Comparative analysis of integrative and conjugative mobile genetic elements in the genus *Mesorhizobium*

Elena Colombi1,2,*, Benjamin J. Perry3, John T. Sullivan3, Amanuel A. Bekuma4†, Jason J. Terpolilli4, Clive W. Ronson3 and Joshua P. Ramsay1,2,*

**Abstract**

Members of the *Mesorhizobium* genus are soil bacteria that often form nitrogen-fixing symbioses with legumes. Most characterised *Mesorhizobium* spp. genomes are ~8 Mb in size and harbour extensive pangenomes including large integrative and conjugative elements (ICEs) carrying genes required for symbiosis (ICESyms). Here, we document and compare the conjugative mobilome of 41 complete *Mesorhizobium* genomes. We delineated 56 ICEs and 24 integrative and mobilizable elements (IMEs) collectively occupying 16 distinct integration sites, along with 24 plasmids. We also demonstrated horizontal transfer of the largest (853,775 bp) documented ICE, the tripartite ICE*MspSym AA22*. The conjugation systems of all identified ICEs and several plasmids were related to those of the paradigm ICESym ICE*SymR7A*, with each carrying conserved genes for conjugative pilus formation (*trb*), excision (*rdfS*), DNA transfer (*rlxS*) and regulation (*fseA*). ICESymS have likely evolved from a common ancestor, despite occupying a variety of distinct integration sites and specifying symbiosis with diverse legumes. We found extensive evidence for recombination between ICEs and particularly ICESymS, which all uniquely lack the conjugation entry-exclusion factor gene *trbK*. Frequent duplication, replacement and pseudogenization of genes for quorum-sensing-mediated activation and antiactivation of ICE transfer suggests ICE transfer regulation is constantly evolving. Pangenome-wide association analysis of the ICE identified genes potentially involved in symbiosis, rhizosphere colonisation and/or adaptation to distinct legume hosts. In summary, the *Mesorhizobium* genus has accumulated a large and dynamic pangenome that evolves through ongoing horizontal gene transfer of large conjugative elements related to ICE*SymR7A*.

**DATA SUMMARY**

Bacterial genomes analysed in this study have been deposited previously in the National Centre for Biotechnology Information (NCBI) Genome database and are listed in Table S1 (available with the online version of this article). Genomes sequences produced in this study are deposited in NCBI, accession number NZ_NSFP00000000 and CP048406. Nine supplementary figures and seven supplementary tables are included in the online version of this article. Details on the programmes and scripts used are available at https://github.com/EC-Rufina/MesICE.

**INTRODUCTION**

Horizontal gene transfer (HGT) is a major force shaping bacterial evolution [1–5]. Transfer of suites of genes encoding metabolic pathways refined by selection in a donor organism can endow recipients of HGT with the ability to colonise...
Gram-negative soil bacteria in the genus *Mesorhizobium* can establish a non-obligatory nitrogen-fixing symbiosis with legumes [25]. Symbiotic *Mesorhizobium* spp. infect legume root nodules, within which the bacteria fix atmospheric N\(_2\) into NH\(_3\), which is assimilated by the plant. In return the host plant supplies carbon, essential nutrients, and a niche for the microsymbiont. In *Mesorhizobium* spp., genes essential to the establishment of N\(_2\)-fixing associations are most often encoded on symbiosis ICEs (ICESyms) [19, 26–28] and occasionally on plasmids [29]. Symbiosis ICEs capable of horizontal transfer have been described in *Mesorhizobium* spp. isolated from nodules of *Lotus, Biserrella pelecinus* (biserrella) and *Cicer* (chickpea) [19, 26–28, 30]. ICESyms have been observed to transfer to non-symbiotic *Mesorhizobium* spp. and into symbiotic elements in prokaryotes are integrative and conjugative elements (ICEs) [9]. ICEs are chromosomally-integrated elements that are passively replicated as a part of the genome but are capable of horizontal transmission facilitated by their encoded excision and conjugation systems [10, 11]. ICEs typically carry a conserved set of ‘backbone’ genes essential for ICE integration, excision, conjugation and regulation of transfer [12]. In addition to backbone genes, ICEs often carry accessory genes of adaptive significance to their hosts, such as genes affecting biofilm formation, pathogenicity, antibiotic and heavy metal resistance, bacteriocin synthesis, iron acquisition, and symbiosis [13–23]. During conjugative transfer, ICEs excise from the chromosome and circularise. ICE DNA is then nicked by a conjugal relaxase, and a single-stranded DNA copy of the ICE is transferred to the new host. Second strand synthesis, recircularization, and integration then occurs in donor and recipient cells [10, 11]. The conjugative machinery of ICEs and conjugative plasmids can also facilitate horizontal transfer of co-resident non-conjugative elements such as mobilizable plasmids and integrative mobilizable elements (IMEs) [24].

ICEs are important components of the bacterial accessory genome and key drivers of bacterial evolution but remain relatively unexplored compared with their plasmid counterparts. In this study, we performed a comparative analysis of mobile genetic elements identified in 41 complete *Mesorhizobium* genomes to gain insight into ICE distribution, evolution and structure. ICEs carrying symbiosis genes (ICESyms) likely evolved from a common ancestor and have subsequently diverged to integrate at a variety of chromosomal sites and specify symbiosis with several distinct legume hosts. Overall, the *Mesorhizobium* mobilome was dominated by ICEs and conjugative plasmids harbouring conjugation-gene clusters related to those of ICESyms and we observed extensive evidence of recombination between elements. Frequent rearrangement and duplication of genes controlling quorum sensing and ICESym transfer regulation suggested ICEs are undergoing extensive evolution to fine-tune their transferability. This study provides a significant advance in our understanding of *Mesorhizobium* genome organisation and ICESym evolution.

and other tripartite ICEs possess three distinct site-specific integrases intS, intM and intG, that sequentially recombine three separate chromosomal regions \(\alpha\), \(\beta\) and \(\gamma\), through three pairs of \(attL\) and \(attR\) sites, to form a single circular DNA element for conjugative transfer [37]. The tripartite \(\alpha\) regions are the largest fragments of tripartite ICEs, and while the \(\alpha\) regions lack integrase genes (which are encoded on fragments \(\beta\) and \(\gamma\)) they appear to carry most genes with known involvement in symbiosis and conjugative transfer.

