ISME

ARTICLE   OPEN

Check for updates

Evolution of rhizobial symbiosis islands through insertion sequence-mediated deletion and duplication

Haruka Arashida [iD][1], Haruka Odake[1], Masayuki Sugawara [iD][1], Ryota Noda[1], Kaori Kakizaki[1], Satoshi Ohkubo [iD][1], Hisayuki Mitsui[1], Shusei Sato [iD][1] and Kiwamu Minamisawa [iD][1] ✉

Symbiosis between organisms influences their evolution via adaptive changes in genome architectures. Immunity of soybean carrying the *Rj2* allele is triggered by NopP (type III secretion system [T3SS]-dependent effector), encoded by symbiosis island A (SymA) in *B. diazoefficiens* USDA122. This immunity was overcome by many mutants with large SymA deletions that encompassed T3SS (*rhc*) and N$_2$ fixation (*nif*) genes and were bounded by insertion sequence (IS) copies in direct orientation, indicating homologous recombination between ISs. Similar deletion events were observed in *B. diazoefficiens* USDA110 and *B. japonicum* J5. When we cultured a USDA122 strain with a marker gene *sacB* inserted into the *rhc* gene cluster, most sucrose-resistant mutants had deletions in *nif/rhc* gene clusters, similar to the mutants above. Some deletion mutants were unique to the *sacB* system and showed lower competitive nodulation capability, indicating that IS-mediated deletions occurred during free-living growth and the host plants selected the mutants. Among 63 natural bradyrhizobial isolates, 2 possessed long duplications (261–357 kb) harboring *nif/rhc* gene clusters between IS copies in direct orientation via homologous recombination. Therefore, the structures of symbiosis islands are in a state of flux via IS-mediated duplications and deletions during rhizobial saprophytic growth, and host plants select mutualistic variants from the resultant pools of rhizobial populations. Our results demonstrate that homologous recombination between direct IS copies provides a natural mechanism generating deletions and duplications on symbiosis islands.

INTRODUCTION
Symbiotic organisms can co-evolve through adaptive changes in the organization of the functional elements in their genomes. For instance, in animal–bacteria symbiosis, the evolution of obligate symbiont bacteria with intercellular lifestyles has been accompanied by a marked reduction in genome size [1–3]. The best-studied plant–bacteria symbiosis involves legume plants and nitrogen-fixing bacteria called rhizobia [3, 4]. Rhizobia have repeated symbiotic phases (in the plant) and free-living phases (in the soil) [4, 5]. As facultative symbionts, rhizobia generally possess the distinct packages of symbiosis genes (symbiosis islands or symbiotic plasmids) within their genome [3, 4, 6].

Major rhizobia in Alphaproteobacteria include species within the genera *Bradyrhizobium*, *Azorhizobium*, *Sinorhizobium*, *Rhizobium*, and *Mesorhizobium* [7]. Among them, *Bradyrhizobium* species are thought to be ancestral rhizobia, because *Bradyrhizobium*, which shows enormous species diversity (>800 species), nodulates primitive leguminous plants [6, 8, 9].

Rhizobial symbiosis islands on the genomes of *Bradyrhizobium* [10–12], *Azorhizobium* [13], and *Mesorhizobium* [14–18] contain strain-specific symbiotic genes. Comparisons of *Mesorhizobium* genomes reveal different structures of symbiosis islands depending on geography and host plants [16–18]. In *Bradyrhizobium*, symbiosis islands include *nod*, *nif*, and *rhc*, which function in nodulation of roots, N$_2$ fixation, and type III secretion system (T3SS), respectively.

The symbiosis islands contain conserved regions relevant to the above symbiotic genes and extremely mosaic regions containing insertion sequences (ISs) [10–12, 16–18]. However, little is known about how the symbiosis island structure changes depending on the strain.

ISs are simple mobile genetic elements that impact bacterial evolution including deleterious, neutral, or beneficial effects in bacteria [19]. Multiple copies of an identical IS element dispersed over a genome can promote various genomic rearrangements such as inversion, deletion, duplication, and fusion of two replicons [20, 21]. However, the evolutionary roles of IS elements have been underestimated due to difficulty in their identification on bacterial genomes; (i) repetition of the same ISs on the genomes, and (ii) peculiar signatures of IS elements [22]. ISs are composed of one or two transposase-encoding genes and two terminal inverted repeats, generating two direct repeated sequences as target duplication at the border during transposition [23].

Cultivar-specific restriction of nodulation in soybean has a long history from the 1960s: several dominant genes (*Rj2*, *Rj3*, *Rj4*, and *Rfg1*) in soybeans restrict nodulation with specific rhizobial strains [24–26]. *Rj2*–genotype soybeans restrict nodulation with *Bradyrhizobium diazoefficiens* USDA122 [24, 25]. Rhizobia secrete through type III secretion system nodule outer proteins (Nops), which impact positive, neutral, and negative
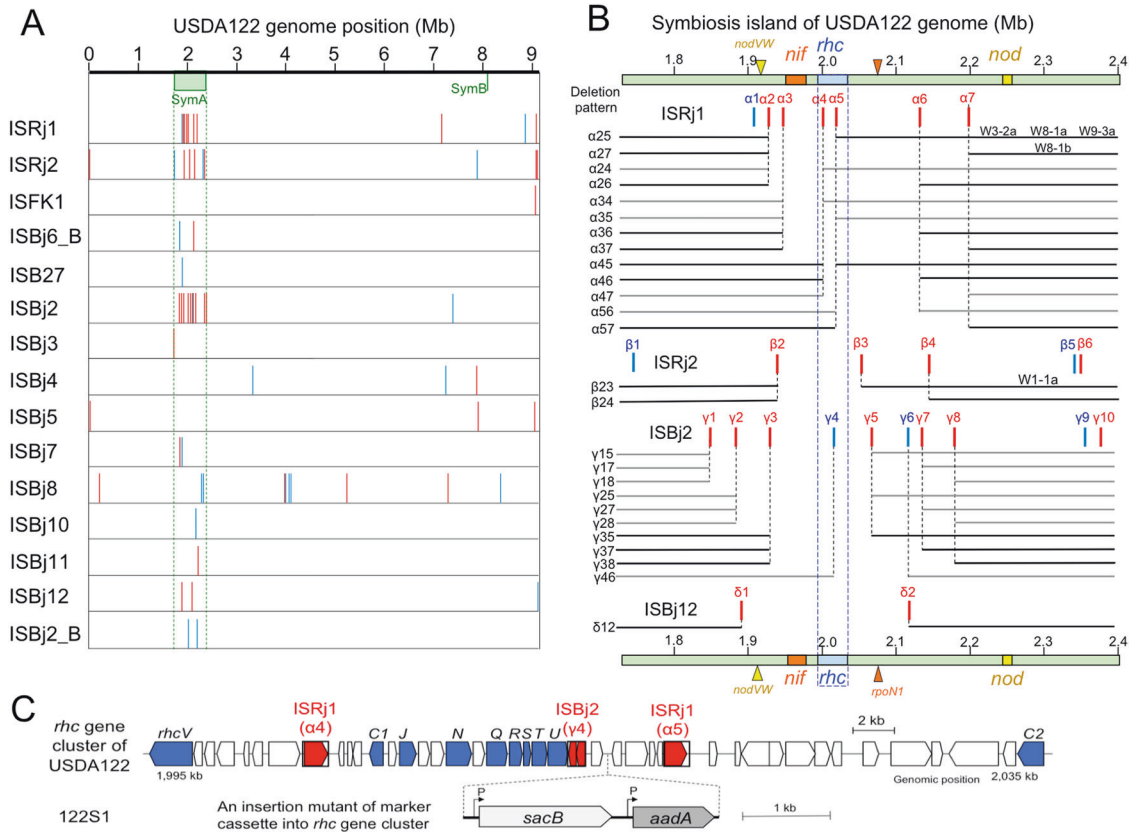
Fig. 1 The position of the insertion sequences (ISs) in the genome (A), symbiosis island A (B), and the *rhc* gene cluster (C) of *Bradyrhizobium diazoefficiens* USDA122. Red and blue vertical lines show ISs with their clockwise and anticlockwise directions on the genome, respectively. **A** IS position of 15 different ISs was identified on USDA122 genome. Symbiosis islands of A and B (SymA and SymB) were estimated based on USDA110 genome [10, 11]. Each IS element was numbered along with the location of the SymA, such as α2. **B** Predicted deletion patterns between the ISs were estimated on the assumption that parts of the *rhc* genes encoding the T3SS machinery were deleted between the same IS elements with identical directions on the SymA of USDA122. The deletion patterns were named based on the IS position, such as α25. We included previous deletion patterns of α25, α27, and β23 from the USDA122 mutants W3-2a, W8-1a, W9-3a, W8-1b, and W1-1a [31]. Bold black lines indicate the deletion patterns that were experimentally obtained. Orange, blue, and yellow squares show *nif*, *rhc*, and *nod* gene clusters on SymA, respectively. Yellow and red triangles indicate the positions of *nodVW* and *rpoN1* genes, respectively (see text). **C** Genetic organization of the *rhc* gene cluster on the genomes of *B. diazoefficiens* USDA122 and its marker cassette-insertion mutant 122S1. The marker cassette containing *sacB* (sucrose sensitive gene) and *aadA* (streptomycin and spectinomycin resistant gene) with its own constitutive promoter (P), was inserted in an intergenic region in the *rhc* gene cluster.

effects on symbiosis depending on combinations between rhizobia and host plants [27–29]. Although most Nops diminish plant defense responses during rhizobial infection [27–29], *Sinorhizobium fredii* NopP oppositely elicited soybean defense responses [30]. Sugawara and co-workers found that bradyrhizobial NopP is a causal T3SS effector to induce symbiotic incompatibility with *Rj2*–genotype soybeans via effector-triggered immunity [31–33]. Zhao et al. [34] reported adaptive evolution of symbiotic compatibility by IS insertion into type III and *nopP* genes by incompatible combinations between wild-type *Sinorhizobium fredii* and soybeans.

