

## ORIGINAL ARTICLE

# Genomic comparison of *Bradyrhizobium japonicum* strains with different symbiotic nitrogen-fixing capabilities and other Bradyrhizobiaceae members

Manabu Itakura<sup>1</sup>, Kazuhiko Saeki<sup>2</sup>, Hirofumi Omori<sup>3</sup>, Tadashi Yokoyama<sup>4</sup>, Takakazu Kaneko<sup>5</sup>, Satoshi Tabata<sup>5</sup>, Takuji Ohwada<sup>6</sup>, Shigeyuki Tajima<sup>7</sup>, Toshiki Uchiumi<sup>8</sup>, Keina Honnma<sup>9</sup>, Konosuke Fujita<sup>9</sup>, Hiroyoshi Iwata<sup>10</sup>, Yuichi Saeki<sup>11</sup>, Yoshino Hara<sup>1</sup>, Seishi Ikeda<sup>1</sup>, Shima Eda<sup>1</sup>, Hisayuki Mitsui<sup>1</sup> and Kiwamu Minamisawa<sup>1</sup>

<sup>1</sup>Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi, Japan; <sup>2</sup>Department of Biological Science, Faculty of Science, Nara Women's University, Nara, Japan; <sup>3</sup>Department of Biology, Graduate School of Science, Osaka University, Toyonaka, Osaka, Japan; <sup>4</sup>Faculty of Agriculture, Tokyo University of Agriculture and Technology, Fuchu, Tokyo, Japan; <sup>5</sup>Kazusa DNA Research Institute, Kisarazu, Chiba, Japan; <sup>6</sup>Department of Agricultural and Life Sciences, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido, Japan; <sup>7</sup>Department of Life Science, Kagawa University, Miki-cho, Kagawa, Japan; <sup>8</sup>Faculty of Science, Department of Chemistry and Bioscience, Kagoshima University, Kagoshima, Japan; <sup>9</sup>Graduate School of Biosphere Science, Hiroshima University, Higashihiroshima, Hiroshima, Japan; <sup>10</sup>National Agricultural Research Center, Tsukuba, Ibaraki, Japan and <sup>11</sup>Department of Biochemistry and Applied Biosciences, Faculty of Agriculture, Miyazaki University, Miyazaki, Japan

**Comparative genomic hybridization (CGH) was performed with nine strains of *Bradyrhizobium japonicum* (a symbiotic nitrogen-fixing bacterium associated with soybean) and eight other members of the Bradyrhizobiaceae by DNA macroarray of *B. japonicum* USDA110. CGH clearly discriminated genomic variations in *B. japonicum* strains, but similar CGH patterns were observed in other members of the Bradyrhizobiaceae. The most variable regions were 14 genomic islands (4–97 kb) and low G + C regions on the USDA110 genome, some of which were missing in several strains of *B. japonicum* and other members of the Bradyrhizobiaceae. The CGH profiles of *B. japonicum* were classified into three genome types: 110, 122 and 6. Analysis of DNA sequences around the boundary regions showed that at least seven genomic islands were missing in genome type 122 as compared with type 110. Phylogenetic analysis for internal transcribed sequences revealed that strains belonging to genome types 110 and 122 formed separate clades. Thus genomic islands were horizontally inserted into the ancestor genome of type 110 after divergence of the type 110 and 122 strains. To search for functional relationships of variable genomic islands, we conducted linear models of the correlation between the existence of genomic regions and the parameters associated with symbiotic nitrogen fixation in soybean. Variable genomic regions including genomic islands were associated with the enhancement of symbiotic nitrogen fixation in *B. japonicum* USDA110.**

*The ISME Journal* (2009) 3, 326–339; doi:10.1038/ismej.2008.88; published online 30 October 2008

**Subject Category:** integrated genomics and post-genomics approaches in microbial ecology

**Keywords:** *Bradyrhizobium japonicum*; Bradyrhizobiaceae; genomic comparison; nitrogen fixation; soybean

Correspondence: K Minamisawa, Graduate School of Life Sciences, Tohoku University, Katahira, Aoba-ku, Sendai 980-8577, Japan.

E-mail: kiwamu@ige.tohoku.ac.jp

Received 12 June 2008; revised 21 August 2008; accepted 26 August 2008; published online 30 October 2008

## Introduction

*Bradyrhizobium japonicum* is a symbiotic nitrogen-fixing soil bacterium that has the ability to form root nodules on soybeans. Genotypic and phenotypic variations of *B. japonicum* strains have been reported in terms of DNA fingerprints (Hartmann *et al.*, 1992; Minamisawa *et al.*, 1999), internal transcribed spacer (ITS) sequences between 16S and

23S rDNA (van Berkum and Fuhrmann, 2000), uptake hydrogenase (van Berkum, 1990), denitrification (Sameshima-Saito *et al.*, 2006), symbiotic associations (Ishizuka *et al.*, 1991) and nitrogen fixation (Basit *et al.*, 1991).

A notable feature of the genome of *B. japonicum* strain USDA110 (9.1 kb) is the existence of a large genomic island (GI), termed a 'symbiosis island' (681 kb), carrying a cluster of symbiotic genes with lower GC contents, which is structurally inserted into a val-tRNA gene on the genome (Kaneko *et al.*, 2002). Besides the symbiosis island, 14 smaller GIs (4–97 kb) were found on the genome (Kaneko *et al.*, 2002). However, little is known about the genomic diversity in the strain variations of *B. japonicum* and its correlation with the presence of GIs.

Genomic islands are horizontally acquired DNA regions that are usually inserted in the vicinity of tRNA genes on chromosomes and flanked by direct repeat sequences (Dobrindt *et al.*, 2004). The GC content of GIs often differs from that of the core genomes of bacteria, and GIs harbor several functional genes that encode for proteins involved in pathogenicity, xenobiotic degradation, iron uptake, antibiotic resistance, secondary metabolism or symbiosis (Dobrindt *et al.*, 2004).

*Bradyrhizobium japonicum* is a member of the family Bradyrhizobiaceae, which belongs to the order Rhizobiales in the Alphaproteobacteria (Gupta and Mok, 2007). Members of Bradyrhizobiaceae include a number of nonsymbiotic bacteria with diverse biochemical functions such as photosynthesis (Molouba *et al.*, 1999; Larimer *et al.*, 2004; Giraud *et al.*, 2007), oligotrophy (Saito *et al.*, 1998; King, 2007), 2,4-dichlorophenoxyacetic acid degradation (Kamagata *et al.*, 1997) and nitrification (Starkenbug *et al.*, 2006).

Comparative genomic hybridizations (CGHs) by DNA micro- and macroarrays have been employed to reveal the evolution and function of pathogenicity in genomic terms in organisms such as *Escherichia coli* (Dobrindt *et al.*, 2003; Carter *et al.*, 2008), *Yersinia pestis* (Hinchliffe *et al.*, 2003), *Yersinia pseudotuberculosis* (Zhou *et al.*, 2004), *Xylella fastidiosa* (Koide *et al.*, 2004), *Campylobacter jejuni* (Pearson *et al.*, 2003), *Streptococcus agalactiae* (Brochet *et al.*, 2006) and *Streptococcus pneumoniae* (Obert *et al.*, 2006). The results of these studies indicate that mobile genetic elements such as phages, transposons and GIs contribute to pathogenicity acquisition and environmental adaptation. The CGH approach is suitable for efficient determination of the global genome variations among target bacterial populations, although the comparison has been limited to the original gene repertoires in template genomes. In rhizobia, DNA micro- and macroarrays have been used mainly for global gene expression analysis of *B. japonicum* (Chang *et al.*, 2007; Pessi *et al.*, 2007; Brechenmacher *et al.*, 2008; Wei *et al.*, 2008), *Mesorhizobium loti* (Uchiumi *et al.*, 2004) and *Sinorhizobium meliloti* (Becker

*et al.*, 2004). CGH analyses of four natural isolates of *S. meliloti* by microarray revealed a significant fraction of variable genes including transposease and unknown genes on pSymA megaplasmid (Giuntini *et al.*, 2005). However, there is no other report for array-based CGH analysis of rhizobia including *B. japonicum*.

Our aims were to clarify the variations in the genomes of *B. japonicum* and other members of the family Bradyrhizobiaceae and to determine the involvement of these genomic variations in the symbiotic phenotypes of *B. japonicum* strains.

## Materials and methods

### *Bacterial strains and media*

The strains used are listed in Table 1. *B. japonicum*, *B. elkanii*, *Agromonas oligotrophica*, G14130 and *Bradyrhizobium* sp. HWK12 and HW13 were grown aerobically at 30 °C in HM salt medium (Nieuwkoop *et al.*, 1987) supplemented with 0.1% arabinose and 0.025% Difco yeast extract (Becton, Dickinson and Company, Sparks, MD, USA). The other strains were grown aerobically at 30 °C in Difco Nutrient broth (Becton, Dickinson and Company). Total bacterial DNA was prepared from cultured cells as described earlier (Minamisawa *et al.*, 2002).

