

Construction of Signature-tagged Mutant Library in *Mesorhizobium loti* as a Powerful Tool for Functional Genomics

Yoshikazu SHIMODA¹, Hisayuki MITSUI², Hiroko KAMIMATSUSE², Kiwamu MINAMISAWA², Eri NISHIYAMA², Yoshiyuki OHTSUBO², Yuji NAGATA², Masataka TSUDA², Sayaka SHINPO¹, Akiko WATANABE¹, Mitsuyo KOHARA¹, Manabu YAMADA¹, Yasukazu NAKAMURA¹, Satoshi TABATA¹, and Shusei SATO^{1,*}

Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan¹ and Graduate School of Life Sciences, Tohoku University, Katahira, Aoba-ku, Sendai 980-8577, Japan²

(Received 30 April 2008; accepted 25 June 2008; published online 25 July 2008)

Abstract

Rhizobia are nitrogen-fixing soil bacteria that establish endosymbiosis with some leguminous plants. The completion of several rhizobial genome sequences provides opportunities for genome-wide functional studies of the physiological roles of many rhizobial genes. In order to carry out genome-wide phenotypic screenings, we have constructed a large mutant library of the nitrogen-fixing symbiotic bacterium, *Mesorhizobium loti*, by transposon mutagenesis. Transposon insertion mutants were generated using the signature-tagged mutagenesis (STM) technique and a total of 29 330 independent mutants were obtained. Along with the collection of transposon mutants, we have determined the transposon insertion sites for 7892 clones, and confirmed insertions in 3680 non-redundant *M. loti* genes (50.5% of the total number of *M. loti* genes). Transposon insertions were randomly distributed throughout the *M. loti* genome without any bias toward G+C contents of insertion target sites and transposon plasmids used for the mutagenesis. We also show the utility of STM mutants by examining the specificity of signature tags and test screenings for growth- and nodulation-deficient mutants. This defined mutant library allows for genome-wide forward- and reverse-genetic functional studies of *M. loti* and will serve as an invaluable resource for researchers to further our understanding of rhizobial biology.

Key words: *Mesorhizobium loti*; signature-tagged mutagenesis; mutant library; reverse genetics

1. Introduction

Nitrogen-fixing rhizobia are of great agronomic benefit, allowing many leguminous crops to be grown without nitrogen fertilizer by forming endosymbiotic relationships with the plants. In the course of the symbiosis, rhizobia elicit formation of specialized organs, 'root nodules', on the roots of compatible host legumes. Inside the nodules, the bacteria convert inert atmospheric dinitrogen into biologically usable ammonia. Since this symbiotic relationship is

established by the results of highly regulated molecular dialogues between host plant and rhizobia, the system of symbiotic nitrogen fixation is therefore a very suitable model for studies of plant–microbe interactions.

The agronomic and biological importance of the rhizobia has accelerated the determination of the full genome sequences of several rhizobial species.^{1–6} The availability of complete genome sequences has drastically changed the strategy for studying rhizobial genetics. Gene identification and functional assignment have been accelerated by utilizing the genome sequences. Simultaneously with the completions of rhizobial genome sequences, post-genomic researches on several rhizobia have rapidly progressed. Comprehensive transcriptome or proteome analyses have been conducted to examine physiological states of rhizobia under a

Edited by Naotake Ogasawara

* To whom correspondence should be addressed. Tel. +81 438-52-3935. Fax. +81 438-52-3934. E-mail: ssato@kazusa.or.jp

© The Author 2008. Kazusa DNA Research Institute.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

variety of conditions, including symbiosis with the host legume or nutrient-depleted conditions.⁷⁻¹² In addition, we previously conducted a large-scale interactome analysis in order to propose functional relationships between known and unknown rhizobial genes.¹³ These advances in rhizobial genetics have generated a vast amount of information about gene function and regulation; however, the advances have simultaneously raised the need for development of novel molecular tools which make use of the results of these functional analyses.

One of the most utilized genetic approaches to reveal gene function is the disruption of particular genes and subsequent phenotypic characterization of the mutants. Comprehensive collections of single gene disruptants have been systematically constructed in several model organisms by transposon random mutagenesis or targeted gene disruption approaches.¹⁴⁻¹⁸ These analyses have contributed to the discovery of essential genes of the organisms and have provided fundamental tools for functional studies of many known and unknown genes. In rhizobia, recent genome-wide analyses have provided various intriguing and useful starting points for more detailed functional studies;⁷⁻¹³ however, there are still large gaps between such systematic analyses and the availability of the corresponding gene mutants. Such gaps have resulted primarily because targeted gene disruption achieved by homologous recombination is not suitable for systematic generation of mutants in rhizobia because of its low efficiency. In addition, screening of mutants in some test conditions is problematic because phenotypes of individual mutants have to be evaluated one-by-one. Therefore, more effective approaches to generate large mutant collections and allow simultaneous screening of many mutants are needed.

In this study, we conducted the first large-scale random mutagenesis of *Mesorhizobium loti*, an endosymbiont of the model legume *Lotus japonicus*, using a signature-tagged mutagenesis (STM) technique. STM is based on transposon insertional mutagenesis that allows large numbers of mutants to be analyzed simultaneously. This is accomplished by tagging each mutant with a unique short DNA sequence so that it can subsequently be distinguished from other mutants by detecting unique signature tags. This system has been used with a variety of bacteria, and, in particular, has been applied to discovery of virulence genes in several microbial pathogens.¹⁹ Recently, a large STM library was constructed in *Sinorhizobium meliloti* strain 1021,²⁰ a symbiont of *Medicago truncatula* and contributed to the discovery of many genes relevant to symbiosis and competitiveness.²¹ However, genome-wide generation of disruption mutants in *M. loti*, in addition to the set of *S. meliloti*, is also very valuable for several reasons. First, *M. loti* and *S. meliloti* exhibit different

aspects as symbiont for legumes. They naturally exhibit distinct host-specificity and form different types of nodules (determinate and indeterminate nodules) on their respective model legume hosts, *L. japonicus* and *M. truncatula*. With respect to this symbiosis, previous studies have discovered several rhizobial factors and metabolic cycles required for nodulation and nitrogen-fixation.²²⁻²⁴ However, comparative verifications of the significance of these factors on respective symbioses have not been fully explored. Second, comparative genomics analyses have revealed the existence of genes that are conserved, as well as many that are specific to each species,^{2,25} but the molecular tools available for the functional comparisons of these genes are extremely limited. Therefore, large collections of disruption mutants on the individual species are necessary to reveal common and unique biological aspects of the two species.

In this report, we developed a large signature-tagged mutant library of *M. loti* in order to provide fundamental tools for functional studies of *M. loti* genes and characterized it with the gene coverage and transposon insertion distribution of the library. Sequencing of transposon insertion sites enabled us to collect defined sets of transposon mutants. Furthermore, pilot experiments showed the utility of the mutants and provide basic experimental conditions that can be used for future mutant screenings. This large signature-tagged mutant library will be an important and powerful resource for future functional genomics in *M. loti*.

2. Materials and methods

2.1. Bacterial strains and growth conditions

Escherichia coli strains DH5 α and MT616 were used as hosts for the transposon plasmids and pRK600 helper plasmid,²⁶ respectively. Transposon insertion mutants were generated in the wild-type *M. loti* strain MAFF303099. *Mesorhizobium loti* was grown at 30°C in tryptone yeast (TY) extract medium²⁷ supplemented with phosphomycin (100 μ g/ml). *Escherichia coli* was grown at 37°C in Luria-Bertani (LB) medium. LB medium was supplemented with streptomycin (100 μ g/ml) and spectinomycin (100 μ g/ml) for DH5 α donor cells harboring transposon plasmids and with chloramphenicol (20 μ g/ml) for MT616 helper cells. Transposon insertion mutants were selected on TY medium supplemented with streptomycin, spectinomycin and phosphomycin at the same concentration as above.