Sugawara et al. [31] also found five mutants that overcame NopP–*Rj2* incompatibility due to partial genome deletions in *Bradyrhizobium diazoefficiens* USDA122. Some of these mutants had deleted T3SS gene (*rhc*) clusters on the symbiosis island; in these mutants the resultant lack of the T3SS machinery prevented NopP secretion [31]. Other mutants, in which both *nif* and *rhc* genes were deleted, showed no N_2-fixing activity [31] and were parasitic mutants [35].

Here, we studied the dynamics of symbiosis island structures based on IS elements and *Rj2* incompatibility to reveal the mechanisms behind the generation of symbiosis island variations in *Bradyrhizobium*, and their involvement in phenotypic drift for mutualistic and parasitic behaviors.

RESULTS

Bioinformatic prediction of IS-mediated deletions

In a previous study, we determined that two ISs in direct orientation on symbiosis island A (SymA) of *B. diazoefficiens* USDA122 were involved in three types of deletions in the USDA122 genome [31]. Here, we examined the IS element distributions in USDA122 SymA to predict and confirm the modes of IS-mediated deletion.

The positions of symbiosis islands on the USDA122 genome [36] were determined using the low G + C content of symbiosis islands and their position relative to the Val-tRNA gene in the genome of strain USDA110 [10, 11]. The major symbiosis island SymA (671 kb), which includes *nif*, *rhc*, and *nod* genes, was located in the USDA122 genome at coordinates 1,730,003–2,401,617 bp (Fig. 1A).

We then searched the USDA122 genome for 21 IS elements (ISRj1, ISRj2, ISFK1, IS*1632*, ISBj6_B, ISB27, ISBj2 to ISBj12, ISBj7_B, ISBj5_B, ISBj2_B, and IS*1631*) that were previously identified in the genomes of strains USDA110 [11] and NK6 [37]. When we subjected the USDA122 genome to a BlastN search with the IS elements as query sequences, A total of 63 copies of 15 different IS elements were found, 37 copies of which were located within SymA (Fig. 1A; Table S1).

Seven copies of ISRj1, named α1–α7, and six copies of ISRj2, named β1–β6, resided on SymA on USDA122 genome (Fig. 1B).
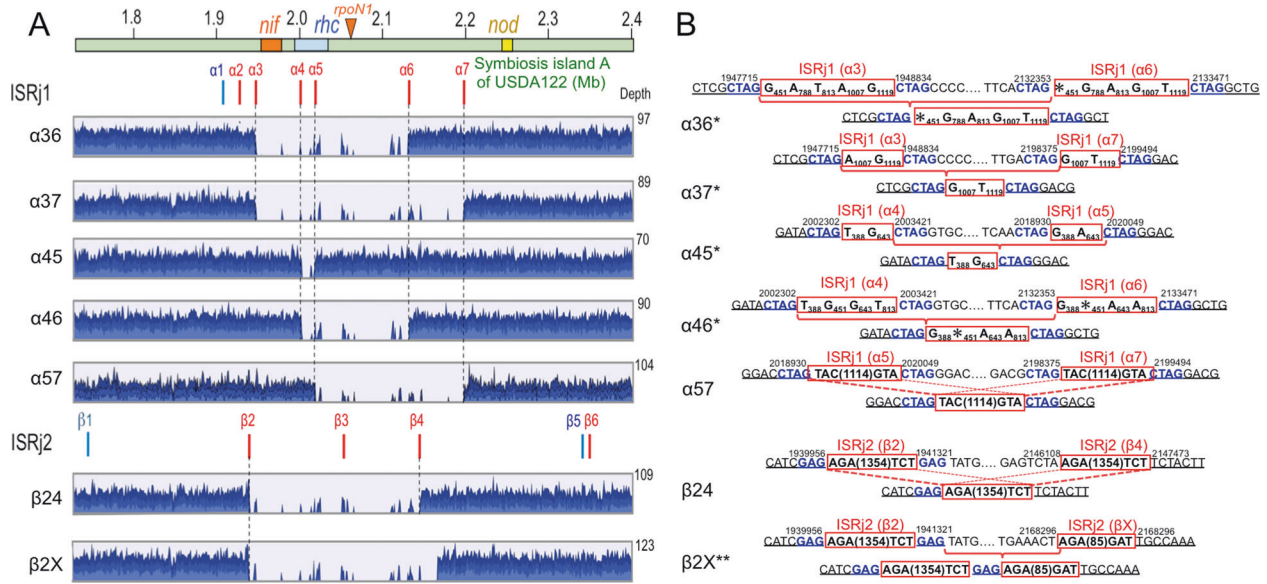
Fig. 2 Mapping profiles of MiSeq reads of *B. diazoefficiens* USDA122 mutant on the parent genome (A) and their junction sequences (B). A Representative mapping profiles of α36, α37, α45, α46, α57, β24, and β2X were derived from derivatives W17, W29, W27, W46b, HG12, W35, and W20, respectively (Table S2). The colors of IS positions and gene clusters were the same in Fig. 1B. **B** Sequence comparisons of the junctions around the IS elements between the parent (above) and mutant (below) for respective deletion patterns of the USDA122. Red boxes (brackets) and blue letters indicate IS elements and putative target duplication sites, respectively. When DNA sequences of the two IS elements were slightly different, the different nucleotides within the IS elements are shown in the red IS boxes, where "*" shows nucleotide deletion. Red brake lines and solid lines show the estimations of the deletion modes of homologous recombination, by identical and slightly different (asterisked deletion pattern) sequences of the IS elements, respectively. Genomic positions on USDA122 are indicated above the DNA sequences.

The transposase genes α2–α7 were oriented clockwise (red in Fig. 1B), whereas α1 was oriented anticlockwise (blue in Fig. 1B). The previously described deletion patterns in USDA122 [31] were designated as α25 (mutants W3-2a, W8-1a, and W9-3a), α27 (W8-1b), and β23 (W1-1a) (Fig. 1B), based on the above nomenclatures for ISRj1 and ISRj2 copies: i.e., α25 is a deletion between α2 and α5, α27 is a deletion between α2 and α7, and β23 is a deletion between β2 and β3.

On the basis of the assumption that the regions harboring *rhc* genes could be deleted between the IS copies in direct orientation by homologous recombination, we found 26 possible deletion patterns mediated by ISRj1, ISRj2, ISBj2, and ISBj12 (Fig. 1B). This suggests that an additional 23 patterns of IS-mediated deletions, other than the previously reported patterns, could occur in USDA122 SymA.

