Comparison of the Transcript Profiles from the Root and the Nodulating Root of the Model Legume *Lotus japonicus* by Serial Analysis of Gene Expression

Erika Asamizu, Yasukazu Nakamura, Shusei Sato, and Satoshi Tabata

Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan

Submitted 7 July 2004. Accepted 6 January 2005.

We performed a comprehensive transcript analysis on the early stage of root nodulation in the model legume *Lotus japonicus* by serial analysis of gene expression (SAGE). SAGE libraries were made from uninfected roots and nodulating roots abundant in nodule primordia, and 85,482 and 80,233 SAGE tags were recovered, respectively. Comparison of the tag frequency identified 407 tag species that appeared in significantly greater numbers in the nodulating root than in the uninfected root, and the converse was found for 428 tag species. Gene identification of the tags was performed by matching them to *L. japonicus* expressed sequence tag sequences. We made several novel findings by applying SAGE to transcript analysis of legume root nodulation. A gene that showed the most significant increase in tag number upon nodulation has not been described previously. Different levels of transcription induction among leghemoglobin gene paralogs were found, indicating the effectiveness of SAGE in discriminating different gene family members. We identified genes for 44 unknown tags by means of reverse SAGE. We found 11 antisense tags that increased during nodulation, indicating that regulation of gene expression by antisense transcripts may occur in an organ-dependent manner.

Additional keywords: RT-PCR, transcriptome.

Lotus japonicus has several characteristics that make it suitable for molecular genetic studies. This legume is diploid ($n = 6$) and has a relatively small genome (approximately 472 Mb) and short generation time (2 to 3 months) (Handberg and Stougaard 1992; Ito et al. 2000; Kawaguchi 2000). Systematic structural analysis of the genome is underway through such efforts as the expressed sequence tag (EST) collection (Asamizu et al. 2000, 2004; Endo et al. 2000, 2002a; Poulsen and Podenphant 2002; Szczygłowski et al. 1997), the generation of a high-density linkage map of the chromosomes (Hayashi et al. 2001; Sandal et al. 2002), and the sequencing of chloroplast (Kato et al. 2000) and nuclear (Asamizu et al. 2003; Kaneko et al. 2003; Kato et al. 2003; Nakamura et al. 2002; Sato et al. 2001) genomes.

Corresponding author: Erika Asamizu; Telephone: +81 438 52 3935; Fax: +81 438 52 3934; E-mail: asamizu@kazusa.or.jp

*The e-Xtra logo stands for “electronic extra” and indicates the HTML abstract available on-line contains four supplemental tables not included in the print edition.

Symbiotic relationships with mycorrhizal fungi and rhizobial bacteria are important phenomena both from the biological and agricultural points of view. These aspects of legume biology have been extensively studied, using mutant strains of *Lotus japonicus* with abnormalities in various symbiotic and nitrogen fixation processes (Imaizumi-Anraku et al. 1997; Schauser et al. 1998; Szczygłowski et al. 1998). Genes responsible for these processes have been characterized by map-based cloning using an established transformation system (Martirani et al. 1999; Schauser et al. 1999; Webb et al. 2000). Some key genes related to the perception of the rhizobial bacteria-secreted Nod factor and the early steps of the signal transduction pathways initiating nodule organogenesis have been identified (Endre et al. 2002; Kistner and Parniske 2002; Krussel et al. 2002; Nishimura et al. 2002; Riely et al. 2004; Searle et al. 2003; Stracke et al. 2002). Recently, a comprehensive survey of genes expressed during the nodulation process was performed, using a cDNA array (Colebatch et al. 2002, 2004; Endo et al. 2002b; Kouchi et al. 2004). These efforts have contributed to our understanding of the process but much remains to be clarified, due to the complex nature of the interaction between the rhizobial bacterium and the host plant, which results in drastic changes in gene expression.

Serial analysis of gene expression (SAGE), first described by Velculescu and associates (1995), is a sequence-based technology by which a genome-wide transcript analysis can be conducted. SAGE generates short cDNA fragments, referred to as SAGE tags, which are unique to each transcript and which provide a number of advantages for transcriptome analyses in many organisms. No prior cloning is necessary to generate the expression profile, so information that includes expression of unknown genes can be obtained. The transcript level, obtained as the SAGE tag frequency, is handled as digital data, so it is easy to compare data within an experiment and between different experiments, even when they are performed in different laboratories. Most SAGE tags are derived from the 3' untranslated region (UTR) of a transcript, thus making it possible to discriminate the expression patterns of gene family members.

One obstacle to analyzing SAGE data is the difficulty in identifying a gene with only a 14-base pair tag sequence. This has restricted the application of SAGE primarily to model organisms for which EST or genome sequence data is available in the public databases. In plants, SAGE studies have been completed mainly for two model plants *Oryza sativa* (Matsumura et al. 1999, 2003a and b) and *Arabidopsis thaliana* (Ekman et al. 2003; Fizames et al. 2004; Jung et al. 2003; Lee and Lee 2003), but no such study has been performed in a legume species.

In this study, we performed SAGE analysis of the nodulation process of *L. japonicus* to obtain a comprehensive transcript profile of the root and the nodulating root rich in nodule primordia. In addition to generating a list of genes whose expression changes upon nodulation, we identified a number of previously unknown genes. Additionally, we found several antisense tags, indicating possible regulation of gene expression by antisense strand transcription.

RESULTS

SAGE data acquisition and gene assignment of the tags.

By sequencing 1,812 clones from the root library and 1,925 clones from the nodulating root library, the sequences of 85,482 and 80,233 SAGE tags were obtained, respectively. We obtained 9,890 tag species that appeared more than twice and 18,547 singleton tags from the root library. From the nodulating root library, 9,811 multiple-frequency tags and 18,046 singleton tags were obtained (Table 1).

In order to assign gene identities to the tags, a virtual tag database was created from *L. japonicus* EST sequences. Using the virtual-tag database, we identified genes that are highly transcribed in the root and the nodulating root. A list of the 20 most abundantly transcribed genes in each organ is presented in Table 2. We found that the most abundant tags in the root corresponded to a metallothionein gene (1,232 tags), an aquaporin protein gene (669 tags), and an extensin protein gene (491 tags). The most abundant tags from the nodulating root were from two members of the leghemoglobin gene family (2,328 and 661 tags) and from a gene without significant similarity to known sequences (559 tags). An extended list of the tag-to-gene assignments in each library is available at the Kazusa *L. japonicus* EST index website.

Comparison of the expression profiles of the root and the nodulating root.

For comparative analysis, normalization of the tag frequency between the two libraries was performed. Tags that appeared more than twice were selected for analysis, because singleton tags might contain erroneous or artifactual sequences. We obtained 8,532 tag species for the comparative analysis. Changes in tag frequencies between the root versus the nodulating root were quantified (Fig. 1). A total of 52 tag species appeared at a frequency 20-fold greater in the nodulating root compared with the root ($N > R$), and 20 tags were more than 50-fold enriched in the nodulating root. In contrast, 22 tag species were more than 20-fold enriched in the root ($R > N$), and only one tag species was more than 50-fold enriched.

We performed a test to identify statistically significant tag frequency differences. As a result, 407 tag species were identified as those that appeared in higher frequency in the nodulating root than in the root at significance level (P -chance) of less than 0.01, and 428 tag species appeared in higher frequency in the root. The tags that showed the greatest differential representation and the complete list of tag species that showed a significant frequency difference between the root and nodulating root are provided in Table 3.

Table 1. Summary of tags accumulated from root (R) and nodulating root (N) libraries

Category	R	N
Sequenced clones	1,812	1,925
Tags	85,482	80,233
Tag species with more than two tags	9,890	9,811
Singleton tags	18,547	18,046
Common tag species with more than two tags	8,532	

The tag species showing the highest tag frequency difference in the $N > R$ direction (>557 -fold) was a tag with no similarity to known sequences in the database, followed by a tag for a leghemoglobin gene (464.2-fold) and a tag for the early nodulin ENOD 18 gene (>448 -fold). On the other hand, the highest tag frequency difference in the $R > N$ direction was found in tags for the asparagine synthetase gene (64-fold), followed by tags for a root iron-transporter protein (>40 -fold) and a nitrate-transporter protein (>40 -fold).

In order to examine the characteristic features of genes that are either up- or down-regulated upon nodulation, a functional classification of candidate genes was performed according to the gene ontology (GO) of *Arabidopsis thaliana* provided by The Arabidopsis Information Resource (TAIR). For this analysis, 178 up-regulated and 159 down-regulated tag species from the nodulating root SAGE library were used. The tags chosen for the analysis were the most reliably annotated tag species that matched accurately to the virtual tag sequence of the most 3' end *Nla*III site of several (not one) EST sequences. The corresponding EST sequences were subjected to a similarity search against 26,000 *A. thaliana* protein sequences annotated by MIPS, using the BLASTX program.

EST sequences corresponding to 98 increased and 112 decreased tag species had significant matches (P -value < 0.00005) in *A. thaliana* protein sequences. Functional categorization was performed based on the GO annotations of the *A. thaliana* sequences. (This service is available at the TAIR website.) The differences in genes represented in the up-regulated and down-regulated nodulating root populations are visualized by bar graphs according to the GO in two principle categories, biological process and molecular function (Fig. 2). In biological process, "signal transduction" and "transcription" were predominantly represented in the up-regulated tags, whereas "cell organization and biogenesis" and "electron transport or energy pathways" were enriched in the down-regulated tags. In molecular function, "DNA or RNA binding," "kinase activity," "transcription factor activity," and "receptor binding or activity" were well represented in the up-regulated tags, while "structural molecule activity" was enriched in the down-regulated tags.