2.2. Construction of signature-tagged transposon plasmids

The backbone of the transposon plasmid was pTnMod-OGM.²⁸ The Ω interposon cassette was

excised from pHP45 Ω ²⁹ and inserted by replacing a *SacI* fragment containing the gentamicin resistance gene cassette of pTnMod-OGm. The resultant construct was designated pTnMod-O Ω . Twenty-seven different 21-base oligonucleotides (Supplementary Table S1) were annealed to complementary molecules to make double-strand tags. Each oligonucleotide tag was inserted individually into the blunted *KpnI* site of pTnMod-O Ω (Fig. 1). These 27 signature-tagged constructs were designated pTnMod-O Ω -tag1 to pTnMod-O Ω -tag41.

2.3. Transposon mutagenesis

Introduction of tagged transposon plasmids into *M. loti* was carried out by tri-parental mating.³⁰ *Escherichia coli* strains DH5 α and MT616 (containing the pRK600 helper plasmid) were used as the conjugation donor and helper, respectively. An overnight culture of donor and helper (0.5 ml each) was mixed with 1 ml of a 2-day culture of *M. loti*. After washing several times with sterilized water, the cell mixture was suspended in 50 μ l of sterilized water and spotted onto a sterilized membrane filter (MILLIPORE, Billerica, MA, USA, pore size 0.45 μ m, Cat. No. HAWG047S3) that was placed on TY agar medium without antibiotics. After an 8 h incubation on the membrane filter at 30°C, all the cells were collected and suspended in 1 ml of sterilized water. Suspended cells were plated on TY selection medium supplemented with phosphomycin, spectinomycin and streptomycin, and grown further

until positive colonies appeared. Finally, individual positive colonies were picked up randomly in 96-well plates and stored at -80°C.

2.4. Determination of transposon insertion sites

Transposon insertion sites were determined by direct sequencing of transposon-genome flanking region. Cell culture and genomic DNA isolation were carried out in 96-well plate format. Five microliter of stocked mutant clones were inoculated into 1 ml of TY medium containing phosphomycin (100 μ g/ml), streptomycin (100 μ g/ml) and spectinomycin (100 μ g/ml), and the cultures were grown in 96-well deep-well plate (Axygen Bioscience, Union City, CA, USA) at 30°C for 2 days. After centrifugation of cultured cells at 4000 rpm for 15 min, genomic DNA of mutants was prepared using AquaPure Genomic DNA isolation kit (Bio-Rad Laboratories, Tokyo, Japan). Collected cells in each well were suspended in 150 μ l of Genomic DNA Lysis Solution and were incubated at 80°C for 15 min. RNaseA solution (0.75 μ l) was added to cell lysate, and the mixture was incubated at 37°C for 60 min. After the incubation on ice for 15 min, 50 μ l of protein precipitation solution was added to each well and the samples were mixed well by vortex. Samples were then transferred individually to 1.5 ml microcentrifuge tube and were centrifuged at 15 000 rpm for 15 min. Supernatants (100–125 μ l) were transferred to new 96-well PCR plate containing 100 μ l of isopropanol and were mixed well by inverting plate. Genomic DNA precipitated by centrifugation (4000 rpm, 30 min) was rinsed three times with 200 μ l of 70% ethanol. Genomic DNA was then dried and dissolved in 20 μ l of DNA hydration solution. After heating the samples at 65°C for 60 min, concentration of genomic DNA was measured and ~2 μ g of genomic DNA was used for direct sequence. Direct sequencing was performed with the STM-seq primer (5'-TTTGCTGGCCTTTTGCTCACATGTTCTTTC-3') using the following PCR conditions: 98°C for 3 min, followed by 70 cycles of 97°C for 15 s and 60°C for 4 min. The composition of reaction mixture for direct sequence was as follows. Two micrograms of genomic DNA, 1 μ l of STM-seq primer (3.2 μ M), 1 μ l of BigDye Terminator premix (Applied Biosystem, Foster City, CA, USA), 3.5 μ l of 5 \times sequencing buffer (Applied Biosystems) and sterilized water were added up to 20 μ l.

For some clones which were not successfully identified by direct sequencing, we determined the transposon integration sites by two-round inverse PCR. For inverse PCR, genomic DNA (~1 μ g) isolated from each mutant was digested with *SalI* and *XhoI*, and then self-ligated using a DNA ligation kit (TAKARA

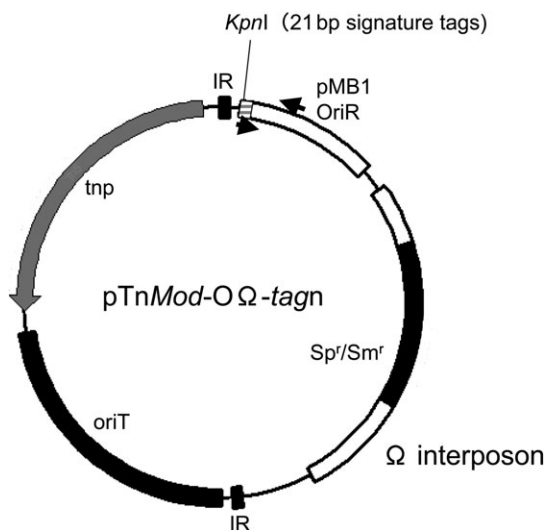


Figure 1. Vector construct used for STM of *M. loti*. The tagged transposon plasmids were constructed from the backbone of pTnMod-OGm.²⁸ The omega-interposon cassette (*Sp'*/*Smr'*) was excised from pHP45 Ω ²⁹ and inserted into the *SacI* site of pTnMod-OGm. Oligonucleotide tags (21 bp) were incorporated into the *KpnI* site. Arrows indicate the locations of the STM common primer and the tag-specific primer (Supplementary Table S1). OriT is an RP4 origin of transfer. IR and tnp represent the inverted repeat and Tn5 transposase, respectively.

Bio, Shiga, Japan). Using a self-ligated DNA template and a primer pair (5'-TTCGCCACCTCTGACTTG AGCGTCG-3' and 5'-GAATTGATCCGGTGGATGAC-3'), first round PCR was performed using the following cycling conditions: 98°C for 2 min, followed by 35 cycles of 97°C for 15 s, 55°C for 45 s, 72°C for 3 min and a final extension at 72°C for 10 min. A 1 µl aliquot from the first round PCR product was added to a secondary PCR reaction mixture containing a nested primer pair (5'-TTTGCTGGCCTTTTGCT CACATGTTCTTTC-3' and 5'-ACGGTTTACAAGCATAA AGC-3'). Second round PCR was performed using the following PCR conditions: 98°C for 2 min, followed by 35 cycles of 97°C for 15 s, 57°C for 45 s, 72°C for 3 min and a final extension at 72°C for 10 min. After the secondary PCR, products were purified by polyethylene glycol precipitation and suspended in 10 µl distilled water. The composition of sequence reaction mixture with amplified PCR product was as follows. Two microliters of PCR product, 1 µl of sequencing primer (5'-ACGGTTTA CAAGCATAAAGC-3') (3.2 µM), 1 µl of BigDye Terminator premix (Applied Biosystems), 3.5 µl of 5× sequencing buffer (Applied Biosystems) and 12.5 µl of sterilized water. Cycle sequencing was performed at 96°C for 1 min and then at 96°C for 10 s, 50°C for 5 s and 60°C for 4 min for 25 cycles. All sequencings were analyzed by ABI 3730 autosequencer. The resulting sequences were subsequently subjected to BlastN search against the *M. loti* genome sequence (<http://www.kazusa.or.jp/rhizobase/Mesorhizobium/index.html>) to determine the transposon insertion sites.