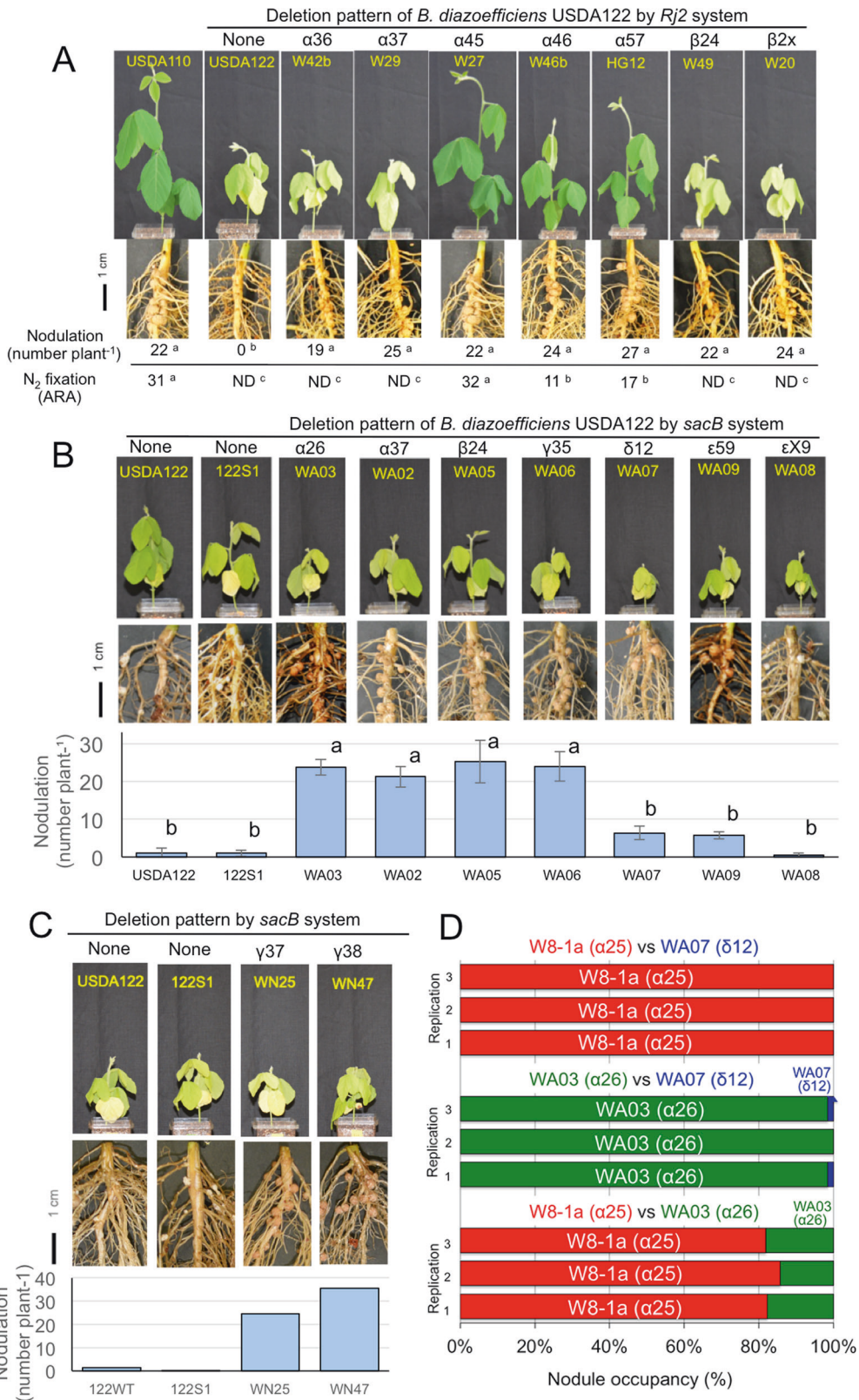### Deletion mutants of *B. diazoefficiens* USDA122

To test whether the predicted deletion patterns were generated, we inoculated USDA122 onto 158 plants of *Rj2*-soybean cultivar 'Hardee'. PCR analyses showed that 30 isolates from 72 spontaneous nodules lost *nifH* and/or *rhcJ* genes but possessed *nodC* gene (Table S2). The remaining 42 isolates showed *nifH/rhcJ/nodC* signals based on PCR analysis, which were not subjected to further genome analyses. However, we continued genome analyses of an isolate HG20 from one nodule of 'Hardee' inoculated with 122GFP (Table S3), although HG20 showed positive PCR signals of *nifH/ rhcJ/nodC* (Table S2). The mapping of MiSeq reads on the USDA122 genome suggested 6 new patterns of genome deletions (α36, α37, α45, α46, α57, and β24) between ISRj1 or ISRj2 copies in direct orientation (clockwise), and a β2X pattern between an 85-bp short fragment (βX) and full copy (β2) of ISRj2 (Fig. 2A), while the remaining 12 mutants showed the previously identified deletion patterns, α25, α27, and β23 (Table S2).

When the deletion junctions of the isolates were sequenced, the patterns α36, α37, α45, α46, α57, and β24 were confirmed as new deletion patterns (Fig. 2A) that were generated by typical homologous recombination between two copies of the same IS elements in direct orientation (Fig. 2B). Because four ISRj1 copies at positions α3, α4, α6, and α7 on USDA122 SymA possessed minor sequence variations and target duplicates, deletion patterns α36, α37, α45, and α46 consistently kept this sequence variation and a target duplicate (5′-CTAG) during homologous recombination (Fig. 2B). The β24 deletion pattern also showed similar homologous recombination between two ISRj2 copies, at positions β2 and β4, on USDA122 SymA (Fig. 2B). All the deletion patterns, except β2X (Fig. 2A), were predicted in the bioinformatic analysis, validating the prediction method (Fig. 1B).

Nodulation and N₂ fixation of USDA122 mutants

Among the seven newly identified deletion patterns (Fig. 2A), *nif* and *rhc* clusters were deleted in α36, α37, β24, and β2X, whereas only the *rhc* cluster was deleted in α45, α46, and α57 (Fig. 2A). To determine the symbiotic phenotypes of representative mutants (W42a, W29, W27, W46b, HG12, W49, and W20) covering the seven deletion patterns were inoculated onto "Hardee". All plants inoculated with the mutants were well nodulated (Fig. 3A). In contrast, wild-type USDA122 did not form nodules on *Rj2*-soybean roots due to *Rj2*-incompatibility (Fig. 3A). Then, we evaluated the N₂-fixing activity of the nodules by measuring acetylene-reducing activity (ARA). No ARA was detected in the nodules infected with the mutants W42b (deletion pattern, α36), W29 (α37), W49 (β24), or W20 (β2X) (Fig. 3A), which lacked *nif* genes as well as *rhc* genes (Fig. 2A). ARA was detected in nodules with W27 (α45), W46b (α46), and HG12 (α57) (Fig. 3A), which had a conserved *nif* cluster but deleted *rhc* cluster (Fig. 1B); however, ARA values of W46b (α46) and HG12 (α57) were significantly lower than those of W27 (α45) (Fig. 3A). This finding indicates that genes for efficient symbiotic N₂ fixation may be located in the regions between α5 and α6 of USDA122 SymA (Fig. 2A). Our survey of this region suggests that *rpoN1* encoding sigma54 of RNA polymerase is a candidate gene for efficient symbiotic N₂ fixation (Fig. 2A), because the *rpoN1* mutant of *B. diazoefficiens* USDA110 shows reduced N₂-fixing activity [38].

A Deletion pattern of *B. diazoefficiens* USDA122 by *Rj2* system

	None	α36	α37	α45	α46	α57	β24	β2x
	USDA110	USDA122	W42b	W29	W27	W46b	HG12	W49	W20
Nodulation (number plant^{-1})	22 [a]	0 [b]	19 [a]	25 [a]	22 [a]	24 [a]	27 [a]	22 [a]	24 [a]
N$_2$ fixation (ARA)	31 [a]	ND [c]	ND [c]	ND [c]	32 [a]	11 [b]	17 [b]	ND [c]	ND [c]

B Deletion pattern of *B. diazoefficiens* USDA122 by *sacB* system

None	None	α26	α37	β24	γ35	δ12	ε59	εX9
USDA122	122S1	WA03	WA02	WA05	WA06	WA07	WA09	WA08

C Deletion pattern by *sacB* system

None	None	γ37	γ38
USDA122	122S1	WN25	WN47

D
W8-1a (α25) vs WA07 (δ12)
WA03 (α26) vs WA07 (δ12)
W8-1a (α25) vs WA03 (α26)

Deletion in *B. diazoefficiens* USDA110 and *B. japonicum* J5

When IS elements of *B. diazoefficiens* USDA110 [11] and *B. japonicum* J5 [39] were searched using the same strategy as that used for USDA122 above, we identified 16 and 14 different IS elements, respectively, in their genomes (Table S1); these IS elements were concentrated on SymA in both genomes (Fig. S1). Based on the location and orientation of these IS elements and the *rhc* cluster, we predicted 23 and 17 patterns for IS-mediated deletions on SymA of the USDA110 and J5 genomes, respectively (Fig. 4). Although J5 possesses a *nopP* that is incompatible with

Fig. 3 Symbiotic phenotypes of *Rj2*-soybean plants inoculated with representative mutants, for the different deletion patterns of *Bradyrhizobium diazoefficiens* USDA122. *Rj2*-soybean plants were inoculated with deletion mutants by the *Rj2* system (**A**) and by the *sacB* system (**B**, **C**). **A**, **B** "Nodulation" is expressed as the average number of nodules per plant. The bar shows standard deviation. N_2-fixing activities were evaluated by acetylene-reducing activity (ARA), which is expressed as μmol C_2H_2 produced h^{-1} g nodule fresh weight^{-1}. ND indicates "not detected" ($<0.08\ \mu$mol $h^{-1}\ g^{-1}$). Values are expressed as averages of 3 or 4 replications, except for the nodulation data in panel C ($n = 2$–3). Average values with the same letter are not significantly different by Tukey's HSD test ($P < 0.05$). **D** Nodule occupancy of *Rj2*-soybean plants inoculated with mixtures of equal amounts of the two deletion mutants. The seeds of *Rj2*-soybean cv. Hardee was inoculated with one-to-one mixed cells of three pairwise combinations: (i) W8-1a (pattern $\alpha25$) and WA07 ($\delta12$), (ii) WA03 ($\alpha26$) and WA07 ($\delta12$), and (iii) W8-1a ($\alpha25$) & WA03 ($\alpha26$). Nodule occupancy was determined by multiplex PCR (Fig. S8).