The aim of this work was to characterise the mobilome of *Mesorhizobium* spp. in order to gain insight into its evolution. The accurate delineation of ICEs and large tripartite ICEs that frequently exceed 0.5 Mb in length requires complete and accurately assembled genomes. In this study, we interrogated 41 complete *Mesorhizobium* genome sequences and conducted a comparative analysis of conjugative and mobilizable elements. The *Mesorhizobium* spp. mobilome was found to be dominated by ICEs and conjugative plasmids with conjugation and regulatory genes related to ICEMSym\(^{R7A}\). The various ICESyms specifying symbiosis with diverse legumes [26–28, 30] have likely evolved from a common ancestor within the larger family of proteobacterial ICEs [34, 38]. Extensive evidence for gene exchange between mobile elements suggests that ICEs, IME and plasmids have frequently recombined, swapped genetic cargo and their
mechanisms of host maintenance. Evidence of frequent rearrangement and duplications of quorum-sensing regulatory genes controlling the induction of ICESym mobilization suggest that the rates of ICE transfer are under continuing selection.

METHODS

Whole genome sequencing

\textit{Mesorhizobium} sp. AA22 was sequenced both using Illumina HiSeq 2×100 bp paired-end reads (MrDNA) and the PacBio RSII platform (Mycogen, South Korea). PacBio raw reads were de novo assembled with Flye 2.6 [39]. The assembly was polished with the Pachio reads five times using Minimap2 2.1.7-r941 [40] and Racon 1.4.3 [41]. The assembly was then further polished five times with the Illumina HiSeq reads using Pilon 1.23 [42]. The starting position of the genome was set with Circulator 1.5.5 [43].

The transconjugant strain \textit{M. japonicum} R7ANSxAA22 was sequenced using Illumina MiSeq. Illumina sequence adapter contamination was removed using nesoniclips (v0.132) (https://github.com/Victorian-Bioinformatics-Consortium/nesoni), and reads were corrected using Lighter (v1.1.1) [44] and assembled with SPAdes V. 3.5.0 [45].

\textit{Mesorhizobium} genus phylogeny

The \textit{Mesorhizobium} genus phylogeny was constructed using a method modified from Prokhorchik \textit{et al.} [46]. Genomes (Table S1) were re-annotated using Prokka [47], and Proteinortho [48] was run with default settings to identify single-copy core genes. The nucleotide sequences of 1,609 single copy orthologues were then aligned with the MAFFT algorithm [49], and core gene alignments were concatenated and stripped of gaps with Goalign (https://github.com/evolbioinfo/goalign). Maximum likelihood phylogeny reconstruction was performed with RAxML v8.2.10 (parameters: -f a -p $RANDOM -x $RANDOM -N 100 m GTRCATX -T 16) [50]. The tree was rooted at midpoint using the midpoint() function from the phangorn package in R [51] and then it was visualised using FigTree v1.4.4 (https://github.com/rambaut/figtree/), only nodes with bootstrap values above 80 were displayed (Fig. S1).

Pairwise average nucleotide identity (ANI) was calculated using FastANI [52] on the entire genome of the strains. Strains were classified as same species if they showed an ANI of 95% or above [53].

Mobile genetic element search and pangenome classification

To identify putative horizontally transmissible elements, relaxase gene sequences were searched with hmmmscan from HMMER [54] using available hidden Markov model (HMM) protein profiles (MOB database) [55] (Table S2) defined for distinct MOB-protein families [56]. For this search, a bit score threshold (-T) of 33 was used. For ambiguous classifications, the MOB-protein family was chosen by the best e-value in the full sequence. For the identification of ICEs and IMEs, genomes were queried using the BLASTN algorithm [57] using \textit{trbB} (EB234_29245) from ICE\textit{Ml}Sym\textit{A} [58], and \textit{traA} (EB235_34495) from ICE\textit{MlMl}Sym\textit{R} [59], here referred as ICE\textit{MlMl}Sym\textit{R}. We used a cut off of 50% pairwise identity and, where applicable, only elements with recognisable \textit{att} sites were selected (Table S3). Four pTONO-1-like sequences were present within chromosomal contig assemblies (CP034446.1, CP034447.1, CP034448.1, and CP034451.1). While it was possible that these plasmids were indeed chromosomally integrated, this seemed unlikely as each of the regions harboured plasmid-like replication genes (\textit{repABC}), and \textit{att} sites were not detected, suggesting the plasmid sequences were artificially joined to the chromosome sequence during assembly (Table S2).

The tool Phaster [60] was used to search all the genomes for potential prophage regions. Regions classified as ‘intact’ were inspected manually for the presence of capsid and tail proteins. Phage boundaries were delineated by the identification of direct sequence repeats (part of \textit{att} core regions) produced following the integration of the phage in the genome (Table S3).

Core and accessory genome calculations were initially performed using iterative incrementation of the minimum percent identity threshold of BLASTP for the assignment in protein ortho-groups from 20–95% amino-acid identity (Fig. S2); to avoid underestimating the number of accessory gene orthologues, Proteinortho was run with default settings (25% identity) and reporting singleton genes.

Backbone identification

Backbone genes were identified as genes present in at least 95% of the elements, a threshold that was introduced to take into account sporadic deletion events. The identification of backbone genes was guided by Roary [61] and Proteinortho [48]. For Roary, a minimum amino-acid identity of 40% was used because with this cut-off the programme identified genes known to be involved in the ICE life cycle. Proteinortho was run with default settings (25% identity) and reporting singleton genes.

ICE comparative analysis

For all the following analyses, only non-redundant ICEs (Table S3) were used. Redundancy was defined based on overall gene content rather than by accumulation of SNPs, i.e. an element was considered redundant if it shared the same accessory genes as another ICE, with the exception of insertion sequences (IS). The nucleotide sequences of strictly conserved and single-copy backbone genes (Table S4) were aligned with the MAFFT algorithm [49], trimmed and concatenated in Geneious [62]. The alignment was used to build phylogeny of the elements using PhylML 3.3 [63] with 100 bootstrap replicates which was rooted at midpoint using the midpoint() function from the phangorn package in R [51]. This alignment was also used to build a neighbor-net tree with SplitTree [64], and to infer recombination with ClonalFrameML [65].
For the integrase and quorum sensing genes trees, nucleotide sequences were aligned with the MAFFT algorithm [49], the tree built with RAxML v8.2.10 (parameters as above) [50], and visualised with FigTree (v.1.4.4, http://tree.bio.ed.ac.uk/software/figtree/).