2.5. PCR and real-time PCR

The specificity of oligonucleotide tags was examined by PCR. Tag-containing DNA fragments (160 bp) were amplified from genomic DNA using the Herculase II Fusion DNA Polymerase (STRATAGENE, La Jolla, CA, USA), and forward (Supplementary Table S1) and reverse (STM-common, 5'-TTCGCCACCTCTGACTTG AGCGTCG-3') primers were used for amplification. PCR reactions were conducted in GeneAmp PCR system 9700 (Applied Biosystems) using the following conditions: initial denaturation at 98°C for 2 min, followed by 25 cycles of 97°C for 15 s, 68°C for 1 min and a final extension at 72°C for 10 min. Amplified products were analyzed on 2% agarose gels.

The population of growth-deficient mutant within a mutant pool was examined by real-time PCR. Genomic DNA isolated from input and output pools was used as the DNA template. Real-time PCR was performed using a DyNAmo HS SYBR Green qPCR kit (Finnzyme, Espoo, Finland) and reactions were detected by a DNA Engine Opticon system (Bio-Rad). The conditions used for real-time PCR are as follows:

15 min at 95°C, followed by 35 cycles of 94°C for 30 s, 58°C for 30 s and 72°C for 30 s and a final extension at 72°C for 10 min. Quantitative and specificity (melting curve) analyses of amplified PCR products were carried out using associated software in the Opticon system (Opticon monitor3) by following the manufacture's protocol.

2.6. Detection of signature tags from plant samples

Seeds of *L. japonicus* (MG-20 Miyakojima) were sterilized and germinated on half-strength B&D agar medium. Six-day-old seedlings were transferred to B&D agar medium or inoculation pots filled with vermiculite and were then inoculated with 1×10^7 cell/ml of mutant mixture.

For tag detection from nodules, detached nodules were washed twice with a solution of 0.5%-SDS-100 mM NaCl and then soaked in 70% ethanol for 5 min. After five times washes with sterilized water, total DNA was isolated using a DNeasy plant mini kit (Qiagen, Hilden, Germany). For amplification of a tag-containing fragment from a single nodule, nodule was washed as above and crushed in 700 µl sterilized water. After boiling the samples at 95°C for 5 min, 5 µl of supernatant was used for PCR. PCR reactions with nodule samples used the following conditions: initial denaturation at 98°C for 2 min, followed by 25 cycles of 97°C for 15 s, 68°C for 1 min and a final extension at 72°C for 10 min. Amplified products were analyzed on 2% agarose gels.

3. Results and discussion

3.1. Construction of a signature-tagged mutant library

In this study, we adapted transposon-based STM techniques to genome-wide mutagenesis of the symbiotic bacterium, *M. loti*. As a transposon construct, the plasposon plasmid pTnMod-OGm which carries a Tn5-based mini-transposon and conditional origin of replication²⁸ was used for mutagenesis (Fig. 1). In order to screen for genes which were inactivated by transposon insertion and those affected polarity effect of the inserted transposon, we introduced the Ω interposon cassette by replacing the gentamycin resistance gene (aacC1) of pTnMod-OGm (Fig. 1). Since pTnMod is a broad host range suicide vector and the insertion of the Ω fragment confers spectinomycin and streptomycin resistance, only mutants in which the transposon fragment is integrated into the *M. loti* genome can be selected effectively by their antibiotic resistance.

In order to distinguish individual transposon mutants, we introduced 21 bp short oligonucleotide tags that have been used in a previous study.³¹ Since these oligonucleotide tags were designed to have similar G+C content and melting temperature, they

Table 1. Summary of experimental results

Mutants collected	29 330
Mutants sequenced for mapping of transposon insertion site	9344
Mapped insertion locations	7892
Unique insertion locations	7586
Insertion between ORFs (intergenic region)	1156
Chromosome	1018
pMLa	80
pMLb	58
Insertion inside ORFs	6430
Chromosome	6000
pMLa	318
pMLb	112
Insertion inside ORFs (non-redundant)	3680
Genes supported by more than two mutant alleles	1592

are feasible for PCR-based mutant screening. Considering the isolation of symbiotic mutants from root nodules in which reduction of variation was expected, number of tag variation need to be restricted to obtain reliable results of screening from a reasonable number of plant materials (root nodules). Therefore, we further selected tags from the original 41 tags³¹ by analyzing their specificity of PCR amplification. We selected a total of 29 unique tags based on the reproducibility of PCR amplification and the plasmid plasmids tagged with each of 29 tags (pTnMod-Ω-tagn) were then individually mobilized into wild-type *M. loti* by tri-parental mating. Among the 29 tags, two tag plasmids which yielded significantly small numbers of transposon mutant colonies were eliminated, and the remaining 27 tags (Supplementary Table S1) were used for further analysis (name of tags are designated according to Hunt et al.³¹). In order to construct independent mutant libraries, tri-parental mating and collection of transposon mutants were carried out separately for each oligonucleotide tag. As the result of 63 independent tri-parental mating experiments, 29 330 independent transposon mutants were obtained (Table 1). Individual mutants were stored in 96-well plates and mutant names (ID) were assigned according to their position in the plate with prefix indicating tag number followed by 'T' (i.e. 07T02d09 is a mutant integrated with tag7 plasmid).

The coverage of the collected mutants is approximately fourfold with respect to the 7281 ORFs of *M. loti*.¹ If the transposon insertions are evenly distributed throughout the *M. loti* genome, the mean distance between independent insertions is estimated to be 259 bp, and thus 91% of *M. loti* genes are expected to carry at least one insertion since *M. loti* genes shorter than 259 bp account for 9% of all genes.

3.2. Utility of STM mutants

The STM technique allows for the simultaneous screening of multiple mutants in mixed populations

by detecting unique DNA tags. However, previous studies have reported that non-specific amplification or failure in tag amplification from mutant chromosomal DNA is a major inherent problem that affects the efficiency and reproducibility of the screening procedure.^{31,32} Therefore, the specificity and efficiency of tag amplification after integration into the genome needs to be tested in the respective targeted organisms. To investigate the possibility of cross-detection among the 27 signature tags used in this study, we conducted test amplifications by PCR. Fig. 2A shows representative results in which PCR was performed with genomic DNA from three individual tagged mutants (Tag2, Tag9 and Tag11). When a tag-specific primer (Supplementary Table S1) and the STM common primer (5'-TTCGCCACCTTGACTTGAGCGT CG-3') that anneals 120 bp downstream of the integrated oligonucleotide tag were used for PCR, specific amplification was observed only in reactions containing the appropriate tag-specific primer. A series of PCR reactions performed on the other oligonucleotide tags produced the same results as Fig. 2A, indicating that there are no cross-amplifications among the designed oligonucleotide tags. In addition, all 27 oligonucleotide tags were amplified specifically when genomic DNA from a mixture of mutants was used as the template (Fig. 2B). These test amplifications demonstrated that the applied oligonucleotide tags can be recovered specifically from the *M. loti* genome and that the 27 individually tagged mutants can be distinguished from one another by detecting these signature tags.