Rj2-soybeans, USDA110 has a *nopP* that is compatible with *Rj2*-soybeans; this gene is slightly different from the USDA122 *nopP* [31]. Therefore, we used a USDA110 derivative that carries a USDA 122–type *nopP* gene (strain 110*nopP*$_{122}$) [31] for the inoculation experiments.

After USDA110 (110*nopP*$_{122}$) and J5 were inoculated onto "Hardee", we obtained 5 and 3 mutants, respectively, from the spontaneous nodules (Table 1, Table S2). Mapping and sequencing data for these mutants indicated that the deleted regions included *rhc* genes and were between two copies of the same IS (ISRj1 or ISBj2) in direct orientation on SymA (Black bold lines in Fig. 4). The deletion junction sequences in USDA110 and J5 suggested homologous recombination events like USDA122 (Fig. S2).

The USDA110 deletions showed four patterns ($\alpha25$D, $\alpha34$D, $\alpha35$D, and $\alpha36$D) (Fig. 4A, Table S2), all of which deleted both *nif* and *rhc* clusters (Fig. 4A). The J5 deletions showed two patterns ($\alpha34$J and $\gamma69$J) (Fig. 4B, Table S2): in the $\alpha34$J pattern, only the *rhc* cluster was deleted; but in the $\gamma69$J pattern, both *nif* and *rhc* clusters were deleted (Fig. 4B). The pattern $\gamma69$J in J5 showed the largest deletion (541 kb), which included 54% of SymA (998 kb) and 5.3% of the entire J5 genome (10.1 Mb) [39].

Symbiotic phenotypes of representative mutants from USDA110 and J5 were also examined by *Rj2*-soybean inoculation (Fig. S3). USDA110 mutants, M7 (deletion pattern, $\alpha35$D), M8 ($\alpha36$D), M11b ($\alpha25$D), and M14 ($\alpha34$D) showed a negative N_2-fixing (Fix−) phenotype (Fig. S3A), as expected from their *nif* deletion genotype (Table 1). J5 mutants J2a ($\gamma69$J) and J9 ($\alpha34$J) showed Fix- and positive N_2-fixing (Fix+) phenotypes, respectively, also in accordance with their *nif* genotypes (Fig. S3B).

IS-mediated deletion assayed using the *sacB* system
We then examined whether IS-mediated deletions of the USDA122 genome can occur in free-living growth without the *Rj2*-soybean host. Thus, we designed an experimental system (*sacB* system) using a negative selection marker, *sacB*, to detect the deletion of the *rhc* gene cluster (Fig. 1C). A marker cassette containing *sacB* and *aadA* genes with constitutive promoters was inserted into the intergenic region of *rhc* gene cluster of USDA122, resulting in strain 122S1 (Fig. 1C). Because the expression of *sacB* encoding levansucrase is lethal for bacteria in the presence of sucrose [40, 41], we expected that sucrose-resistant colonies would have a deleted or mutagenized *sacB* gene and that a large subset of these colonies would display a deleted *rhc* genes. In contrast, since *aadA* gene confers resistance to spectinomycin and streptomycin (hereafter, Sp/Sm) [42], we expected that colonies resistant to Sp/Sm would have an intact *aadA* gene indicating that the *rhc* cluster was not deleted.

After cultivation of strain 122S1 in HM broth for 5 days, we obtained 32 sucrose-resistant mutants from independent colonies on HM agar plates supplemented with 10% (w/v) sucrose. Of the 32 mutants, 22 were also sensitive to Sp/Sm and produced no *sacB* PCR products (Table 2, Fig. S4), indicating that the marker cassette was likely deleted. Mapping analyses of these 22 mutants on USDA122 genome (Fig. S5A), indicated that 21 mutants were represented in ten patterns of IS-mediated deletions ($\alpha26$, $\alpha36$, $\alpha37$, $\beta23$, $\beta24$, $\gamma35$, $\gamma37$, $\gamma38$, $\delta12$, and $\epsilon59$) and one mutant, WA08,

was not (ϵX9) (Table S2). Sequencing verified that the deletion events involved homologous recombination via ISs in the 21 mutants, although the sequence (1081 bp) for homologous recombination of WA09 was not a full sequence of ISFK1 (2592 bp) (Fig. S5B).

Of the 32 sucrose-resistant mutants, ten were resistant to Sp/Sm (Table 2), suggesting the existence of an intact *aadA* gene in *sacB/aadA* cassette (Fig. 1C). Subsequent analyses of *sacB* indicated that 8 of these mutants possessed single nucleotide polymorphisms (SNPs) on *sacB* (Table 2, Table S4), which may lose their levansucrase activity. Combining the two bioassays for sucrose resistance and Sp/Sm sensitivity of 122S1 greatly facilitated the efficient detection of the 21 mutants with IS-mediated deletions. In addition, the frequency of IS-mediated deletion mutants reached 2.2×10^{-4}, based on total colony-forming units in 5-day-old cultures of 122S1 (Table 2).

To examine the real diversity in the deletion patterns of mutants from free-living cultures, we further isolated an additional 96 sucrose-resistant mutants from 4-day-old cultures of 122S1. Sp/Sm sensitivity assay suggested that 59 mutants had deleted the marker cassette. The IS-PCR analysis indicated that 49 of these 59 mutants (WN01, WN03–WN24, WN26–WN46, WN48, WN49, and WN51–WN53) had already-known deletion patterns ($\alpha25$, $\alpha26$, $\alpha27$, $\alpha36$, $\alpha37$, $\alpha45$, $\beta23$, $\beta24$, $\gamma35$, and $\delta12$) (Table S2). MiSeq reads of the remaining 10 mutants were mapped on the USDA122 genome, and 4 mutants (WN02, WN25, WN47, and WN50) were identified as having new ISBj2-mediated deletion patterns ($\gamma37$ and $\gamma38$) on USDA122 SymA (Fig. S5. Table 1, Table S2).

The remaining six mutants (WN101–WN106) showed complicated mapping profiles based on the USDA122 genome (Table S2, Fig. S6). The deletion events likely occurred between a full IS copy (ISRj2 or ISBj2) and a shorter fragment of the corresponding IS in three mutants, WN101, WN102, and WN103 (Fig. S6). MiSeq reads of mutant WN105 showed two characteristic loci of heavily (74 kb) and null (145 kb) mapped regions that were adjacent to each other (Fig. 5A). The heavily mapped region (mapping depth, ~4 times the basal level) had an ISFK1 fragment at each border. The results for mutant WN105 suggest that multiple duplication events likely occurred between two ISFK1 fragments, as well as a deletion event (detected as the 145 kb null mapped region), that included *nif/rhc* gene clusters (Fig. 5A).

