Assembly and transfer of tripartite integrative and conjugative genetic elements

Timothy L. Haskett^a, Jason J. Terpolilli^a, Amanuel Bekuma^a, Graham W. O'Hara^a, John T. Sullivan^b, Penghao Wang^a, Clive W. Ronson^b, and Joshua P. Ramsay^{c,1}

^aCentre for Rhizobium Studies, Murdoch University, Perth, WA 6150, Australia; ^bDepartment of Microbiology and Immunology, University of Otago, Dunedin 9016, New Zealand; and ^cSchool of Biomedical Sciences and Curtin Health Innovation Research Institute, Curtin University, Perth, WA 6102, Australia

Edited by Frederick M. Ausubel, Harvard Medical School, Massachusetts General Hospital, Boston, MA, and approved September 9, 2016 (received for review August 10, 2016)

Integrative and conjugative elements (ICEs) are ubiquitous mobile genetic elements present as “genomic islands” within bacterial chromosomes. Symbiosis islands are ICEs that convert nonsymbiotic mesorhizobia into symbionts of legumes. Here we report the discovery of symbiosis ICEs that exist as three separate chromosomal regions when integrated in their hosts, but through recombination assemble as a single circular ICE for conjugative transfer. Whole-genome comparisons revealed exconjugants derived from nonsymbiotic mesorhizobia received three separate chromosomal regions from the donor *Mesorhizobium ciceri* WSM1271. The three regions were each bordered by two nonhomologous integrase attachment (*att*) sites, which together comprised three homologous pairs of *attL* and *attR* sites. Sequential recombination between each *attL* and *attR* pair produced corresponding *attP* and *attB* sites and joined the three fragments to produce a single circular ICE, ICEM_{Sym}¹²⁷¹. A plasmid carrying the three *attP* sites was used to recreate the process of tripartite ICE integration and to confirm the role of integrase genes *intS*, *intM*, and *intG* in this process. Nine additional tripartite ICEs were identified in diverse mesorhizobia and transfer was demonstrated for three of them. The transfer of tripartite ICEs to nonsymbiotic mesorhizobia explains the evolution of competitive but suboptimal N₂-fixing strains found in Western Australian soils. The unheralded existence of tripartite ICEs raises the possibility that multipartite elements reside in other organisms, but have been overlooked because of their unusual biology. These discoveries reveal mechanisms by which integrases dramatically manipulate bacterial genomes to allow cotransfer of disparate chromosomal regions.

integrative and conjugative elements | integrase | recombination | conjugation | symbiosis

Horizontal gene transfer plays an instrumental role in prokaryotic evolution because it facilitates the rapid acquisition of complex phenotypic traits required for pathogenicity, symbiosis, metabolism, fitness, and antimicrobial resistance (1–8). Integrative and conjugative elements (ICEs) are an abundant class of conjugative elements in bacteria, but they are also the most recently recognized and least characterized (8, 9). An ICE integrates site-specifically within the chromosome of its host and is flanked by a direct repeat sequence that demarcates the insertion site. Before transfer, site-specific recombination between the flanking sequences results in excision and circularization of the ICE and restoration of the host chromosome. A single-stranded DNA copy of the circularized ICE is then formed via rolling-circle replication and transferred to recipient cells via an ICE-encoded conjugative type IV secretion system. Following delivery of the ICE to the recipient cell, the second DNA strand of the ICE is synthesized and the circularized ICE integrates site-specifically into the recipient genome (10).

Symbiosis islands are the largest documented ICEs and their transfer converts nonsymbiotic mesorhizobia into nitrogen (N₂)-fixing symbionts of leguminous plants. The symbiosis island of *Mesorhizobium loti* strain R7A was discovered following the use of R7A as an agricultural inoculant for *Lotus corniculatus* at a

New Zealand field site devoid of compatible native symbiotic mesorhizobia (11, 12). After a 7-y period, only ~20% of *L. corniculatus* root nodules contained R7A; the other 80% of nodules contained a diverse mix of native mesorhizobia that had acquired symbiosis genes from R7A (11). Molecular analyses revealed these new symbionts had acquired a 502-kb symbiosis ICE, subsequently named ICEM_{Sym}^{R7A} (12–14).

The mechanism and regulation of ICEM_{Sym}^{R7A} integration and excision have been investigated in detail (12, 15–18). Integration of ICEM_{Sym}^{R7A} is catalyzed by the integrase IntS, which recombines the attachment (*att*) site *attP* located on the circularized ICEM_{Sym}^{R7A} with the *attB* site located at the 3' end of the sole *phe*-tRNA gene present in *Mesorhizobium* genomes. This recombination integrates ICEM_{Sym}^{R7A} and produces flanking attachment sites *attL* and *attR*. Excision of ICEM_{Sym}^{R7A} is stimulated by a recombination directionality factor RdfS, which reverses the favored direction of IntS-mediated recombination toward formation of *attP* and *attB* (12, 17).

Horizontal transfer of a second symbiosis island, here named ICEM_{Sym}¹²⁷¹, was discovered during field trials of *Biserrula pelecinus* inoculated with *Mesorhizobium ciceri* bv. *biserrulae* WSM1271 at a Western Australia field site (19, 20). *B. pelecinus* nodules sampled 6 y after inoculation contained genetically distinct strains, including the novel species *Mesorhizobium australicum*

Significance

Integrative and conjugative elements (ICEs) are one of the most prevalent but least-characterized families of mobile genetic elements in bacteria. We identified a family of ICEs that exists as three separate parts integrated within the single chromosomes of symbiotic mesorhizobia. These “tripartite ICEs,” through a series of chromosomal recombinations mediated by integrase proteins, assemble into a single circular ICE. Following transfer to nonsymbiotic mesorhizobia, tripartite ICEs integrate and disassemble into three parts in the recipient genome and exconjugant mesorhizobia gain the ability to form a symbiosis with legumes. These discoveries expand our appreciation of the potential for gene transfer in bacteria and demonstrate how mobile genetic elements can dramatically manipulate the bacterial genome.