In order to test the utility of our mutant sets, we carried out additional pilot experiments using both in vitro and in vivo conditions. In vitro experiments were conducted with a pool of mutants that contains a mutant with an attenuated growth on rich medium. To make a mutant set for the test experiment, one mutant was selected randomly from each 27-tag library and cultured individually. One growth-deficient clone (07T02d09) was included in the pool (Fig. 3A). Independently cultured clones were then mixed at same concentration to make an input pool. As for an output pool, the input pool was diluted 500-fold and cultured further for 2 days. After isolation of genomic DNA from both pools, the tag population was examined by real-time PCR. Real-time PCR was performed using the SYBR Green detection system with the same primer set used to test the cross-reactivity (Fig. 2). Fig. 3 illustrates the results of real-time PCR of the input (Fig. 3B) and output (Fig. 3C) pool samples. Consistent with the growth on agar medium (Fig. 3A), the quantity of PCR product derived from the growth-deficient mutant (07T02d09) was significantly lower than the other tags which were all amplified to similar levels

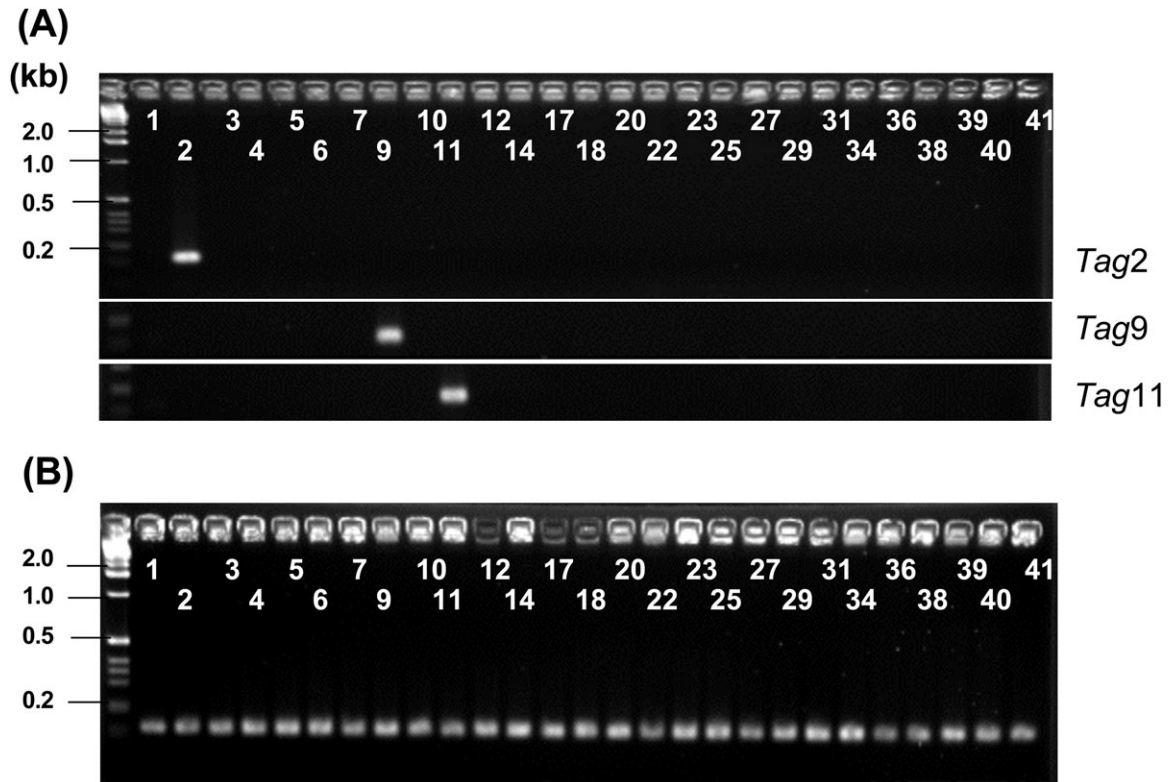


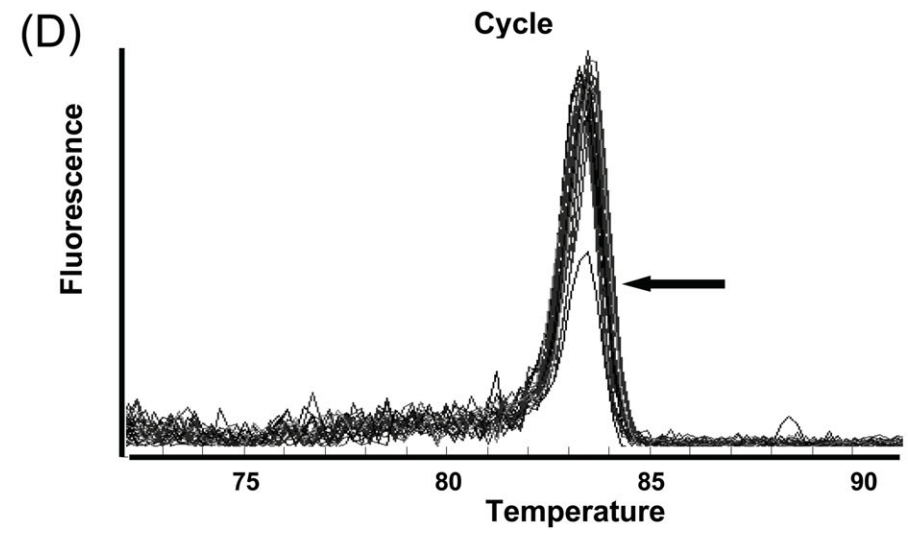
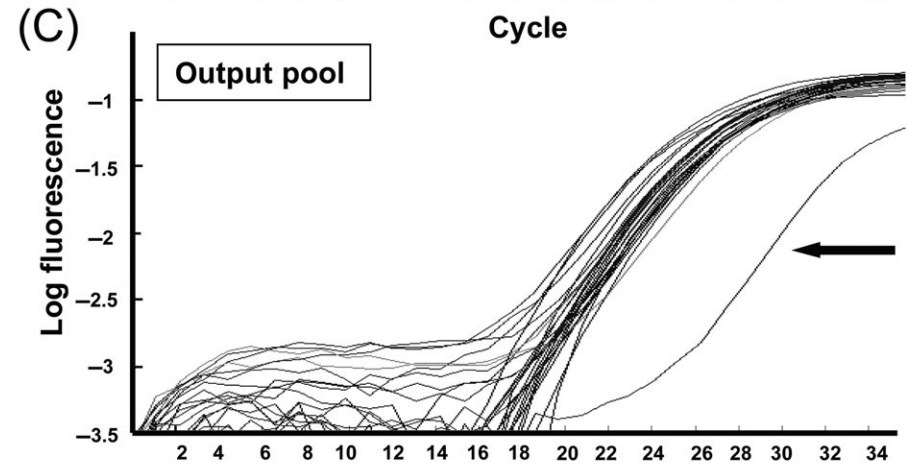
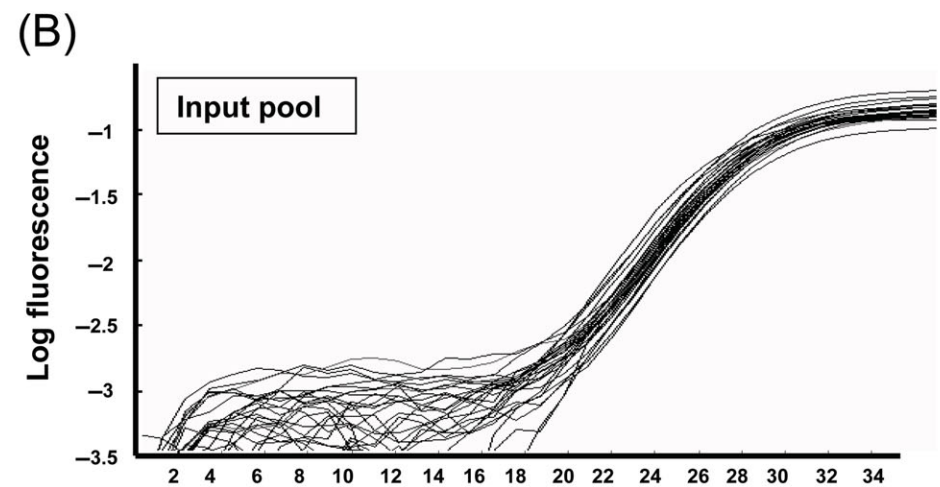
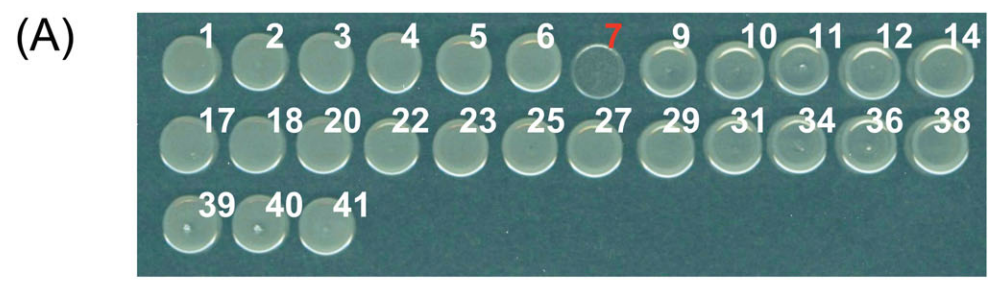
Figure 2. Representative gels of control experiments testing the specificity of signature oligonucleotide tags. PCR was performed with genomic DNA isolated from mutants with pTnMod-O Ω -Tag2 (upper), pTnMod-O Ω -Tag9 (middle) or pTnMod-O Ω -Tag11 (bottom) alone (A), and using genomic DNA isolated from cultures containing a mixture of 27 mutants (B). The STM-common primer was used with one of the 27 different tag-specific primers (Supplementary Table S1) in each reaction. The numbers indicate the combination of primers.