Symbiotic phenotype of deletion mutants isolated by the *sacB* system
Representative mutants from *sacB* system were inoculated onto "Hardee" (Fig. 3B, C). The mutants WA03 (deletion pattern, $\alpha26$), WA02 ($\alpha37$), WA05 ($\beta24$), and WA06 ($\gamma35$) produced large numbers of nodules, in contrast to the parent strain, 122S1. However, WA07 ($\delta12$) and WA09 ($\epsilon59$) produced significantly lower numbers of nodules than the above mutants (Fig. 3B). In a survey for genes relevant to nodulation, *nodVW* gene, which encodes a flavonoid sensor and regulator [43] was found in the deleted regions of WA07 ($\delta12$) and WA09 ($\epsilon59$) (Fig. S5A). Thus, it is possible that these mutants were not obtained in the experiments involving inoculation of *Rj2*-soybean 'Hardee' plants (hereafter, *Rj2*-system), due to the lack of *nodVW*.
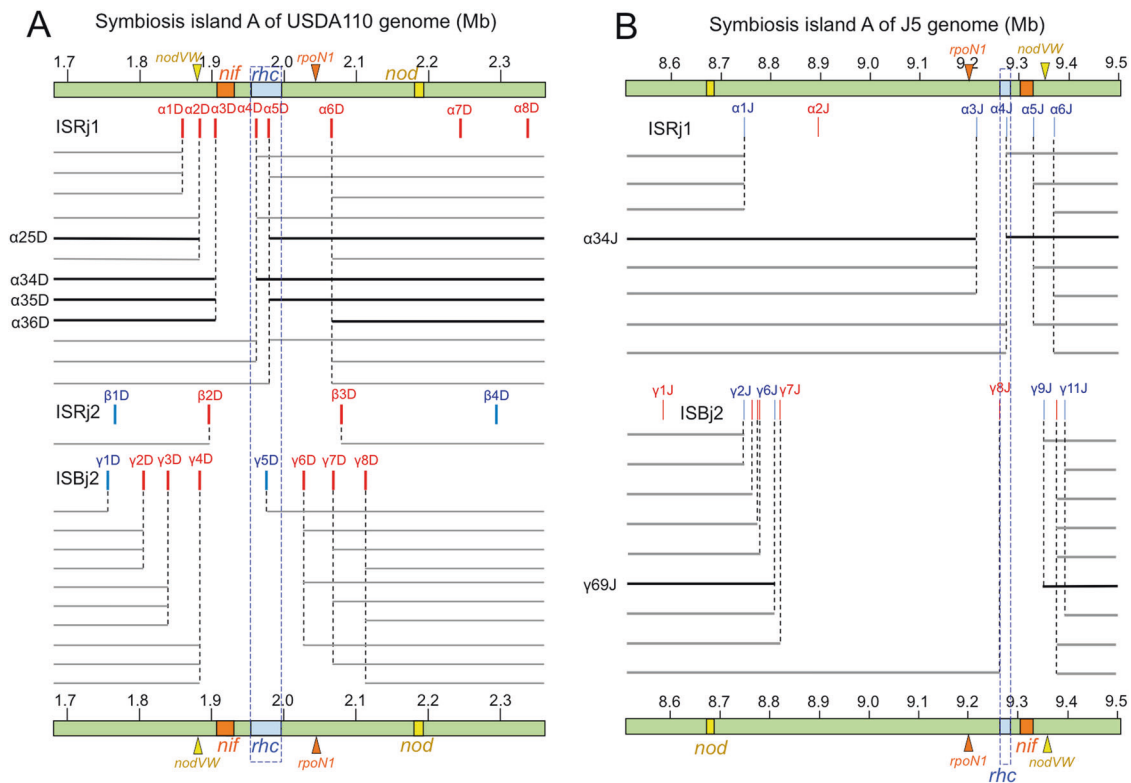
Fig. 4 The positions of insertion sequences (ISs) and the predicted IS-mediated deletion patterns within symbiosis island A of *Bradyrhizobium diazoefficiens* USDA110 (A) and *B. japonicum* J5 (B). Red and blue vertical lines show ISs with clockwise and anticlockwise directions on the genomes, respectively. Predicted deletion patterns between the ISs were estimated on the same assumption as in Fig. 1A and B. The deletion patterns were named based on the IS positions, such as α25D for USDA110 and α34J for J5. Other indications are identical to those in Fig. 1.

Comparisons of deletion patterns obtained by the *Rj2* and *sacB* systems

When deletion patterns of USDA122 mutants obtained using *Rj2* system were compared with those from *sacB* system, both systems were found to share common deletion patterns of α25, α27, α36, α37, α45, β23, and β24 (Table 1). However, *sacB* system generated unique deletion patterns (α26, γ35, γ37, γ38, δ12, and ε59) that were not obtained from *Rj2* system (Table 1); 11 and 25 mutants showed the deletion patterns of α26 and δ12, respectively, accounting for approximately half of the 74 IS-meditated mutants from *sacB* system (Table 1). Therefore, we designed an inoculation experiment to compare the competitive nodulation ability of *sacB* system-specific deletion mutants WA03 (α26) and WA07 (δ12) with that of a mutant detected by both systems, W8-1a (α25) (Fig. S7, Table 1).

When one-to-one mixed cells of mutants W8-1a (α25) and WA07 (δ12) were inoculated onto "Hardee", the nodule occupancy of W8-1a (α25) was extremely dominant against WA07 (δ12) (Fig. 3D). When one-to-one mixed cells of mutants WA03 (α26) and WA07 (δ12) were inoculated, the nodule occupancy of WA03 (α26) was also dominant against WA07 (δ12) (Fig. 3D). These results are expected, due to the low nodulation capability of mutant WA07 (δ12) even in single inoculation experiments (Fig. 3B). When one-to-one mixed cells of mutants W8-1a (α25) and WA03 (α26) were inoculated, the nodule occupancy of W8-1a (α25) was 80%, indicating that nodulation by mutant WA03 (α26) was inferior to that by W8-1a (α25) (Fig. 3D). These results suggest that the mutants with *sacB*-specific deletion patterns α26 and δ12 failed to efficiently nodulate soybean roots under competitive conditions during the selection process of *Rj2* system, thus explaining the lack of mutants with deletion patterns α26 and δ12 obtained from *Rj2* system (Table 1).

PCR verification of IS-mediated deletion

To obtain more direct evidence for IS-mediated deletions during the cultivation of free-living cells of USDA122 and 122S1, we developed two sets of PCR reactions to detect α25 and α26 deletion events mediated by ISRj1 (Fig. S8). When total DNAs from full growth cultures of USDA122 and 122S1 were used as template DNAs, we detected PCR products of 1.8-kb for the α25 deletion and 2.2-kb for the α26 deletion, in both strains (Fig. S8). This result indicates that free-living cells of USDA122 and 122S1 stochastically generated ISRj1-mediated deletions during simple cultivation.

IS-mediated duplications on symbiosis islands

Our *sacB* and *Rj2* systems are not designed to detect duplication events on symbiosis islands; however, by chance, a mutant in which both duplication and deletion happened simultaneously, mutant WN105, was identified (Fig. 5A). Therefore, we explored IS-mediated duplications on symbiosis islands by using a culture collection of *B. diazoefficiens* from soybean fields in Hokkaido, Japan. The DNAs of 62 strains from the culture collection were sequenced on the MiSeq platform and mapped to the reference genome of *B. diazoefficiens* USDA110 [11, 44]. Two field isolates, HF07 and HH12, had heavy mapping depths in a region of SymA in the USDA110 genome, when compared with closely related isolates HF08 and HH15 (Fig. S9). When HF07 and HH12 were further purified by single colony isolation on HM agar medium to produce isolates F07S3 and H12S4, respectively, the re-isolates still showed heavy mapping depth in the same region (Fig. 5B, C). Magnification of F07S3 and H12S4 profiles showed that the heavy mapping region spanned 263 and 230-kb of SymA, respectively, and included *nif* and *rhc* genes but not the *nod* genes (Fig. 5B). Interestingly, in both F07S3 and H12S4, the borders of the heavily

Table 1. Summary of the deletion patterns in mutants of *B. diazoefficiens* USDA122, *B. diazoefficiens* USDA110, and *B. japonicum* J5.

Pattern[a]	Back-ground	IS	Number of mutants		Deleted length (% in SymA)[b]	Symbiotic genes deletion		Plant phenotype[c]	
			Rj2 system	*sacB* system		*nif/rhc*	Others	Nodulation	N$_2$-fixation
α25*	USDA122	ISRj1	12	5	91 kb (14%)	*nif, rhc*		Normal	Fix−
α26	USDA122	ISRj1		11	204 kb (30%)	*nif, rhc*	*rpoN1*	Normal	Fix−
α27*	USDA122	ISRj1	3	3	270 kb (40%)	*nif, rhc*	*rpoN1*	Normal	Fix−
α36	USDA122	ISRj1	9	4	185 kb (27%)	*nif, rhc*	*rpoN1*	Normal	Fix−
α37	USDA122	ISRj1	1	4	251 kb (37%)	*nif, rhc*	*rpoN1*	Normal	Fix−
α45	USDA122	ISRj1	3	1	17 kb (2%)	*rhc*		Normal	Fix+
α46	USDA122	ISRj1	1		130 kb (19%)	*rhc*	*rpoN1*	Normal	Fix+/−
α57	USDA122	ISRj1	1		179 kb (27%)	*rhc*	*rpoN1*	Normal	Fix+/−
β23*	USDA122	ISRj2	2	4	113 kb (17%)	*rhc*		Normal	Fix+
β24	USDA122	ISRj2	3	7	206 kb (31%)	*rhc*	*rpoN1*	Normal	Fix−
γ35	USDA122	ISBj2		5	137 kb (20%)	*nif, rhc*		Normal	Fix−
γ37	USDA122	ISBj2		3	206 kb (31%)	*nif, rhc*	*rpoN1*	Normal	Fix−
γ38	USDA122	ISBj2		1	250 kb (37%)	*nif, rhc*	*rpoN1*	Normal	Fix−
δ12	USDA122	ISBj12		25	227 kb (34%)	*nif, rhc*	*nodVW*	Low	Fix−
ε59	USDA122	Partial ISFK1		1	130 kb (19%)	*nif, rhc*	*nodVW*	Low	Fix−
α34D	USDA110	ISRj1	1		57 kb (8%)	*nif, rhc*		Normal	Fix−
α35D	USDA110	ISRj1	1		74 kb (11%)	*nif, rhc*		Normal	Fix−
α36D	USDA110	ISRj1	1		161 kb (24%)	*nif, rhc*	*rpoN1*	Normal	Fix−
α25D	USDA110	ISRj1	2		96 kb (14%)	*nif, rhc*		Normal	Fix−
α34J	J5	ISRj1	2		61 kb (6%)	*rhc*		Normal	Fix+
γ69J	J5	ISBj2	1		541 kb (54%)	*nif, rhc*	*rpoN1*	Normal	Fix−

[a]Asterisks indicate previous deletion patterns [31].
[b]Percentage of deletion length against symbiosis island A (SymA) of USDA122 (672 kb), USDA110 (681 kb), and J5 (998 kb).
[c]Fix phenotypes were based on Fig. 3 and Fig. S3.