The nod box site motifs were identified as described in Perry et al. [28]. nhmmscan [66] was run using the training sets of previously characterised nod box sequences from ICEMsp. SymR7A and ICEMspSymCFP007 [58, 67] as models.

Alfy [68] was used to identify chimaerism among ICEs and IMEs. Alfy was run selecting only matches with a P value <0.05 within a sliding window of 1000bp. Cytoscape [69] was then used to visualise the percentage of recombined sequences among ICEs, IMEs, and plasmids.

A distance matrix was calculated with the dist.gene() function from the ape package in R [70] using as input the gene presence-absence data produced by Roary run with a minimum amino-acid identity of 40%. The distance matrix was then used to build a neighbor-joining tree using the nj() function from the ape package in R [70]. eggNOG mapper 5.0 [71] was used for the functional annotation of accessory genes. Pangenome-wide association analyses of the ICEs were performed with Scoary [72]. Mesorhizobium sp. M1D.F.Ca. ET043.01.1.1 harbours two ICEs that carry redundant symbiosis genes and one of the two ICEs (ICEMsp. SymMID-1) appeared to be undergoing pseudogenization; for example it carried a truncated copy of nodC [27]. ICEMsp. SymMID-1 was therefore removed to avoid misleading the association analysis. To identify genes associated with symbiosis, Scoary was run on the output files produced by Roary that was run on all elements with minimum percentage identity for BLASTP of 40% and with paralogs not split. For the identification of genes associated with the host plant, Roary was run on only ICESyms using a minimum percentage identity from the ape package in R [70] using as input the distance matrix calculated with the dist.gene() function.

RESULTS AND DISCUSSION

Identification of complete Mesorhizobium genomes

Forty complete or nearly complete Mesorhizobium spp. genomes (Table S1) available from NCBI (updated April 2020) were analysed. The complete genome sequence of the B. pelecinus spp. pelecinus symbiont Mesorhizobium sp. AA22 [75] was additionally completed in this study. A maximum-likelihood phylogeny inferred from a concatenated alignment of single-copy core genes revealed the 41 strains represented a genetically diverse group of mesorhizobia, and their ANI score showed they comprise 29 genospecies (Fig. S1). Most genomes analysed here are from strains isolated from root nodules of legume hosts Cicer, Lotus or Biserrula but also included are single strains isolated from Astragalus sinicus and Robinia pseudoacacia. Seven strains are from soil or aquatic environments.

Previous large-scale genome comparisons have revealed mesorhizobia harbour large and plastic pangenomes [27, 28]. Iterative pangenome analyses were carried out here with the 41 complete genomes using Proteinortho [48], initially using protein identity cut-offs ranging from 20–95% (Fig. S2). The estimated number of core genes remained largely unchanged with cut-off values between ~20–60%, but decreased above 60%, suggesting values <60% include most core genes without counting core genes from more distantly-related species as distinct orthologues. Using a 25% cut-off, the average genome in this dataset encoded 6386 proteins of which 1670 (26%) were part of the core genome.

Identification of ICEs, IMEs, conjugative plasmids and mobilizable plasmids in complete Mesorhizobium genome sequences

Conjugative experiments

M. japonicum R7ANS pFAJ1708-GFP [26] was used as recipient strain and Mesorhizobium sp. AA22 as donor strain. R7ANS is auxotrophic for biotin, nicotinate and thiamine, and the genes for the biosynthesis of biotin and nicotinate are present on ICEMsp.SymA22, so exconjugants were selected on defined (G/RDM) medium containing thiamine (1 µg ml⁻¹) and tetracycline (2 µg ml⁻¹), but lacking biotin and nicotinate. Donor and recipient were separately grown to stationary phase in liquid TY at 28 °C with shaking at 200 r.p.m. One millilitre of culture of donor and recipient strains were pelleted by centrifugation at 14,000 r.p.m. for 1 min, resuspended together in 50 µl of TY, spotted onto TY agar plates, and incubated at 28 °C for 48 h. Spot matings were resuspended in sterile deionized water and serial dilutions were plated onto G/RDM supplemented with tetracycline and thiamine to select transconjugant strains, and onto G/RDM supplemented with thiamine to count donor strains. Plates were incubated at 28 °C for 7–10 days. The rate of conjugative transfer was calculated as the number of transconjugants per donor cell.
Agrobacterium (not to be confused with VirD2 relaxases) and relaxases encoded by several rhizobial plasmids. Potential conjugation and conjugative mobilization systems in the Mesorhizobium genomes were identified by searching for each of the MOB-family proteins using hidden-Markov models [55]. Eighty-two MOBP subtype 1 (MOB P1) relaxase genes were identified on plasmids and chromosomes, including the ICE\textsuperscript{Ml} Sym R7A relaxase RlxS (Table S2, Fig. 1). blastn analysis performed using the ICE\textsuperscript{Ml} Sym R7A type IV secretion system (T4SS) conjugation gene \textit{trbB} (EB234_29245) revealed 60 chromosomal regions and five plasmids carrying ICE\textsuperscript{Ml} Sym R7A-like T4SS conjugation genes, which resemble the \textit{trb} conjugation-gene clusters of plasmid RP4 [58, 77, 78]. Thirteen MOBP\textsubscript{1} genes were located on plasmids with distinct \textit{vir}-like conjugative T4SS gene clusters resembling those of the \textit{Mesorhizobium} sp. TONO plasmid pTONO-1. One of the MOBP\textsubscript{1} plasmids (pMaB5P-1) carried a T4SS coupling-protein (T4CP) gene \textit{virD4} and two \textit{mobP C} loci, but no T4SS genes, suggesting it was potentially mobilizable by other conjugative elements.