(Fig. 3B and C). In addition, after each round of PCR amplification, a melting curve was generated for each run of 27 samples (Fig. 3D). The melting curve of all reactions exhibited a sharp peak at $\sim 83.5^{\circ}\text{C}$ and a low melting peak was observed in the reactions with the tag7 primer (Fig 3D arrow), indicating that all tag-containing fragments were amplified specifically and that the quantity of tag7-containing fragment was lower than the other tags. This test experiment demonstrated that we can distinguish differences in populations of each tagged mutant within pools by quantifying the signature tags.

In consideration of future screening experiments that would be explored *in planta*, we next examined whether tagged mutants can also be detected from

plant samples (root nodules). For these test experiments, one mutant from each of the 27 tag libraries was mixed in roughly equal quantities and inoculated onto *L. japonicus* roots to form nodules. Four weeks after inoculation, the nodules were detached and used for mutant detection. In order to determine how many mutants exist in a single nodule, we examined 22 nodules formed by the inoculation of two different pools of mutants. This confirmation enables us to estimate how many nodules should be analyzed for mutant screenings. When the presence of mutants in a single nodule was examined by the procedures described in the Materials and methods section, most nodules (18 of 22 tested nodules) were occupied by a mutant with a single tag

Figure 3. Test experiments for identifying a growth-deficient mutant within a pool of mutants. (A) Growth of 27-tagged mutants on rich TY medium. Same concentration of individual mutants was spotted on TY broth medium. Mutant 07T2d09, integrated with pTnMod-O Ω -Tag7, exhibited reduced growth compared with other mutants. Quantitation of PCR amplifications of the input pool (B) and output pool (C). X and Y axes show the number of PCR cycles and log fluorescence of PCR products, respectively. Arrow in C indicates the quantitative curve from PCR reaction with Tag7 primer. The amounts of amplified products in all reactions of input pool were similar (ranging from 0.37- to 1.7-fold compare with average). In output pool, the amount of amplified products of all but tag7 were similar to that of input pool (ranging from 0.25- to 1.6-fold compared with that of input pool), whereas the amount in reaction of tag7 primer was 90-fold lower than that of input pool. (D) Melting curve of the output pool showing the change in fluorescence from 72°C to 91°C . All melting curves are shown as relative intensities among all reactions. Arrow in D indicates the melting curve of PCR reaction with Tag7 primer.



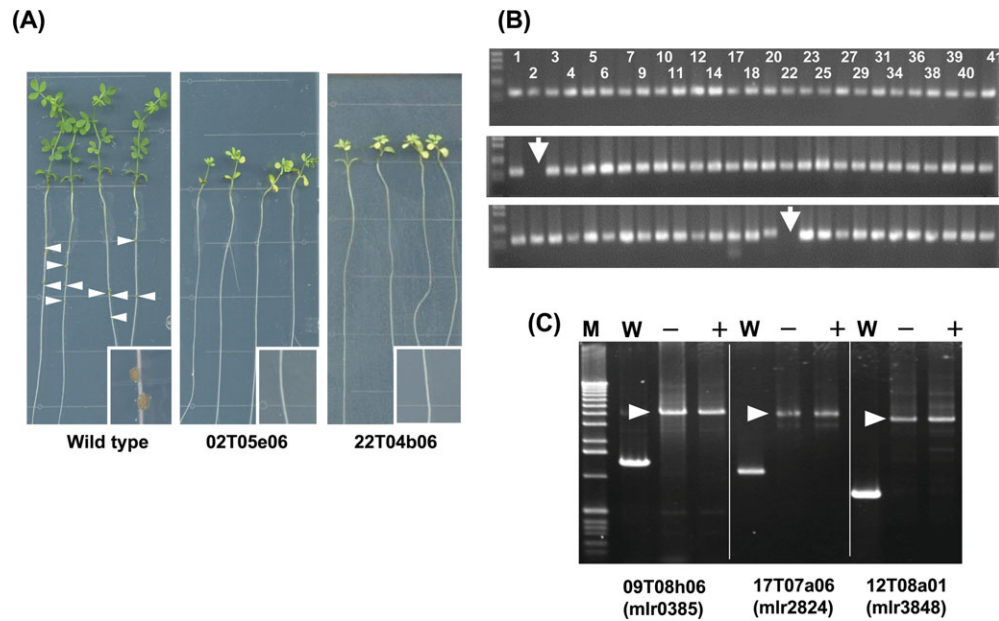


Figure 4. Pilot experiments identifying mutants with a nodulation defective phenotype. **(A)** Phenotype of *L. japonicus* after inoculation with wild-type *M. loti* or with two nodulation-deficient mutants (02T05e06 and 22T04b06). Arrowheads indicate the nodules formed on the roots of *L. japonicus* and the inset show the enlarged image of root and nodules. **(B)** Tag amplification from genomic DNA isolated from input pool (upper) and output pool (middle and bottom). The numbers indicate the combination of primers used. The arrows indicate the absence of PCR products in reactions with the Tag2 or Tag22 primer. **(C)** Representative PCR amplification showing the stability of the integrated transposon fragment in three different mutants before (-) and after (+) inoculation. Arrowheads indicate the fragment containing the integrated transposon element. W, PCR amplification with genomic DNA from wild-type *M. loti*.

(Supplementary Fig. S1). This result was consistent with a previous study in *S. meliloti*,²¹ indicating that the colonization or infection process may have similar features for both determinate and indeterminate nodules, and competition of infection may already occur at the colonization stage on host root hairs. And, as far as we tested, all tags were recovered in a screen of 70 nodules, suggesting that at least 70 nodules should be analyzed for mutant screenings in nodules.

Additionally, we tested whether nodulation mutants can be screened properly for in nodules formed by pools of mutants. Two mutants (02T05e06 and 22T04b06) that exhibit defective in nodulation (Fig. 4A) were included in two distinct test sets (input samples) and inoculated onto *L. japonicus* roots. When we examined the presence of mutants in the total DNA isolated from 120 nodules (output samples), tag2 and tag22 containing fragments were not recovered, whereas all the other tags were detected (Fig. 4B). These results are consistent with the defective nodulation phenotype of the mutants (02T05e06 and 22T04b06), as tested individually (Fig. 4A). During the infection process, integrated transposon elements were stable since identical amplification patterns were obtained before and after infection (Fig. 4C, three representative results are shown).