Table 2. Mutation types of 32 sucrose-resistant mutants in free-living cultures of *B. diazoefficiens* 122S1.

Antibiotics test (Sp/Sm)(Fig. S4)[a]	*sacB* PCR (Fig. S4)	Mutation type	Recombination	Number of mutants	Frequency (10^{-4})[b]	Original data
Sensitive (22 mutants)[c]	Negative (22 mutants)	Marker cassette deletion	Homologous	21 (66 %)	2.2	Fig. S5, Table S4
		Marker cassette deletion	Non-homologous	1 (3 %)	0.1	Table S4
Resistant (10 mutants)	Positive (8 mutants)	*sacB* SNP[d]	None	8 (25 %)	0.8	Table S5
	Negative (2 mutants)	Unknown	Unknown	2 (6 %)	0.2	

[a]Spectinomycin (Sp) and Streptomycin (Sm)-sensitive/resistant test on agar plates (Fig. S4).
[b]Frequency calculated from the corresponding mutant number ratio and total frequency of sucrose-resistant colonies of 122S1 that appeared at a frequency of 3.3 ×10^{-4}, based on CFU on HM agar plates without sucrose.
[c]Sp/Sm sensitive mutants from the 32 sucrose-resistant mutants: WA01–WA10, WA12–WA19, and WA21–WA24 (Table S4). Among them, 21 were generated from deletions mediated by the full length of ISs, and 1 for a Partial ISFK1 (Table 1).
[d]Single nucleotide polymorphisms (SNP) of the *sacB* gene (Table S5).

mapped regions were adjacent to copies of ISBj2: γ2D and γ7D in F070S3 and γ4D and γ8D in H12S4 (Fig. 4A; Fig. 5B). These results suggest that the partial duplication of the symbiosis island on the F07S3 and H12S4 genomes was meditated by ISBj2 copies.

The results of hybrid assembly of MiSeq and Nanopore reads suggest that there was tandem duplication of part of SymA: 260 kb in F07S3 (Fig. 5D) and 357 kb in H12S4 (Fig. 5E). This notion was confirmed by PCR and Sanger sequence analyses targeting the junctions of tandem duplications (Fig. 5D, E; Figs. S10 and S11). The determination of the complete genome sequences showed that the genome sizes of F07S3 (9,432,644 bp) and H12S4 (9,535,585 bp) were larger than those of HF08 (9,109,292 bp) and

HH15 (9,177,979 bp), which were obtained from the Otofuke and Kamishihoro fields, respectively, and that this increase in size was due to the tandem duplications, including *nif* and *rhc* genes on SymA (Fig. 5D, E).

DISCUSSION

The horizontal transfer events of symbiosis islands were mediated by integrative and conjugative elements (ICEs) in *Azorhizobium* and *Mesorhizobium* under laboratory conditions [13, 45, 46]. However, these studies do not address how the variations in symbiosis islands were generated during their evolution between
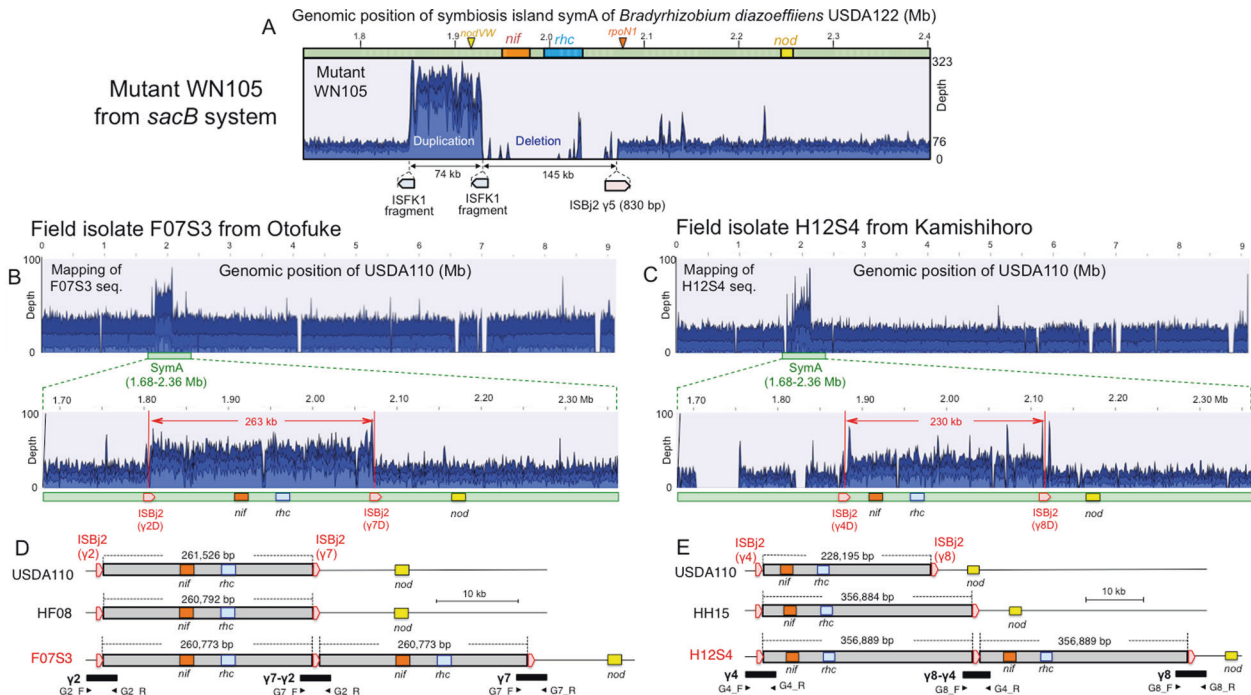
Fig. 5 IS-mediated duplication on symbiosis islands of *Brdadyrhizobium diazoefficiens*. A Mapping profile of the mutant WN105 from 122S1 on USDA122 symbiosis island, that suggested *nif/rhc* deletions and duplications. (BCDE) Complete genome determination of *B. diazoefficiens* field isolates, carrying IS-mediated tandem duplications, including *nif* and *rhc* genes on the symbiosis island. **B**, **C** Mapping profiles of MiSeq reads of *B. diazoefficiens* isolates of F07S3, from Otofuke field (**B**) and H12S4 from Kamishihoro field (**C**), on the *B. diazoefficines* USDA110 genome. **D**, **E** Structures of tandem duplication in F07S3 (**D**) and H12S4 (**E**) mediated by different copies of ISBj2 on their symbiosis islands, which were verified by PCR sequence analyses (Figs. S10 and S11).

rhizobia and legumes [5]. Symbiosis islands of individual strains of *Bradyrhizobium* have become adapted to leguminous plants [47], suggesting the importance of symbiosis island evolution with host plants [5].

Recombination between IS elements in rhizobia has been known as a mechanism for generating variation in rhizobial genomes. In *Sinorhizobium meliloti* carrying a symbiotic plasmid, cointegrates generated IS-mediated replicon fusion, which did not demonstrate selection for the rearrangements [48]. Zhao et al. [34] reported adaptive evolution of symbiotic compatibility of *Sinorhizobium fredii* by inactivation of *rhc* and *nopP* genes by IS insertion. In contrast, our results demonstrate that homologous recombination between direct IS copies on bradyrhizobial symbiosis island provides a natural mechanism generating deletions, which were verified by the *Rj2* and *sacB* systems on laboratory time scales. Our sequence analysis of the boundary regions of the IS-mediated deletions suggests that the major deletion events occurred via typical homologous recombination [20] between the already existing full IS copies (Fig. 2B, Fig. S2, Fig. S5B). The deletion events also occurred between the fragments of an IS element, ISFK1 (Table 1, Fig. S5B). Thus, IS-mediated deletion does not always require IS transposition by transposase activity. However, we regard the deletion pattern β2X to be the trace of two successive events: ISRj2 transposition and subsequent ISRj2-mediated deletion, based on border sequence analysis (Fig. 2B, Fig. S12).

Most of the IS-mediated mutants with deleted *nif* and *rhc* genes exhibited a Fix- phenotype under the genomic backgrounds of *B. diazoefficiens* USDA122 and USDA110 and *B. japonicum* J5 (Table 1). The incidence of IS-mediated deletion mutants ranged from 2.2×10^{-4} (Tables 2) to 8.1×10^{-4} in the experiments where 122S1 was grown free-living for 4–5 days. In addition, major deletion events following the patterns α25 and α26 (Table 1) were directly detected in both USDA122 and 122S1 cultures under no

selection pressure by PCR (Fig. S8). Thus, these mutants defective in *nif* genes (non-N$_2$ fixing mutants) were formed in subpopulations during saprophytic growth (Table 2, Fig. S8) and are likely generated even in soil environments as parasitic bradyrhizobia. However, field isolates of *Bradyrhizobium* from soybean nodules consistently symbiotically fixed N$_2$ [49–51]. As for this discrepancy, one explanation may be a host sanction hypothesis to eliminate the parasitic mutants with deleted *nif* genes [5, 52–56]. However, little effect of host sanction was also observed by inoculation tests and modeling in the symbiosis between soybeans and *Bradyrhizobium*, suggesting unknown mechanisms to eliminate non-N$_2$ fixing bradyrhizbia in field soils [57].