### *Comparative genomic hybridization*

Total DNA was sonicated for 30 s with an Ultrasonic Cleaner Vs-25 (As One, Osaka, Japan). Twenty-five nanograms of fragmented DNA was labeled with [ $\alpha$ -<sup>33</sup>P] dCTP (2500 Ci mmol<sup>-1</sup>; Amersham Biosciences, Pittsburgh, PA, USA) by using a Rediprime II random primer labeling system (Amersham Biosciences). DNA macroarray of *B. japonicum* USDA110 (Ito *et al.*, 2006; Wei *et al.*, 2008) was used for CGH analyses. The macroarray contains 3960 spots on a nylon membrane, on which 2.7-kb DNA segments (on average) of USDA110 brb libraries (Kaneko *et al.*, 2002) were mainly spotted. Hybridization was carried out as described earlier (Uchiumi *et al.*, 2004; Ito *et al.*, 2006; Wei *et al.*, 2008). The membrane was washed twice for 15 min at 55 °C in 2 × saline sodium citrate (SSC) containing 0.1% sodium dodecyl sulfate and twice for 15 min at 55 °C in 0.1 × SSC (0.2 × SSC was used for bacteria other than *B. japonicum*) containing 0.1% sodium dodecyl sulfate (1 × SSC is 0.15 M NaCl plus 0.015 M sodium citrate). Image acquisition and data analysis were performed as described earlier (Uchiumi *et al.*, 2004; Ito *et al.*, 2006; Wei *et al.*, 2008). At least three sets of biologically independent array analyses were performed for each strain. Signal values were normalized against total signal values. After normalization, the signals of the tested strains were calculated as ratios of the signal of USDA110. Principal component analysis of the CGH profile was performed with CANOCO (version 4.5

**Table 1** Bacterial strains used

| Strain                            | Accession no. of DNA sequence |                  | Host plant<br>or source       | Relevant characteristics <sup>a</sup>                                    | Reference                            |
|-----------------------------------|-------------------------------|------------------|-------------------------------|--|--------------------------------------|
|                                   | 16S rRNA<br>gene              | ITS <sup>b</sup> |                               |  |                                      |
| <i>Bradyrhizobium japonicum</i>   |                               |                  |                               |  |                                      |
| USDA110                           | AP005940                      | AB100749         | <i>Glycine max</i>            | Soybean bradyrhizobia, <i>hup</i> <sup>+</sup> , <i>nos</i> <sup>+</sup> | Kaneko <i>et al.</i> (2002)          |
| NC6                               | AB231918                      | AB278126*        | <i>G. max</i>                 | Soybean bradyrhizobia, <i>hup</i> <sup>+</sup> , <i>nos</i> <sup>+</sup> | Sameshima-Saito <i>et al.</i> (2006) |
| USDA122                           | AF208503                      | AB100751         | <i>G. max</i>                 | Soybean bradyrhizobia, <i>hup</i> <sup>+</sup> , <i>nos</i> <sup>+</sup> | van Berkum and Fuhrmann (2000)       |
| NK2                               | AB070567                      | AB278127*        | <i>G. max</i>                 | Soybean bradyrhizobia, <i>hup</i> <sup>+</sup> , <i>nos</i> <sup>+</sup> | Sameshima <i>et al.</i> (2003)       |
| NC4                               | AB231917                      | AB278125*        | <i>G. max</i>                 | Soybean bradyrhizobia, <i>hup</i> <sup>+</sup> , <i>nos</i> <sup>+</sup> | Sameshima-Saito <i>et al.</i> (2006) |
| USDA124                           | AF208505                      | AB100753         | <i>G. max</i>                 | Soybean bradyrhizobia, <i>hup</i> <sup>-</sup> , <i>nos</i> <sup>-</sup> | van Berkum and Fuhrmann (2000)       |
| USDA6 <sup>T</sup>                | U69638                        | AB100741         | <i>G. max</i>                 | Soybean bradyrhizobia, <i>hup</i> <sup>-</sup> , <i>nos</i> <sup>-</sup> | van Berkum and Fuhrmann (2000)       |
| T7                                | AB231929                      | AB278128*        | <i>G. max</i>                 | Soybean bradyrhizobia, <i>hup</i> <sup>-</sup> , <i>nos</i> <sup>-</sup> | Sameshima-Saito <i>et al.</i> (2006) |
| T9                                | AB231930                      | AB278129*        | <i>G. max</i>                 | Soybean bradyrhizobia, <i>hup</i> <sup>-</sup> , <i>nos</i> <sup>-</sup> | Sameshima-Saito <i>et al.</i> (2006) |
| <i>Agromonas oligotrophica</i>    |                               |                  |                               |  |                                      |
| S58                               | D78366                        | —                | Rice rhizosphere              | Slow-growing oligotrophic bacteria                                       | Saito <i>et al.</i> (1998)           |
| G14130                            | AB003457                      | —                | Pasture soil                  | Slow-growing oligotrophic bacteria                                       | Saito <i>et al.</i> (1998)           |
| <i>Bradyrhizobium</i> sp.         |                               |                  |                               |  |                                      |
| BTAi1                             | D86354                        | NC_009485        | <i>Aeschynomene sensitiva</i> | Photosynthetic stem-nodulating bacteria                                  | Giraud <i>et al.</i> (2007)          |
| ORS278                            | AJ133779                      | NC_009445        | <i>Aeschynomene indica</i>    | Photosynthetic stem-nodulating bacteria                                  | Giraud <i>et al.</i> (2007)          |
| HW13                              | D89027                        | —                | Hawaiian soil                 | 2,4-D-degrading bacteria   | Kamagata <i>et al.</i> (1997)        |
| HWK12                             | D89028                        | —                | Hawaiian soil                 | 2,4-D-degrading bacteria   | Kamagata <i>et al.</i> (1997)        |
| <i>Rhodopseudomonas palustris</i> |                               |                  |                               |  |                                      |
| CGA009                            | BX572607                      | BX572607         | —                             | Photosynthetic bacteria  | Larimer <i>et al.</i> (2004)         |
| <i>Bradyrhizobium elkanii</i>     |                               |                  |                               |  |                                      |
| USDA76 <sup>T</sup>               | U35000                        | —                | <i>G. max</i>                 | Soybean bradyrhizobia  | van Berkum and Fuhrmann (2000)       |

Abbreviations: ITS, internal transcribed spacer; 2,4-D, 2,4-dichlorophenoxyacetic acid.

<sup>a</sup>*hup* and *nos* indicate the presence and absence of structural genes for uptake hydrogenase and nitrous oxide reductase, respectively.

<sup>b</sup>Asterisked DNA sequences were determined in this study.

for Windows; Microcomputer Power, Ithaca, NY, USA). The default parameters were used except for intersample scaling to generate ordination plots based on the scores of the first two principal components. Signal differences among respective spots were confirmed by *t*-test ( $P \leq 0.05$ ) between the tested strain and USDA110.

#### PCR amplification and sequence analysis

Twelve primer pairs were designed for PCR amplification of the boundary regions of the GIs observed in the USDA110 genome (Supplementary Table 1). LA Taq polymerase (Takara, Osaka, Japan) was used for PCR amplification. The reaction mixture was first incubated at 94 °C for 1 min; then subjected to 14 cycles of 98 °C for 20 s and 68 °C for 20 min; 16 cycles of 98 °C for 20 s and 68 °C for 20–24 min (15-s increase per cycle); and finally to incubation at 72 °C for 10 min. Amplified DNA fragments were separated on agarose gels and purified by using a Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA). Direct sequencing was carried out by using the PCR products as templates and the

PCR primers with an ABI PRISM 310 DNA sequencer and Big Dye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). PCR amplification and sequence analysis of 16S-23S rDNA ITS sequences were performed as described by Saeki *et al.* (2005, 2006).

#### Phylogenetic analysis

Table 1 lists the DDBJ/GenBank/EMBL accession numbers for the 16S rRNA genes and ITS sequences between 16S- and 23S-rRNA that are used in this study. For the phylogenetic analysis, the neighbor-joining method and Clustal W were used as described earlier (Saito *et al.*, 2008).

#### Determination of fixed nitrogen

Soybean (*Glycine max* (L.) Merr. 'Enrei') was cultivated in a greenhouse at Hiroshima University. Pots (3l) containing granitic regosol, perlite and peatmoss at 2:1:1 (v v<sup>-1</sup>) were sterilized by autoclave, and then *B. japonicum* cells cultured in yeast extract-mannitol liquid medium (Jordan, 1984) at

30 °C for 7 days were mixed into the pots (10<sup>6</sup> cells per seed). One week later, five surface-sterilized soybean seeds were planted in each pot. For cultivation, sterilized inorganic nutrient solution was supplied as described earlier (Masuda *et al.*, 1989). Plants were harvested 56 days after seed planting. Plants were carefully sampled from the pots and dissected into leaves, stem, roots and nodules. The plant samples were dried individually at 80 °C, weighed and then ground with a vibration sample mill (Model T1-100; Heiko Co. Ltd, Iwaki, Fukushima, Japan) for nitrogen analysis. The powdered samples were digested with sulfuric acid and acid mixture to quantify total nitrogen (TN) by the Kjeldahl method. The amount of TN was calculated from the sum of nitrogen amounts in each part (leaves, stem, roots and nodules). Fixed nitrogen per plant (FN) was calculated by subtracting the amount of TN in uninoculated plants from that in the plants inoculated with each strain.

#### Correlation and multiple regression analyses

Correlation analyses between genomic variable regions and host plant parameters were performed with Excel software (Microsoft Corporation, Redmond, WA, USA). To examine the effect of each profile on the phenotypes of the inoculated soybeans, we conducted multiple regression analyses by using R 2.0.1 (Ihaka and Gentleman, 1996; available at <http://www.R-project.org>). Profiles A, B, C and D were all included in the analyses as candidate-independent variables (for ease of notation, we assigned the numbers 1, 2, 3 and 4 to profiles A, B, C and D, respectively). For each profile, we tested a model wherein the profile was included as an independent variable and a model where the profile was excluded. That is, the observed phenotypic value of strain *i* (*i* = 1, 2, ..., 9), *y<sub>i</sub>*, can be described by the linear model

$$y_i = b_0 + \sum_{j=1}^4 \delta_j b_j x_{ij} + e_i,$$

where *b*<sub>0</sub> is an intercept, *b<sub>j</sub>* is the effect (that is, coefficient) associated with profile *j*, *e<sub>i</sub>* is an error term that follows a normal distribution and *x<sub>ij</sub>* denotes the results of genomic Southern analysis of profile *j* for strain *i* and is defined by 1 for a positive result and 0 for a negative result.  $\delta_j$  is an indicator variable where  $\delta_j=1$  corresponds to the case in which the profile *j* (*j* = 1, 2, ..., 4) is included in the model and  $\delta_j=0$  implies exclusion. All possible models were tested (that is, a total of 2<sup>4</sup> = 16). Models were selected in accordance with the Akaike Information Criterion (Akaike, 1973). That is, the model that showed the minimum Akaike Information Criterion was selected as the best model. The significance of the best model and the significance of each effect included in the best model were evaluated by *F*-test and *t*-test, respectively.