Author contributions: T.L.H., J.J.T., A.B., G.W.O., J.T.S., and J.P.R. designed research; T.L.H., A.B., and J.T.S. performed research; P.W. contributed new reagents/analytic tools; T.L.H., J.J.T., J.T.S., P.W., C.W.R., and J.P.R. analyzed data; and T.L.H., J.J.T., J.T.S., C.W.R., and J.P.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The strains and sequences reported in this paper have been deposited in the NCBI (accession nos. CP016079, CP015064.1, LYTO00000000, LYTN00000000, LYTK00000000, LYTL00000000, AZUV00000000.1, JAFG00000000.1, CAUM00000000.1, LZTK00000000, LZTH00000000, LZTJ00000000, LZTL00000000).

¹To whom correspondence should be addressed. Email: josh.ramsay@curtin.edu.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613358113/-DCSupplemental.

WSM2073 and *Mesorhizobium opportunistum* WSM2075. These strains and WSM1271 carried identical copies of an ICE M Sym^{R7A} *intS* homolog adjacent to the *phe*-tRNA gene in each strain (21). By analogy with ICE M Sym^{R7A}, it was concluded that WSM1271 also harbored a mobile symbiosis ICE. Interestingly, glasshouse experiments revealed symbiotic proficiency varied markedly for the ICE M Sym¹²⁷¹ exconjugants (20, 21). Compared with WSM1271 on *B. pelecinus*, WSM2073 was only partially effective in N₂ fixation, whereas WSM2075 did not fix N₂.

In this work, we compared genome sequences of WSM1271, WSM2073, and WSM2075 (22, 23), and confirmed the presence of a near-identical ~476-kb complement of WSM1271-derived DNA in both of the field-isolated exconjugants. However, the comparison also revealed that three distinct regions of identical DNA—each separated by chromosomal DNA—were present in all three strains, suggesting they had transferred together from WSM1271. Here we show that the three regions indeed cotransfer in laboratory conjugation experiments and that they accomplish this by assembling into a single circular “tripartite” ICE.

Results

Three Separate Cotransferable Regions in WSM1271. BLASTN analysis of the WSM1271 genome with the genomes of WSM2073 and WSM2075 revealed three regions of near-perfect nucleotide identity (in addition to near-identical rRNA genes) (Fig. S1). The first region, named α , was 445,220 bp in WSM1271 and was identical in WSM2073 and WSM2075, aside from point mutations within a single putative transposase gene *Mesci_5575*. The second largest region, named β , was 22,971 bp, and smallest region, named γ , was 7,760 bp. The β and γ regions were identical in each strain. Genome assemblies were confirmed as correct by PCR-amplification and sequencing of all six junctions between the core chromosome and acquired regions.

To determine if regions α , β , and γ were transferred together, conjugation experiments using strains WSM1271, WSM2073, and WSM2075 as donors and the nonsymbiotic *M. loti* strain R7ANS (17) as the recipient were conducted. Transfer occurred at a frequency of $4.65 \times 10^{-8} \pm 7.89 \times 10^{-9}$ (SE) from WSM1271, $8.5 \times 10^{-9} \pm 8.5 \times 10^{-10}$ from WSM2073, and $3.0 \times 10^{-9} \pm 6.0 \times 10^{-10}$ from WSM2075 donors. In comparison, the rate of transfer

Fig. 1. Recombination of ICE M Sym¹²⁷¹. The α , β , and γ regions are colored dark blue, brown, and yellow, respectively. IntG and associated att sites are colored magenta, IntM and associated att sites are colored green, IntS and associated att sites are colored cyan. Schematics are not to scale. (A) Schematic of ICE M Sym¹²⁷¹ regions α , β , and γ , and predicted att site core sequences. (B) Recombination states *i*–*viii* of the WSM1271 chromosome obtained through the actions of IntS, IntG, and IntM. Chromosomal DNA is in gray and is fixed at the WSM1271 origin of replication, indicated by an “O.” The orientation of each att site is indicated by an arrow (5′–3′ direction) at each region boundary. (C) Relative abundance of each attP and attB site present as a percentage of WSM1271 chromosomes in liquid cultures grown over 50 h, presented as average of three qPCR experimental replicates. Samples taken at 60, 70, and 80 h produced similar averages to 50-h samples. (D) DNA gels of PCR products amplified from the 12 att sites (rows) in each of the eight possible recombination states (columns) from *M. loti* R7ANS carrying the pMINI3 plasmid depicted in B. Each dashed-line arrow represents a pathway successfully demonstrated using pMINI3. *Recombination states *iii* and *v* were not obtained. An example PCR profile from a single isolate is shown for each of these recombination states; recombination state *iii* showed the presence of an unexpected PCR product for attB _{α} (†), whereas recombination state *v* lacked an expected attR _{α} PCR product (‡).

of ICE $McSym^{R7A}$ from *M. loti* R7A is $\sim 1 \times 10^{-7}$, but increases to $\sim 1 \times 10^{-4}$ in derepressed strains (15–17). Exconjugants from 16 independent experiments were screened by PCR for genes located on regions α , β , γ , and a region specific for the R7ANS chromosome (Table S1). All amplicons were detected in all exconjugants. The genome of an exconjugant (R7ANS $\times$ WSM1271) from mating of WSM1271 with R7ANS was draft sequenced. BLASTN comparison of the de novo assembled R7ANS $\times$ WSM1271 genome with WSM1271 confirmed complete transfer of all three regions and an integration pattern resembling the tripartite configuration of WSM1271 (Fig. S1).