Previous phylogenetic analyses of a natural population of *Bradyrhizobium* suggested that loss-of-nodulation-capability events were potentially driven by mutations or deletions of symbiosis loci [53, 54]. Here we found that pairs of IS copies in direct orientation were distributed around *nod* gene cluster on SymA of USDA122 (Fig. 1B), USDA110 (Fig. 4A), and J5 (Fig. 4B). Thus, it is possible that IS-mediated deletions of *nod* gene clusters may be an underlying mechanism for the loss of nodulation capabilities in natural bradyrhizobial populations [53, 54].

The symbiosis island structures of isolates F07S3 and H12S4 demonstrated tandem duplications of symbiosis island regions (260–357 kb) via two different copies of ISBj2 in direct orientation (Fig. 5B–E); this process appears similar to IS-mediated deletion in that the duplicated regions contained *nif/rhc* gene clusters between two copies of the same IS element in direct orientation. The above two isolates were found among 63 isolates of *B. diazoefficiens* in two independent fields in Hokkaido, Japan. Thus, the duplication events happened in nature, and their incidence was as high as 3% (2/63). Studies of the evolutionary role of IS elements in symbiotic bacteria have focused on genome reduction in obligate symbionts via IS-mediated deletion and host dependence [1–3]. On the basis of our findings, the IS-

H. Arashida et al.

mediated duplication may play an important role in the evolution of rhizobia as facultative symbionts. Tandem duplications on genomes are of great importance in evolutionary genetics and cancer biology because they can dramatically alter gene functions via subsequent genome remodeling [58–60]. In addition, tandem duplications and deletions are often generated by similar recombination mechanisms [58–60]. Since we found IS-mediated tandem duplications on SymA in two field isolates (Fig. 5B–E), we consider that the structures of bradyrhizobial symbiosis islands are likely in a state of flux, and variants could be generated via tandem duplications and subsequent remodeling, including deletions and genetic exchange within rhizobial populations [58–60]. The structures of large tandem duplications in symbiotic *Bradyrhizobium* species have been previously overlooked due to conventional short DNA sequencing rather than long-read sequencing technologies (ideally > 400 kb).

On the basis of our findings of experimental deletions and natural duplications, we propose a comprehensive hypothesis for symbiosis island evolution: (i) active ISs transpose to form IS-rich loci within low G + C symbiosis islands; (ii) the ISs stochastically induce deletions and duplications that generate variations in the symbiosis islands, and (iii) host plants select the variants of the symbiosis islands in rhizobial populations for improved adaptations.

Extensive attention has been paid to the suppression of plant immunity to establish rhizobia–legume symbiosis [35, 61–64]. However, incompatible NopP protein in rhizobia strongly induces plant immunity via the host *Rj2* allele, leading the host to reject the rhizobial infection [31, 64]. A foliar systemic resistant regulator, glycerol-3-phosphate, is required for *Rj2* incompatibility through root-shoot-root signaling [65]. In addition, Zhang et al. [66] reported that a new resistant protein of *NNL1* gene in soybean accessions directly interacts with NopP effector from *B. diazoefficiens* USDA110 to inhibit nodulation through root hair infection. Taken together with our results, such incompatibility might facilitate co-evolution of symbiosis islands in bradyrhizobia, as a host driving force in nature.

MATERIALS AND METHODS
Bacterial materials, marker cassette construction, nodulation assay, nitrogen fixation assay, IS identification, PCR primers (Table S5), deletion profile assay, and complete genome determination are described in Supplementary materials.

REFERENCES
1. Lo WS, Huang YY, Kuo CH. Winding paths to simplicity: genome evolution in facultative insect symbionts. FEMS Microbiol Rev. 2016;40:855–74.
2. Toft C, Andersson SG. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat Rev Genet. 2010;11:465–75.
3. Batut J, Andersson SG, O'Callaghan D. The evolution of chronic infection strategies in the alpha-proteobacteria. Nat Rev Microbiol. 2004;2:933–45.
4. Poole P, Ramachandran V, Terpolilli J. Rhizobia: from saprophytes to endosymbionts. Nat Rev Microbiol. 2018;16:291–303.
5. Remigi P, Zhu J, Young JPW, Masson-Boivin C. Symbiosis within symbiosis: evolving nitrogen-fixing legume symbionts. Trends Microbiol. 2016;24:63–75.
6. Ormeño-Orrillo E, Martínez-Romero E. A genomotaxonomy view of the *Bradyrhizobium* genus. Front Microbiol. 2019;10:1334.
7. Garrido-Oter R, Nakano RT, Dombrowski N, Ma KW, The AgBiome Team, McHardy AC, et al. Modular traits of the Rhizobiales root microbiota and their evolutionary relationship with symbiotic rhizobia. Cell Host Microbe. 2018;24:155–67.
8. Avontuur JR, Palmer M, Beukes CW, Chan WY, Coetzee MPA, Blom J, et al. Genome-informed *Bradyrhizobium* taxonomy: where to from here? Syst Appl Microbiol. 2019;42:427–39.
9. Grönemeyer JL, Reinhold-Hurek B. Diversity of bradyrhizobia in subsahara Africa: a rich resource. Front Microbiol. 2018;9:2194.
10. Kaneko T, Maita H, Hirakawa H, Uchiike N, Minamisawa K, Watanabe A, et al. Complete genome sequence of the soybean symbiont *Bradyrhizobium japonicum* strain USDA6T. Genes (Basel). 2011;2:763–87.
11. Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S, et al. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Res. 2002;9:189–97.
12. Siqueira AF, Ormeño-Orrillo E, Souza RC, Rodrigues EP, Almeida LG, Barcellos FG, et al. Comparative genomics of *Bradyrhizobium japonicum* CPAC 15 and *Bradyrhizobium diazoefficiens* CPAC 7: elite model strains for understanding symbiotic performance with soybean. BMC Genomics. 2014;15:420.
13. Ling J, Wang H, Wu P, Li T, Tang Y, Naseer N, et al. Plant nodulation inducers enhance horizontal gene transfer of *Azorhizobium caulinodans* symbiosis island. Proc Natl Acad Sci USA. 2016;113:13875–80.
14. Uchiumi T, Ohwada T, Itakura M, Mitsui H, Nukui N, Dawadi P. et al. Expression islands clustered on the symbiosis island of the *Mesorhizobium loti* genome. J Bacteriol. 2004;186:2439–48.
15. Kasai-Maita H, Hirakawa H, Nakamura Y, Kaneko T, Miki K, Maruya J, et al. Commonalities and differences among symbiosis islands of three *Mesorhizobium loti* strains. Microbes Environ. 2013;28:275–8.
16. Sullivan JT, Trzebiatowski JR, Cruickshank RW, Gouzy J, Brown SD, Elliot RM, et al. Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. J Bacteriol. 2002;184:3086–95.
17. Perry BJ, Sullivan JT, Colombi E, Murphy RJT, Ramsay JP, Ronson CW. Symbiosis islands of Loteae-nodulating *Mesorhizobium* comprise three radiating lineages with concordant nod gene complements and nodulation host-range groupings. Microb. Genom. 2020;6:mgen000426.
18. Greenlon A, Chang PL, Damtew ZM, Muleta A, Carrasquilla-Garcia N, Kim D, et al. Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. Proc Natl Acad Sci USA. 2019;116:15200–9.
19. Consuegra J, Gaffé J, Lenski RE, Hindré T, Barrick JE, Tenaillon O, et al. Insertion-sequence-mediated mutations both promote and constrain evolvability during a long-term experiment with bacteria. Nat Commun. 2021;12:980.
20. Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev. 2014;38:865–91.
21. Vandecraen J, Chandler M, Aertsen A, Van Houdt R. The impact of insertion sequences on bacterial genome plasticity and adaptability. Crit Rev Microbiol. 2017;43:709–30.
22. Couchoud C, Bertrand X, Valot B, Hocquet D. Deciphering the role of insertion sequences in the evolution of bacterial epidemic pathogens with panISa software. Microb. Genom. 2020;6:e000356.
23. Siguier P, Gourbeyre E, Varani A, Ton-Hoang B, Chandler M. Everyman's guide to bacterial insertion sequences. Microbiol. Spectr. 2015;3: MDNA3-0030-2014.
24. Caldwell BE. Inheritance of a strain-specific ineffective nodulation in soybeans. Crop Sci. 1966;6:427.
25. Hayashi M, Saeki Y, Haga M, Harada K, Kouchi H, Umehara Y. *Rj (rj)* genes involved in nitrogen-fixing root nodule formation in soybean. Breed Sci. 2012;61:544–53.
26. Shobudani M, Htwe AZ, Yamakawa T, Ishibashi M, Tsurumaru H. Mutants disrupted in the type III secretion system of *Bradyrhizobium elkanii* BLY3-8 overcame nodulation restriction by Rj3-genotype soybean. Microbes Environ. 2020;35:ME19151.
27. Staehelin C, Krishnan HB. Nodulation outer proteins: double-edged swords of symbiotic rhizobia. Biochem. J. 2015;470:263–74.
28. Miwa H, Okazaki S. How effectors promote beneficial interactions. Curr Opin Plant Biol. 2017;38:148–54.
29. Ratu STN, Teulet A, Miwa H, Masuda S, Nguyen HP, Yasuda M, et al. Rhizobia use a pathogenic-like effector to hijack leguminous nodulation signalling. Sci Rep. 2021;11:2034.
30. López-Baena FJ, Monreal JA, Pérez-Montaño F, Guasch-Vidal B, Bellogín RA, Vinardell JM, et al. The absence of Nops secretion in *Sinorhizobium fredii* HH103 increases GmPR1 expression in Williams soybean. Mol Plant Microbe Interact. 2009;22:1445–54.
31. Sugawara M, Takahashi S, Umehara Y, Iwano H, Tsurumaru H, Odake H, et al. Variation in bradyrhizobial NopP effector determines symbiotic incompatibility with *Rj2*-soybeans via effector-triggered immunity. Nat Commun. 2018;9:3139.
32. Tsukui T, Eda S, Kaneko T, Sato S, Okazaki S, Kakizaki-Chiba K, et al. The type III secretion system of *Bradyrhizobium japonicum* USDA122 mediates symbiotic incompatibility with *Rj2* soybean. Appl Enviro Microbiol. 2013;79:1048–51.
33. Sugawara M, Umehara Y, Kaga A, Hayashi M, Ishimoto M, Sato S, et al. Symbiotic incompatibility between soybean and *Bradyrhizobium* arises from one amino acid determinant in soybean *Rj2* protein. PLoS ONE. 2019;14: e0222469.
34. Zhao R, Liu LX, Zhang YZ, Jiao J, Cui WJ, Zhang B, et al. Adaptive evolution of rhizobial symbiotic compatibility mediated by co-evolved insertion sequences. ISME J. 2018;12:101–11.
35. Benezech C, Doudement M, Gourion B. Legumes tolerance to rhizobia is not always observed and not always deserved. Cell Microbiol. 2020;22:e13124.