Twenty-nine MOBP relaxase genes were identified in chromosomal sequences and two on plasmids. The majority of the MOBP genes (25 IMEs and one plasmid across 17 genomes) were part of a Ti-plasmid-like \textit{traACD} locus containing genes for the TraA relaxase and the relaxasome accessory factors TraC and TraD [79]. The \textit{traA} and \textit{traC} genes on each element were divergently orientated, as they are on the Ti plasmid where they flank the origin of transfer sequence (\textit{oriT}). This \textit{traACD} locus was previously identified on an adhesin-encoding IME located downstream of ICE\textsuperscript{Ml} Sym R7A in \textit{Mesorhizobium} sp. R88B [59, 80] (renamed here ICE\textsuperscript{R88B}, consistent with its classification as an IME). Each of the ICE\textsuperscript{R88B}-like \textit{traACD} regions additionally carried a putative recombination directionality factor (RDF) gene downstream of \textit{traA}, suggesting that RDF-gene transcription and chromosomal excision of the IME may be transcriptionally coupled with relaxase-gene expression. The remaining four chromosomal MOBP relaxases were associated with unrecognised elements. One IME (IME\textsuperscript{M} sp. M2A.F.02-2) carrying a lone MOBP gene was also identified. Only five MOB\textsubscript{Q} relaxase-genes identified, four were located across three plasmids and one on a putative IME (Table S2, Fig. 1). All putative mobilizable plasmids and IMEs that lacked \textit{traACD} genes carried one or more copies of a \textit{virD4}-like T4CP gene. No IME carried complete set of \textit{nod}-factor and nitrogenase-biosynthesis genes required for nitrogen-fixing symbiosis.

To delineate the chromosomal boundaries of identified ICEs and IMEs, sequence regions surrounding the MOB genes and/or conjugation-gene clusters were inspected for integrase genes and flanking DNA sequence repeats commonly present within integrase attachment (\textit{att}) sites. For tripartite ICEs, the conserved \textit{att} site core sequence repeats corresponding with the integrases IntS, IntG and IntM were used both to identify the presence of tripartite ICEs and delineate the three tripartite
ICE regions in the chromosome. In total, 56 ICES across 38 genomes were delineated (Table S3). The att sites were not identified for four ICES (Table S2). Delineated ICES ranged in length from 65,889 bp for ICEMP2202 [3] to 853,775 bp for the tripartite ICE ICEMSPSymM [4]. All ICES identified carried ICEMISymR7-related conjugation-gene clusters. Genes for nitrogen fixation and symbiosis were identified on 34 ICES and these are hereafter collectively referred to as ICESyms. Thirteen of the ICESyms were tripartite in structure and carried the same three pairs of att sites and integrase genes as the paradigm tripartite ICE ICEMCSym1271. Monopartite ICESyms of Mesorhizobium spp. isolated from Lotus were all integrated adjacent to the tRNA-Phe gene, while monopartite Cicer ICESyms were integrated adjacent to a tRNA-Ser gene (Table S3) as documented for M. ciceri CC1192 [5]. ICES lacking symbiosis genes were identified adjacent to tRNA-Met, tRNA-Arg, tRNA-Gly, tRNA-His, and tRNA-Pro genes, as well as gualA, radC and dusa. Chromosomal boundaries for 24 putative MOB- carrying IMEs (Table S3) were delineated. IMEs ranged in size from 5,814 bp (IMEMsp.M2A.F05.1) to 271,605 bp (IMEMsp.N2P2077.1). Three IMEs were integrated adjacent to a tRNA-Asn gene, while the remaining 21 utilised the same spectrum of att sites as ICES (radC, tRNA-Met, tRNA-Arg, tRNA-Phe and tRNA-Ser genes). IMEMSP.M2A.F82.2 (40,024 bp) and the MOB- carrying IME, IMEMsp.M2A.F9284 (98929 bp), occupied unique sites adjacent to tRNA-Thr and tRNA-Trp genes, respectively. In total, the 79 identified ICES and IMEs collectively occupied 16 distinct integration sites within the 41 Mesorhizobium genomes (Table S7). Seventy-three of the elements carried integrase genes encoding tyrosine recombinases and serine recombinases/integrase genes were present on five IMEs and one ICE.

Five of the 13 conjugative plasmids identified carried conjugation-gene clusters related to that of ICEMISymR7 and the remaining eight carried conjugation-gene clusters resembling the vir-like T4SS system of pTONO-1 (Fig. 1). Conjugative plasmids with ICEMISym-like T4SS genes ranged in size from 208,315 bp (pMlb of MAFF303099) to 648,231 bp (pMc1192 of CC1192) and none of these carried a complete set of nodulation or nitrogen-fixation genes. pMc1192 was recently shown to be not essential for N2 fixation with chickpea [6]. Overall, most genomes contained at least one ICE (1.22±0.85 ICE per genome) and less frequently IMEs (0.5±0.7IME per genome). Twenty-five genomes contained no plasmids, nine contained a single plasmid and seven harboured multiple plasmids. M. oceanicum sp. B7, M. sp. 8, M. amorphae CCNWGS0123 and M. sp. 7653R were the only strains lacking any detectable ICES or IMEs and, interestingly, CCNWGS0123 and 7653R were also the only strains carrying a complete complement of genes required for nodulation and nitrogen fixation on plasmids. Only the symbiosis plasmid of CCNWGS0123 carried a conjugation-gene cluster (vir-like).

The identified mobile genetic elements (MGE) carry approximately 12% of the identified pangenome. This suggests a major fraction of the pangenome is either located on unidentified MGE or that most of the pangenome genes have lost mobility and now form a large but stable fraction of each chromosome. For the pangenome genes identified on MGEs, i.e. the mobile pangenome or mobilome, ~72% were present on ICES and of these, 78% were present on ICESyms (Fig. S3). Plasmids and IMEs carried 20.4% and 5.9% of the mobilome, respectively, while bacteriophages carried only 4.6%.

**ICES in Mesorhizobium share a conserved core-gene organisation**

Proteinortho [7] and Roary [8] were used to identify protein-coding genes conserved across 95% of ICES and plasmids carrying an ICEMISymR7-like conjugation-gene cluster. Both tools identified 17 backbone genes including the ICEMISymR7 conjugation-gene cluster (T4SS) genes msi031, trbBCDEJFG, and msi021, the T4CP gene traG, the relaxase rlxS, the putative peptidoglycan transglycosylase gene msi107, the traF preplin peptidase gene and the RDF gene rdfs. Genes encoding the DUF736 domain (msi110 on ICEMISymR7) were present on all elements (Fig. S4), often in duplicate. Concatenated DNA alignments of the single-copy backbone genes (Table S4) were used to generate a maximum likelihood tree (Fig. 2a). Despite the association of ICESyms with a variety of integrate genes and chromosomal integrational sites, the tree grouped all monopartite and tripartite ICESyms into a single clade that exhibited lower average sequence divergence (88% pairwise nucleotide identity) than other ICES (69% pairwise nucleotide identity). Only one ICE lacking symbiosis genes (ICEMsp.M2A.F826) grouped with the ICESyms. ICESyms present in strains isolated from the same plant genus grouped together (with the exception of ICEMspSymAA24), even though alignments used to produce the tree only included genes with transfer-related functions. Interestingly, all identified ICESyms (and ICEMsp.M2A.F826) lack the conjugation entry exclusion gene trbK [9, 10], genes encoding a putative Rep_3 superfamily replication initiator, and DUF2840-domain protein located between rdfs and traF (Figs 1 and S4), which were present on all other ICES identified. These genes are present on distantly related ICES and plasmids found throughout the proteobacteria, including Tn4371 [3, 11], suggesting gene loss occurred in a common ancestor of the ICESyms. Conjugative plasmids pMlb and pCC1192 clustered closely with ICES present in separate branches of the tree, suggesting the conjugation-gene clusters on these elements may have switched between ICE and plasmid modes of host maintenance. All plasmids carrying an ICEMISymR7-like conjugation system retained the rdfs gene, which in ICEMISymR7 and ICEMCSym1271 stimulates integrase-mediated ICE excision [32, 37]. Rdfs is also a transcriptional regulator [36, 37], so it may have additional regulatory roles in transfer, which may explain its retention on plasmids [9].