Three Integrases and Three Pairs of Integrase Attachment Sites. Analysis of gene content on the three WSM1271 regions (Table S2) revealed that *intS* was located on region γ , downstream of the *phe*-tRNA gene, marking one boundary of this region (Fig. 1A). Region γ also carried a second integrase gene, *intM*, located adjacent to the *met*-tRNA gene, marking the other region γ boundary. Region β was adjacent to the GMP-synthase gene *guaA* and harbored a third integrase gene, *intG*, the product of which resembled integrases associated with mobile elements that integrate into *guaA* (24). Region α did not contain an integrase gene.

During integrase-mediated ICE integration, recombination between the core regions of the ICE *attP* and the chromosomal *attB* produce *attL* and *attR* sites, which flank the integrated element. *attL* sites contain chromosomal DNA 5' of the core and ICE DNA 3' of the core, whereas *attR* sites contain ICE DNA 5' of the core and chromosomal DNA 3' of the core. Therefore, all four *att* site types are structurally distinct and can be distinguished by the origin of flanking DNA and the relative orientation of the core sequence. Integrase genes are often encoded adjacent to the *attL* site (25). Because the *phe*-tRNA, *guaA*, and *met*-tRNA genes were located adjacent to *intS*, *intG*, and *intM*, respectively, these were likely *attL* sites; however, it was not clear where each corresponding *attR* site was located. The 17-bp core sequence associated with the ICE $McSym^{R7A}$ integrase *IntS* is 5'-TCCGCCTCTGGGCACCA-3'. The same sequence was identified at the 3' end of the γ -region boundary within the WSM1271 *phe*-tRNA gene, which we named *attL_S* (Fig. 1A). We located another copy of the *IntS* core sequence at the 3' boundary of the α region, which we named *attR_S*. The conserved core sequence targeted by *guaA*-associated integrases is 5'-GAGTGGGA-3' (24). We identified two sites containing this sequence, *attL_G* within the *guaA* gene at the 5' end of the β -fragment, and *attR_G* at the 3' end of the α -fragment. Finally, we identified a perfect duplication of a 16-bp sequence 5'-CCCTCCGGGCCACCA-3', *attL_M*, at the 5' end of region γ within the end of the *met*-tRNA gene, and *attR_M* at the 3' of region β . In summary, regions α , β , and γ were each bordered by putative core sites associated with two different integrases (Fig. 1A and B, i).

Excision and circularization of DNA located between *attL* and *attR* requires that their core regions form a directly orientated repeat. However, *attR_S* on region α was inverted relative to *attL_S*, indicating recombination of *attL_S* and *attR_S* to produce *attP_S* and *attB_S* would result in an inversion and juxtaposition of regions α and γ (Fig. 1B, ii). Similarly, *IntG*-mediated recombination of convergently oriented *attL_G* and *attR_G* would produce *attP_G* and *attB_G* and juxtapose fragments α and β (Fig. 1B, iv). Finally, *attL_M* and *attR_M* were directly oriented, so their recombination would excise DNA between them, leaving behind *attB_M* and juxtaposing regions β and γ on a circular 248-kb DNA fragment carrying *attP_M* and 218 kb of chromosomal DNA (Fig. 1B, iii).

Coordinated Formation of Each Pair of *attP* and *attB* Sites. Quantitative PCR (qPCR) was previously used to calculate the excision frequency of ICE $McSym^{R7A}$ by measuring *attP* and *attB* abundance relative to the chromosomal gene *melR* (17). We adapted this assay to detect and measure *attP* and *attB* formation in WSM1271. Six pairs of primers for each of the 3 *attP* and *attB* sites and primers for WSM1271 *melR* were designed. Tryptone yeast (TY) broths seeded with a 1/250 dilution of stationary-phase

WSM1271 culture were harvested at 10-h intervals for 80 h (Fig. 1C). All *attP* and *attB* products were detected in DNA extracted from all samples and sequencing of qPCR products confirmed recombination had occurred within each predicted core site. The relative abundance of each *attP* and *attB* pair was comparable for *att* sites of the same type, consistent with the interdependent production of *attP* and *attB* sites from corresponding *attL* and *attR* sites. The *IntS* and *IntG* *attP* and *attB* sites were detected in $\sim 0.01\%$ of cells in log-phase growth (20 h) and this increased to 0.1% of cells in stationary-phase cultures (50 h onwards). *attP_M* and *attB_M* were ~ 10 -fold less abundant than *IntS* and *IntG* *attP* and *attB* sites in both phases of growth, although the abundance of these products also increased ~ 10 -fold in stationary-phase cells.

Reconstruction of ICE $McSym^{1271}$ Integration and Disassembly Pathways.

We hypothesized that the concerted action of *IntS*, *IntG*, and *IntM* would lead to recombination of α , β , and γ regions into a single circular ICE in the donor before conjugative transfer. To define the potential recombination pathways, a network diagram was created using the position and orientation of each *att* site and the predicted products of each recombination (Fig. 1B). In this network, eight possible recombination states were predicted to be possible (i to viii in Fig. 1B), with each state able to transition to three other states through the action of one of the three integrases. This model indicated that sequential action of each integrase in any order to form three pairs of *attB* and *attP* sites would result in excision of a single circular ICE $McSym^{1271}$ element and restoration of the ancestral WSM1271 chromosome. The model also suggested that the three reverse reactions (forming *attL* and *attR* sites) in combination would integrate ICE $McSym^{1271}$ and separate it into the tripartite configuration observed in all ICE $McSym^{1271}$ exconjugants.