Recently, Kouchi and associates (2004) reported a cDNA array analysis of the early stages of *L. japonicus* nodule formation and identified genes showing accelerated expression during this process. We compared these genes to the list of root-nodule-induced transcripts identified in our SAGE analysis. Among the 52 most significantly increased transcripts detected

Fig. 1. The number of tag species is plotted relative to the ratio of tag frequency between the root and the nodule libraries. The center of the X-axis (ratio 1) indicates the number of tag species that appeared with an equal frequency in the two libraries. Tag species to the left of center appeared more frequently in the nodule than the root, and tag species to the right of center were more frequent in the root than in the nodule root.

by SAGE (Table 3), 37 transcripts were contained on the cDNA arrays. In the array analysis, target mRNA was extracted from uninfected roots and nodule samples 2, 4, 7, and 12 days after inoculation with *Mesorhizobium loti*, and up-regulation of gene expression relative to the uninfected root was investigated. Of the 37 arrayed transcripts present in our SAGE data, 26, 2, and 6 of them were up-regulated 4, 7, and 12 days after the inoculation, respectively, and the remaining three showed no significant change.

To confirm the authenticity of the SAGE data, we performed semiquantitative reverse transcription-polymerase chain reaction (RT-PCR) for 22 randomly selected transcripts. We tested

11 transcripts that were represented only in the nodulating root SAGE library and another 11 transcripts that were represented in both libraries (Table 4). Of the former transcripts (Fig. 3A through K), RT-PCR detected increased expression in the nodulating root in all cases and, with one exception (Fig. 3K), no band was detected in the root (Fig. 3A through J). Among the transcripts identified from both SAGE libraries (Fig. 3L through V), the tag corresponding to the translation initiation factor 5A (Fig. 3V) was used as a control, since the tag frequency difference between root and nodulating root was not significant (P -chance = 0.02). The expression levels detected by RT-PCR were consistent with the tag number increases

Table 2. The 20 most abundant tags represented in the root (A) or the nodulating root (B) library

Tag count	Tag	EST	Putative gene product	Organism	Accession	E-value
Root library						
1232	CATGAGCTATGGTT	MWM200f03_f	Metallothionein-like protein class II	<i>Fagus sylvatic</i>	emb CAA10232.1	2E-26
669	CATGTGTGATTTAC	MPD006h02_f	Aquaporin protein PIP1;1	<i>Medicago truncat</i>	gb AAK66766.1 AF386739_1	1E-128
491	CATGTGTATCGCAA	MWM209g09_f	Extensin-like protein	<i>Glycine max</i>	gb AAM75351.1	6E-09
455	CATGTGATCTGGTG	SSQS106c02_f	Plasma membrane integral protein ZmPIP2-7	<i>Zea mays</i>	gb AAK26763.1	2E-34
382	CATGATGAGCGTTA	SPD080h04_f	Glyceraldehyde 3-phosphate dehydrogenase, cytosolic	<i>Pisum sativum</i>	sp P34922 G3PC_PEA	1E-15
316	CATGCTGTGTAATT	CB826748	Protease inhibitor/seed storage/lipid transfer protein family protein	<i>Arabidopsis thaliana</i>	ref NP_190966.1	2E-29
297	CATGTAACTTTAGT	MR025e04_f				
271	CATGCTTGTCCAAA	MWM071e03_f	Cytosolic ascorbate peroxidase	<i>Cucumis sativu</i>	dbj BAA13671.1	0.0002
260	CATGGTGAATGGAA	SPD099g05_f				
229	CATGGTTGGATGGA	MR042d03_f	Putative ripening related protein	<i>Cicer arietinum</i>	emb CAD31716.1	2E-12
209	CATGGCCGATGTTG	BG662376.1	Glycine-rich RNA binding protein 1	<i>Pelargonium x hortorum</i>	gb AAB63581.1	7E-38
207	CATGAGTGATGTAC	GENLf040f08	PR10-1 protein	<i>Medicago truncatula</i>	emb CAA69931.1	6E-24
199	CATGGGTGTTTTCA	BG662154.1	Pathogenesis-related protein 1 (PvPR1)	<i>Phaseolus vulgaris</i>	gi 21044 emb CAA43637.1	1E-20
170	CATGTAAGTTATGT	MSQL062b06_f	Water-selective transport intrinsic membrane protein 1; LIMP1	<i>Lotus japonicus</i>	gb AAF82790.1 AF275315_1	1E-131
167	CATGGAAGGACTAC	GNf008e01	Elongation factor 1-alpha	<i>Pisum sativum</i>	sp Q41011 EF1A_PEA	7E-27
165	CATGATGTTCTGTT	SPD096d09_f	Aquaporin protein PIP1;1	<i>Medicago truncatula</i>	gb AAK66766.1	2E-52
162	CATGTGCTGGCAAT	SPDL081e02_f	Repetitive proline-rich cell wall protein 2 precursor	<i>Glycine max</i>	sp P13993 PRP2_SOYBN	6E-25
162	CATGTTGTAATCAT	MWM233f05_f	Asparagine synthetase [glutamine-hydrolyzing] 2	<i>Lotus japonicus</i>	sp P49093 ASN2_LOTJA	3E-36
160	CATGTTGTGCTTCT	GNf056h09	Indole-3-acetic acid induced protein ARG2	<i>Vigna radiata</i>	sp P32292 ARG2_PHA AU	1E-14
157	CATGTGTTTTGCCT	MFB082a12_f	MtN5	<i>Medicago truncatula</i>	emb CAA75593.1	5E-25
Nodulating root library						
2328	CATGTTTCAGAAAGC	GNf096c12	Leghemoglobin	<i>Lotus japonicus</i>	dbj BAB18106.1	3E-62
661	CATGCTGAAAAGGT	GNf092e01	Leghemoglobin	<i>Lotus japonicus</i>	sp Q9FEP8 LGB_LOTJA	6E-44
559	CATGCGAATATATT	GNf058f11	Metallothionein-like protein class II	<i>Fagus sylvatica</i>	emb CAA10232.1	2E-26
546	CATGAGCTATGGTT	MWM200f03_f				
450	CATGGTATGTTGTG	GNf005g08	Early nodulin ENOD18	<i>Vicia faba</i>	emb CAC18556.1	1E-36
399	CATGTTGTAATCAT	MWM233f05_f	Asparagine synthetase [glutamine-hydrolyzing] 2	<i>Lotus japonicus</i>	sp P49093 ASN2_LOTJA	3E-36
279	CATGTTTTTGATGAA	MR066h04_f	Expressed protein	<i>Arabidopsis thaliana</i>	ref NP_564277.1	3E-17
275	CATGAATGGATGTG	MWM130e08_f	Thiamin biosynthetic enzyme	<i>Glycine max</i>	dbj BAA88226.1	3E-15
274	CATGTGGTTCTCGT	BI417968.1	Early nodulin 36B	<i>Glycine max</i>	sp Q02919 N36B_SOYBN	0.0001
250	CATGTGTGTATGGT	AW719605.1				
224	CATGATGAGCGTTA	SPD080h04_f	Glyceraldehyde 3-phosphate dehydrogenase, cytosolic	<i>Pisum sativum</i>	sp P34922 G3PC_PEA	1E-15
223	CATGGGTTTGCCAT	MWM194h12_f	Glutamate-rich protein	<i>Lotus japonicus</i>	dbj BAD11133.1	6E-33
212	CATGAATTGTTGAG	GNLf020f04	Metallothionein	<i>Glycine max</i>	pir JQ2128	8E-39
199	CATGCAGATCTTCG	BI419880.1	Polyubiquitin	<i>Phaseolus vulgaris</i>	gi 1684855 gb AAB36545.1	1E-75
190	CATGGCTTAAGAAA	MR018b09_f	Putative G-box binding protein	<i>Oryza sativa</i>	gi 18644696 gb AAL76334.1	2E-23
187	CATGAGCAACGTTT	GNf098b03	Lectin-related polypeptide	<i>Robinia pseudoacacia</i>	dbj BAA36416.1	8E-84
182	CATGTAATAGTGT	SPD077a01_f	Auxin-repressed protein	<i>Robinia pseudoacacia</i>	gb AAG33924.1	4E-28
170	CATGGCTGTGGTGT	MFB048e07_f	Ubiquitin conjugating enzyme	<i>Pisum sativum</i>	gb AAA64427.1	1E-67
156	CATGGTGAATGGAA	SPD099g05_f	Cytosolic ascorbate peroxidase	<i>Cucumis sativus</i>	dbj BAA13671.1	0.0002
147	CATGTGCTGGCAAT	SPDL081e02_f	Repetitive proline-rich cell wall protein 2 precursor	<i>Glycine max</i>	sp P13993 PRP2_SOYBN	6E-25

Table 3. List of tags that showed the most significant frequency differences between the root and the nodulating root

