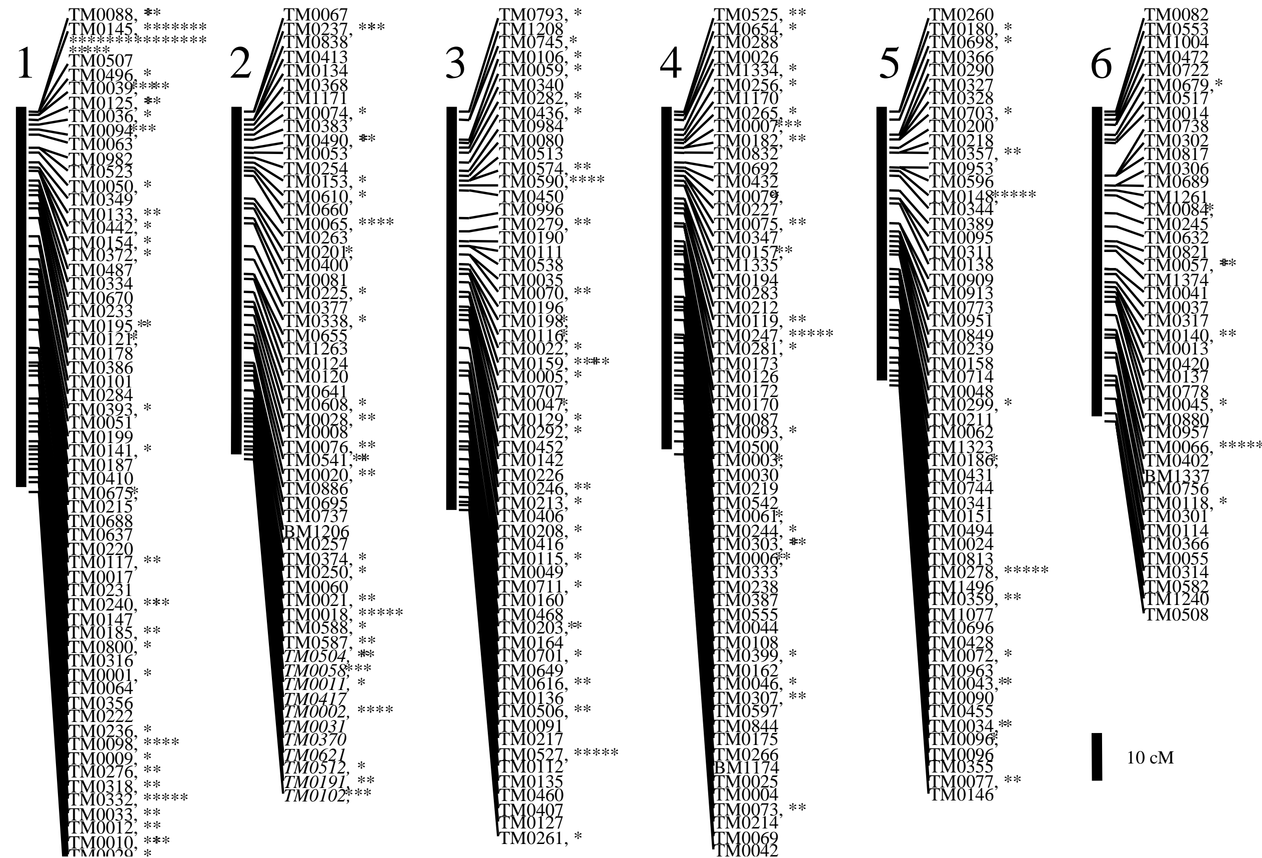

Figure 1. Relative positions of the SSR and dCAPS markers on the genetic linkage map. Asterisks indicate redundant markers.

A.J. Márquez (Editorial Director). 2005. *Lotus japonicus* Handbook. pp. 135-138.
http://www.springer.com/life+sci/plant+sciences/book/978-1-4020-3734-4

writing, a total of 714 DNA markers including SSLP and some dCAPS (derived cleaved amplified polymorphic sequence) markers have been mapped onto the linkage map. These PCR-based markers, almost all of which are co-dominant markers, with the surrounding DNA sequences will accelerate the process map-based cloning in Lotus. The marker information is available through our web database at www.kazusa.or.jp/lotus/.

REFERENCES

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

Asamizu E, Nakamura Y, Sato S, and Tabata S. (2004) **Characteristics of *Lotus japonicus* Gene Repertoire Deduced from Large-scale Expressed Sequence Tags (EST) Analysis**. *Plant Molecular Biology* 54, 405-414.

Endo M, Kokubun T, Takahata Y, Higashitani A, Tabata S, and Watanabe M. (2000) **Analysis of Expressed Sequence Tags of Flower Buds in *Lotus japonicus***. *DNA Research* 7,213-216.

Sato S, Kaneko T, Nakamura Y, Asamizu E, Kato T, and Tabata S. (2001) **Structural analysis of a *Lotus japonicus* genome. I: Sequence features and mapping of fifty-six TAC clones which cover the 5.4 Mb regions of the genome**. *DNA Research* 8, 311-318.

Kato T, Sato S, Nakamura Y, Kaneko T, Asamizu E, and Tabata S. (2003) **Structural analysis of a *Lotus japonicus* genome. V: Sequence features and mapping of sixty-four TAC clones which cover the 6.4 Mb regions of the genome**. *DNA Research* 10, 277-285.