To test this model (Fig. 1B), we constructed a nonreplicative mini-ICE $McSym^{1271}$ plasmid, pMINI3, that contained each of the three *attP* sites arranged in the same order and orientation as on the circular ICE $McSym^{1271}$ predicted in state viii (Fig. 1B). pMINI3 conferred gentamicin resistance but could not replicate in *Mesorhizobium*, so it needed to recombine with the chromosome to be maintained. We also constructed three expression plasmids (pSacIntS, pSacIntG, and pSacIntM) containing *intS*, *intG*, and *intM*, respectively, cloned downstream of the *lac* promoter on plasmid pSac, which carries a copy of the *Bacillus subtilis* *sacB* gene (26), enabling selection for loss of each pSac plasmid by growing on medium containing sucrose.

Sequence analysis of R7ANS revealed it carried *attB_S*, *attB_G*, and *attB_M* in the same relative position and orientation as predicted for the WSM1271 chromosome when cured of ICE $McSym^{1271}$ (state viii in Fig. 1B) and lacked genes for *intS*, *intG*, and *intM*. Each pSac plasmid was separately introduced into R7ANS. pMINI3 was then conjugated into each of the three strains and colonies harboring integrated pMINI3 were selected on medium containing gentamicin. Integration of pMINI3 was observed in each strain carrying an integrase-expressing pSac plasmid, but not in a strain carrying an empty pSac vector, confirming dependence of pMINI3 integration on the presence of an integrase gene. Using PCR screens (Fig. 1D), we confirmed the isolation of *attB_S::pMINI3* and *attB_M::pMINI3* insertions in R7ANS(pSacIntS) and R7ANS(pSacIntM), which represented recombination states vii and vi, respectively (Fig. 1B). However, although we were able to isolate pMINI3 integrants in R7ANS(pSacIntG), their PCR profiles did not match those predicted for state v, as individual colonies lacked either *attL_G* or *attR_G* (Fig. 1D).

We further manipulated *attB_M::pMINI3* and *attB_S::pMINI3* by curing them of the pSacInt plasmid used to create them and introducing each of the two other pSacInt plasmids. Following isopropyl- β -D-thiogalactopyranoside (IPTG) induction, randomly selected single colonies were isolated, cured of the pSacInt plasmid, and screened for recombination state. From state vi we derived states iv and ii and from state vii we derived state iv, but we were unable to derive state iii from state vii. For prospective strains in state vii, an unexpected *attB_G* PCR product was detected (Fig. 1D). Finally to stimulate the formation of

state *i*, each previously unintroduced pSacInt plasmid was introduced into strains in states *ii* and *iv*. Following IPTG induction and plasmid curing, PCR screens confirmed the conversion of strains in states *ii* and *iv* to state *i*. Sequencing of PCR amplicons of all *attL* and *attR* junctions amplified from the two independently derived state *i* strains confirmed the predicted pMINI3-chromosome recombination junctions. In summary, we isolated strains in six of the eight predicted recombination states (Fig. 1 *B* and *D*).

***M. loti* R7ANS Exconjugants Carrying ICEMcSym¹²⁷¹ Form Partially Effective N₂-Fixing Symbioses with *B. pelecinus*, Irrespective of the Donor Strain.** Strains WSM1271, WSM2073, and WSM2075 all harbor identical symbiosis gene complements. However, WSM1271 is an effective N₂-fixing microsymbiont on *B. pelecinus*, whereas WSM2073 is only partially effective and WSM2075 nodulates but does not fix N₂ (20). To assess the symbiotic properties of R7ANS exconjugants carrying ICEMcSym¹²⁷¹, we inoculated *B. pelecinus* with nine R7ANS exconjugants, one derived from each of three independent matings with each of WSM1271, WSM2073, and WSM2075. Plants were grown for 8 wk before recording shoot dry weights (Fig. 2). All R7ANS exconjugants yielded weights comparable to that of the partially effective strain WSM2073, irrespective of the symbiotic proficiency of the donor strain from which their ICEMcSym¹²⁷¹ originated.

Tripartite Symbiosis Islands Are Present in Diverse Symbiotic Mesorhizobia. To determine whether tripartite ICE elements might be present in other mesorhizobia, we carried out BLASTP searches against sequenced mesorhizobial genomes using IntS, IntG, and IntM sequences as queries. Tripartite ICEs were identified in the *B. pelecinus* symbiont *M. ciceri* bv. *biserrulae* WSM1284 (27), the *Anthyllus vulneraria* symbiont *Mesorhizobium metallidurans* STM2683 (28), the *Bituminaria bituminosa* symbiont *M. ciceri* WSM4083, and the *Lotus* sp. symbionts *M. loti* NZP2037 (29) and *M. loti* WSM1293. We also draft-sequenced and identified putative tripartite ICEs in *B. pelecinus* symbionts isolated from Ethiopia and Greece, *Mesorhizobium* sp. AA22 and *M. ciceri* bv. *biserrulae* WSM1497, and in two additional *Lotus* sp. symbionts, *M. loti* NZP2042 and *M. loti* SU343. We were able to identify all three pairs of *attL* and *attR* core sites corresponding to IntS, IntG and IntM in all but two strains (Dataset S1).

We investigated the putative tripartite ICEs identified in *Lotus* sp. symbionts NZP2037, NZP2042, and SU343. Previous analysis of the 7.5-Mbp scaffold of the sequenced NZP2037 genome indicated the assembly lacked a distinct contig for plasmid pRlo2037 (29–31). We carried out long-read single-molecule real-time (SMRT)-cell sequencing and combined reads with short-read paired-end sequences in a hybrid de novo assembly. Two circular contigs were assembled, corresponding to the NZP2037 chromosome and pRlo2037, respectively. The chromosome contained regions corresponding to α , β , and γ in the same relative position, order, and orientation as located in WSM1271. To determine whether ICEMcSym²⁰³⁷, ICEMcSym²⁰⁴², and ICEMcSym³⁴³ were mobile, mating experiments were carried out using NZP2037, NZP2042, and SU343 as donors and R7ANS as recipient. Exconjugants were isolated from all three matings and confirmed to nodulate the host *Lotus pedunculatus*. The genomes of strains reisolated from nodules were draft-sequenced. Whole-genome BLASTN comparisons of the de novo assembled exconjugant genomes R7ANSxNZP2037, R7ANSxNZP2042, and R7ANSxSU343 with the corresponding donor genome sequences confirmed transfer of regions α , β , and γ from all three donors (Fig. S2).