#### Expression analysis

For expression analysis by macroarray, surface-sterilized soybean seeds (*G. max* 'Enrei') were germinated in sterile vermiculite for 2 days at 25 °C and transplanted into a Leonard jar (Leonard, 1943; Trung and Yoshida, 1983) that contained sterile vermiculite and nutrient solution (Minamisawa *et al.*, 2002). *B. japonicum* was inoculated at 10<sup>9</sup> cells per seed. Plants were grown in a phytotron (Koito type KC; Koito Industries, Tokyo, Japan) for 28 days with a day temperature of 25 °C for 16 h and a night temperature of 23 °C for 8 h; daylight was supplied at a photon flux density of 277 μmol m<sup>-2</sup> s<sup>-1</sup>. Total RNA of bacteroids (isolated from 2 g of 'Enrei' nodules) and free-living cells (cultured in 100 ml of Yeast Mold medium) was prepared by the hot phenol method as described earlier (Ditta *et al.*, 1987; Uchiumi *et al.*, 2004). Bacteroid preparation, cDNA labeling, hybridization, image acquisition and data analysis were performed as described earlier (Uchiumi *et al.*, 2004; Ito *et al.*, 2006; Wei *et al.*, 2008).

#### Nucleotide sequence accession number and microarray data

DNA sequences determined in this study were deposited under the following accession numbers in the DDBJ DNA database: boundary regions around GIs, AB282934–AB282955; and ITS sequences, AB278125 (NC4), AB278126 (NC6), AB278127 (NK2), AB278128 (T7) and AB278129 (T9). The CGH and expression data are available at the MacroArray Analysis of *Bradyrhizobium japonicum* website (<http://orca10.bio.sci.osaka-u.ac.jp/array02/>).

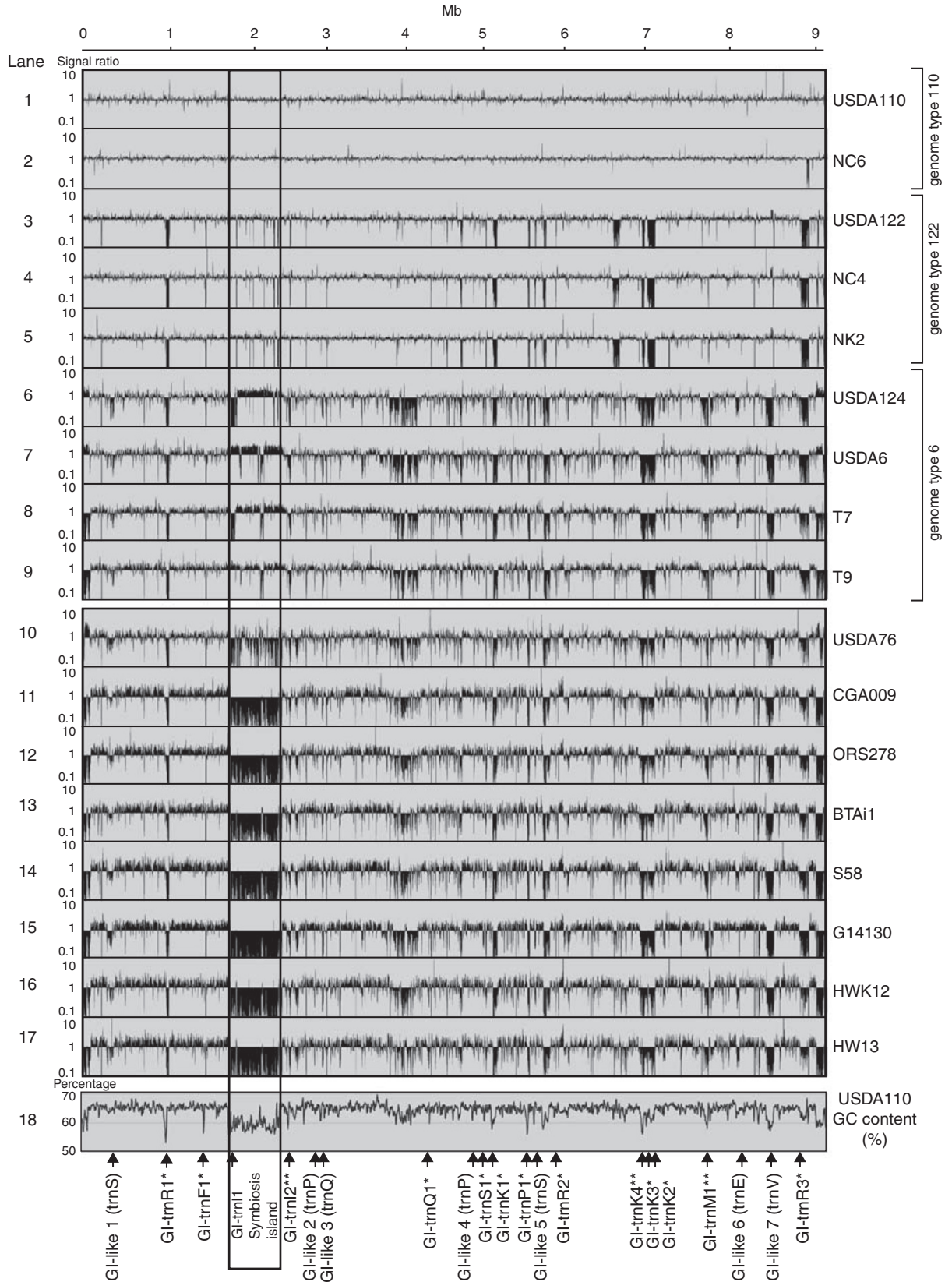
## Results

#### Conditions and presentation of CGH analysis

Total DNAs extracted from the respective strains (USDA110, USDA122, USDA124, USDA6, NC4, NC6, NK2, T7, T9, S58, G14130, BTAi1, ORS278, HW13, HWK12, CGA009 and USDA76 in Table 1) were labeled with radioactive <sup>33</sup>P and were hybridized to DNA macroarray membranes of USDA110. Unlike in the nine strains of *B. japonicum*, weak signals were detected in CGH analysis of the other eight strains, S58, G14130, BTAi1, ORS278, HW13, HWK12, CGA009 and USDA76 (referred to here as the 'non-Bj' strains). This is probably because the genomes of the non-Bj strains were less similar to the USDA110 genome than were the genomes of the *B. japonicum* strains. Therefore, a low-stringency second wash (0.2 × SSC) was adopted for the non-Bj strains after hybridization, as described in the Materials and Methods, whereas *B. japonicum* strains were subjected to a normal second wash (0.1 × SSC). The signal ratios of 3960 probes in the tested strains were calculated as compared with

those of USDA110, and the ratios along with the genomic positions of USDA110 were plotted as CGH profiles (Figure 1).

*Intraspecies comparison of B. japonicum CGH profiles*  
When *B. japonicum* NC6 was compared with USDA110, the profiles of both signal intensities

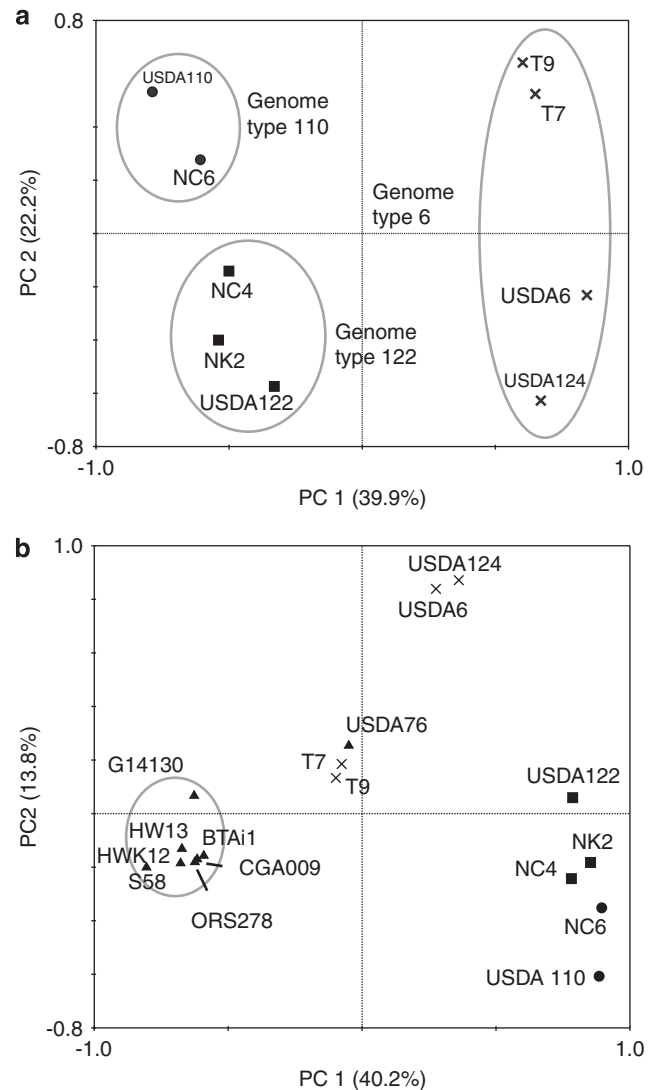


were almost the same, except at the 8.9-Mb position (lanes 1 and 2; Figure 1). In contrast, in the other strains of *B. japonicum* (USDA122, NK2, NC4, USDA124, USDA6, T7 and T9), we found many genomic regions with lower signal ratios (LSR) in clusters, the positions of which were likely shared by these strains (lanes 3–9; Figure 1). Interestingly, the LSR regions corresponded to those of low G + C content in the USDA110 genome (lane 18; Figure 1), and they included 14 GIs previously reported by Kaneko *et al.* (2002) ('GI-like' in Figure 1). These results suggest that LSR regions are commonly missing on the genomes of many *B. japonicum* strains. When the DNA sequences in and around the LSR regions were surveyed on the USDA110 genome, we found seven additional GI-like regions that flanked t-RNA genes around the LSR regions on the USDA110 genome, although the direct repeat sequences of the t-RNA genes were not structurally observed as genomic islands ('GI-like' in Figure 1).

To compare the CGH profiles of *B. japonicum* strains, a principal component analysis was conducted based on the signal ratios (Figure 2). Plots of the nine strains of *B. japonicum* were divided into three places on the principal component analysis chart, where 62% variance was explained by principal components PC1 and PC2 (Figure 2a). In this way, we classified and termed genome type 110 (USDA110 and NC6), genome type 122 (USDA122, NC4 and NK2) and genome type 6 (USDA6, USDA124, T7 and T9). This grouping of the genome types of *B. japonicum* strains was supported by the CGH profiles (Figure 1).