Tag	R	N	Fold change	EST	Putative gene product	Organism	Accession	E-value	cDNA array probe ^a	Timing ^b
Nodulating root > Root										
CATGCGAATATATT	0	557	>557	GNf058f11	No similarity				GENLf076f10	N4
CATGTTTCAGAAAGC	5	2321	464.2	GNf096c12	Leghemoglobin	<i>Lotus japonicus</i>	dbj BAB18106.1	3E-62	GENLf016c06	N4
CATGGTATGTTGTG	0	448	>448	GNf005g08	Early nodulin ENOD18	<i>Vicia faba</i>	emb CAC18556.1	1E-36	GENf079d10	N4
CATGGCTTAAGAAA	0	189	>189	MR018b09_f	Putative G-box binding protein	<i>Oryza sativa</i>	gi 18644696 gb AAL76334.1	2E-23	MR018b09_f	N4
CATGCTGAAAAGGT	4	659	164.8	GNf092e01	Leghemoglobin	<i>Lotus japonicus</i>	sp Q9FEP8 LGB_LOTJA	6E-44	GENLf039f11	N4
CATGTCTTTAGCAG	0	145	>145	GENLf091e03	Leghemoglobin	<i>Lotus japonicus</i>	dbj BAB18107.1	1E-61	GENLf016c06	N4
CATGTCGTAATAAA	0	132	>132	GNf098a01	Carbonic anhydrase	<i>Lotus japonicus</i>	emb CAD66064.1	2E-10	GNf069g12	N4
CATGCTTAGGCAAA	0	107	>107	GNf065d10	Unnamed protein product	<i>Arabidopsis thaliana</i>	dbj BAA97184.1	3E-18	GNf065d10	N4
CATGTTGTGTACCA	0	100	>100	GNf038c11	Putative protein; protein id: At5g19600.1	<i>Arabidopsis thaliana</i>	ref NP_568377.1	3E-07	GENf058d12	N12
CATGCTATTTATGT	0	99	>99	GNf026e05	No similarity				GENLf076f10	N4
CATGTCTTTTCTT	0	90	>90	GNf042f11	Nlj21	<i>Lotus japonicus</i>	gb AAB69643.1	2E-16	GNf042f11	N7
CATGCATCAGGTGG	0	85	>85	GNf091e07	No similarity				GNf073f04	N4
CATGAGGGGGTGT	0	66	>66	GNf099g07	IDS3	<i>Hordeum vulgare</i> subsp. <i>vulgare</i>	dbj BAB07798.1	2E-07	GENf004c03	N4
CATGTGAAAATCCT	0	58	>58	GNf088f02	Similar to chloroplast nucleoid DNA binding protein	<i>Arabidopsis thaliana</i>	ref NP_671924.1	9E-12	GENf077c01	N4
CATGTTTAGGAAGC	1	56	56	BI420986.1	No similarity					
CATGAACAGCACCT	0	54	>54	GENLf091e03	Leghemoglobin	<i>Lotus japonicus</i>	dbj BAB18107.1	1E-61	GENLf016c06	N4
CATGAAGATGCAAT	0	53	>53	GENLf089e12	No similarity				GENLf076f10	N4
CATGTTATTTATAT	0	53	>53	GNf011c11	No similarity				GENLf076f10	N4
CATGTTTATTACAC	0	52	>52	GNf096d11	Isoliquiritigenin 2'-O-methyl-transferase	<i>Medicago sativa</i>	sp P93324 CHMT_ME_DSA	3E-34	GENf042b10	N4
CATGTCTAAGTGTG	0	48	>48	No EST						
CATGAGGAGATCCT	0	46	>46	BI418139.1	No similarity					
CATGTCTCTCACGG	1	41	41	No EST						
CATGCAACAGATAC	0	38	>38	GNf091c07	No similarity				GNf043a03	N4
CATGCAGCACACTT	0	37	>37	GNf066a11	No similarity				GNf099e05	N4
CATGGGGAGAATAT	0	37	>37	GNf100b07	GA 2-oxidase	<i>Phaseolus coccineus</i>	emb CAB41036.1	3E-58	GNf087h07	N12
CATGCTCCACCACC	0	37	>37	MWM169a12	No similarity					
CATGTTTGTGAGA	0	34	>34	GNf097a02	No similarity				GENf051e05	N7
CATGTGTATATTAT	1	34	34	GNf060e08	No similarity				GENLf026c12	N12
CATGAAGGGACACC	0	32	>32	GNf100d06	Hypothetical protein ARG10	<i>Vigna radiata</i>	pir T07820	2E-40	GNf052d08	No change
CATGCAGACAGATG	0	32	>32	No EST						
CATGTATGTGTGTT	2	62	31	GENf078d09	No similarity					
CATGATCCTTCATT	0	31	>31	GNf078a01	Chorimate mutase	<i>Lycopersicon esculentum</i>	gb AAD48923.1 L47356_1	8E-09	GENf002g12	N4
CATGACGCCAAGAA	1	31	31	GNf050e08	Phosphatidylinositol transfer-like protein IV	<i>Lotus japonicus</i>	gb AAK63248.1 AF367434_1	5E-65	GENf020a06	N4
CATGAACCATCCCG	0	30	>30	GNf099c02	Elongation factor 1-alpha	<i>Manihot esculenta</i>	sp O49169 EF1A_MANES	2E-67		
CATGGTTGTAATAA	0	29	>29	GENf086b11	Hypothetical protein	<i>Oryza sativa</i> (<i>japonica</i> cv. group)	emb CAE04574.1	3E-17	GENf070c03	N4
CATGTCTTTTGCAG	0	29	>29	GENLf070f10	Leghemoglobin	<i>Lotus japonicus</i>	sp Q9FEP8 LGB_LOTJA	1E-62	GENLf039f11	N4
CATGCCGGTCACAA	0	28	>28	No EST						

Continued on next page

^a Nucleotide sequences of arrayed cDNA clone and tag-assigned EST are identical.^b Timing of induction. cDNA array data is available on the Kazusa DNA Research Institute *Lotus japonicus* EST index website.

Table 3. Continued from preceding page.

Tag	R	N	Fold change	EST	Putative gene product	Organism	Accession	E-value	cDNA array probe ^a	Timing ^b
Nodulating root > Root (<i>continued</i>)										
CATGTGTATGTAAA	0	27	>27	GENf038d06	Expressed integral membrane protein common family	<i>Arabidopsis thaliana</i>	ref NP_565891.1	2E-16	GENf080d09	N4
CATGTAAGTGTG	0	27	>27	No EST						
CATGGTGATACTGA	0	25	>25	GNf069e03	Putative peptide transporter protein	<i>Arabidopsis thaliana</i>	gb AAL36413.1	7E-12	GNf004f04	no change
CATGTGGTTTTGTT	0	25	>25	SPDL099b03_f	No similarity				GENLf076f10	N4
CATGAGGTGGCGGT	0	25	>25	GENf059g09	No similarity					
CATGTAAGTTAGGT	0	25	>25	No EST						
CATGGCAGCAACTG	0	24	>24	GNf085c03	Coproporphyrinogen III oxidase, chloroplast precursor	<i>Glycine max</i>	sp P35055 H EM6_SOY BN	2E-45	GNf097d06	N4
CATGGGCTTGGAGT	0	24	>24	GENLf091e03	Leghemoglobin	<i>Lotus japonicus</i>	dbj BAB18107.1	7E-62	GENLf016c06	N4
CATGGCCCTGGTGC	1	22	22	GNf099f11	Hypothetical protein	<i>Cicer arietinum</i>	emb CAA10289.1	7E-34	MWM054c10_r	no change
CATGGCTCAGGTGG	0	22	>22	No EST						
CATGGTGACAGAAAT	4	80	20	MWM053b03_f	Expressed protein	<i>Arabidopsis thaliana</i>	ref NP_566108.1	7E-22	MWM245g08_r	N12
CATGAGCAGTGGGG	0	20	>20	GNf093c08	No similarity				GENf083f12	N12
CATGAGCAGTGGGG	0	20	>20	MWM153f04_f	DNA binding with one finger 7 protein	<i>Pisum sativum</i>	dbj BAC81664.1	2E-15		
CATGTAAGTACTTT	0	20	>20	GNf099d02	Dioxygenase	<i>Marah macrocarpus</i>	emb CAA70330.1	5E-29	GNf091b09	N12
CATGCACATTTTAG	0	20	>20	AF000408.1	No similarity					
Root > Nodulating root										
CATGAATGGAAAT	64	1	64	SPDL042c03_f	Asparagine synthetase 1	<i>Lotus japonicus</i>	sp P49092 AS N1_LOTJA	6E-40		
CATGCGTTTGCAAA	40	0	>40	MR057e08_f	Root iron transporter protein	<i>Pisum sativum</i>	pir T06385	5E-34		
CATGGTATATTGGT	40	0	>40	MR088e08_f	Putative component of high affinity nitrate transporter	<i>Arabidopsis thaliana</i>	gb AAM62807.1	8E-17		
CATGGCAGAGGAGC	36	1	36	No EST						
CATGAAGCGTTTTG	35	1	35	No EST						
CATGGTGTCTCCTT	30	0	>30	GENLf069d10	Proline-rich protein, 14K	<i>Phaseolus vulgaris</i>	pir S70586	2E-35		
CATGTCAAAAGTTG	30	1	30	MWL080h10_f	Tubulin α -3 chain	<i>Eleusine indica</i>	sp O22349 T BA3_ELEIN	1E-58		
CATGACACATTTTT	29	0	>29	MR081b04_f	Ferredoxin-nitrite reductase	<i>Lotus japonicus</i>	emb CAC06095.1	1E-07		
CATGGTGTCTGCTT	28	0	>28	MR019c09_f	No similarity					
CATGTGTTATGTAT	56	2	28	MR100h11_f	Transmembrane channel protein	<i>Cicer arietinum</i>	emb CAA06745.1	5E-32		
CATGCCTAAGGATA	27	1	27	No EST						
CATGTGGAAGTTTA	73	3	24.3	SPDL089h03_f	Ferredoxin-NADP reductase	<i>Pisum sativum</i>	pir T06773	6E-48		
CATGTCCCTATAAG	24	0	>24	No EST						
CATGAGAATCGGTC	22	1	22	AI967408.1	Hypothetical protein	<i>Oryza sativa</i>	dbj BAD07869.1	1E-16		
CATGTGTATCGCAA	45	21	21.9	MWM209g09_f	Extensin-like protein	<i>Glycine max</i>	gb AAM75351.1	6E-09		
CATGTGTAGTGTTC	21	0	>21	MRL040h08_f	Ferredoxin-NADP reductase, root isozyme, chloroplast precursor	<i>Pisum sativum</i>	sp Q41014 F ENS_PEA	1E-09		
CATGGACTGAGATT	41	2	20.5	SPD050c11_f	No similarity					
CATGTTGGTTTTGC	102	5	20.4	MR048d02_f	Membrane transporter	<i>Lotus japonicus</i>	emb CAC35729.1	7E-45		
CATGCAGAATTTTG	20	0	>20	MR018f08_f	Type IIIa membrane protein cp-wap13	<i>Vigna unguiculata</i>	pir T11577	2E-15		
CATGTGAGCACCTA	20	0	>20	No EST						
CATGATGAGGGTCC	20	1	20	No EST						
CATGTATGTTTCAG	20	1	20	No EST						