Discussion

In this study we show that the symbiosis island of WSM1271 is composed of three separate DNA regions α , β , and γ that assemble into a single transferrable element capable of converting nonsymbiotic mesorhizobia into N₂-fixing symbionts. The assembly is the endpoint of the sequential action of three ICEMcSym¹²⁷¹-encoded integrases—IntS, IntG, and IntM—which

Fig. 2. Effectiveness of ICEMcSym¹²⁷¹ exconjugants on *B. pelecinus*. *B. pelecinus* plants grown in nitrogen-limited conditions were inoculated with indicated strains and grown for 8 wk. Uninoculated and nitrogen-fed (supplied as KNO₃) plants were included as negative and positive controls, respectively. Each bar represents the mean shoot dry weight for 20 plants split between five position-randomized pots. Shoot dry weights were compared using one-way ANOVA followed by Tukey's honest significant difference post hoc test at 5% significance. Treatments that share a letter are not significantly different. R7ANS did not nodulate *B. pelecinus*.

each facilitate recombination between specific pairs of *attL*/*attR* sites to form corresponding pairs of *attB*/*attP* sites. Following transfer, ICEMcSym¹²⁷¹ is likely able to integrate into any one of the three *attB* sites in a naïve mesorhizobial chromosome and disassemble into the three regions.

Tyrosine recombinases like the ICEMcSym^{R7A} integrase IntS catalyze the reversible recombination of *attL* + *attR* $\leftrightarrow$ *attP* + *attB*. For a monopartite ICE, the forward reaction excises and circularizes the ICE, whereas the reverse reaction integrates the ICE (17). For ICEMcSym¹²⁷¹, the recombination reaction substrates and products are similar, but the macromolecular rearrangement depends on the relative positions and orientations of three pairs of *att* sites. The action of any single integrase is inadequate for excision of the ICEMcSym¹²⁷¹, but the combined forward actions of the three integrases excise ICEMcSym¹²⁷¹. It therefore follows that the forward reactions are likely to be coregulated. qPCR analysis revealed the abundance of all three pairs of *attP* and *attB* sites increased ~10-fold in stationary-phase cultures. A caveat of our qPCR assay is that it averages the ensemble of recombination states in a population, so further single-cell experiments are necessary to confirm that the three reactions occur concurrently in the same cell. Nevertheless, cotransfer of all three ICEMcSym¹²⁷¹ fragments by conjugation (and concomitant rolling-circle replication) strongly indicates that the three forward reactions must occur together in single cells before transfer to facilitate excision and circularization.

The direction of the recombination catalyzed by an integrase is often determined by a recombination directionality factor, also known as an excisionase (32–34). Excisionases are noncatalytic DNA-binding proteins that promote formation of *attP* and *attB*. ICEMcSym^{R7A} excision is stimulated by RdfS and in its absence, IntS activity favors formation of *attL* and *attR* (17). Expression of IntS, IntG, and IntM stimulated recombination of the pMINI3 *attP* sites with each cognate *attB* site in R7ANS, producing *attL* and *attR*, suggesting equilibrium reactions favor *attL* and *attR* production in the absence of other ICEMcSym¹²⁷¹ genes for all three ICEMcSym¹²⁷¹ integrases. ICEMcSym¹²⁷¹ contains a homolog of *rdiS* and two other putative excisionase genes located adjacent to *intG* and *intM*, *rdiG* and *rdiM* (Table S2) that, like *rdiS*, encode predicted products belonging to the AlpA excisionase family (35). It seems likely that expression of the *rdiS*, *rdiG*, and *rdiM* genes is coregulated to promote excision of ICEMcSym¹²⁷¹.

Using pMINI3 and sequential expression of each integrase, we demonstrated formation of six of the eight predicted recombination states (Fig. 1 *B* and *D*). However, states *iii* and *v* were not

reproduced. When pMINI3 was introduced into R7ANS(pSacBIntG) to produce state *v*, colonies isolated had lost both *attP_G* and *attB_G*, suggesting recombination had occurred as expected. However, individual colonies were positive for either *attL_G* or *attR_G*, but not both (Fig. 1D, † symbol). Further inspection of the R7ANS chromosome revealed the presence of an additional copy of the *attB_G* core sequence within Meslo_RS0109425 (NZ_KI632510). This second *attB_G* (not present in WSM1271, WSM2073, or WSM2075), together with strong overexpression of IntG from pSacBIntG, may have led to additional IntG-mediated recombination events, destroying one of the *attL_G* or *attR_G* sites in each isolate. Interestingly, these secondary recombination events were not apparent in the genomes of the sequenced exconjugants WSM2073, WSM2075, and R7ANS×WSM1271, so this phenomenon could be limited to our artificial system.