Comparisons of ICE backbone gene phylogeny and ICE integration site suggested there was no strict association between the two. The phylogenies of tripartite ICE integrase genes (Fig. S5) and their ICE backbone genes (Fig. 2a) were also incongruent. While the tripartite ICES carry closely related IntG and IntM integrases that group separately from similar integrase genes on monopartite ICES, the transfer genes of
monopartite and tripartite ICEs present in Lotus-associated mesorhizobia are more closely related to each other than they are to those of tripartite ICEs present in Cicer-associated mesorhizobia and vice versa.

**Gene duplications and pseudogenization have shaped quorum sensing and transfer regulation on ICESymS**

Excision and conjugal transfer of ICEM/Sym$^{87A}$ and the tripartite ICE, ICEMcSym$^{127I}$, are activated by quorum sensing; however, they carry different arrangements of $traR$ and $traI$ genes. ICEM/Sym$^{87A}$ carries a single copy of the QS regulator gene $traR$ (or $msi174$, referred to here as $traR2$) and two acyl-homoserine lactone synthase genes named $traI1$ and $traI2$. While the TraI2 sequence shares 65% amino-acid identity with TraI1, the $traI2$ gene is not essential for AHL production and an in-frame deletion of $traI2$ is not affected for ICEM/Sym excision or transfer [33]. Conversely, ICEMcSym$^{127I}$ lacks $traI2$ altogether, but carries two functional copies of $traR$ ($traR1$ and $traR2$) encoding proteins with 54% amino-acid identity, and a single copy of $traI$ ($traI1$). Here we searched for TraR/TraI homologues.

Fig. 2. ICE phylogeny and regulatory-gene organisation. (a) PhyML tree based on the concatenation of alignments of 15 single-copy backbone genes, scale bar indicates substitutions per site. (b) Organization of regulatory genes colour-coded as in (c), lighter boxes are likely pseudogenes, numbers in $traR$ and $traI$ indicate the relatedness between the copies as described in main text, * indicates the gene is truncated. (c) Regulatory network activating ICE transfer in ICEM/Sym$^{87A}$ and ICEMcSym$^{127I}$.
encoded on ICESyms, which revealed further variations in copy number and gene organisation of the \textit{traR} and \textit{traI} genes (Figs 2b and S6). While additional genes encoding TraR/LuxR-family regulators were also identified both on ICESyms and other ICEs and plasmids, the predicted proteins were distantly related (~25\% amino-acid identity). All but one ICESyms and ICE \textit{M} \textit{sp. M2A.F .046} carried \textit{traI1} but mostly only \textit{Lotus}-associated ICESyms carried a copy of \textit{traI2}. No \textit{Lotus}-associated ICESyms carried a copy of \textit{traR1}, but closer inspection revealed that a ~100 bp remnant of the 3′ end of \textit{traR1} was present directly upstream of each \textit{traI1}, confirming that an ancestral copy of \textit{traR1} had been deleted in this ICE lineage. Nucleotide alignments of the eight \textit{traI2} and eight \textit{traI1} genes on \textit{Lotus}-ICESyms revealed that \textit{traI2} sequences collectively carried 147 unique SNPs, 92 amino-acid changes and two frame-shifting deletions. In contrast, \textit{traI1} genes exhibited lower divergence with only 23 unique SNPs (13 synonymous). This suggests the \textit{traI2} alleles, while conserved on \textit{Lotus}-associated ICESyms, are at various stages of pseudogenization. These observations support our previous proposal \cite{37} that gene duplication of the \textit{traR- traI} locus occurred in an ancestor of the ICESym clade and subsequently different ICESyms have variably lost one of the duplicates.

On ICE\textit{M} \textit{Sym} \textit{M2A.F}, both \textit{TraR2} and the activator of \textit{rdfS} expression, FseA, are inhibited by the antiactivator QseM \cite{35}. \textit{qseM} expression is controlled by an adjacently-encoded helix-turn-helix protein QseC (Fig. 2c). Searches for \textit{qseM} and \textit{qseC} genes on ICEs revealed numerous duplications and rearrangements. ICESyms of \textit{Lotus} and \textit{Biserrula} each harboured one copy of \textit{qseM} and multiple variants of \textit{qseC}. In some instances, \textit{qseC} variants carried mutations and/or deletions likely rendering them non-functional (Fig. 2b). All \textit{Cicer} lineage ICEs harboured at least two distinct copies of paired \textit{qseM}-\textit{qseC} genes, with the exception of ICE\textit{M} \textit{sp. Sym M2A.F .02}. The presence of two copies of \textit{qseM} in the \textit{Cicer}-associated ICESyms suggests that there may be distinct inputs controlling each \textit{qseM} gene, or perhaps the two QseM proteins may target the two TraR proteins also encoded by these ICEs with differing affinities. In summary, both positive and negative regulators controlling QS and conjugative transfer have been frequently duplicated, pseudogenized and rearranged on ICESyms, suggesting that natural selection is constantly adjusting the regulation of ICESym transfer.