36. Sugawara M, Tsukui T, Kaneko T, Ohtsubo Y, Sato S, Nagata Y, et al. Complete genome sequence of *Bradyrhizobium diazoefficiens* USDA 122, a nitrogen-fixing soybean symbiont. Genome Announc. 2017;5:e01743–16.

37. Iida T, Itakura M, Anda M, Sugawara M, Isawa T, Okubo T, et al. Symbiosis island shuffling with abundant insertion sequences in the genomes of extra-slow-growing strains of soybean bradyrhizobia. Appl Environ Microbiol. 2015;81: 4143–54.

38. Kullik I, Fritsche S, Knobel H, Sanjuan J, Hennecke H, Fischer HM. *Bradyrhizobium japonicum* has two differentially regulated, functional homologs of the σ54 Gene (rpoN). J Bacteriol. 1991;173:1125–38.

39. Kanehara K, Minamisawa K. Complete genome sequence of *Bradyrhizobium japonicum* J5, isolated from a soybean nodule in Hokkaido, Japan. Genome Announc. 2017;5:e01619–16.

40. Hynes MF, Quandt J, O'Connell MP, Pühler A. Direct selection for curing and deletion of *Rhizobium* plasmids using transposons carrying the *Bacillus subtilis* sacB gene. Gene 1989;78:111–20.

41. Gay P, Le Coq D, Steinmetz M, Ferrari E, Hoch JA. Cloning structural gene *sacB*, which codes for exoenzyme levansucrase of *Bacillus subtilis:* expression of the gene in *Escherichia coli.* J Bacteriol. 1983;153:1424–31.

42. Hollingshead S, Vapnek D. Nucleotide sequence analysis of a gene encoding a streptomycin/spectinomycin adenylyltransferase. Plasmid. 1985;13:17–30.

43. Göttfert M, Grob P, Hennecke H. Proposed regulatory pathway encoded by the *nodV* and *nodW* genes, determinants of host specificity in *Bradyrhizobium japonicum*. Proc Natl Acad Sci USA. 1990;87:2680–4.

44. Hara S, Morikawa T, Wasai S, Kasahara Y, Koshiba T, Yamazaki K. et al. Identification of nitrogen-fixing *Bradyrhizobium* associated with roots of field-grown sorghum by metagenome and proteome analyses. Front Microbiol. 2019;10:407.

45. Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. Proc Natl Acad Sci USA. 1995;92:8985–9.

46. Haskett TL, Terpolilli JJ, Bekuma A, O'Hara GW, Sullivan JT, Wang P, et al. Assembly and transfer of tripartite integrative and conjugative genetic elements. Proc Natl Acad Sci USA. 2016;113:12268–73.

47. Parker MA. Legumes select symbiosis island sequence variants in *Bradyrhizobium*. Mol Ecol. 2012;21:1769–78.

48. Guo X, Flores M, Mavingui P, Fuentes SI, Hernández G, Dávila G, et al. Natural genomic design in *Sinorhizobium meliloti:* novel genomic architectures. Genome Res. 2003;13:1810–7.

49. Itakura M, Saeki K, Omori H, Yokoyama T, Kaneko T, Tabata S, et al. Genomic comparison of *Bradyrhizobium japonicum* strains with different symbiotic nitrogen-fixing capabilities and other Bradyrhizobiaceae members. ISME J. 2009;3:326–39.

50. Yuan K, Reckling M, Ramirez MDA, Djedidi S, Fukuhara I, Ohyama T, et al. Characterization of rhizobia for the improvement of soybean cultivation at cold conditions in central Europe. Microbes Environ. 2020;35:1–13.

51. Habibi S, Ayubi AG, Ohkama-Ohtsu N, Sekimoto H, Yokoyama T. Genetic characterization of soybean rhizobia isolated from different ecological zones in northeastern Afghanistan. Microbes Environ. 2017;32:71–9.

52. Sachs JL, Ehinger MO, Simms EL. Origins of cheating and loss of symbiosis in wild *Bradyrhizobium*. J Evol Biol. 2010;23:1075–89.

53. Sachs JL, Skophammer RG, Regus JU. Evolutionary transitions in bacterial symbiosis. Proc Natl Acad Sci USA. 2011;108:10800–7.

54. Sachs JL, Russell JE, Hollowell AC. Evolutionary instability of symbiotic function in *Bradyrhizobium japonicum*. PLoS ONE. 2011;6:e26370.

55. Kiers ET, Rousseau RA, West SA, Denison RF. Host sanctions and the legume-rhizobium mutualism. Nature 2003;425:78–81.

56. Oono R, Anderson CG, Denison RF. Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. Proc Biol Sci. 2011;278:2698–703.

57. Marco DE, Carbajal JP, Cannas S, Pérez-Arnedo R, Hidalgo-Perea Á, Olivares J, et al. An experimental and modelling exploration of the host-sanction hypothesis in legume–rhizobia mutualism. J Theor Biol. 2009;259:423–33.

58. Reams AB, Roth JR. Mechanisms of gene duplication and amplification. Cold Spring Harb Perspect Biol. 2015;7:a016592.

59. Anderson RP, Roth JR. Tandem genetic duplications in phage and bacteria. Annu Rev Microbiol. 1977;31:473–505.

60. Bobay LM, Ochman H. The evolution of bacterial genome architecture. Front Genet. 2017;8:72.

61. Gourion B, Berrabah F, Ratet P, Stacey G. Rhizobium-legume symbioses: the crucial role of plant immunity. Trends Plant Sci. 2015;20:186–94.

62. Cao Y, Halane MK, Gassmann W, Stacey G. The role of plant innate immunity in the legume-rhizobium symbiosis. Annu Rev Plant Biol. 2017;68:535–61.

63. Nelson MS, Sadowsky MJ. Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. Front Plant Sci. 2015;6:491.

64. Yang S, Tang F, Gao M, Krishnan HB, Zhu H. R gene-controlled host specificity in the legume-rhizobia symbiosis. Proc Natl Acad Sci USA. 2010;107:18735–40.

65. Shine MB, Gao QM, Chowda-Reddy RV, Singh AK, Kachroo P, Kachroo A. Glycerol-3-phosphate mediates rhizobia-induced systemic signaling in soybean. Nat Commun. 2019;10:5303.

66. Zhang B, Wang M, Sun Y, Zhao P, Liu C, Qing K, et al. Glycine max NNL1 restricts symbiotic compatibility with widely distributed bradyrhizobia via root hair infection. Nat Plants. 2021;7:73–86.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41396-021-01035-4.

Correspondence and requests for materials should be addressed to K.M.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.