From the results of these pilot experiments, we concluded that the signature-tagged mutants can be used

for future STM screenings both in vitro and in vivo using our experimental conditions.

3.3. Mapping of transposon insertion sites

Along with the collection of transposon mutants, we determined the transposon insertion sites of individual mutants. Transposon insertion sites were determined by sequencing of transposon-flanking genomic regions, followed by comparison with the reference genome sequence. Determination of transposon insertions revealed that the inverted repeat of both the left and right ends of the integrated transposon were followed by the individual 9 bp direct repeat of the target site, indicating that the transposition were occurred properly by Tn5-transposon integration.³³

The distribution of transposon insertion sites in the *M. loti* genome is illustrated in Fig. 5. Among 9344 mutants examined, insertion sites of 7892 were determined (success rate, 84.5%) that were mapped to 7586 unique insertion sites (Table 1). This proportion indicates that ~96% of mutants carry a unique transposon insertion in their genome and there is no significant redundancy of a clone arising from identical parent clones in the same tri-parental mating experiment. Among the unique insertions, 7018 (92.5%), 398 (5.2%) and 170 (2.3%) were mapped onto the chromosome and two plasmids of *M. loti* (pMLa and pMLb), respectively (all insertion

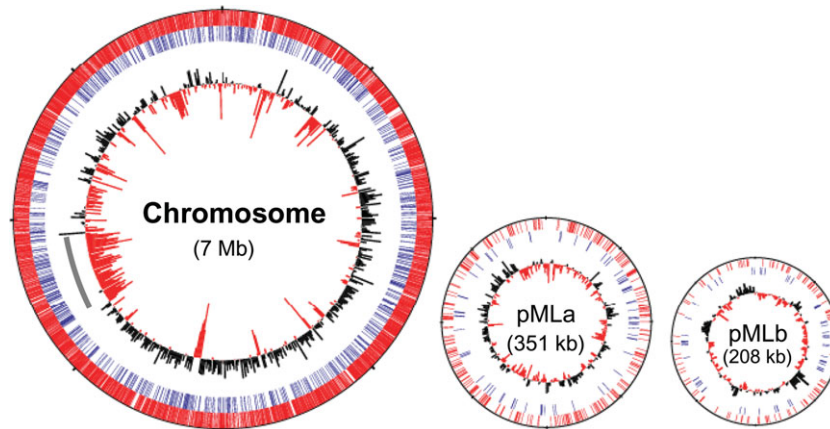


Figure 5. Distribution of transposon insertions in the *M. loti* genome. The bars in the first and second circles show the position of transposon insertion sites within ORFs (red bars) and within intergenic regions (blue bars), respectively. The innermost circles represent the average G+C content calculated in each 10 kb. A positive deviation in G+C content from the average (62.7%) is shown by black bars pointing outward and a negative deviation by red bars pointing inward. Grey bar in chromosome indicates the region of symbiosis island.

sites are listed in Supplementary Table S2 and S3). The proportion of transposon distribution was consistent with the proportion of total size of each *M. loti* replicon (Chr, 92.6%, pMLa 4.6% and pMLb 2.7%). Furthermore, among the unique insertion sites, 6430 (84.8%) were within ORFs, corresponding to the overall proportion (85.6%) of coding sequences within the entire *M. loti* genome.¹

Previous studies have demonstrated that the distribution patterns of transposon insertion into genomes vary depending on the types of transposon and organisms used for mutagenesis.^{34–36} To examine whether the transposon insertions were distributed randomly throughout the *M. loti* genome, we first investigated potential target site preferences for transposon insertion. Transposon insertions occurred in both intragenic and intergenic regions and the orientations of integrated transposons, with respect to the ORF directions of transcription, are present in nearly equal proportions (3274 clones in the same direction as transcription and 3397 clones in the opposite direction) (Supplementary Table S2). In addition, according to a previous report,³⁷ G+C content was calculated from the 9 nucleotides of the insertion-target sites and 10 nucleotides up- and down-stream of target site of 50 randomly selected clones. It showed a high variance ranging from 30% to 88% (Supplementary Table S4), and the average G+C content of the target site (62.7%) was consistent with that of *M. loti* chromosome (62.7%).¹ Furthermore, 50 randomly selected target sites showed no obvious sequence pattern, including palindromic or symmetrical sequences, when these were analyzed by TEIRESIAS, the sequence pattern search algorithm (<http://cbcsrv.watson.ibm.com/Tspd.html>).³⁸ These features indicate that there are no clear preferences in insertion with regard to G+C content or structural features of the target sites.

In order to verify the random distribution of transposon insertion, we additionally investigated the relationship between the number of genes carrying an inserted transposon and the size of the corresponding genes. Fig. 6 shows the distribution of transposon insertion within ORFs. As expected, more transposon insertions occurred in larger ORFs and the number of disrupted genes clearly correlate with the number of *M. loti* ORFs in each size range. In addition, as shown in Fig. 5, there is no obvious correlation between the distribution of transposon insertion and local G+C content of the *M. loti* genome. The symbiosis island, which exhibits features of horizontally transferred genomic element of low G+C content, also contained transposon insertions at rates similar to other chromosomal region (Fig. 5 grey line). Furthermore, this random distribution was observed for all tagged plasmid constructs.

Several previous studies have provided sets of essential genes of organisms predicted from comprehensive random mutagenesis or targeted gene disruption.^{36,39,40} Although here we collected a large number of transposon mutants of *M. loti*, it is currently impossible to estimate the precise number of *M. loti* essential genes because the number of mutants for which transposon insertion sites have been determined is not sufficient to cover all non-essential genes. However, we can predict essential gene candidates among the relatively large size *M. loti* genes since large size *M. loti* genes (>3000 bp) were disrupted at high coverage (83.1%). Among 77 genes whose size is larger than 3000 bp, transposon insertions were not detected in 13 genes, whereas an average of 4.7 insertions occurred in other genes in this size range (Supplementary Table S5). Considering the random distribution of transposon insertion, these 13 genes are candidate *M. loti* essential genes required for growth on rich media. Among the genes,