#### Relationships between phylogenetic analysis and CGH profile among *B. japonicum* strains

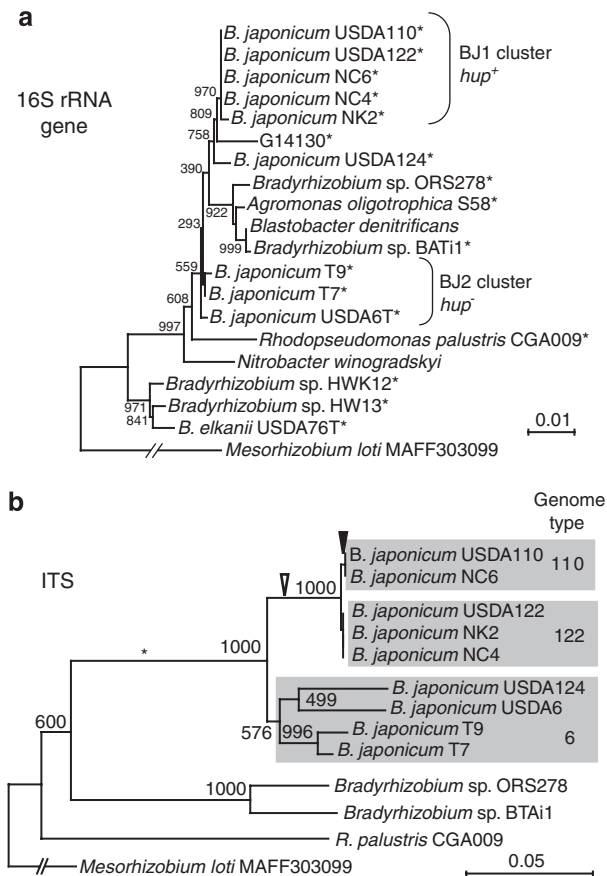
To examine whether the CGH profiles of *B. japonicum* strains reflected the phylogeny of the strains, phylogenetic trees were constructed based on the 16S rRNA gene and the ITS sequences between the 16S and 23S rRNA genes (Figure 3). The tree based on the 16S rRNA gene divided the *B. japonicum* strains into two main clusters, BJ1 and BJ2, as described earlier (Sameshima *et al.*, 2003; Sameshima-Saito *et al.*, 2006) (Figure 3a). *B. japonicum* strains belonging to the two genome types 110 and 122, as defined by the CGH profile (Figures 1 and 2), still formed a single clade, BJ1 (Figure 3a). The ITS analysis tree enhanced the resolution of the phylogenetic relationships, as



**Figure 2** Principal component analysis of CGH profiles. (a) Principal component plots generated from all the CGH profiles of seven strains of *Bradyrhizobium japonicum*. (b) Principal component plots generated from the genome core of CGH profiles without the symbiosis island of *B. japonicum* and from non-Bj strains. CGH, comparative genomic hybridization.

reported by van Berkum and Fuhrmann (2000); the *B. japonicum* strains were subdivided into genome types 110, 122 and 6, with high bootstrap values (Figure 3b). These results showed that the CGH profiles certainly reflected the phylogeny within *B. japonicum*.

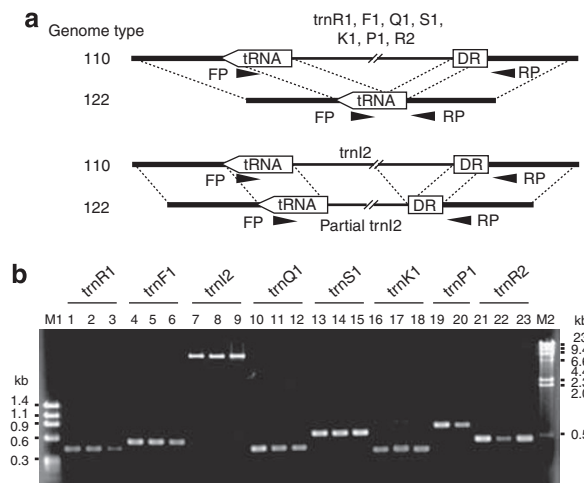
**Figure 1** Comparative genomic hybridization (CGH) analysis of 17 strains of Bradyrhizobiaceae. Horizontal axis indicates array probes arranged in order on the USDA110 genome, and the vertical axis in the CGH window (lanes 1–17) shows the signal ratio for each strain (signal of tested strain versus signal of USDA110). Lanes 1–9 show the CGH profiles of *Bradyrhizobium japonicum* strains, and lanes 10–17 show the profiles of other members of the Bradyrhizobiaceae (see Table 1). Lane 18 shows the GC content of *B. japonicum* strain USDA110, calculated by using a 10-kb window. \*GIs with lower signals in strains USDA122, NC4 and NK2 of genome type 122 as compared with those of USDA110. \*\*GIs that appear to be conserved among strains of genome types 110 and 122. The names of the trn elements are those given as GIs by the whole-genome sequencing of *B. japonicum* USDA110 (Kaneko *et al.*, 2002). The position of the symbiosis island in *B. japonicum* USDA110 (1.88–2.29 Mb) is boxed. GI, genomic island.



**Figure 3** Phylogenetic relationships of *Bradyrhizobium japonicum* and other Bradyrhizobiaceae members based on (a) 16S rRNA gene sequences and (b) internal transcribed spacer (ITS) sequences. (a) Tree constructed on the basis of the 16S rRNA gene sequences of the 17 strains that we tested (asterisked), together with those of *Blastobacter denitrificans* (AF338176) and *Nitrobacter winogradskyi* (CP000115), by the neighbor-joining (NJ) method. (b) Tree constructed on the basis of ITS sequences by the NJ method. The black arrowhead shows the estimated positions of horizontal transfer of seven GIs (trnR1, trnF1, trnQ1, trnS1, trnK1 and trnP1 and trnR2) during microevolution of *B. japonicum*. The white arrowhead shows the estimated positions of horizontal transfer of trnM1 and trnK4 (see Discussion). Asterisk shows the estimated position of acquisition of an ancestral symbiosis island associated with soybeans (see Discussion). For both trees, *Mesorhizobium loti* MAFF303099 was used as the outgroup. Numbers at the nodes are bootstrap values from 1000 replications. Bars show base substitutions per nucleotide. GI, genomic island.

#### PCR and sequence analysis of variable regions

The intraspecies comparisons of CGH analyses suggested that the variable genomic regions of seven *B. japonicum* strains (lanes 3–9; Figure 1) corresponded to the positions of GIs on the USDA110 genome (lane 18; Figure 1). To examine whether GIs were structurally missing on the genomes of these strains, PCR amplification was conducted targeting the GIs of *B. japonicum* strains. PCR primers were designed in the sequences flanking the GIs in the genome of USDA110 for the respective GI targets (Figure 4a, Supplementary Table 1).



**Figure 4** PCR amplification and sequence analysis of genomic islands. (a) Schematic presentation of PCR primer and variable regions around GIs in genome types 110 and 122, as revealed by DNA sequencing of the PCR products. Bold lines indicate the genome core and 'DRs' are direct repeats derived from target tRNA genes. FP, forward primer; RP, reverse primer. (b) Gel electrophoresis of PCR products of USDA122 (lanes 1, 4, 7, 10, 13, 16, 19, 21), NC4 (lanes 2, 5, 8, 11, 14, 17, 20, 22) and NK2 (lanes 3, 6, 9, 12, 15, 18, 23). trnR1, trnF1, trnI2, trnQ1, trnS1, trnK1, trnP1 and trnR2 are the genomic islands targeted by using the primer sets in Supplementary Table 1. M1 and M2 are DNA size markers of Lambda DNA *Hind*III digest and fX174 DNA *Hae*III digest, respectively. No PCR product was observed for *Bradyrhizobium japonicum* strains, USDA124, USDA6, T7 and T9 belonging to genome type 6 (Figures 1 and 2).

When total DNA of strain USDA110 or NC6 was used as a template (Figure 4a), no PCR products for trnR1, trnI1, trnI2, trnK1, trnP1 and trnR2 were observed, but PCR products (6–9 kb) for trnF1, trnQ1 and trnS1 were detected (Supplementary Table 1). No detection of most PCR products is probably due to the large target sizes of the GIs on the USDA110 genome. However, when total DNA of strains USDA122, NC4 or NK2 of genome type 122 was used as a template, short PCR products could be detected for eight GIs (trnR1, trnF1, trnI2, trnQ1, trnS1, trnK1, trnP1 and trnR2), except with a combination of NK2 and trnP1 primer sets (Figure 4b). The sizes of the PCR products suggested the occurrence of direct connections, without GIs (Supplementary Table 1). Examination of the DNA sequences of these amplified DNA fragments revealed that at least seven GIs (trnR1, trnF1, trnQ1, trnS1, trnK1, trnP1 and trnR2) were indeed absent in genome type 122 (USDA122, NC4 and NK2) with direct connections of the genome core; the exception was trnP1 in strain NK2 (Figure 4b, Supplementary Table 1). Analysis of the DNA sequence of the 7-kb PCR product of the use of the trnI2 primer set and the NC4 DNA template (Figure 4b) suggested partial truncation of the structure of the original trnI2 on the USDA110 genome (Figure 4a). With strains USDA124, USDA6, T7 and T9 of genome type 6, no PCR product was observed (data not shown),

probably because of sequence divergence around the GIs in these strains. These results indicate that at least seven GIs (*trnR1*, *trnF1*, *trnQ1*, *trnS1*, *trnK1*, *trnP1* and *trnR2*) are missing in the *B. japonicum* strains belonging to genome type 122.