\textbf{Cargo genes of conjugative and mobilizable elements}

Functional annotation by eggNOG mapper \cite{71} of the accessory genes on ICESyms indicated 56\% had putative metabolic roles. Of these, 27\% were likely involved in amino-acid metabolism and transport, 24\% in energy production, and 16\% in coenzyme metabolism. Similarly, a major portion of the accessory genes (43\%) carried by other ICEs and
related plasmids were predicted to be involved in metabolism, including 28% potentially involved in amino-acid metabolism and transport, and 25% in energy production and conversion. Collectively ICESyms carried 11,726 genes, with 2,200 (19%) genes present on one ICE only. Only 66 accessory genes were present on all ICESyms, while no accessory genes were universally conserved ICES lacking symbiosis genes, which collectively encoded 4,037 genes, of which 1,830 were present on one ICE only (45%). A similar pattern was observed for IMEs, which collectively carried 3,195 accessory genes, 47% of which were encoded by one IME only. No antibiotic-resistance genes, bacteriocins, non-ribosomally synthesized or post-translationally modified peptides were detected on any of the ICES and IMEs (ResFinder v3.1, Bagel4 [73]).

ICESym genes based on their host-strain legume association. Comparative analysis of ICESym gene cargo with that of other ICES identified genes specific to ICESyms (Table S5). As expected, these included genes coding for the synthesis and maturation of active nitrogenase (nif genes), for a bacteroid electron transfer flavoprotein complex and for a low O₂-cbb₃-type cytochrome oxidase (fix genes), genes participating in the biosynthesis and secretion of Nod factors (nodABC and nodII) and the recognition of plant flavonoids (nodD2), fixV, the key upstream regulator of the nif genes in R7A [67], was also present on all ICESyms. Conserved genes uninvestigated for roles in Lotus, Cicer and Biserrula symbiosis were also identified. A putative isosine isomerase with similarity to Sinorhizobium meliloti pSymB protein Smb20711 [82] was present in all ICESyms encoded adjacent to nodD2. Genes required for production of the queuosine precursor PreQ₀, queCDE, were present on all ICESyms but not on other ICES. Queuosine is a modified guanosine nucleotide base present on some tRNAs. In S. meliloti, queC appears to be required for an effective symbiosis with Medicago trunculata [83]. All Lotus-associated ICESyms also carry a second PreQ₀ synthesis cluster with genes ordered queEDC. All ICESyms carried at least one copy of dctA, coding for a C₃-dicarboxylate transport protein essential for nitrogen fixation [84]. In the Lotus lineage ICES, a single copy of dctA was paired with cognate regulatory genes dctBD, while in the other ICESyms only dctA was present. The Cicer-associated ICESyms and ICEMspSym³A22 harboured two copies of dctA. In S. meliloti pSymB, the expression of dctA is dependent on dctBD in free-living conditions but during symbiosis, dctA is efficiently expressed even in the absence of dctBD [85].

An association analysis was carried out comparing ICESym genes based on the host of isolation to identify genes that might be of specific adaptive significance for each host (Table S6). ICESym genes specifically associated with Cicer-associated mesorhizobia included the type III secretion system (T3SS) effector gene nopP and two hypothetical genes. These two hypothetical protein-coding genes are likely to code for T3SS effectors as they are preceded by T3SS boxes [86] and they were predicted to be secreted (SecP score >0.73) by SecretomeP [87]. In Bradyrhizobium spp., variations in nopP alleles control soybean host compatibility [88]. All Cicer-associated ICESyms carry the nodulation gene nodH and a type II secretion system (T2SS) gene cluster. Using nhmmscan [66] and a hidden Markov model profile generated from characterised nod box sequences [28], we identified a putative nod-box motif upstream of the T2SS gene cluster in CC1192 (CP0150621 coordinates: 4,306,615–4,306,661). Although the potential role of T2SS in symbiosis is not known, numerous phytopathogens employ T2SS systems for secretion of plant-tissue degrading enzymes [89]. It is possible that ICESym T2SS systems have a role in root colonization or infection. The ICESyms of Lotus-isolated mesorhizobia uniquely carry nodulation genes nodS (msi313), noef (msi383) and noeK (msi384), and an uncharacterised operon (msi322 to msi327, and previously not-annotated EB234_30895 between msi323 and msi324) encoding putative asparate and asparagus metabolism-related enzymes. No ICESym genes were significantly associated with strains specifically isolated from Biserrula hosts (Bonferroni P<0.05).

**Gene flux between MGEs**

ICES can exchange genes with other MGEs through several mechanisms, including IS-mediated transposition and inter-ICE/IME/plasmid recombination [23, 90–92]. To detect evidence for inter-ICE recombination we used the neighbor-net algorithm [93] to create a network representing the relationship between the backbone genes present on ICEs and plasmids (Fig. S8). Patterns of reticulation within the network supported inter-ICE recombination even between highly conserved and contiguous ICE backbone genes (Phi test for recombination [94] was P<0.0001). Analyses with ClonalFrameML [65] also indicated recombination events between backbone genes. The alignment-free programme Alfy was used to extend the analysis to include entire gene content of ICEs, IMEs, and plasmids sequences [68]. Extensive evidence of gene exchange between ICEs and IMEs was detected. Clustering of ICEs based on evidence for gene exchange separated ICESyms from other ICEs and IMEs, suggesting greater gene exchange between ICESyms and little recombination between ICESyms and other ICEs or IMEs (Fig. 3). When this analysis was repeated using only contiguous gene clusters present on all ICESyms (Fig. S9), most of the connections indicative of DNA exchange between ICESyms reduced from 172 to 79, suggesting a large proportion of the detected recombination events represent homologous recombination between conserved ICESym regions.

Given the extensive inter-ICESym recombination detected in the previous analyses, we interrogated some of the sequences in more detail. A comparison of the QS loci of the Lotus-associated ICESyms ICEMjspSym³AFF30099, ICEMSym⁴MZP2037 and ICEMSym⁴R7A, revealed several disruptions in average nucleotide identity between the three regions (Fig. 4a). In
ICE\textsubscript{Mj Sym MAFF303099}, the nucleotide sequence between \textit{traR2} and the second copy of \textit{qseC} was 96\% identical to tripartite ICE\textsubscript{Ml Sym NZP2037}, however, downstream of the copy of \textit{qseC} the nucleotide identity reduced to 82\% but shared 96\% identity to a region on ICE\textsubscript{Ml Sym R7A}. The QS locus in ICE\textsubscript{Msp Sym M4B} appeared to be a chimaera of the ICE\textsubscript{Mc Sym M1D-2} and tripartite ICE\textsubscript{Msp Sym M1A}, carrying alternating regions of 100\% pairwise nucleotide identity shared with these two ICEs (Fig. 4b). We also identified clear evidence for ancestral recombination between the \textit{fix}-gene loci of ICE\textsubscript{Ml Sym NZP2037}, ICE\textsubscript{Ml Sym NZP2042} and ICE\textsubscript{Ml Sym R7A} (Fig. 4c).