Recombination state *iii* is the only state that splits the chromosome. The smaller portion (248,280 bp) harbors regions β and γ along with the *guaA* and *phe*-tRNA genes, but appears to lack an origin of replication. Presumably, state *iii* is not viable, because postsegregational loss of the excised region would result in loss of the sole *phe*-tRNA gene. In our attempt to recombine pMINI3 from state *ii* to state *vi* using pSacBIntG, secondary recombination events mediated by IntG may have reintegrated this fragment into the main chromosome, resulting in the rescue of these recombined cells and the unexpected PCR profile in Fig. 1D, ‡ symbol. Eckhardt gel DNA electrophoresis did not identify an episomal fragment in the 250-kbp size range (Fig. S3). Interestingly, the IntM-mediated excision products *attP_M* and *attB_M* were the lowest-abundance products detected by our qPCR assay. This finding implies that *attL_M* + *attR_M* → *attP_M* + *attB_M* may be the last or lowest-rate reaction, or that nonviable cells in state *iii* are lost from cell populations. If *attP_M* + *attB_M* formation is the final step in excision of the assembled ICEMcSym¹²⁷¹ (transition *vi* → *viii* in Fig. 1B), then recombination state *iii* would be avoided during the recombination pathway that produces circularized ICEMcSym¹²⁷¹.

The tripartite configuration of ICEMcSym¹²⁷¹ seems unnecessarily complex compared with monopartite ICEs, but it is present in diverse and widely distributed mesorhizobia. How could tripartite ICEs have evolved and what has ensured their success? Despite the complexity of the *att*-site arrangement on ICEMcSym¹²⁷¹, this configuration could have evolved in only a few steps (Fig. 3) if we assume that ICEMcSym¹²⁷¹ evolved from three independent elements integrated within the *attB_S*, *attB_G*, and *attB_M* *att* sites. If a genomic inversion between an IntS-associated element and an IntM-associated element was followed by an inversion between a resulting hybrid element and an IntG-associated element, then a tripartite ICE like ICEMcSym¹²⁷¹ would be formed (Fig. 3). Such inversions could have easily been mediated by one of the numerous transposable elements found on mesorhizobial ICEs (14), either as part of the transposition process or through RecA-mediated recombination between repetitive elements. Why the tripartite arrangement has persisted in nature is a difficult question to answer. It could be that the combined gene complement of the three archetypal integrative elements (Fig. 3A) was more beneficial than that of any single element and that the tripartite configuration ensured their cotransfer. Our preliminary bioinformatic analyses indicate that genes on region β encode enzymes (Mesci_2556, Mesci_2561, Mesci_2562) resembling those involved in melanin biosynthesis (36–38). However, the β region of ICEMcSym²⁰³⁷, although similarly sized to that of ICEMcSym¹²⁷¹, lacks this cluster, suggesting the specific gene complement on each region is not a conserved feature of tripartite ICEs. Alternatively, because no single integrase-mediated recombination event leads to the loss of any single fragment, the tripartite configuration could be a selfish mechanism to stabilize the ICE in the genome.

Fig. 3. Model of tripartite ICE evolution. The arrangement of *att* sites on tripartite ICEs may have evolved through two chromosomal inversions between three separate elements flanked by distinct *attL* and *attR* sites. Colors correspond with those in Fig. 1A. The dashed lines segmenting the chromosome indicate where the inversions may have occurred. (A) The ancestral chromosome configuration. (B) Configuration of the chromosome following the first inversion. (C) The final tripartite ICE structure following the second inversion.

Transfer of ICEMcSym¹²⁷¹ to *M. loti* R7ANS conferred on recipients the ability to nodulate *B. pelecinus*. All R7ANS exconjugants harboring ICEMcSym¹²⁷¹ formed a partially effective symbiosis, even when ICEMcSym¹²⁷¹ was transferred from a strain unable to fix N₂ (WSM2075). This finding indicates that although ICEMcSym¹²⁷¹ in WSM2075 is functional, other factors in WSM2075 are incompatible with forming an effective symbiosis with *B. pelecinus*. The evolution of poor N₂-fixing rhizobia is a significant problem associated with legume inoculation in agriculture, because ineffective strains can dominate soil populations and reduce crop productivity (11, 19, 20, 39). These experiments provide insight into how ineffective rhizobia can evolve through transfer of ICEs.

Other integrative elements have been found to harbor multiple sets of *att* sites capable of site-specific inversion (40). ICEMcSym¹²⁷¹, however, obligatorily requires chromosomal inversions to facilitate excision and transfer. In this study we identified putative tripartite ICEs in genetically and geographically diverse mesorhizobia, indicating these elements are common in this genus. Bioinformatic analyses have revealed that chromosomally integrated elements dominate the mobilomes of many prokaryotes, although the vast majority of elements remain uncharacterized (8, 9). The obscure nature of tripartite ICEs makes their detection nontrivial, so it is plausible that multipartite elements have been overlooked in other organisms. If this is the case, it may be that many of the presumed immobilized genetic elements identified in diverse organisms could actually be mobile.

Methods

Strains and Growth Conditions. Strains and accession numbers are listed in Table S3. Strains were cultured as previously described (16, 17, 41, 42). *Mesorhizobium* conjugation was carried out using a method modified from that previously described (17), which uses the auxotrophy of R7ANS to select for vitamin synthesis genes carried by symbiosis islands.

Molecular Biology. Details of plasmids and primers are listed in Tables S1 and S3.

DNA Sequencing. All de novo genome assemblies were carried out using SPAdes (43). Procedures are outlined in SI Methods.

Glasshouse Procedures. *Biserrula* were grown as per established methods (44) and outlined in SI Methods.

ACKNOWLEDGMENTS. The authors thank Matthew McBirney and Shea Devlin for helpful discussions on the tripartite recombination reactions; Professor Jens Stougaard and the Centre for Carbohydrate Recognition and Signalling (funded by the Danish National Research Foundation); and Curtin Health Innovation Research Institute for facilities and technical support. This work was supported in part by the Grains Research and Development Corporation of Australia (T.L.H., J.J.T., and G.W.O.) and Australia Awards Africa (A.B.).

- Finan TM (2002) Evolving insights: Symbiosis islands and horizontal gene transfer. *J Bacteriol* 184(11):2855–2856.
- Schmidt H, Hensel M (2004) Pathogenicity islands in bacterial pathogenesis. *Clin Microbiol Rev* 17(1):14–56.