(Fig. 3L through R and T) and decreases (Fig. 3U) detected between the root and nodulating root SAGE libraries, with one exception (Fig. 3S). In four cases, in which only a small number of tags were acquired from the root (two to seven), no band was detected in the RT-PCR analysis (Fig. 3L, O, Q, and R). These results indicate that in 20 out of the 21 cases tested, the statistically significant tag number differences were confirmed by the RT-PCR analysis.

Characterization of the antisense and the unknown tags.

To complete our analysis of transcript regulation during the nodulation process, further characterization of tags that had no matches in the virtual tag database was performed. We searched *L. japonicus* EST sequences and found 11 N > R tags and 14 R > N tags that matched the noncoding strand of a cDNA (Table 5). To confirm the presence of antisense transcripts, we performed strand-specific RT-PCR analysis for the 11 N > R tags. In this method, a gene-specific primer in the antisense orientation was used for first-strand cDNA synthesis (Table 6), allowing selective amplification of the antisense transcript in the subsequent PCR. The results of strand-specific RT-PCR of the root and the nodulating root tissues are shown in Figure 4. Among the six cases in which the antisense tag count was 0 in the root, no band was detected by strand-specific RT-PCR in four of the six cases. In five cases in which tags were obtained from the root library, a band corresponding to the transcript was observed in four of the five cases. In the nodulating root, the RT-PCR results validated the presence of antisense transcripts in nine of the 11 cases.

A powerful advantage of SAGE is that information on unknown genes can be obtained. As a result of the comparison of tag frequencies between the root and the nodulating root, we

obtained 104 nodulation-increased tag species with no match to either the virtual tag database or the antisense strand of the *L. japonicus* EST sequences (Table 5). Characterization of these unknown tags was performed by generating 3'-extended cDNA fragments by the reverse (r)SAGE method. The amplified cDNA fragments were cloned and sequenced. The authenticity of the sequence generated from a gene-specific primer of 14 bases in length was determined by comparison with corresponding genome sequences. We identified 30 novel 3' cDNA sequences, nine paralogous sequences that showed 80 to 90% identity with known EST sequences, and five tags with a sequence error. The remaining 54 tags resulted in no amplified product or nonspecific amplification. The authenticity of the remaining six rSAGE products has not been confirmed to date, since the entire genome sequence is not yet available. The list of 30 novel 3' cDNA species with search results of corresponding genome sequences is shown in Table 7.

DISCUSSION

To analyze the transcription profile of the *L. japonicus* early nodulation process on a genome-wide scale, we performed SAGE of uninfected root and nodulating root abundant in nodule primordia and compared the SAGE tag frequency. Large-scale expression analyses of the *L. japonicus* nodulation process using cDNA array technology have been reported previously (Colebatch et al. 2002, 2004; Endo et al. 2002b; Kouchi et al. 2004). We chose SAGE in order to obtain a more comprehensive expression profile of this process by obtaining information on previously uncharacterized genes. This is the first report of SAGE technology applied to transcript profiling of the legume nodulation process.

We first compared the transcript levels of genes within each library and found that a metallothionein (MT) gene is the most abundantly transcribed gene in the root and a highly transcribed gene in the nodulating root (Table 2). Plant MT genes have been isolated from various species and are suggested to function in many biological processes, including heavy metal detoxification, metal ion transport, detoxification of reactive oxygen species, stress response to plant pathogens, and may have additional roles at specific developmental stages. A high level of expression of an MT gene in the root and the nodulating root was also demonstrated in the actinorhizal tree *Casuarina glauca* (Laplaze et al. 2002). The MT gene may transport metal ions required for enzyme activities in both organs, and one possible example is nitrogenase activity in the nodule.

Our SAGE data also revealed that genes encoding water channel and transport proteins (aquaporin and plasma membrane intrinsic protein), and a cell wall protein (extensin-like protein) are highly transcribed in the root. These tags are not included in the nodule-abundant class (Table 2), indicating that transport and cell wall biosynthesis predominates in roots. In the nodulating root, leghemoglobin genes were most abundantly transcribed, and some early nodulins (ENOD18 and ENOD36B) were also observed in this abundant class. Leghemoglobin is an oxygen-binding hemoprotein that acts to control the oxygen concentration in nitrogen-fixing nodules and is highly accumulated in the nodule (about 30 to 40% of the total soluble protein) (Wittenberg et al. 1975). We constructed the nodulating root SAGE library from relatively early-stage nodules that were collected four days after inoculation with a rhizobial bacterium, *Mesorhizobium loti*. It is possible that our sample contained nodules exhibiting nitrogen-fixing activity; however, we consider it more likely that leghemoglobin gene transcription is initiated at an early stage of the nodulation process, before the onset of the nitrogen fixation. The result of the cDNA array analysis supports this notion. As indicated in Table 3, up-regu-

Fig. 2. The result of gene ontology (GO) categorization of expressed sequence tag sequences corresponding to tag species significantly up-regulated in the root (R) or in the nodulating root (N). **A.** The percentage of hits to GO categories with “biological process” annotations. **B.** The percentage of “molecular function” annotations.

lation of the leghemoglobin gene was detected at 4 days after the *M. loti* inoculation (N4). It has been reported that the expression of a symbiotic leghemoglobin gene, *VsLb1*, in *Vicia sativa* is observed as early as 1 h after Nod-factor addition (Heidstra et al. 1997).

We conducted a statistical comparison of the tag frequencies in the two libraries and found 407 tags that appeared significantly more frequently in the nodulating root than in the root and 428 tags that appeared significantly more frequently in the root than in the nodulating root. The largest frequency differences were found in the nodulating root up-regulated transcripts, indicating a dramatic change in gene expression occurs during the nodulation process. In Table 3, 52 and 22 tags that displayed a greater than 20-fold tag frequency difference in the nodulating root versus the root are listed. The tag showing the greatest increase (>557-fold) in the nodulating root corre-

sponded to an EST that lacked similarity to known protein sequences. We first hypothesized that the 3' EST was too small to provide enough coding sequence for protein sequence comparison, as we could identify no open reading frame in the 269-bp EST sequence. We compared the sequence with the annotated *L. japonicus* genomic sequence and found that the EST matched a predicted gene, LjT10c05.7, which would encode a 187-aa protein. However, the sequence of this putative protein did not show any significant similarity to any registered sequences. We performed an InterPro scan (Mulder et al. 2003) to identify any conserved functional protein domains but none were detected. These data suggest that this gene may code for a novel, nodulation-specific gene.

The list in Table 3 contains an additional 17 tags that match EST sequences without similarity to known protein sequences and seven tags that do not match any EST present in the se-

Table 4. The list of primers used in the reverse transcription-polymerase chain reaction analysis