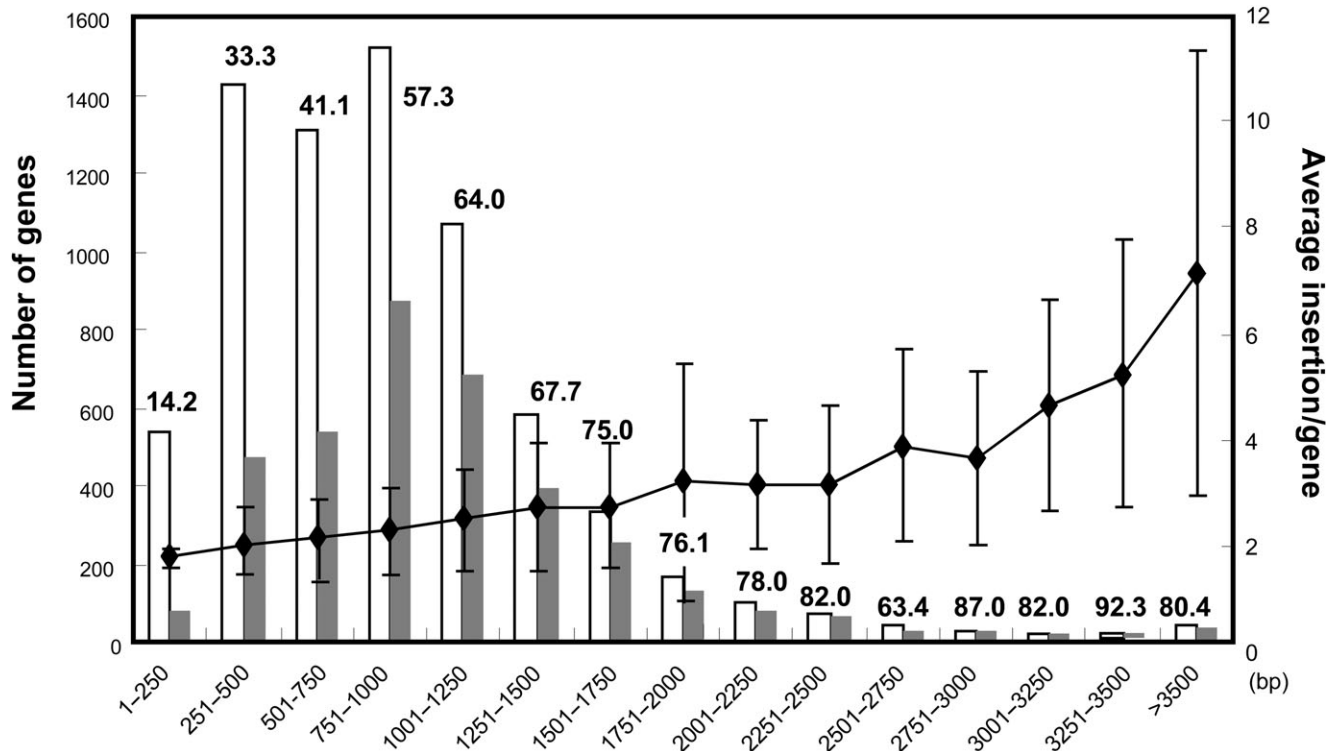


Figure 6. Distribution of transposon insertion within *M. loti* ORFs. The white bars represent the number of *M. loti* genes in each size range and the grey bars represent the number of genes with transposon insertions. The line chart represents the average number of insertion in genes of each size range with standard deviation. Percentages in each bar indicate the coverage of disrupted genes in each size range.

11 genes (excluding *mlr9704* and *mll1090*) are highly conserved in the genome of completely sequenced rhizobia and four genes, *mlr8250* (isoleucyl-tRNA synthetase), *mll0870* (DNA polymerase III alpha subunit), *mlr0276* (RNA polymerase beta subunit) and *mlr0277* (RNA polymerase, beta prime subunit), showed high similarity to essential genes of *E. coli* (*ileS*, *dnaE*, *rpoB* and *rpoC*, respectively). Furthermore, *mlr2517* is orthologous to the essential gene of *Acinetobacter baylyi* ADP1 (*carB*, carbamoyl-phosphate synthase, large subunit).⁴¹ Other essential gene candidates could be those with insertions only near 3'-terminus of the ORF. Mutants for *mll1426*, *mlr0796* and *mlr0585* could be the candidate of essential genes since relatively small number of insertions, all of which were in their 3'-terminus, were detected as listed in Supplementary Table S5.

In bacteria, essential genes are more evolutionarily conserved than non-essential genes.⁴² We compared *E. coli* essential genes and *M. loti* genes [BlastP search was carried out against *E. coli* essential genes listed in the PEC data base (www.shigen.nig.ac.jp/ecoli/pec/index.jsp)] and revealed that a total of 196 *M. loti* ORFs showed high similarity to *E. coli* essential genes. Transposon mutants were obtained for only 21 of these 196 genes (10.7%). This ratio is significantly low since the result of sequence analysis transposon insertion sites identified mutants for

~50% of *M. loti* genes. Among the 21 genes with transposon insertion, eight genes have putative paralogous genes which are considered to complement the function of the disrupted genes. These results indicate that significant overlap exists among sets of essential genes in *E. coli* and *M. loti*. However, saturating collections of transposon mutants under different selective conditions are, of course, needed to confirm these genes cannot actually be disrupted.

3.4. Mutant database

To provide resources for depositing and retrieving information on *M. loti* mutants, we developed a relational database, RhizoGenes (<http://www.kazusa.or.jp/rhizobase/Mesorhizobium/genes2/index.html>). This database operates as a public web database that provides general information about mutants, e.g. mutant IDs, transposon insertion site, list of mutant alleles and experimental results (which can be deposited by users). In this database, individual mutants can be retrieved through keyword search by clone ID, name and functional annotation of disrupted genes. Therefore, when genes have attracted interest on the basis of bioinformatic or functional genomic analyses, appropriate mutants can be searched immediately. The transposon mutants collected in this study will be available from the National Bioresource project of

Japan (<http://www.legumebase.agr.miyazaki-u.ac.jp/index.jsp>).

Furthermore, we have integrated this database into the genome information database, RhizoBase (<http://bacteria.kazusa.or.jp/rhizobase/Mesorhizobium/index.html>), so as to link directly to more detailed information about *M. loti* ORFs, such as operon structure, predicted protein domains and orthologous (paralogous) protein groups.

3.5. Concluding remarks

In the present study, we have described a large-scale transposon mutagenesis in the nitrogen-fixing symbiotic bacterium, *M. loti*. The STM technique was adapted to barcode individual transposon mutants and a large mutant library containing >29 000 independent clones was created using a set of 27 signature tags. Sequencing of the transposon insertion sites of one-third of the mutants revealed a random distribution of transposon insertion and ~50% of *M. loti* genes have been confirmed to be disrupted. Since such a large number of disruption mutants have not previously been available for *M. loti* and the low efficiency of gene-disruption mediated by homologous recombination has been a serious technical bottleneck for functional studies of this symbiont, our mutant collections should be a powerful tool for future functional studies of *M. loti* genes.

Applied STM technique and PCR-based screening allow for the simultaneous discovery of mutants with subtle phenotypic differences, such as reduced competitiveness or survival under various conditions. Additionally, random distribution of transposon insertion ensures the screenings of mutants over the entire *M. loti* genome and reduces the number of clones to be screened. Determination of transposon insertion sites also enables users to select defined, non-redundant sets of mutants depending on individual needs and facilitates rapid identification of causal genes of the mutant phenotypes. Furthermore, individual transposon mutants can be used as single gene disruptants. The availability of single gene disruptants allows functional verifications of individual or paralogous *M. loti* genes as well as comparative functional studies by examining with the orthologous genes of other rhizobium species. Additionally, our resource will be useful for *in planta* studies in the host legume. Actually, our STM mutants have recently contributed to reveal the importance of purine metabolisms in *Lotus-M. loti* symbiosis.²⁴ Use of the mutant collections in *Lotus-M. loti* symbiosis will contribute to discovery of novel rhizobial factors or host plant recognition systems required for the symbiosis. By providing this mutant resource to the research

communities, we hope to contribute to various functional analyses that lead to novel findings and understandings of rhizobial biology.

Supplementary data: Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

This work was supported by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas “Comparative Genomics” from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and the Kazusa DNA Research Institute Foundation.

References

1. Kaneko, T., Nakamura, Y., Sato, S., et al. 2000, Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*, *DNA Res.*, **7**, 331–338.
2. Kaneko, T., Nakamura, Y., Sato, S., et al. 2002, Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110, *DNA Res.*, **9**, 189–197.
3. Galibert, F., Finan, T. M., Long, S. R., et al. 2001, The composite genome of the legume symbiont *Sinorhizobium meliloti*, *Science*, **293**, 668–672.
4. González, V., Santamaría, R. I., Bustos, P., et al. 2006, The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons, *Proc. Natl. Acad. Sci. USA*, **103**, 3834–3839.
5. Young, J. P., Crossman, L. C., Johnston, A. W., et al. 2006, The genome of *Rhizobium leguminosarum* has recognizable core and accessory components, *Genome Biol.*, **7**, R34.
6. Giraud, E., Moulin, L., Vallenet, D., Barbe, V., et al. 2007, Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia, *Science*, **316**, 1307–1312.
7. Bergès, H., Lauber, E., Liebe, C., et al. 2003, Development of *Sinorhizobium meliloti* pilot macroarrays for transcriptome analysis, *Appl. Environ. Microbiol.*, **69**, 1214–1219.
8. Djordjevic, M. A., Chen, H. C., Natera, S., et al. 2003, A global analysis of protein expression profiles in *Sinorhizobium meliloti*: discovery of new genes for nodule occupancy and stress adaptation, *Mol. Plant Microbe Interact.*, **16**, 508–524.
9. Barnett, M. J., Toman, C. J., Fisher, R. F. and Long, S. R. 2004, A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote–host interaction, *Proc. Natl. Acad. Sci. USA*, **101**, 16636–16641.
10. Hoa, L. T. P., Nomura, M. and Tajima, S. 2004, Characterization of bacteroid proteins in soybean nodules formed with *Bradyrhizobium japonicum* USDA110, *Microbiol. Environ.*, **19**, 71–75.
11. Uchiumi, T., Ohwada, T., Itakura, M., et al. 2004, Expression islands clustered on symbiosis island of *Mesorhizobium loti* genome, *J. Bacteriol.*, **45**, 2439–2448.