#### Determination of regions missing in *B. japonicum* strains

Strains of *B. japonicum* are divided into two genotypes for the presence and absence of *hup* structural genes encoding uptake hydrogenase (van Berkum, 1990) and *nos* structural genes encoding nitrous oxide reductase (Sameshima-Saito *et al.*, 2006; Table 1). These genes were used to determine a criterion for the absence and presence of certain genomic regions in *B. japonicum* strains. When *hup*<sup>+</sup> strains were used as templates to array probe *brb11967* including *hupSL* (*bll6941* and *bll6942*), the resultant signal ratios ranged from 0.8 to 1.5, whereas the signal ratios ranged from 0.2 to 0.5 in *hup*<sup>-</sup> strains. When the *nos*<sup>+</sup> strains were used as templates to array probe *brb00265* including a nitrous oxide reductase gene (*nosZ*), the signal ratios ranged from 0.9 to 1.1, whereas the signal ratios ranged from 0.2 to 0.3 in *nos*<sup>-</sup> strains. Probably, cross-hybridization of different genomic regions gives rise to higher background signals because of the large size and complexity of the *B. japonicum* genome (Kaneko *et al.*, 2002; Ito *et al.*, 2006).

Therefore, if the signal ratio of a certain probe (spot) was less than 0.5 and differed significantly by *t*-test ( $P \leq 0.05$ ) between the tested strain and USDA110, we considered that the test strain lacked the corresponding region on the genome and produced a matrix consisting of 0 ('missing') or 1 ('present') in the tested strains, along with the array probe positions (Figure 5a). Similar criteria have been adopted in microarray CGH analysis of *Streptococcus* species (Brochet *et al.*, 2006).

When the determined missing regions of nine *B. japonicum* strains were aligned based on USDA110 genome position, it is likely that strains USDA124, USDA6, T7 and T9, belonging to genome type 6, lack *trnM1* and *trnK4* and the vicinity of 4 Mb position (Figure 5a). In other words, these regions appear to be conserved among strains of genome types 110 and 122.

#### Strain variations of symbiotic phenotype in *B. japonicum*

The above results strongly indicate that *B. japonicum* USDA110 has acquired many DNA fragments, such as GIs, in its genome (Figures 1 and 3). When we surveyed the possible functions of genes on GIs and other variable regions of the USDA110 genome, there were many candidate genes (approximately 2138), 57% of which were unknown earlier. This situation was similar to CGH result of *S. meliloti* (Giuntini *et al.*, 2005). These results beg the question of how the acquired foreign regions function in *B.*

*japonicum* USDA110. Strain USDA110 of *B. japonicum* is highly adapted to modern soybean cultivars and often has superior symbiotic nitrogen-fixation capability (Israel *et al.*, 1986; Basit *et al.*, 1991). These facts prompted us to examine whether the variable regions containing GIs enhance symbiotic nitrogen fixation in soybeans.

When soybean plants were inoculated with the nine strains used in this study and cultivated in a greenhouse from seed for 56 days, we observed strain differences in TN, FN and total dry weight (TDW) of the inoculated soybean cultivar Enrei (Table 2). USDA110 showed the highest mean values in TN, FN and TDW as compared with other *B. japonicum* strains. Statistical analysis indicated that these indexes of USDA110 were significantly higher than those of USDA122, NC4, USDA124, USDA6, T7 and T9 (genome types 122 and 6), although NK2 had high errors (s.d.) in TN, FN and TDW values. On the other hand, T7 and T9 (genome type 6) showed markedly low values of TN and FN, which were statistically significant as compared with USDA110 and NC6 (genome type 110).

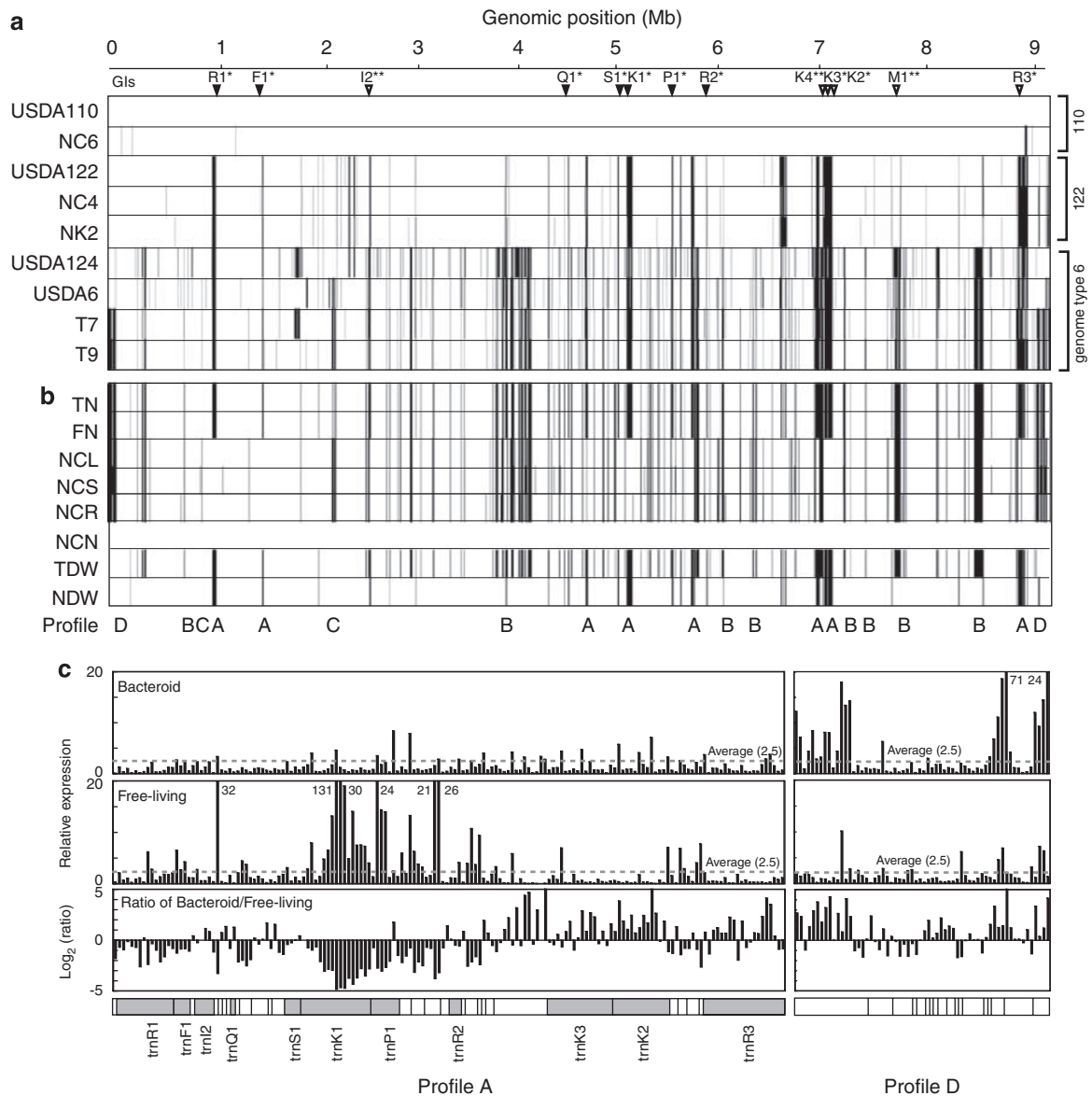
#### Correlation of region absence and symbiotic phenotype

By using data on the nine strains of *B. japonicum*, we calculated the correlation coefficients between the existence of certain genomic regions ( $X = 0$  or 1) (lines were scored in missing regions ( $X = 0$ ) in Figure 5a) and the values of the eight parameters (Figure 5b) of inoculated host plants including TN, FN, TDW, nodule dry weight, nitrogen contents of leaves, stem, roots and nodules along with 3960 array probes. When significant positive correlations ( $P < 0.01$ ) were detected, lines were scored on the positions on USDA110 genome (Figure 5b).

The positions where significant correlations were detected generally corresponded to the missing regions in strains USDA122, NC4, NK2, USDA124, USDA6, T7 and T9 belonging to genome types 122 and 6 (Figures 5a and b). Correlations were found for indexes of nitrogen fixation (TN and FN), plant growth (TDW), nitrogen contents of leaf, stem and root (Figure 5b). On the other hand, no and weak correlations were observed in nodule nitrogen content and nodule dry weight, respectively (Figure 5b). In particular, many missing regions (Figure 5a) were heavily correlated with indexes of nitrogen fixation (TN and FN) and TDW (Figure 5b). When similar patterns of the missing regions (Figure 5a) and the positive correlation with the indexes (Figure 5b) were sorted irrespective of the probe positions on the USDA110 genome, four major variable regions appeared; we designated them as profiles A, B, C and D (Figure 5b below, Supplementary Figure 1).

Profile A on the USDA110 genome was completely missing in genome type 122 (strains USDA122, NC4 and NK2) and genome type 6 (strains USDA6,





**Figure 5** Correlation and expression analysis of genome structure and symbiotic nitrogen fixation phenotype among *Bradyrhizobium japonicum* strains. **(a)** Missing genomic regions in nine strains of *B. japonicum*, as determined from CGH profiles. Black lines indicate missing regions (signal ratio  $\leq 0.5$ ;  $P \leq 0.05$ ) on the genome in tested strains. Arrowheads indicate positions of the GIs with asterisks, whose explanations are the same in Figure 1. **(b)** Correlation analysis between the existence of certain genomic regions and symbiosis parameters. The symbiosis parameters are total nitrogen per plant (TN), fixed nitrogen per plant (FN), nitrogen content of leaves (NCL), nitrogen content of stem (NCS), nitrogen content of roots (NCR), nitrogen content in nodules (NCN), total dry weight of plant (TDW) and nodule dry weight per plant (NDW). Black lines indicate significant positive correlations between each array probe and the symbiosis parameter. **(c)** Expression in profiles A and D under bacteroid and free-living cells of *B. japonicum* USDA110. Dotted line shows average relative expression throughout all the genomic regions. CGH, comparative genomic hybridization; GI, genomic island.

USDA124, T7 and T9). A high rate of occurrence of GIs was found in profile A: 10 GIs were concentrated in profile A among the 14 confirmed GIs on the USDA 110 genome (71%). Profile D had missing regions only in strains T7 and T9. Profiles C and D did not include GIs.