ID	Tag	R ^a	N ^b	EST clone	Putative gene product	Primer	Product ^c	P-chance ^d
A	CATGCGAATATATT	0	557	GNf058f11	Unkown	TGCAACGAGCTTGAATTGAC CTTTAGAAAATATATTGCGATG	191	0
B	CATGGTATGTTGTG	0	448	GNf005g08	Early nodulin ENOD18	CTCTATTGAAGATCTAAAGC TGAAGCCACAACATACCATG	253	0
C	CATGGCTTAAGAAA	0	189	GENLf083c09	B12D protein	CTGAGCATGGCTTAAGAAAAG AGACGATTTTCATATTTTATG	270	0
D	CATGCTTAGGCAAA	0	107	GNf065d10	Putative protein	CATGCTTAGGCAAAATCTGC TTCTCTCATTTCTAACTTG	236	0
E	CATGTCGTAATAAA	0	132	GNf093h12	Carbonic anhydrase	TGACATGACTATCTAAATCC CAATCTTTTATTACGACATG	237	0
F	CATGCTATTTATGT	0	99	GNf026e05	Unknown	TGAATTGACATGCTATTTATG TATTCGCATGAAACAACAAC	168	0
G	CATGTCTTTTCTT	0	90	GNf042f11	Nlj21	AAGGAAGAGGTTGTTAATAC CAAAGAAAAAGACATGGAGC	215	0
H	CATGAGGGGGTGTG	0	66	GNf099g07	Iron deficiency protein Ids3	ATCAATCATATTCTAGACAG ATAAGGACAACCCCTCATG	229	0
I	CATGTGAAAATCCT	0	58	GNf088f02	Chloroplast nucleoid DNA binding protein	TCGCCATCTTTGGGAATTG GTCTGAAGGATTTTCACATG	239	0
J	CATGGGGAGAATAT	0	37	GNf100b07	Gibberellin 2-oxidase	ATCAGACTTCTTTCTTCATC TTTCTTATATTCTCCCATG	227	0
K	CATGTTATTTATAT	0	53	GENf011c11	Unkown	AATTATCGAAAGTACCATCC ATATAAATAACATGTCAATTCAAGC	236	0
L	CATGTTTCAGAAAGC	5	2,321	GNf096c12	Leghemoglobin	GCTGTTTCATGTTTCAGAAAGC CCATTATGGCTATATATTC	196	0
M	CATGTTGTAATCAT	151	397	GNf096c06	Asparagine synthetase	GCCTGCTGTTTATTTTATCC CTCAAGATGATTACAACATG	223	0
N	CATGAATGGATGTG	24	274	GNf081e01	Thiamin biosynthetic enzyme	TAATGACAATAAGCTTCTTG CACATCCATTTCATGTTATTAAC	157	0
O	CATGGTGCAGAAAT	4	80	MWM053b03_f	Universal stress protein	AAGCCGTGGAGAAGTATCAC ATTTCTGCACCATGTACACG	188	0
P	CATGTTATGAACTA	27	122	GNf086f03	Thiamin biosynthetic enzyme	TCTGCTGAAATTGCAGATGC CTGCGATAGTTCATAACATG	283	0
Q	CATGTATGTGTGTT	2	62	MWM026f10_f	Universal stress protein	TATGGTGTTCATCAAGAGGGC ACGAAAAACACACATACATG	205	0
R	CATGAGTGCTATAG	7	42	GNLf002a04	Hypersensitive-induced response protein	GAAGGTTTTCTGGAATGTAG TAGCAACTATAGCACTCATG	191	0
S	CATGCAGATCTTCG	122	198	GENf086a06	Polyubiquitin	GCATGCAGATCTTCGTGAAG TCAGAGCTCTCAACCTCAAG	291	3E-06
T	CATGGAAGGCTGTT	8	26	MR075a06_f	Proline-rich protein	TGCATGGAAGGCTGTTGGC AAAAAACACAGTGAACAAGG	193	8E-04
U	CATGGAAGGACTAC	156	94	GNf095a12	Elongation factor 1 alpha	ATTTGTCATGGAAGGACTAC TTCTCAAAAGACACGACAGC	223	4E-04
V	CATGGGGGAGGAGC	110	74	GENLf014b05	Translation initiation factor 5A	CCATAACTGTGATATCCCTC AATCTGCTCCTCCCCATGG	217	0.02

^a The number of tags obtained from the root library.

^b The number of tags obtained from the nodulating root library.

^c Product measured in bp.

^d Significance of the tag frequency difference calculated by the SAGE2000 program.

quence banks. These results show that the SAGE analysis enabled us to identify novel ‘nodulin’ genes with unknown functions.

There were six tags corresponding to leghemoglobin genes. *L. japonicus* has three symbiotic leghemoglobin genes, *LjLb1*, *LjLb2*, and *LjLb3* (Uchiyama et al. 2002). *LjLb1* and *LjLb2* are almost identical in their nucleotide sequences (98% identity) and are located in tandem on the genome. *LjLb3* is also located on the same chromosome (chromosome 5) and shares 87 and 86% nucleotide sequence identity, respectively, with the aforementioned genes. We found that the most abundant leghemoglobin tag, CATGTTTCAGAAAGC, corresponds to *LjLb1* and *LjLb2*, and the second most abundant leghemoglobin tag, CATGCTGAAAAGGT, corresponds to *LjLb3*. This illustrates one example of the advantage of SAGE, which is that it can discriminate different gene family members that may not be necessarily resolved by hybridization analysis. Another leghemoglobin tag, CATGGGCTTGGAGT, corresponded to *LjLb1* and *LjLb2*, but the tag was derived from the second most 3' end *Nla*III site, probably due to inadequate digestion by the anchoring enzyme, *Nla*III. We found three other leghemoglobin tags that match leghemoglobin mRNA in the antisense orientation. The sequences indicated that these antisense tags, CATGTCTTTAGCAG and CATGAACAGCACCT, were derived from either *LjLb1* or *LjLb2*, and the third antisense tag, CATGTCTTTTGCAG, was derived from *LjLb3*. The antisense transcription of *LjLb3* was confirmed by strand-specific RT-PCR analysis (Fig. 4), suggesting the possibility that *LjLb1* and *LjLb2* are also transcribed in the antisense orientation.

Table 5. Expressed sequence tag (EST) matching of differentially represented tags

EST match	Nodulating root > root	Root > nodulating root
Sense strand	292	313
Antisense strand	11	14
No match	104	101
Total	407	428

GO functional classification analysis revealed several trends that point to differences in the transcriptome of the root and the nodulating root. Genes more abundant in the root were primarily classified under general cellular processes. These included ribosomal protein genes categorized into cell organization and biogenesis and structural molecular activity, and genes categorized into electron transport or energy pathways, including those encoding ferredoxin NADP oxidoreductase and cytochrome P450. These trends suggest that cellular activities such as protein synthesis and energy metabolism are less active in nodulating root than in the root. In contrast, genes related to signal transduction and regulation predominated in the nodulating root. These included receptor-like protein kinase genes classified into signal transduction and kinase activity, and transcription factors categorized into DNA- or RNA-binding and transcription factor activity. The up-regulation of receptor kinase and transcription factor genes upon nodulation suggests that these genes may play key roles as regulators of the nodulation program. The function of these genes is to be elucidated in future studies.

To confirm the authenticity of the SAGE data, we first compared the SAGE data with the cDNA array results reported by Kouchi and associates (2004). The results correlated quite well in that 26 out of 37 up-regulated transcripts found by SAGE analysis of nodulating roots at 4 days postinfection (Table 3) corresponded to arrayed probes that showed up-regulation at 4 days after *M. loti* inoculation (N4). It is notable that when the difference in the frequency of a tag between the root and nodulating root was relatively low, up-regulation was either detected at later stages (N7 or N12) or was not detected in the array experiment. Further expression analysis of individual transcripts will be necessary to interpret these latter results that suggest SAGE analysis can detect much smaller changes in gene expression level.

We also performed a semiquantitative RT-PCR for 22 randomly selected transcripts, 21 of which showed a statistically significant (P -chance < 0.01) tag frequency difference and one of which was selected as a control. In almost all cases (20 out of

Fig. 3. The results of semiquantitative reverse transcription-polymerase chain reaction (PCR) analysis. A-V corresponds to genes listed in Table 4. Upper panels show the root, and lower panels the nodulating root. Each panel shows the ethidium bromide-staining of agarose gel electrophoresed PCR products generated using 1, 10, 100, and 1,000 ng of total RNA as a template (from left to right). The size marker at the left is a 100-bp ladder ranging from 100 to 600 bp.

21), RT-PCR detected expression differences that confirmed the SAGE analysis data (Table 4, Fig. 3). This result indicates that the statistically significant tag frequency difference reliably reflects true differences in expression level. Our RT-PCR analysis failed to detect transcripts when only a small number of tags were found in the SAGE library (Fig. 3L, O, Q, and R), indicating that the sensitivity of detection is high in the SAGE method.

One of the advantages of SAGE is that information on previously uncharacterized transcripts is obtained. For example, a number of tags derived from the noncoding (antisense) strand of a transcript were detected in this study. Detection of antisense SAGE tags has previously been reported in several organisms (Jones et al. 2001; Patankar et al. 2001). In a genome-wide transcript analysis of *Arabidopsis thaliana*, approximately 7,600 annotated genes (30% of all annotated genes) showed significant antisense transcription (Yamada et al. 2003). In human, tissue-specific expression of sense and antisense transcripts has been detected, suggesting a possible role in the regulation of gene expression (Yelin et al. 2003). In *L. japonicus*, a member of the phosphatidylinositol transfer protein gene family is bidirectionally transcribed (Kapranov et al. 2001). It is speculated that an antisense transcript forms a sense-antisense RNA duplex, but the

mechanism by which these duplexes may regulate gene expression is not clearly established (Carmichael 2003). Our SAGE data also suggest that antisense transcription is a general phenomenon in eukaryotes. The results of the strand-specific RT-PCR analysis demonstrated that antisense transcription is induced in a nodulation-dependent manner, suggesting the possibility that some of the antisense transcripts may be responsible for gene regulation in the nodulation process. One reason for the discrepancy between the observed tag number and the results of the strand-specific RT-PCR analysis may be the unstable nature of the antisense transcript or variability in its structure.