12. Sarma, A. D. and Emerich, D. W. 2005, Global protein expression pattern of *Bradyrhizobium japonicum* bacteroids: a prelude to functional proteomics, *Proteomics*, **5**, 4170–4184.
13. Shimoda, Y., Shinpo, S., Kohara, M., Nakamura, Y., Tabata, S. and Sato, S. 2008, A large scale analysis of protein–protein interactions in the nitrogen-fixing bacterium *Mesorhizobium loti*, *DNA Res.*, **15**, 13–23.
14. Hutchison, C. A., Peterson, S. N., Gill, S. R., et al. 1999, Global transposon mutagenesis and a minimal mycoplasma genome, *Science*, **286**, 2165–2169.
15. Kobayashi, K., Ehrlich, S. D., Albertini, A., et al. 2003, Essential *Bacillus subtilis* genes, *Proc. Natl. Acad. Sci. USA*, **100**, 4678–4683.
16. Kang, Y., Durfee, T., Glasner, J. D., et al. 2004, Systematic mutagenesis of the *Escherichia coli* genome, *J. Bacteriol.*, **186**, 4921–4930.
17. Baba, T., Ara, T., Hasegawa, M., et al. 2006, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection, *Mol. Syst. Biol.*, **2**, 0008.
18. Glass, J. I., Assad-Garcia, N., Alperovich, N., et al. 2006, Essential genes of a minimal bacterium, *Proc. Natl. Acad. Sci. USA*, **103**, 425–430.
19. Saenz, H. L. and Dehio, C. 2005, Signature-tagged mutagenesis: technical advances in a negative selection method for virulence gene identification, *Curr. Opin. Microbiol.*, **8**, 612–619.
20. Pobigaylo, N., Wetter, D., Szymczak, S., et al. 2006, Construction of a large signature-tagged Mini-Tn5 transposon library and its application to mutagenesis of *Sinorhizobium meliloti*, *Appl. Environ. Microbiol.*, **72**, 4329–4337.
21. Pobigaylo, N., Szymczak, S., Nattkemper, T. M. and Becker, A. 2008, Identification of genes relevant to symbiosis and competitiveness in *Sinorhizobium meliloti* using signature-tagged mutants, *Mol. Plant Microbe Interact.*, **21**, 219–231.
22. Frayse, N., François, C. and Véréna, P. 2003, Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis, *FEBS J.*, **270**, 1365–1380.
23. D'Antuono, A. L., Casabuono, A., Couto, A., Ugalde, R. A. and Lepek, V. C. 2005, Nodule development induced by *Mesorhizobium loti* mutant strains affected in polysaccharide synthesis, *Mol. Plant Microbe Interact.*, **18**, 446–457.
24. Okazaki, S., Hattori, Y. and Saeki, K. 2007, The *Mesorhizobium loti* purB gene is involved in infection thread formation and nodule development in *Lotus japonicus*, *J. Bacteriol.*, **189**, 8347–8352.
25. Guerrero, G., Peralta, H., Aguilar, A., et al. 2005, Evolutionary, structural and functional relationships revealed by comparative analysis of syntenic genes in Rhizobiales, *BMC Evol. Biol.*, **5**, 55.
26. Finan, T. M., Kunkel, B., De Vos, G. F. and Signer, E. R. 1986, Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes, *J. Bacteriol.*, **167**, 66–72.
27. Beringer, J. E. 1974, R factor transfer in *Rhizobium leguminosarum*, *J. Gen. Microbiol.*, **84**, 188–198.
28. Dennis, J. J. and Zylstra, G. J. 1998, Plasposon: modular self-clonig minitransposon derivatives for rapid genetic analysis of Gram-negative bacterial genomes, *Appl. Environ. Microbiol.*, **64**, 2710–2715.
29. Prentki, P. and Kirsch, H. M. 1984, *In vitro* insertional mutagenesis with a selectable DNA fragment, *Gene*, **29**, 303–313.
30. Simon, R., Priefer, U. and Pühler, A. 1983, A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria, *Bio/Technology*, **1**, 784–791.
31. Hunt, T. A., Kooi, C., Sokol, P. A. and Valvano, M. A. 2004, Identification of *Burkholderia cenocepacia* genes required for bacterial survival in vivo, *Infect. Immun.*, **72**, 4010–4022.
32. Autret, A. and Charbit, A. 2005, Lessons from signature-tagged mutagenesis on the infectious mechanisms of pathogenic bacteria, *FEMS Microbiol. Rev.*, **29**, 703–717.
33. Reznikoff, W. S. 2003, Tn5 as a model for understanding DNA transposition, *Mol. Microbiol.*, **47**, 1199–1206.
34. Groh, J. L., Luo, Q., Ballard, J. D. and Krumholz, L. R. 2005, A method adapting microarray technology for signature-tagged mutagenesis of *Desulfovibrio desulfuricans* G20 and *Shewanella oneidensis* MR-1 in anaerobic sediment survival experiments, *Appl. Environ. Microbiol.*, **71**, 7064–7074.
35. Suzuki, N., Okai, N., Nonaka, H., Tsuge, Y., Inui, M. and Yukawa, H. 2006, High-throughput transposon mutagenesis of *Corynebacterium glutamicum* and construction of a single-gene disruptant mutant library, *Appl. Environ. Microbiol.*, **72**, 3750–3755.
36. Gallagher, L. A., Ramage, E., Jacobs, M. A., Kaul, R., Brittnacher, M. and Manoil, A. 2007, A comprehensive transposon mutant library of *Francisella novicida*, a bio-weapon surrogate, *Proc. Natl. Acad. Sci. USA*, **104**, 1009–1014.
37. Mormann, S., Lömker, A., Rückert, C., et al. 2006, Random mutagenesis in *Corynebacterium glutamicum* ATCC 13032 using an IS6100-based transposon vector identified the last unknown gene in the histidine biosynthesis pathway, *BMC Genomics*, **7**, 205.
38. Rigoutsos, I. and Floratos, A. 1998, Combinatorial pattern discovery in biological sequences: the TEIRESIAS algorithm, *Bioinformatics*, **14**, 55–67.
39. Jacobs, M. A., Alwood, A., Thaipisuttikul, I., et al. 2003, Comprehensive transposon mutant library of *Pseudomonas aeruginosa*, *Proc. Natl. Acad. Sci. USA*, **100**, 14339–14344.
40. Liberati, N. T., Urbach, J. M., Miyata, S., et al. 2006, An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants, *Proc. Natl. Acad. Sci. USA*, **103**, 2833–2838.
41. de Berardinis, V., Vallenet, D., Castelli, V., et al. 2008, A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1, *Mol. Syst. Biol.*, **4**, 174.
42. Gerdes, S. Y., Scholle, M. D., Campbell, J. W., et al. 2003, Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655, *J. Bacteriol.*, **185**, 5673–5684.