#### Multiple regression analysis

Linear models were used to conduct a multiple regression analysis of the correlations between the

four profiles and the parameters measured in the inoculated soybeans. For TN, the model that included all the profiles had the minimum Akaike information criterion (Akaike, 1973) among 16 models and was selected as the best model as 'A+B+C+D' (Table 3). This model explained about 95% of the total variations observed among the nine strains (that is,  $R^2 = 0.951$ ). Among the four profiles, A and D had highly significant (that is,  $P < 0.01$ ) positive effects on this trait. The coefficients of profiles B and C did not show significant

**Table 2** Total nitrogen content, fixed nitrogen, total dry weight and nodule weight in soybean plants inoculated with various *Bradyrhizobium japonicum* strains<sup>1</sup>

| Inoculated strain      | TN (mg N per plant)         | FN (mg N per plant) <sup>2</sup> | TDW (g per plant)              | NDW (g per plant)            |
|------------------------|-----------------------------|----------------------------------|--------------------------------|------------------------------|
| <i>Genome type 110</i> |                             |                                  |                                |                              |
| USDA110                | 239 ± 12 <sup>a</sup>       | 197 ± 12 <sup>a</sup>            | 8.8 ± 0.5 <sup>a</sup>         | 0.60 ± 0.06 <sup>a</sup>     |
| NC6                    | 222 ± 46 <sup>a,b</sup>     | 179 ± 46 <sup>a,b</sup>          | 8.0 ± 1.6 <sup>a,b</sup>       | 0.63 ± 0.14 <sup>a,b</sup>   |
| <i>Genome type 122</i> |                             |                                  |                                |                              |
| USDA122                | 161 ± 25 <sup>b</sup>       | 118 ± 25 <sup>b,c</sup>          | 6.6 ± 1.1 <sup>b</sup>         | 0.55 ± 0.01 <sup>a,b</sup>   |
| NC4                    | 156 ± 29 <sup>b</sup>       | 113 ± 29 <sup>c</sup>            | 5.9 ± 1.2 <sup>b,c</sup>       | 0.37 ± 0.00 <sup>c</sup>     |
| NK2                    | 181 ± 81 <sup>a,b,c</sup>   | 139 ± 81 <sup>a,b,c,d</sup>      | 7.4 ± 3.5 <sup>a,b,c,d</sup>   | 0.39 ± 0.21 <sup>a,b,c</sup> |
| <i>Genome type 6</i>   |                             |                                  |                                |                              |
| USDA124                | 145 ± 47 <sup>b,c,d</sup>   | 103 ± 47 <sup>b,c,d,e</sup>      | 5.4 ± 1.3 <sup>b,c,d</sup>     | 0.47 ± 0.12 <sup>a,b,c</sup> |
| USDA6                  | 165 ± 35 <sup>b,c,d,e</sup> | 122 ± 35 <sup>b,c,d,e,f</sup>    | 6.3 ± 1.4 <sup>b,c,d,e</sup>   | 0.50 ± 0.11 <sup>b,c</sup>   |
| T7                     | 90 ± 28 <sup>c,d,f</sup>    | 48 ± 28 <sup>d,e</sup>           | 5.8 ± 2.0 <sup>b,c,d,e,f</sup> | 0.40 ± 0.16 <sup>a,b,c</sup> |
| T9                     | 80 ± 26 <sup>c,e,f</sup>    | 38 ± 25 <sup>d,f</sup>           | 5.2 ± 1.4 <sup>b,d,e,f</sup>   | 0.34 ± 0.10 <sup>c</sup>     |
| Uninoculated           | 42 ± 9 <sup>f</sup>         | —                                | 4.1 ± 0.3 <sup>c,d,f</sup>     | 0                            |

Abbreviations: FN, fixed nitrogen per plant; NDW, nodule dry weight; TDW, total dry weight; TN, total nitrogen.

<sup>1</sup>Values are expressed as mean ± s.d. Columns followed by the same letters (a–g) do not differ significantly by *t*-test ( $P < 0.05$ ).

<sup>2</sup>FN was estimated by subtracting the amount of TN in uninoculated plants from that in the plants inoculated with each strain.

**Table 3** Regression analyses with various models

| Parameter <sup>a</sup> | Best model <sup>b</sup> | AIC <sup>c</sup> | Intercept <sup>d</sup> | Profile A                |                   | Profile B                |                   | Profile C                |                   | Profile D                |                   |
|------------------------|-------------------------|------------------|------------------------|--------------------------|-------------------|--------------------------|-------------------|--------------------------|-------------------|--------------------------|-------------------|
|                        |                         |                  |                        | Coefficient <sup>e</sup> | s.e. <sup>f</sup> | Coefficient <sup>e</sup> | s.e. <sup>f</sup> | Coefficient <sup>e</sup> | s.e. <sup>f</sup> | Coefficient <sup>e</sup> | s.e. <sup>f</sup> |
| TN                     | A*+B+C+D*               | 74.46            | 85                     | 64.5*                    | 10.65             | 21                       | 13.47             | −20                      | 16.5              | 80*                      | 14.29             |
| FN                     | A*+D*                   | 75.13            | 43                     | 69*                      | 10.33             | —                        | —                 | —                        | —                 | 76*                      | 10.33             |
| NCL                    | B+D                     | 50.39            | 25                     | —                        | —                 | 3.5                      | 2.61              | —                        | —                 | 6.5                      | 3.12              |
| NCS                    | C+D                     | 34.34            | 13.5                   | —                        | —                 | —                        | —                 | 3.67                     | 1.38              | 6.5                      | 1.57              |
| NCR                    | D                       | 18.72            | 16.5                   | —                        | —                 | —                        | —                 | —                        | —                 | 2.64                     | 0.82              |
| NCN                    | A+C+D                   | 31.89            | 46.5                   | 2                        | 0.95              | —                        | —                 | −8.5                     | 1.22              | 7.5                      | 1.34              |
| TDW                    | A*+B                    | 20.67            | 5.68                   | 1.77*                    | 0.55              | 0.96                     | 0.46              | —                        | —                 | —                        | —                 |
| NDW                    | A*+D                    | −19.44           | 0.37                   | 0.16*                    | 0.05              | —                        | —                 | —                        | —                 | 0.09                     | 0.05              |

<sup>a</sup>Plant parameters: FN, fixed nitrogen (mg N) per plant; NCL, nitrogen content in leaf (mg N per g dry weight); NCN, nitrogen content in nodules (mg N per g dry weight); NCR, nitrogen content in roots (mg N per g dry weight); NCS, nitrogen content in stem (mg N per g dry weight); NDW, dry weight of nodules (g per plant); TDW, total dry weight of plant (g per plant); TN, total nitrogen (mg N) per plant. Asterisks indicate statistical significance ( $P < 0.01$ ).

<sup>b</sup>Best model that minimizes AIC.

<sup>c</sup>Akaike information criterion (AIC; Akaike, 1973) of the best model.

<sup>d</sup>Estimate of intercept of the best model.

<sup>e</sup>Estimate of regression coefficient of independent variable.

<sup>f</sup>Standard error of the estimate.

effects of these profiles by the *t*-test. For FN, ‘A + D’ was selected as the best model (Table 3). Both profiles A and D had highly significant (that is,  $P < 0.01$ ) positive effects on FN. For TDW, ‘A + B’ was selected as the best model (Table 3). Profile A had highly significant (that is,  $P < 0.01$ ) positive effects on TDW.

#### Gene expression in variable regions under symbiotic conditions

The above results suggested that symbiotic nitrogen fixation is increased by the genes present in profiles A and D. It is reasonable to expect that the genes highly expressed in nodule bacteroids would be the genes in profiles A and D that enhance nitrogen fixation. Transcriptome analysis indicated that

several parts of profiles A and D were selectively expressed in bacteroids and free-living cells of USDA110 (Figure 5c). Comparison of our expression data with microarray results (Chang *et al.*, 2007) revealed that *blr6420* (one of three *pobA* homologs; 4-hydroxybenzoate hydroxylase) in GI *trnK2* and *bll6377* (putative acyl-CoA dehydrogenase) and *blr6378* (transcriptional regulatory protein LacI family) in *trnK3* were symbiotically upregulated genes with known functions.

#### CGH profiles of non-Bj strains

The patterns of the CGH profiles of non-Bj strains were similar to each other (Figure 1), although the washing conditions for non-Bj strains were slightly different from those for *B. japonicum* strains. In

particular, the low hybridization signals on the symbiosis island (Figure 1) suggest that the symbiosis island was missing in non-Bj strains, with the exception of *B. elkanii* USDA76, a soybean endosymbiont. Indeed, no symbiosis island containing *nif/fix* and *nod* genes has been found on the genomes of *Rhodopseudomonas palustris* (Larimer *et al.*, 2004) or *Bradyrhizobium* sp. BTAi1 and ORS278 (Giraud *et al.*, 2007), although BTAi1 and ORS278 are stem-nodulating and nitrogen-fixing bacteria (Giraud *et al.*, 2007).

When we examined the relatedness of the core genomic regions outside the symbiosis island by principal component analysis, non-Bj strains, except for *B. elkanii* USDA76, formed a compact cluster (Figure 2b). In contrast, *B. japonicum* USDA110, *R. palustris* CGH009 and *Bradyrhizobium* sp. BTAi1 and ORS278 differed from each other in terms of genome synteny (Supplementary Figure 2A). Therefore, CGH profiles based on the *B. japonicum* USDA110 array were not useful for evaluating the genomes of the non-Bj strains, probably because of the overall lower level of synteny than that of the USDA110 genome.

## Discussion

Horizontal gene transfer is an evolutionary phenomenon that involves the occurrence of genetic exchanges between different evolutionary lineages (Jordan and Koonin, 2004). We found that the GIs on the genome of *B. japonicum* USDA110 were variable genome regions that were not conserved universally on the genomes of the Bradyrhizobiaceae (Figure 1). We demonstrated that at least seven GIs (trnR1, trnF1, trnQ1, trnS1, trnK1, trnP1 and trnR2) on the USDA110 genome did not exist on the strains of genome type 122 within *B. japonicum* (Figure 4). The trnK3-, trnK2- and trnR3-missing regions were commonly observed among *B. japonicum* strains of genome types 122 and 6 (Figures 1 and 5a), although their boundary sequencing could not be always determined by our PCR strategy. CGH profile analysis suggested that the absence of GIs on the genome types 122 and 6 in *B. japonicum* extends to non-Bj strains (Figure 1). So far, the GI sequences have proven specific for *B. japonicum* USDA110 by our BLAST search (<http://www.ncbi.nlm.nih.gov/blast>) of data on the genomes of 10 strains of the family Bradyrhizobiaceae (Gupta and Mok, 2007), although we found partial homologous sequences within the GIs in *Bradyrhizobium* sp. BTAi1 and ORS278, *R. palustris* CGA009, BisA53, BisB18, BisB5 and HaA2, and *Rhizobium leguminosarum* bv. *viciae* pRL plasmids.