Additional examples of uncharacterized transcripts that can be identified by SAGE are previously uncloned genes. We characterized 104 unknown tags whose frequency was significantly increased in the nodulating root by the rSAGE method. Although the authenticity of the rSAGE cDNA fragment has not been confirmed for six tag species, we successfully characterized 45% of the rSAGE tags and found 30 novel sequences, nine paralogous sequences, and five error-containing sequences. We were unable to specifically amplify the remaining 54 unknown tags, primarily due to the low melting temperature of the 14-bp tag primer. We will be able to characterize them by mapping the tag se-

Table 6. List of antisense tags and primers used in the strand-specific reverse transcription-polymerase chain reaction analysis

No.	Tag	R ^a	N ^b	EST clone	Putative gene product	Primer type ^c	Primer sequence
1	CATGAAATTGCAAC	0	7	GNf056e05	Unknown	Fwd	CTTCTCATGTGCGCGGTAGTTC
						Rev	TTTCTCCAAAACCAAGCAGCAG
2	CATGAAGGGACACC	0	32	GNf100d06	Unknown	Fw	GGTTCGTTGAGAGCAAGGCATC
						Rev	CAACAACATGATTGGCAGGGTAG
3	CATGAGGTGGCGGT	0	25	GENf059g09	Unknown	Fwd	AACGAGCCATTGTGCGGAAC
						Rev	ATCGCAAAACATTTTCGACAG
4	CATGCAAAAAGTCT	1	11	GNf099a01	Elongation factor 1	Fwd	CGGTTACGACTGCGTATTCTCTG
						Rev	TCTTCACATCGTCAATAGTGTGCC
5	CATGCAGACAACCTG	6	37	GNf097g10	Unknown	Fwd	CACCAGTTTACACTTACCCTCACC
						Rev	AACACAACAGTCCTCTCGATCCTC
6	CATGCTCAATCCCT	0	7	GNf094a04	<i>o</i> -Methyltransferase	Fwd	TGGTCAGACGAGAAATGCTTGG
						Rev	GTGAAACTCGATCACTCCCAAAAC
7	CATGGTGAGACACA	30	55	SPD047g08_f	Calmodulin	Fwd	TTGATGAGATGATCCGTGAGGC
						Rev	TCAAACTTATCTGCCCTGTCCC
8	CATGTAGATATCAC	1	10	GNf096c06	Asparagine synthetase	Fwd	GGAAGTTTCTCCTTGGAGTTGC
						Rev	GGCTTTCAGGTGAACACTATTTGG
9	CATGTAGGTATACC	10	33	GENf010c02	Unknown	Fwd	AATGGTCCCAGTTTCTGGTATCAG
						Rev	AACAGTAACACACAAGGGCAAACC
10	CATGTCTTTTGCAG	0	29	GENLf070f10	Leghemoglobin	Fwd	AGTGACACTTGCAGATGCCAGC
						Rev	CCCATTGCCTTCTTAATTGCAG
11	CATGTTATTTAAGG	0	8	SPD087c07_f	Unknown	Fwd	AAACGCTAAGGAGTTGGGTTTACC
						Rev	AATTCACCAGCCTCAAGGGAAG

^a The number of tags obtained from the root library.

^b The number of tags obtained from the nodulating root library.

^c Rev primers complementary to the mRNA antisense strand were used in the reverse transcription.

Fig. 4. The results of strand-specific reverse transcription-polymerase chain reaction (PCR) analysis. Lanes 1 through 11 correspond to genes listed in Table 6 from which antisense tags were obtained. The left panel shows the results from the root, and the right panel shows the nodulating root samples, respectively. Each panel shows the ethidium bromide-staining of agarose gel electrophoresed PCR products generated using 100 ng of total RNA as a template (the right lane). In the left lane, 100-bp ladder size markers are loaded.

quences to the annotated genome in the future. The results of our SAGE analysis and the availability of sophisticated sequence information from the *L. japonicus* EST and genome databases clearly demonstrate that this organism is a suitable model to perform large-scale transcriptome analysis to study biological processes characteristic of leguminous plants.

MATERIALS AND METHODS

Plant-growth conditions and RNA preparation.

Lotus japonicus, accession Gifu B-129, was used to prepare root and nodulating root samples. The plants were grown on soil in a 20-h photoperiod of fluorescent illumination (100 $\mu\text{E m}^{-2}\text{s}^{-1}$) at 22°C. The sample abundant in nodule primordia was collected in the laboratory of M. Kawaguchi at the University of Tokyo by harvesting the nodulated zone of hypernodulating roots from the *har1-3* mutant (Schauser et al. 1998) 4 days after inoculation with *Mesorhizobium loti* MAFF303099. Total RNA was extracted by LiCl precipitation (Vicent and Delseny 1999). Purification of polyadenylated RNA was performed by Oligotex-dT30 Super (Takara Bio Inc., Otsu, Japan). Conversion to the biotinylated oligo-dT primed cDNA was performed with SuperScript II reverse transcriptase (Stratagene, La Jolla, CA, U.S.A.), using 5.0 and 6.8 μg of polyadenylated RNA from root and hypernodulating root, respectively.

SAGE library construction.

SAGE libraries were constructed according to the original method (Velculescu et al. 1995, 1997), with minor modifications. Briefly, the synthesized, biotinylated cDNA was digested with *Nla*III, and the most 3' end fragments were collected by

separation with streptavidin-coated magnetic beads. A linker that harbors a recognition site for the type II restriction enzyme *Bsm*FI was ligated to the 5' end of the fragments. Cleavage by *Bsm*FI results in the release of 13-bp tags. The 3' ends of the tags were blunt-ended and ditags were formed by ligating those ends. The ditags were amplified by PCR with linker-specific primers (5'-TACAAGTACGGGGACA-3' and 5'-TCTAACGATGTACGGGGACA-3'). After the amplification, linkers were removed by *Nla*III digestion and were separated by polyacrylamide gel electrophoresis. The ditag fragments ranging from 26 to 30 bp were purified from the gel and were concatenated by ligation. The concatemers were size-selected (>500 bp) by agarose gel electrophoresis prior to cloning into the *Sph*I site of the pZER0 vector (Invitrogen, Carlsbad, CA, U.S.A.).

SAGE tag sequencing and data analysis.

Plasmid DNAs were used as sequence templates. The sequencing reaction was performed using the BigDye Terminator cycle sequencing ready reaction kit (Applied Biosystems, Foster City, CA, U.S.A.), and reactions were run on an automated DNA sequencer, ABI PRISM 377XL (Applied Biosystems). The sequence data analysis was performed with the SAGE2000 program provided by K. Kinzler at Johns Hopkins University (Baltimore). Using this program, tag sequences are extracted after eliminating "duplicate ditags" resulting from distortions in the PCR step, based on an assumption that a same-tag pair would not be created by chance.

The statistical significance of the tag frequency difference between the root and the nodulating root was calculated by Monte Carlo simulation in the same program (Zhang et al.

Table 7. List of novel 3' cDNA-generated tags by the reverse SAGE (rSAGE) method

Tag	R ^a	N ^b	Putative gene product	Organism	Accession	E-value
CATGAAGTCTACTG	0	17	Chalcone isomerase	<i>Pisum sativum</i>	gi 432404 gb AAA50174.1	1E-32
CATGAATTGTAGGC	1	9	No similarity			
CATGAGAGTGAGAT	5	19	Submergence induced protein 2-like	<i>Vitis vinifera</i>	gb AAO12871.1	1E-15
CATGAGCTACCCCT	0	12	Putative alcohol dehydrogenase	<i>Oryza sativa</i>	gb AAN04147.1	3E-17
CATGAGGGTTAATA	0	9	Putative 4-coumarate CoA ligase	<i>Oryza sativa</i>	gb AAG46175.1 AC018727_27	9E-11
CATGAGTTATCGAT	0	9	No similarity		No hits found	
CATGATGATGCAGC	6	36	Wound-induced protein	<i>Lycopersicon esculentum</i>	gi 19320 emb CAA42537.1	5E-16
CATGATGATGGGAG	0	13	Phosphatidic acid phosphatase beta	<i>Vigna unguiculata</i>	gb AAF89745.3 AF171230_1	4E-20
CATGCAAAAGAGGGT	0	10	Histone H3	<i>Medicago sativa</i>	gi 829279 emb CAA31967.1	9E-48
CATGCAACTGAAAA	0	19	Transcription factor IIB	<i>Homo sapiens</i>	ref NP_037374.1	3E-35
CATGCAATGTTAGA	1	13	Nam-like protein 10	<i>Petunia x hybrida</i>	gb AAM34773.1 AF509873_1	1E-08
CATGCCAACGGAAG	0	10	Receptor protein kinase-like protein	<i>Arabidopsis thaliana</i>	gb AAM65586.1	5E-33
CATGCCCCAAAAAT	1	18	Transcription factor	<i>Arabidopsis thaliana</i>	emb CAA74049.1	5E-12
CATGCCTTACACAA	0	10	Putative protein; protein id: At4g32870.1	<i>Arabidopsis thaliana</i>	ref NP_195013.1	2E-04
CATGCGACAAGATG	1	10	Homocysteine S-methyltransferase AtHMT-1	<i>Arabidopsis thaliana</i>	gi 6685161 gb AAF23821.1 AF219222_1	7E-20
CATGCTGTTTCCCC	6	32	Chalcone flavanone isomerase	<i>Pueraria montana</i> var. <i>lobata</i>	gi 944815 dbj BAA09795.1	3E-28
CATGGAAAAGGTCTC	12	48	EF-hand Ca ²⁺ -binding protein CCD1	<i>Arabidopsis thaliana</i>	gb AAM61056.1	3E-38
CATGGAATGTGAAT	0	7	No similarity			
CATGGCTTTCATCA	1	9	No similarity			
CATGGTAGAATAAA	0	14	Nramp1 protein(natural-resistance-associated macrophage protein 1)	<i>Oryza sativa</i>	dbj BAC21327.1	6E-11
CATGGTGTGATCGC	18	112	Histone H1, drought-inducible	<i>Lycopersicon pennellii</i>	gi 2147479 pir S65059	3E-29
CATGGTTACCAGAT	4	34	Early nodulin	<i>Glycine max</i>	gi 218256 dbj BAA02721.1	3E-05
CATGTAAAAACTTC	0	12	No similarity			
CATGTAACTTAGGT	0	25	No similarity			
CATGTATGAGATAT	0	16	MADS-box protein; protein id: At5g60440.1	<i>Arabidopsis thaliana</i>	ref NP_200852.1	7E-39
CATGTCTAAGTGTG	0	48	Transcription factor	<i>Arabidopsis thaliana</i>	emb CAA74049.1	0.022
CATGTTATAGGAAC	0	14	Syringolide-induced protein 13-1-1	<i>Glycine max</i>	dbj BAB86896.1	9E-65
CATGTTATGTTATA	0	14	Iron-superoxide dismutase	<i>Glycine max</i>	gb AAQ13492.1	1E-10
CATGTTCTCCCCAG	29	67	Eukaryotic initiation factor 5A (2)	<i>Nicotiana glauca</i>	gi 19702 emb CAA45104.1	8E-11
CATGTTTTTAGCCT	1	12	No similarity			

^a The number of tags obtained from the root library.