- Juhas M, et al. (2009) Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 33(2):376–393.
- Dobrindt U, Hochhut B, Hentschel U, Hacker J (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2(5):414–424.

5. Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405(6784):299–304.
6. Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu Rev Microbiol* 55:709–742.
7. Hacker J, Carniel E (2001) Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep* 2(5):376–381.
8. Guglielmini J, Quintais L, Garcillán-Barcia MP, de la Cruz F, Rocha EPC (2011) The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet* 7(8):e1002222.
9. Bi D, et al. (2012) ICEberg: A web-based resource for integrative and conjugative elements found in bacteria. *Nucleic Acids Res* 40(Database issue):D621–D626.
10. Wozniak RA, Waldor MK (2010) Integrative and conjugative elements: Mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* 8(8):552–563.
11. Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW (1995) Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc Natl Acad Sci USA* 92(19):8985–8989.
12. Sullivan JT, Ronson CW (1998) Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a *phe*-tRNA gene. *Proc Natl Acad Sci USA* 95(9):5145–5149.
13. Sullivan JT, Brown SD, Yocum RR, Ronson CW (2001) The bio operon on the acquired symbiosis island of *Mesorhizobium* sp. strain R7A includes a novel gene involved in pimeloyl-CoA synthesis. *Microbiology* 147(Pt 5):1315–1322.
14. Sullivan JT, et al. (2002) Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. *J Bacteriol* 184(11):3086–3095.
15. Ramsay JP, et al. (2013) A widely conserved molecular switch controls quorum sensing and symbiosis island transfer in *Mesorhizobium loti* through expression of a novel antiactivator. *Mol Microbiol* 87(1):1–13.
16. Ramsay JP, et al. (2009) A LuxRI-family regulatory system controls excision and transfer of the *Mesorhizobium loti* strain R7A symbiosis island by activating expression of two conserved hypothetical genes. *Mol Microbiol* 73(6):1141–1155.
17. Ramsay JP, Sullivan JT, Stuart GS, Lamont IL, Ronson CW (2006) Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. *Mol Microbiol* 62(3):723–734.
18. Ramsay JP, et al. (2015) Ribosomal frameshifting and dual-target antiactivation restrict quorum-sensing-activated transfer of a mobile genetic element. *Proc Natl Acad Sci USA* 112(13):4104–4109.
19. Nandasena KG, O'Hara GW, Tiwari RP, Howieson JG (2006) Rapid in situ evolution of nodulating strains for *Biserrula pelecinus* L. through lateral transfer of a symbiosis island from the original mesorhizobial inoculant. *Appl Environ Microbiol* 72(11):7365–7367.
20. Nandasena KG, O'Hara GW, Tiwari RP, Sezmiş E, Howieson JG (2007) In situ lateral transfer of symbiosis islands results in rapid evolution of diverse competitive strains of mesorhizobia suboptimal in symbiotic nitrogen fixation on the pasture legume *Biserrula pelecinus* L. *Environ Microbiol* 9(10):2496–2511.
21. Nandasena KG, O'Hara GW, Tiwari RP, Willems A, Howieson JG (2009) *Mesorhizobium australicum* sp. nov. and *Mesorhizobium opportunistum* sp. nov., isolated from *Biserrula pelecinus* L. in Australia. *Int J Syst Evol Microbiol* 59(Pt 9):2140–2147.
22. Nandasena K, et al. (2013) Complete genome sequence of *Mesorhizobium ciceri* bv. *biserrulae* type strain (WSM1271(T)). *Stand Genomic Sci* 9(3):462–472.
23. Reeve W, et al. (2013) Complete genome sequence of *Mesorhizobium opportunistum* type strain WSM2075(T.). *Stand Genomic Sci* 9(2):294–303.
24. Song L, Pan Y, Chen S, Zhang X (2012) Structural characteristics of genomic islands associated with GMP synthases as integration hotspot among sequenced microbial genomes. *Comput Biol Chem* 36:62–70.
25. Groth AC, Calos MP (2004) Phage integrases: Biology and applications. *J Mol Biol* 335(3):667–678.
26. Hynes MF, Quandt J, O'Connell MP, Pühler A (1989) Direct selection for curing and deletion of *Rhizobium* plasmids using transposons carrying the *Bacillus subtilis* *sacB* gene. *Gene* 78(1):111–120.
27. Haskett T, et al. (2016) Complete genome sequence of *Mesorhizobium ciceri* bv. *biserrulae* strain WSM1284, an efficient nitrogen-fixing microsymbiont of the pasture legume *Biserrula pelecinus*. *Genome Announc* 4(3):e00514–e00516.
28. Vidal C, et al. (2009) *Mesorhizobium metallidurans* sp. nov., a metal-resistant symbiont of *Anthyllis vulneraria* growing on metalcolous soil in Languedoc, France. *Int J Syst Evol Microbiol* 59(4):850–855.
29. Kelly S, et al. (2014) Genome sequence of the *Lotus* spp. microsymbiont *Mesorhizobium loti* strain N2P2037. *Stand Genomic Sci* 9(7):7.
30. Pankhurst CE, Macdonald PE, Reeves JM (1986) Enhanced nitrogen fixation and competitiveness for nodulation of *Lotus pedunculatus* by a plasmid-cured derivative of *Rhizobium loti*. *Microbiology* 132(8):2321–2328.
31. Pankhurst CE, Broughton WJ, Wieneke U (1983) Transfer of an indigenous plasmid of *Rhizobium loti* to other rhizobia and *Agrobacterium tumefaciens*. *J Gen Microbiol* 129(8):2535–2543.
32. Lewis JA, Hatfull GF (2001) Control of directionality in integrase-mediated recombination: Examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res* 29(11):2205–2216.