The largest documented horizontally mobile ICE, ICE\textsubscript{Msp Sym AA22}

\textit{Mesorhizobium} sp. AA22 was isolated from \textit{Biserrula pelecinus} and carries a tripartite ICE with a total length of 853,775 bp, making ICE\textsubscript{Msp Sym AA22} the largest ICE identified in any bacterium to date. The α, β and γ fragments are 766,834, 70,134 and 16,815 bp in length respectively. The other tripartite ICEs have α, β and γ fragments sized between 443,551 and 563,069 bp, between 19,361 and 74,018 bp, and between 5,412 bp and 25,603 bp respectively, indicating that most of the expansion in ICE\textsubscript{Msp Sym AA22} size has occurred within the α fragment. To determine if ICE\textsubscript{Msp Sym AA22} was mobile, we carried out conjugation experiments using the ICE-cured \textit{M. japonicum} R7A derivative R7ANS as a recipient [32]. ICE\textsubscript{Msp Sym AA22} transferred with an average frequency of 6.0×10^{-8} exconjugants per donor (standard deviation 2.3×10^{-8}). An exconjugant was isolated and sequenced using short-read sequencing. De novo assembly (NZ_NSFP0000000) confirmed each of the \textit{attL} and \textit{attR} site junctions were present and mapping of the exconjugant reads to the AA22 sequence confirmed the entire ICE\textsubscript{Msp Sym AA22} had transferred and recombined with the R7ANS chromosome as expected [26].

Pangenome analysis highlighted several gene clusters unique to ICE\textsubscript{Msp Sym AA22}. For instance, ICE\textsubscript{Msp Sym AA22} harbours a complete \textit{nuo} (NADH:ubiquinone oxidoreductase) gene cluster located at the beginning of the α fragment. The ICE\textsubscript{Msp Sym AA22} \textit{nuo} cluster is distinct from that coded for on the AA22 chromosome both in gene order and nucleotide identity (maximum identity observed of 59.6\% for \textit{nuoM}). Interestingly, the ICE\textsubscript{Msp Sym AA22} \textit{nuo} cluster shares the same gene arrangement and 81.7\% average pairwise nucleotide identity to that encoded on the symbiosis plasmid pSymA of \textit{S. meliloti} [95].

ICE\textsubscript{Msp Sym AA22} is also the only ICE in \textit{Mesorhizobium} to carry a CRISPR-Cas system. The type I-C CRISPR-Cas system (CP048406.1, coordinates: 6,182,785–6,192,544) harbours \textit{cas3}, \textit{cas5}, \textit{cas8} and \textit{cas7}; however it lacks \textit{cas4}, \textit{cas1} and \textit{cas2}.

\textbf{Fig. 4.} Chimaeric ICESym loci. Genes coding for QseC proteins are coloured in azure, \textit{qseM} is displayed in red, \textit{fseA} in brown, \textit{traI} in purple, \textit{traR} in yellow, and insertion sequences in white. Black indicates backbone genes, other genes are shown in grey, light-grey boxes indicate pseudogenes. Blue-shaded areas between genes on different ICESyms indicate homologous regions and the pairwise nucleotide identity is indicated. The triangle indicates tripartite ICE. (a) Comparison of the quorum-sensing-gene loci on ICE\textsubscript{Mj Sym MAFF303099}, ICE\textsubscript{Mj Sym MAFF303099} and ICE\textsubscript{Mj Sym MAFF303099}. (b) Comparison of the quorum-sensing-gene loci on ICE\textsubscript{Mc Sym M1D-2}, ICE\textsubscript{Msp Sym M4B} and ICE\textsubscript{Msp Sym M4B}. (c) Comparison of the \textit{fix}-gene loci on ICE\textsubscript{Mj Sym MAFF303099}, ICE\textsubscript{Mj Sym MAFF303099} and ICE\textsubscript{Mj Sym MAFF303099}.
involved in spacer acquisition [96, 97]. The CRISPR locus harbours over 40 spacers interspersed with non-identical palindromic repeats. The targets of the spacers (identified via BLASTn and CRISPRTarget [98]) appear similar to putative phage and plasmid sequences, some of which are found in other soil bacteria such as *Mesorhizobium*, *Rhizobium*, *Streptomyces*, *Azospirillum*, *Agrobacterium* and *Paraburkholderia*.

**CONCLUSIONS**

In this work we analysed 41 complete *Mesorhizobium* spp. genomes to enumerate the conjugative and mobilizable elements present. Of the pangenome genes present on an identifiable MGE (12%), more than 70% were present across 56 ICEs related to ICEMisym[^7A], including the largest mobile ICE characterised to date, ICEMisym[^7A]. The remaining mobilome spanned 24 plasmids, 27 IMEs and 17 bacteriophages. On average, each genome carried at least one ICE (1.22±0.85) and 50% of strains carried an IME. In contrast, of the 41 genomes examined, only 16 carried plasmids and only seven genomes harboured multiple plasmids.

The ICESymS formed a clear monophyletic group distinct from the other identified mesorhizobial ICEs. Presumably, following the capture of symbiosis genes by an archetypal ICESym, descendants diverged into ICESymS that specify symbiosis with different legumes. It is unclear if this archetypal ICESym evolved within the *Mesorhizobium* genus or was acquired from outside the genus. It is also unclear why symbiosis plasmids, which are the most common form of symbiosis MGE in other rhizobia, are rarely identified in *Mesorhizobium* spp. In *Rhizobium* spp. symbiosis plasmids predominantly carry vir-like conjugation systems [99] and interestingly, the only conjugative mesorhizobial symbiosis plasmid identified in this analysis (in CCNWGS0123) also carries a vir-like conjugation system. Thus there seems to be no intrinsic barrier to establishment of these conjugative symbiosis plasmids in *Mesorhizobium*. Broader genomic comparisons suggest ICEs generally have a larger host range than their plasmid counterparts [100], so it may be that once established in *Mesorhizobium* spp., ICESymS have been more successful infiltrating diverse members of the genus. Regardless of the mechanistic or evolutionary reasons, it is clear from these analyses that ICEs dominate the mobilome of nitrogen-fixing legume symbionts of the *Mesorhizobium* genus.