^b The number of tags obtained from the nodulating root library.

1997). It calculates a relative likelihood that a difference would be seen by chance. Given the null hypothesis that the frequency of each tag is same for two samples being compared, simulations are performed to determine the probability of obtaining equal or greater numbers of the tag than the actually observed differences by chance. When the probability was less than 0.01, we considered the difference to be significant.

Gene assignment of SAGE tags.

Lotus japonicus EST sequences were searched for virtual SAGE tags to create a reference database. Our collection of 3' EST sequences, described previously (Asamizu et al. 2004), was the primary component of the database. The 3' EST sequences with poly(T) signal were selected to extract virtual 10-bp SAGE tags downstream from the *Nla*III recognition site (CATG).

RT-PCR.

The RT-PCR analysis was performed using the mRNA selective PCR kit (Takara Bio Inc.). For each primer set, 1 µg of total RNA from the root or the nodulating root and its .1, .01, and .001 dilutions were used as the templates. We used the following protocol: 30 min at 50°C for the RT reaction with an oligo (dT) primer, followed by 30 cycles of 1 min at 85°C, 1 min at 50°C, and 2 min at 72°C. The PCR products were electrophoresed on an agarose gel and were detected by ethidium bromide staining. The result was validated by repeating the RT-PCR reaction once. This system uses dNTP analogs for substrates in the RT and subsequent PCR reactions. This allows denaturation at 85°C and selectively amplifies the RNA-derived cDNA fragments while preventing amplification of contaminating genomic DNA.

Strand-specific RT-PCR for the detection of the antisense transcript was performed, also using the mRNA selective PCR kit (Takara Bio Inc.) with changes as follows. First-strand cDNA was synthesized using 100 ng of total RNA as a template with a gene-specific primer that specifically hybridizes to the antisense strand of an mRNA rather than an oligo (dT) primer. After the completion of the reverse transcriptase reaction, primer pairs were added to perform PCR amplification of the cDNA fragments.

rSAGE.

Tags for which no matching EST was found were identified using the rSAGE method (Polyak et al. 1997; Yu et al. 1999). Briefly, cDNA was synthesized with a biotinylated oligo(dT) primer that has a M13F primer site and an *Asc*I recognition site. After digestion with *Nla*III, the cDNA fragments were collected by streptavidin-coated magnetic beads. A linker was ligated to the 5' ends, followed by release of the 3' proximal cDNA pool by *Asc*I digestion. This pool was enriched by PCR amplification with the linker-specific primer and M13F primer. Using this amplified product as a template, PCR screening was performed with a tag-specific primer and the M13F primer.

ACKNOWLEDGMENTS

We are grateful to M. Kawaguchi (University of Tokyo) for the nodulating root sample preparation, K. Kinzler (Johns Hopkins University) for providing the SAGE protocol and the SAGE analysis software, and H. Matsumura (Iwate Biotechnology Research Center) for technical advice. We thank M. Kohara and A. Watanabe for excellent technical assistance.

LITERATURE CITED

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

2003. Structural analysis of a *Lotus japonicus* Genome. IV. sequence features and mapping of seventy-three TAC clones which cover the 7.5 Mb regions of the genome. *DNA Res.* 10:115-122.
Asamizu, E., Nakamura, Y., Sato, S., and 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.
Carmichael, G. G. 2003. Antisense starts making more sense. *Nat. Biotechnol.* 21:371-372.
Colebatch, G., Kloska, S., Trevaskis, B., Freund, S., Altmann, T., and Udvardi, M. K. 2002. Novel aspects of symbiotic nitrogen fixation uncovered by transcript profiling with cDNA arrays. *Mol. Plant Microbe Interact.* 15:411-420.
Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Montanari, O., Kloska, S., Kopka, J., and Udvardi, M. K. 2004. Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *Plant J.* 39:487-512.
Ekman, D. R., Lorenz, W. W., Przybyla, A. E., Wolfe, N. L., and Dean, J. F. 2003. SAGE analysis of transcriptome responses in *Arabidopsis* roots exposed to 2,4,6-trinitrotoluene. *Plant Physiol.* 133:1397-1406.
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 Res.* 7:213-216.
Endo, M., Hakozaiki, H., Kokubun, T., Masuko, H., Takahata, Y., Tsuchiya, T., Higashitani, A., Tabata, S., and Watanabe, M. 2002a. Generation of 919 expressed sequence tags from immature flower buds and gene expression analysis using expressed sequence tags in the model plant *Lotus japonicus*. *Genes Genet. Syst.* 77:277-282.
Endo, M., Matsubara, H., Kokubun, T., Masuko, H., Takahata, Y., Tsuchiya, T., Fukuda, H., Demura, T., and Watanabe, M. 2002b. The advantages of cDNA microarray as an effective tool for identification of reproductive organ-specific genes in a model legume, *Lotus japonicus*. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 514:229-37.
Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kalo, P., and Kiss, G. B. 2002. A receptor kinase gene regulating symbiotic nodule development. *Nature* 417:962-966.
Fizames, C., Munos, S., Cazettes, C., Nacry, P., Boucherez, J., Gaymard, F., Piquemal, D., Delorme, V., Commes, T., Doumas, P., Cooke, R., Marti, J., Sentenac, H., and Gojon, A. 2004. The *Arabidopsis* root transcriptome by serial analysis of gene expression. Gene identification using the genome sequence. *Plant Physiol.* 134:67-80.
Handberg, K., and Stougaard, J. 1992. *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *Plant J.* 2:487-496.
Hayashi, M., Miyahara, A., Sato, S., Kato, T., Yoshikawa, M., Taketa, M., Hayashi, M., Pedrosa, A., Onda, R., Imaizumi-Anraku, H., Bachmair, A., Sandal, N., Stougaard, J., Murooka, Y., Tabata, S., Kawasaki, S., Kawaguchi, M., and Harada, K. 2001. Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F2 population. *DNA Res.* 8:301-310.
Heidstra, R., Nilsen, G., Martinez-Abarca, F., van Kammen, A., and Bisseling, T. 1997. Nod factor-induced expression of leghemoglobin to study the mechanism of NH₄NO₃ inhibition on root hair deformation. *Mol. Plant Microbe Interact.* 10:215-220.
Imaizumi-Anraku, H., Kawaguchi, M., Koiwa, H., Akao, S., and Syono, K. 1997. Two ineffective-nodulating mutants of *Lotus japonicus*—Different phenotypes caused by the blockage of endocytotic bacterial release and nodule mutation. *Plant Cell Physiol.* 38:871-881.
Ito, M., Miyamoto, J., Mori, Y., Fujimoto, S., Uchiumi, T., Abe, M., Suzuki, A., Tabata, S., and Fukui, K. 2000. Genome and chromosome dimensions of *Lotus japonicus*. *J. Plant Res.* 113:435-442.
Jones, S. J., Riddle, D. L., Pouzyrev, A. T., Velculescu, V. E., Hillier, L., Eddy, S. R., Stricklin, S. L., Baillie, D. L., Waterston, R., and Marra, M. A. 2001. Changes in gene expression associated with developmental arrest and longevity in *Caenorhabditis elegans*. *Genome Res.* 11:1349-1352.
Jung, S. H., Lee, J. Y., and Lee, D. H. 2003. Use of SAGE technology to reveal changes in gene expression in *Arabidopsis* leaves undergoing cold stress. *Plant Mol. Biol.* 52:553-567.
Kapranov, P., Rount, S. M., Bankaitis, V. A., de Bruijn, F. J., Szczyglowski, K. 2001. Nodule-specific regulation of phosphatidylinositol transfer protein expression in *Lotus japonicus*. *Plant Cell* 13:1369-1382.
Kawaguchi, M. 2000. *Lotus japonicus* "Miyakojima" MG-20: An early flowering accession suitable for indoor handling. *J. Plant Res.* 113:507-509.
Kaneko, T., Asamizu, E., Kato, T., Sato, S., Nakamura, Y., and Tabata, S. 2003. Structural analysis of a *Lotus japonicus* genome. III. Sequence features and mapping of sixty-two TAC clones which cover the 6.7 Mb regions of the genome. *DNA Res.* 10:27-33.
Kato, T., Kaneko, T., Sato, S., Nakamura, Y., and Tabata, S. 2000. Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. *DNA Res.* 7:323-330.