These lines of evidence strongly suggest that the GIs were horizontally inserted into the ancestral genome of genome type 110. As CGH profiles on USDA110 symbiosis island in *B. japonicum* USDA124, USDA6, T7 and T9 (genome type 6) were

still conserved rather than those in *B. elkanii* USDA76 and non-Bj strains (Figure 1), the ancestor of *B. japonicum* might diverge into two 110–122 and six lineages after the acquisition of a symbiosis island for soybean associations (asterisk in Figure 3b). The *B. japonicum* strains of genome type 6 have diverged in terms of phylogeny (Figures 3a and b), CGH profiles (Figures 1 and 2a) and no PCR product by using USDA110 primers (Figure 4). Thus, the genome information of USDA6 and its array system would be also required to cover all *B. japonicum* strains with associated soybeans for precise CGH analysis.

From an examination of the ITS sequence phylogeny (Figure 3b) and the distribution of missing regions among *B. japonicum* strains (Figures 1 and 5a), we can speculate that there were two steps of GI acquisition by the ancestral genome: (1) first, two GIs, trnK4 and trnM1, were acquired before the divergence of strains of genome types 110 and 122 (white arrowhead in Figure 3b); (2) then trnR1, trnF1, trnQ1, trnS1, trnP1, trnR2, trnK3, trnK2 and trnR3 were acquired as GIs after the divergence of strains of genome types 110 and 122 (black arrowhead in Figure 3b). Although we do not know how the truncated GI trnI2 (Figure 4a) was generated through partial deletion or insertion, GIs might be highly dynamic entities on *B. japonicum* genomes.

The recent completion of genome-sequencing projects in several rhizobial and bradyrhizobial species enables us to analyze global postgenomic approaches to bacteria, including the analysis of transcriptomes, systematic mutant libraries, OR-Feomes, proteomes, transportomes (Mauchline *et al.*, 2006) and metabolomes, toward the further development of functional genomics (MacLean *et al.*, 2007). In this study, we aimed to reveal bradyrhizobial genomic regions relevant to symbiotic nitrogen fixation by using statistical model analysis to correlate variable genomic regions with symbiotic phenotypes of the host plants. The multiple regression model analysis showed that the variable regions of profiles A and D were significantly correlated with symbiotic nitrogen fixation (Figure 5b, Table 3).

Profile A included 10 of the 14 GIs on the USDA110 genome (Figure 5c). The horizontal transfer of complete genes raises the important issue of how the expression of these genes is regulated in the new host. H-NS protein, a pleiotropic repressor of transcription, has a key role in selective silencing of the transcription of large numbers of horizontally acquired AT-rich genes, including pathogenicity islands (Dorman, 2007). It was recently found that Rho termination is required to suppress similar silencing (Cardinale *et al.*, 2008). If horizontally transferred genes function to enhance symbiotic nitrogen fixation, we can expect them to be expressed or upregulated in bacteroids. Our expression data suggest that, overall, their expression in

profiles A and D in bacteroids is lower than the average expression levels, but some regions are upregulated in symbiosis (Figure 5c). Therefore, one possible approach to verify the involvement of profiles A and D into symbiotic nitrogen fixation is to examine the symbiotic phenotypes of *B. japonicum* USDA110 mutants lacking these expressed regions. An alternative approach is a fine mapping of these profiles by using many strains of *B. japonicum* that share similar genome backgrounds within genome types 110 and 122.

Traditionally, the ability of rhizobial inoculum strains to promote symbiotic nitrogen fixation has been improved. The best-characterized example of this is the uptake hydrogenase (Hup) system, which takes up hydrogen generated through nitrogenase and enhances the energy efficiency of nitrogen fixation (Maier and Triplett, 1996; Baginsky *et al.*, 2005). As genes for the Hup system were included in GI trnM in profile B (Figures 5a and b), we first expected to detect this region by regression model analysis. However, profiles A and D were found to be variable genomic regions, suggesting the existence of novel systems other than the Hup system that enhance symbiotic nitrogen fixation.

In this study, soybean cultivar Enrei, a modern cultivar of soybeans in Japan, was consistently used for the determination of symbiotic phenotypes (Table 3) and expression analysis in bacteroids (Figure 5). However, symbiotic effectiveness for nodulation and nitrogen fixation was sometimes dependent on host legume genotypes (Israel *et al.*, 1986; Nautiyal *et al.*, 1988; Ishizuka *et al.*, 1991; Sadowsky *et al.*, 1991). Thus, care should be taken for soybean genotypes to compare our results with others.

Although rhizobial postgenome studies have provided valuable insights into rhizobial–legume symbioses, marked limitations remain within these types of studies (MacLean *et al.*, 2007). On the basis of the diversity of *E. coli* O157 (Ishii and Sadowsky, 2008), Manning *et al.* (2008) analyzed the association between SNP clades and severe disease, and they suggested that the presence of particular clade lineage was a critical determinant of severe disease. Little is known about the factors in symbiotic bacterial that contribute to variation in symbiotic phenotypes. Our approach combining CGH, symbiotic phenotype and expression analyses should open a new window on how rhizobia have developed efficient systems for symbiotic nitrogen fixation.

Recently, a model system of *nodC*- and *nec1*-targeted monitoring was developed to diagnose rhizobial function (Bontemps *et al.*, 2005) and disease control (Koyama *et al.*, 2007). Our CGH profiles are available to examine overall genotypic diversity, including the *hup* and *nos* genotypes of *B. japonicum* field isolates. In addition, the genomic phylogeny and relatedness of field isolates could be

evaluated within *B. japonicum* (Figure 2). We therefore expect that this method will be applicable to the diagnosis of field populations of *B. japonicum* for soybean production and environmental conservation.

## Acknowledgements

This study was supported in part by a grant-in-aid for Scientific Research on Priority Area 'Comparative Genomics', by grants-in-aid for Scientific Research (no. 17380046), by a grant from PROBRAIN, by Special Coordination Funds for Promoting Science and Technology and by a grant from the Tokachi Federation of Agricultural Cooperatives. We thank H Kouchi (National Institute of Agrobiological Sciences), C Harwood (The University of Iowa) and E Giraud (French National Institute for Agricultural Research) for array spotting, provision of *R. palustris* CGA009 and provision of *Bradyrhizobium* sp. strains ORS278 and BTAi1, respectively.

## References

- Akaike H. (1973). Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds). *2nd International Symposium on Information Theory*. Akadimial Kiado: Budapest. pp 267–281.
- Baginsky C, Brito B, Imperial J, Ruiz-Argüeso T, Palacios JM. (2005). Symbiotic hydrogenase activity in *Bradyrhizobium* sp. (Vigna) increases nitrogen content in *Vigna unguiculata* plants. *Appl Environ Microbiol* **71**: 7536–7538.
- Basit HA, Angle JS, Salem S, Gewaily EM, Kotob SI, van Berkum P. (1991). Phenotypic diversity among strains of *Bradyrhizobium japonicum* belonging to serogroup 110. *Appl Environ Microbiol* **57**: 1570–1572.
- Becker A, Berges H, Krol E, Bruand C, Ruberg S, Capela D *et al.* (2004). Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. *Mol Plant Microbe Interact* **17**: 292–303.
- Bontemps C, Golfier G, Gris-Liebe C, Carrere S, Talini L, Boivin-Masson C. (2005). Microarray-based detection and typing of the *Rhizobium* nodulation gene *nodC*: potential of DNA arrays to diagnose biological functions of interest. *Appl Environ Microbiol* **71**: 8042–8048.
- Brechenmacher L, Kim MY, Benitez M, Li M, Joshi T, Calla B *et al.* (2008). Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*. *Mol Plant Microbe Interact* **21**: 631–645.
- Brochet M, Couve E, Zouine M, Vallaeys T, Rusniok C, Lamy MC *et al.* (2006). Genomic diversity and evolution within the species *Streptococcus agalactiae*. *Microbe Infect* **8**: 1227–1243.
- Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E. (2008). Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* **320**: 935–938.
- Carter B, Wu G, Woodward MJ, Anjum MF. (2008). A process for analysis of microarray comparative genomics hybridisation studies for bacterial genomes. *BMC Genomics* **9**: 53.