33. Echols H, Guarneros G (1983) Control of integration and excision. *Lambda II*, eds Hendrix R, Roberts J, Stahl F, Wiesberg R (Cold Spring Harbor Lab Press, Cold Spring Harbor, New York), pp 73–93.
34. Weisberg RA, Gottesman ME (1971) The stability of Int and Xis functions. *The bacteriophage Lambda*, ed Hershey AD (Cold Spring Harbor Lab Press, Cold Spring Harbor, New York), pp 489–500.
35. Trempey JE, Kirby JE, Gottesman S (1994) Alp suppression of Lon: Dependence on the *slpA* gene. *J Bacteriol* 176(7):2061–2067.
36. Mercado-Blanco J, García F, Fernández-López M, Olivares J (1993) Melanin production by *Rhizobium meliloti* GR4 is linked to nonsymbiotic plasmid pRmeGR4b: Cloning, sequencing, and expression of the tyrosinase gene *mepA*. *J Bacteriol* 175(17):5403–5410.
37. Cubo MT, Buendia-Claveria AM, Beringer JE, Ruiz-Sainz JE (1988) Melanin production by *Rhizobium* strains. *Appl Environ Microbiol* 54(7):1812–1817.
38. Piñero S, et al. (2007) Tyrosinase from *Rhizobium etli* is involved in nodulation efficiency and symbiosis-associated stress resistance. *J Mol Microbiol Biotechnol* 13(1–3):35–44.
39. Barcellos FG, Menna P, da Silva Batista JS, Hungria M (2007) Evidence of horizontal transfer of symbiotic genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs *Sinorhizobium* (Ensifer) *fredii* and *Bradyrhizobium elkanii* in a Brazilian Savannah soil. *Appl Environ Microbiol* 73(8):2635–2643.
40. Dorgai L, Oberto J, Weisberg RA (1993) Xis and Fis proteins prevent site-specific DNA inversion in lysogens of phage HK022. *J Bacteriol* 175(3):693–700.
41. Beringer JE (1974) R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* 84(1):188–198.
42. Ronson CW, Nixon BT, Albright LM, Ausubel FM (1987) *Rhizobium meliloti ntrA* (*rpoN*) gene is required for diverse metabolic functions. *J Bacteriol* 169(6):2424–2431.
43. Bankevich A, et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455–477.
44. Yates R, et al. (2016) Authentication of rhizobia and assessment of the legume symbiosis in controlled plant growth systems. *Working with Rhizobia*, eds Howieson JG, Dilworth MJ (Australian Centre for International Agricultural Research, Canberra), pp 73–108.
45. Gibson DG (2011) Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol* 498:349–361.
46. Reeve WG, Tiwari RP, Meline V, Poole PS (2016) Fundamental molecular techniques for rhizobia. *Working with Rhizobia*, eds Howieson JG, Dilworth MJ (Australian Centre for International Agricultural Research, Canberra), pp 221–242.
47. Song L, Florea L, Langmead B (2014) Lighter: Fast and memory-efficient sequencing error correction without counting. *Genome Biol* 15(11):509.
48. Kolmogorov M, Raney B, Paten B, Pham S (2014) Ragout—a reference-assisted assembly tool for bacterial genomes. *Bioinformatics* 30(12):i302–i309.
49. Terpolilli JJ, O'Hara GW, Tiwari RP, Dilworth MJ, Howieson JG (2008) The model legume *Medicago truncatula* A17 is poorly matched for N₂ fixation with the sequenced microsymbiont *Sinorhizobium meliloti* 1021. *New Phytol* 179(1):62–66.
50. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA (2011) BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics* 12:402.
51. Reeve W, et al. (2013) Complete genome sequence of *Mesorhizobium australicum* type strain (WSM2073(T)). *Stand Genomic Sci* 9(2):410–419.
52. Hubber A, Vergunst AC, Sullivan JT, Hooykaas PJJ, Ronson CW (2004) Symbiotic phenotypes and translocated effector proteins of the *Mesorhizobium loti* strain R7A VirB/D4 type IV secretion system. *Mol Microbiol* 54(2):561–574.
53. Thoma S, Schobert M (2009) An improved *Escherichia coli* donor strain for diparental mating. *FEMS Microbiol Lett* 294(2):127–132.
54. Jarvis BDW, Pankhurst CE, Patel JJ (1982) *Rhizobium loti*, a new species of legume root nodule bacteria. *Int J Syst Evol Microbiol* 32(3):378–380.
55. Bailey RW, Greenwood RM, Craig A (1971) Extracellular polysaccharides of *Rhizobium* strains associated with *Lotus* species. *Microbiology* 65(3):315–324.
56. Reeve W, et al. (2015) A genomic encyclopedia of the root nodule bacteria: Assessing genetic diversity through a systematic biogeographic survey. *Stand Genom Sci* 10(14):1–8.
57. Khan SR, Gaines J, Roop RM, 2nd, Farrand SK (2008) Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti plasmid quorum sensing. *Appl Environ Microbiol* 74(16):5053–5062.
58. Dombrecht B, Vanderleyden J, Michiels J (2001) Stable RK2-derived cloning vectors for the analysis of gene expression and gene function in gram-negative bacteria. *Mol Plant Microbe Interact* 14(3):426–430.
59. Miller WG, Leveau JH, Lindow SE (2000) Improved *gfp* and *inaZ* broad-host-range promoter-probe vectors. *Mol Plant Microbe Interact* 13(11):1243–1250.
60. Antoine R, et al. (2000) New virulence-activated and virulence-repressed genes identified by systematic gene inactivation and generation of transcriptional fusions in *Bordetella pertussis*. *J Bacteriol* 182(20):5902–5905.
61. Quandt J, Hynes MF (1993) Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria. *Gene* 127(1):15–21.