- 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 Res.* 10:277-285.
- Kistner, C., and Parniske, M. 2002. Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci.* 7:511-518.
- Kouchi, H., Shimomura, K., Hata, S., Hirota, A., Wu, G. J., Kumagai, H., Tajima, S., Suganuma, N., Suzuki, A., Aoki, T., Hayashi, M., Yokoyama, T., Ohyama, T., Asamizu, E., Kuwata, C., Shibata, D., and Tabata, S. 2004. Large-scale analysis of gene expression profile during early stages of root nodule formation in a model legume, *Lotus japonicus*. *DNA Res.* 11:263-274.
- Krusell, L., Madsen, L. H., Sato, S., Aubert, G., Genua, A., Szczyglowski, K., Duc, G., Kaneko, T., Tabata, S., de Bruijn, F., Pajuelo, E., Sandal, N., and Stougaard, J. 2002. Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature* 420:422-426.
- Laplaze, L., Gherbi, H., Duhoux, E., Pawlowski, K., Auguy, F., Guernache, F., Franche, C., and Bogusz, D. 2002. Symbiotic and non-symbiotic expression of cgMT1, a metallothionein-like gene from the actinorhizal tree *Casuarina glauca*. *Plant Mol. Biol.* 49:81-92.
- Lee, J. Y., and Lee, D. H. 2003. Use of serial analysis of gene expression technology to reveal changes in gene expression in *Arabidopsis* pollen undergoing cold stress. *Plant Physiol.* 132:517-529.
- Martirani, L., Stiller, J., Mirabella, R., Alfano, F., Lamberti, A., Radutoiu, S. E., Iaccarino, M., Gresshoff, P. M., and Chiurazzi, M. 1999. T-DNA tagging of nodulation- and root-related genes in *Lotus japonicus*: Expression patterns and potential for promoter trapping and insertional mutagenesis. *Mol. Plant Microbe Interact.* 12:275-284.
- Matsumura, H., Nirasawa, S., and Terauchi, R. 1999. Technical advance: Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). *Plant J.* 20:719-726.
- Matsumura, H., Nirasawa, S., Kiba, A., Urasaki, N., Saitoh, H., Ito, M., Kawai-Yamada, M., Uchimiyu, H., and Terauchi, R. 2003a. Overexpression of Bax inhibitor suppresses the fungal elicitor-induced cell death in rice (*Oryza sativa* L.) cells. *Plant J.* 33:425-434.
- Matsumura, H., Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P., Kahl, G., Reuter, M., Kruger, R., and Terauchi, R. 2003b. Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *Proc. Natl. Acad. Sci. U.S.A.* 100:15718-15723.
- Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., Bucher, P., Copley, R. R., Courcelle, E., Das, U., Durbin, R., Falquet, L., Fleischmann, W., Griffiths-Jones, S., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lopez, R., Letunic, I., Lonsdale, D., Silventoinen, V., Orchard, S. E., Pagni, M., Peyrcy, D., Ponting, C. P., Selengut, J. D., Servant, F., Sigrist, C. J. A., Vaughan, R., and Zdobnov, E. M. 2003. The InterPro database, 2003 brings increased coverage and new features. *Nucleic Acids Res.* 31:315-318.
- Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Sato, S., and Tabata, S. 2002. Structural analysis of a *Lotus japonicus* genome. II. Sequence features and mapping of sixty-five TAC clones which cover the 6.5-mb regions of the genome. *DNA Res.* 9:63-70.
- Nishimura, R., Hayashi, M., Wu, G. J., Kouchi, H., Imaizumi-Anraku, H., Murakami, Y., Kawasaki, S., Akao, S., Ohmori, M., Nagasawa, M., Harada, K., and Kawaguchi, M. 2002. HAR1 mediates systemic regulation of symbiotic organ development. *Nature* 420:426-429.
- Patankar, S., Munasinghe, A., Shoaibi, A., Cummings, L. M., and Wirth, D. F. 2001. Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. *Mol. Biol. Cell* 12:3114-3125.
- Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W., and Vogelstein, B. 1997. A model for p53 induced apoptosis. *Nature* 389:300-305.
- Poulsen, C., and Podenphant, L. 2002. Expressed sequence tags from roots and nodule primordia of *Lotus japonicus* infected with *Mesorhizobium loti*. *Mol. Plant Microbe Interact.* 15:376-379.
- Riely, B. K., Ane, J. M., Penmetas, R. V., Cook, D. R. 2004. Genetic and genomic analysis in model legumes bring Nod-factor signaling to center stage. *Curr. Opin. Plant Biol.* 7:408-413.
- Sandal, N., Krusell, L., Radutoiu, S., Olbryt, M., Pedrosa, A., Stracke, S., Sato, S., Kato, T., Tabata, S., Parniske, M., Bachmair, A., Ketelsen, T., and Stougaard, J. 2002. A genetic linkage map of the model legume *Lotus japonicus* and strategies for fast mapping of new loci. *Genetics* 161:1673-1683.
- 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 Res.* 8:311-318.
- Schauser, L., Handberg, K., Sandal, N., Stiller, J., Thykjaer, T., Pajuelo, E., Nielsen, A., and Stougaard, J. 1998. Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of *Lotus japonicus*. *Mol. Gen. Genet.* 259:414-423.
- Schauser, L., Roussis, A., Stiller, J., and Stougaard, J. 1999. A plant regulator controlling development of symbiotic root nodules. *Nature* 402:191-195.
- Searle, I. R., Men, A. E., Laniya, T. S., Buzas, D. M., Iturbe-Ormaetxe, I., Carroll, B. J., and Gresshoff, P. M. 2003. Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. *Science* 299:109-112.
- Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Szczyglowski, K., and Parniske, M. 2002. A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* 417:959-962.
- Szczyglowski, K., Hamburger, D., Kapranov, P., and de Bruijn, F. J. 1997. Construction of a *Lotus japonicus* late nodulin expressed sequence tag library and identification of novel nodule-specific genes. *Plant Physiol.* 114:1335-1346.
- Szczyglowski, K., Shaw, R. S., Wopereis, J., Copeland, S., Hamburger, D., Kasiborski, B., Dazzo, F. B., and de Bruijn, F. J. 1998. Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. *Mol. Plant Microbe Interact.* 11:684-697.
- Uchiumi, T., Shimoda, Y., Tsuruta, T., Mukoyoshi, Y., Suzuki, A., Senoo, K., Sato, S., Kato, T., Tabata, S., Higashi, S., and Abe, M. 2002. Expression of symbiotic and nonsymbiotic globin genes responding to micro-symbionts on *Lotus japonicus*. *Plant Cell Physiol.* 43:1351-1358.
- Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler, K. W. 1995. Serial analysis of gene expression. *Science* 270:484-487.
- Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, B., Basrai, M. A., Bassett, D. E. Jr., Hieter, P., Vogelstein, B., and Kinzler, K. W. 1997. Characterization of the yeast transcriptome. *Cell* 88:243-251.
- Vicient, C. M., and Delseny, M. 1999. Isolation of total RNA from *Arabidopsis thaliana* seeds. *Anal. Biochem.* 268:412-413.
- Webb, K. J., Skot, L., Nicholson, M. N., Jorgensen, B., and Mizen, S. 2000. *Mesorhizobium loti* increases root-specific expression of a calcium-binding protein homologue identified by promoter tagging in *Lotus japonicus*. *Mol. Plant Microbe Interact.* 13:606-616.
- Wittenberg, J. B., Appleby, C. A., Bergersen, F. J., and Turner, G. L. 1975. Leghemoglobin: The role of hemoglobin in the nitrogen-fixing legume root nodule. *Ann. N.Y. Acad. Sci.* 244:28-34.
- Yamada, K., Lim, J., Dale, J. M., Chen, H., Shinn, P., Palm, C. J., Southwick, A. M., Wu, H. C., Kim, C., Nguyen, M., Pham, P., Cheuk, R., Karlin-Newmann, G., Liu, S. X., Lam, B., Sakano, H., Wu, T., Yu, G., Miranda, M., Quach, H. L., Tripp, M., Chang, C. H., Lee, J. M., Toriumi, M., Chan, M. M., Tang, C. C., Onodera, C. S., Deng, J. M., Akiyama, K., Ansari, Y., Arakawa, T., Banh, J., Banno, F., Bowser, L., Brooks, S., Carninci, P., Chao, Q., Choy, N., Enju, A., Goldsmith, A. D., Gurjal, M., Hansen, N. F., Hayashizaki, Y., Johnson-Hopson, C., Hsuan, V. W., Iida, K., Karnes, M., Khan, S., Koesema, E., Ishida, J., Jiang, P. X., Jones, T., Kawai, J., Kamiya, A., Meyers, C., Nakajima, M., Narusaka, M., Seki, M., Sakurai, T., Satou, M., Tamse, R., Vaysberg, M., Wallender, E. K., Wong, C., Yamamura, Y., Yuan, S., Shinozaki, K., Davis, R. W., Theologis, A., and Ecker, J. R. 2003. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* 302:842-846.
- Yelin, R., Dahary, D., Sorek, R., Levanon, E. Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., Nemzer, S., Pinner, E., Walach, S., Bernstein, J., Savitsky, K., and Rotman, G. 2003. Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.* 21:379-386.
- Yu, J., Zhang, L., Hwang, P. M., Rago, C., Kinzler, K. W., and Vogelstein, B. 1999. Identification and classification of p53-regulated genes. *Proc. Natl. Acad. Sci. U.S.A.* 96:14517-14522.
- Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B., and Kinzler, K. W. 1997. Gene expression profiles in normal and cancer cells. *Science* 276:1268-1272.

AUTHOR-RECOMMENDED INTERNET RESOURCES

- Kazusa DNA Research Institute *Lotus japonicus* EST index:
www.kazusa.or.jp/en/plant/lotus/EST
- The Arabidopsis Information Resource (TAIR) database:
www.arabidopsis.org
- MIPS *Arabidopsis thaliana* database: mips.gsf.de/proj/thal/db/index.html
- Johns Hopkins School of Medicine SAGE website:
www.sagenet.org/protocol/index.htm