- Chang WS, Franck LW, Cytryn E, Jeong S, Joshi T, Emerich WD *et al.* (2007). An oligonucleotide microarray resource for transcriptional profiling of *Bradyrhizobium japonicum*. *Mol Plant Microbe Interact* **20**: 1298–1307.
- Ditta G, Virts E, Palomares A, Kin C. (1987). The *nifA* gene of *Rhizobium meliloti* is oxygen regulated. *J Bacteriol* **169**: 3217–3223.
- Dobrindt U, Agerer F, Michaelis K, Janka A, Buchruesser C, Samuelson M *et al.* (2003). Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *J Bacteriol* **185**: 1831–1840.
- Dobrindt U, Hochhut B, Hentschel U, Hacker J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**: 414–424.
- Dorman CJ. (2007). H-NS, the genome sentinel. *Nat Rev Microbiol* **5**: 157–161.
- Giraud E, Moulin L, Vallet D, Barbe V, Cytryn E, Avarre J *et al.* (2007). Legumes symbioses: absence of *nod* genes in photosynthetic bradyrhizobia. *Science* **316**: 1307–1312.
- Giuntini E, Mengoni A, Filippo CD, Cavalieri D, Aubin-Horth N, Landry CR *et al.* (2005). Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of *Sinorhizobium meliloti* natural strains. *BMC Genomics* **6**: 158.
- Gupta SR, Mok A. (2007). Phylogenomics and signature proteins for the alpha proteobacteria and its main groups. *BMC Microbiol* **7**: 106.
- Hartmann A, Catroux G, Amarger N. (1992). *Bradyrhizobium japonicum* strains identification by RFLP analysis using the repeated sequence RS $\alpha$ . *Lett Appl Microbiol* **15**: 15–19.
- Hinchliffe SJ, Isherwood KE, Stabler RA, Prentice MB, Rakin A, Nichols RA *et al.* (2003). Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis*. *Genome Res* **13**: 2018–2029.
- Ihaka R, Gentleman R. (1996). R: a language for data analysis and graphics. *J Comput Graph Stat* **5**: 299–314.
- Ishii S, Sadowsky MJ. (2008). *Escherichia coli* in the environment: implications for water quality and human health. *Microbes Environ* **23**: 101–108.
- Ishizuka J, Suemasu Y, Mizogami K. (1991). Preference of *Rj*-soybean cultivars for *Bradyrhizobium japonicum* for nodulation. *Soil Sci Plant Nutr* **37**: 15–21.
- Israel DW, Mathis JN, Barbour WM, Elkan GH. (1986). Symbiotic effectiveness and host-strain interactions of *Rhizobium fredii* USDA 191 on different soybean cultivars. *Appl Environ Microbiol* **51**: 898–903.
- Ito N, Itakura M, Eda S, Saeki K, Oomori H, Yokoyama T *et al.* (2006). Global gene expression in *Bradyrhizobium japonicum* cultured with vanillin, vanillate, 4-hydroxybenzoate and protocatechuate. *Microbes Environ* **21**: 240–250.
- Jordan DC. (1984). Rhizobiaceae. In: Krieg NR, Holt JG (eds). *Bergey's Manual of Systematic Bacteriology*. Williams & Wilkins: Baltimore. pp 235–244.
- Jordan IK, Koonin EV. (2004). Horizontal gene transfer and prokaryotic genome evolution. In: Miller RV, Day MJ (eds). *Microbial Evolution. Gene Establishment, Survival, and Exchange*. ASM Press: Washington, DC, pp 319–338.
- Kamagata Y, Fulthorpe RR, Tamura K, Takami H, Forney LJ, Tiejfe JM. (1997). Pristine environments harbor a new group of oligotrophic 2,4-dichlorophenoxyacetic acid-degrading bacteria. *Appl Environ Microbiol* **63**: 2266–2272.
- Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S *et al.* (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* **9**: 225–256.
- King G. (2007). Chemolithotrophic bacteria: distribution, functions and significance in volcanic environments. *Microbes Environ* **22**: 309–319.
- Koide T, Zaini PA, Moreira LM, Vencio RZN, Matsukuma AY, Durham AM *et al.* (2004). DNA microarray-based genome comparison of a pathogenic and a nonpathogenic strain of *Xylella fastidiosa* delineates genes important for bacterial virulence. *J Bacteriol* **186**: 5442–5449.
- Koyama O, Manome A, Okubo M, Yokomaku T, Tanaka H. (2007). Necrosis gene-based monitoring and control of potato scab disease. *Microbes Environ* **22**: 123–127.
- Larimer W, Chain FP, Hauser L, Lamerdin J, Malfatti S, Do L *et al.* (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodospseudomonas palustris*. *Nat Biotechnol* **22**: 55–61.
- Leonard LT. (1943). A simple assembly for use in the testing of cultures of rhizobia. *J Bacteriol* **45**: 523–525.
- MacLean AM, Finan TM, Sadowsky MJ. (2007). Genomes of the symbiotic nitrogen-fixing bacteria of legumes. *Plant Physiol* **144**: 615–622.
- Maier RJ, Triplett EW. (1996). Toward more productive, efficient, and competitive nitrogen-fixing symbiotic bacteria. *Crit Rev Plant Sci* **15**: 191–234.
- Manning SD, Motiwala AS, Sprigman AC, Qi W, Lacher DW, Oullette LM *et al.* (2008). Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proc Natl Acad Sci USA* **105**: 4868–4873.
- Masuda T, Fujita K, Kogure K, Ogata S. (1989). Effect of CO<sub>2</sub> enrichment and nitrate application on vegetative growth and dinitrogen fixation of wild and cultivated soybean varieties. *Soil Sci Plant Nutr* **35**: 357–366.
- Mauchline TH, Fowler JE, East AK, Sartor AL, Zaheer R, Hosie AHF *et al.* (2006). Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. *Proc Natl Acad Sci USA* **47**: 17933–17938.
- Minamisawa K, Itakura M, Suzuki M, Ichige K, Isawa T, Yuhashi K *et al.* (2002). Horizontal transfer of nodulation genes in soil and microcosms from *Bradyrhizobium japonicum* to *B. elkanii*. *Microbes Environ* **17**: 82–90.
- Minamisawa K, Nakatsuka Y, Isawa T. (1999). Diversity and field site variation of indigenous populations of soybean bradyrhizobia in Japan by fingerprints with repeated sequence RS $\alpha$  and RS $\beta$ . *FEMS Microbiol Ecol* **29**: 171–178.
- Molouba F, Lorquin J, Willems A, Hoste B, Giraud E, Dreyfus B *et al.* (1999). Photosynthetic bradyrhizobia from *Aeschynomene* spp. are specific to stem-nodulated species and form a separate 16S ribosomal DNA restriction fragment length polymorphism group. *Appl Environ Microbiol* **65**: 3084–3094.
- Nautiyal CS, Hegde SV, van Berkum P. (1988). Nodulation, nitrogen fixation, and hydrogen oxidation by pigeon pea *Bradyrhizobium* spp. in symbiotic association with pigeon pea, cowpea, and soybean. *Appl Environ Microbiol* **54**: 94–97.
- Nieuwkoop AJ, Banfalvi Z, Deshmane N, Garhold D, Schall MG, Sirotkin KM *et al.* (1987). A locus

- encoding host range is linked to the common nodulation genes of *Bradyrhizobium japonicum*. *J Bacteriol* **169**: 2631–2638.
- Obert C, Sublett J, Kaushal D, Hinojosa E, Barton T, Tuomanen EI *et al.* (2006). Identification of a candidate *Streptococcus pneumoniae* core genome and regions of diversity correlated with invasive pneumococcal disease. *Infect Immun* **74**: 4766–4777.
- Pearson BM, Pin C, Wright J, l'Anson K, Humphrey T, Wells JM. (2003). Comparative genome analysis of *Campylobacter jejuni* using whole genome microarrays. *FEBS Lett* **554**: 224–230.
- Pessi G, Ahrens HC, Rehrauer H, Lindemann A, Hauser F, Fischer HM *et al.* (2007). Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. *Mol Plant Microbe Interact* **20**: 1353–1363.
- Sadowsky MJ, Cregan PB, Gottfert M, Sharma A, Gerhold D, Rodriguez-Quinones F *et al.* (1991). The *Bradyrhizobium japonicum nola* gene and its involvement in the genotype-specific nodulation of soybeans. *Proc Natl Acad Sci USA* **88**: 637–641.
- Saeki Y, Aimi N, Tsukamoto S, Yamakawa T, Nagamoto Y, Akao S. (2006). Diversity and geographical distribution of indigenous soybean-nodulating bradyrhizobia in Japan. *Soil Sci Plant Nutr* **52**: 418–426.
- Saeki Y, Kaneko A, Hara T, Suzuki K, Yamakawa T, Nguyen MT *et al.* (2005). Phylogenetic analysis of soybean-nodulating rhizobia isolated from alkaline soils in Vietnam. *Soil Sci Plant Nutr* **51**: 1043–1052.
- Saito A, Kawahara M, Ikeda S, Ishimine M, Akao S, Minamisawa K. (2008). Broad distribution and phylogeny of anaerobic endophytes of cluster XIVa clostridia in plant species, including crops. *Microbes Environ* **23**: 73–80.
- Saito A, Mitsui H, Hattori R, Minamisawa K, Hattori T. (1998). Slow-growing and oligotrophic soil bacteria phylogenetically close to *Bradyrhizobium japonicum*. *FEMS Microbiol Ecol* **25**: 277–286.
- Sameshima R, Isawa T, Sadowsky MJ, Hamada T, Kasai H, Shutsrirung A *et al.* (2003). Phylogeny and distribution of extra-slow-growing *Bradyrhizobium japonicum* harboring high copy numbers of RS $\alpha$ , RS $\beta$  and IS1631. *FEMS Microbiol Ecol* **44**: 191–202.
- Sameshima-Saito R, Chiba K, Minamisawa K. (2006). Correlation of denitrifying capability with the existence of *nap*, *nir*, *nor* and *nos* genes in diverse strains of soybean bradyrhizobia. *Microbes Environ* **21**: 174–184.
- Starkenburger RS, Chain SGP, Sayavedra-Soto AL, Hauser L, Land LM, Larimer WF *et al.* (2006). Genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium *Nitrobacter winogradskyi* Nb-255. *Appl Environ Microbiol* **72**: 2050–2063.
- Trung BC, Yoshida S. (1983). Improvement of Leonard jar assembly for screening of effective *Rhizobium*. *Soil Sci Plant Nutr* **29**: 97–100.
- Uchiumi T, Ohwada T, Itakura M, Mitsui H, Nukui N, Dawadi P *et al.* (2004). Expression islands clustered on the symbiosis island of the *Mesorhizobium loti* genome. *J Bacteriol* **186**: 2439–2448.
- van Berkum P. (1990). Evidence for a third uptake hydrogenase phenotype among the soybean bradyrhizobia. *Appl Environ Microbiol* **56**: 3835–3841.
- van Berkum P, Fuhrmann JJ. (2000). Evolutionary relationships among the soybean bradyrhizobia reconstructed from 16S rRNA gene and internally transcribed spacer region sequence divergence. *Int J Syst Evol Microbiol* **50**: 2165–2172.
- Wei M, Yokoyama T, Minamisawa K, Mitsui H, Itakura M, Kaneko T *et al.* (2008). Soybean seed extracts preferentially express genomic loci of *Bradyrhizobium japonicum* in the initial interaction with soybean *Glycine max* (L.) Merr. DNA Research advance online publication 29 May 2008. DNA Research advance online publication 29 May 2008; doi:10.1093/dnares/dsn012.
- Zhou D, Han Y, Song Y, Tong Z, Wang J, Guo Z *et al.* (2004). DNA microarray analysis of genome dynamics in *Yersinia pestis*: insights into bacterial genome microevolution and niche adaptation. *J Bacteriol* **186**: 5138–5146.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)