The ICEs and IMEs identified in this work were found integrated within 16 distinct sites in the chromosome and carried an overlapping set of integrase genes and integration sites. The phylogeny of ICE backbone genes suggests some ICEs have switched between integration sites during evolution and some ICEs even appear to have become plasmids in several branches of the tree (e.g. pMLb and pCC1192). Overall these observations mirror those made from broader analyses of ICEs, IME and plasmids in bacteria [100]. Given the ubiquity of conjugal elements identified in these genomes, we suspect that the frequent switching of integration mechanisms by these elements reflects fierce competition for chromosomal integration sites amongst ICEs and IME in this genus.

There is clear evidence that some ICESymS have likely switched from a monopartite to a tripartite structure. The tripartite-ICE integrase genes specifying integration at tRNA-Met(a) and guaA are distinct from their closest counterparts on monopartite ICEs (Fig. S5) suggesting the tripartite ICE recombination system evolved once. The observation that the tripartite ICEs structure appears a multiple positions in the ICE backbone tree suggests that tripartite ICEs have likely recombined with monopartite ICESymS and replaced the recombination systems on these ICEs.

We identified numerous instances of gene duplications, pseudogenization, rearrangements and recombination between loci involved in quorum sensing (traR and tral) and its inhibition (*qseM* and *qseC*) (Figs 2 and 4). In contrast, the gene organisation of conjugation-gene clusters and *rdfs*, *traF* and *rlxS* genes remained remarkably conserved even between very distantly related plasmids and ICEs. The frequent rearrangements of top-level regulators of transfer may reflect the conflicting evolutionary forces faced by ICEs, which must balance their evolutionary trajectories between vertical and horizontal descent. High frequency ICE transfer may increase the likelihood that an ICE will arrive in a new host with superior attributes but high transfer rates could also tax the bacterial host resources and hence be subject to negative selection. Dysregulation of quorum sensing resulting from two ICESymS in the same cell might also drive the observed rearrangements. Consistent with this possibility, introduction of an extra copy of *traR* and the preceding promoter region on a low-copy plasmid is enough to activate AHL production, ICEMisp[^7A] excision and conjugative transfer in *M. japonicum* R7A [32, 33]. Interestingly, in the only strains in this dataset carrying two ICESymS (*Mesorhizobium* sp. M1D.F.Ca.ET.043.01.1.1 and M2A.F.Ca.ET.046.03.2.1) one of these ICESymS, ICEMsp. Sym[^7A]+, is also the only ICESym lacking a copy of the *traI* AHL-synthase gene (Fig. 2b).

Our analysis of gene flux between MGE suggests a substantial amount of homologous recombination occurs between the ICESymS and in particular, ICESymS specifying symbiosis with the same legume host. There are several explanations why recombination may occur more readily between the ICESymS than other ICEs. More closely related ICESymS are more likely to utilise the same integration sites and therefore integrate in tandem in the same locus, a situation known to promote ICE recombination [90]. The frequency of homologous recombination is exponentially more likely to occur between the closely-related sequences of the ICESymS (88% backbone pairwise nucleotide identity) than with less related ICEs (69% backbone pairwise nucleotide identity) [101]. Mesorhizobia that participate in symbiosis with the same legume host may occupy the same niche, providing greater opportunity for ICESymS to encounter each other in the same cell through HGT. An interesting observation from the comparisons of conserved ICE backbone genes is that all ICESym and ICEMsp. M2A.F.046 lack the *trbK*, tentatively indicating ICESymS have lost one mechanism to inhibit entry of competing ICESymS. This
could suggest that homologous recombination has been an integral part of ICESym evolution and that loss of trbK has contributed by increasing the opportunity for two ICESymS to meet in the same cell.

**Funding information**
J.P.R. is the recipient of an Australian Research Council Future Fellowship (Project ID FT170100235) funded by the Australian Government. This work was supported by Australia Awards Africa for AAB.

**Author contributions**
Conceptualization: E.C., J.P.R. Funding: J.P.R. Methodology: E.C., J.P.R. Formal analysis: E.C., J.P.R. Investigation: E.C., J.P.R., A.B., J.T.S., A.B., J.J.T., C.W.R. Supervision: J.P.R., C.W.R.

**Conflicts of interest**
The authors declare that there are no conflicts of interest.

**References**
1. Mirkin BG, Fenner TJ, Galperin MY, Koonin EV. Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. BMC Evol Biol 2003;3:2.
2. Kunin V, Goldovsky L, Darzentas N, Ouzounis CA. The net of life: Reconstructing the microbial phylogenetic network. Genome Res 2005;15:954–959.
3. Dagan T, Martin W. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. Proc Natl Acad Sci U S A 2007;104:870–875.
4. Halary S, Leigh JW, Cheaib B, Lopez P, Baptiste E. Network analyses structure genetic diversity in independent genetic worlds. Proc Natl Acad Sci U S A 2010;107:127–132.
5. Kloesges T, Popa O, Martin W, Dagan T. Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. Mol Biol Evol 2011;28:1057–1074.
6. Caro-Quintero A, Konstantinidis KT. Inter-phyllum HGT has shaped the metabolism of many mesophilic and anaerobic bacteria. ISME J 2015;9:958–967.
7. Fondi M, Karkman A, Tamminen MV, Bosi E, Virta M, et al. “Every gene is everywhere but the environment selects”: global geology of gene sharing in environmental samples through network analysis. Genome Biol Evol 2016;8:1388–1400.
8. Lawrence JG. Gene transfer in bacteria: Speciation without species? Theor Popul Biol 2002;61:449–460.
9. Guglielmini J, Quintais L, Garciláñ-Barcia MP, de la Cruz F, Rocha EFC. The repertoire of ice in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. PLoS Genet 2011;7:e1002222.
10. Woźniak RAF, Walord MK. Integrative and conjugative elements: Mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol 2010;8:552–563.
11. Johnson CM, Grossman AD. Integrative and conjugative elements (ICes): What they do and how they work. Annu Rev Genet 2015;49:577–601.
12. Delavat F, Miyazaki R, Carraro N, Pradervand N, van der Meer JR. The hidden life of integrative and conjugative elements. FEMS Microbiol Rev 2017;41:512–537.
13. Peters SE, Hobman JL, Strike P, Ritchie DA. Novel mercury resistance determinants carried by INd3 plasmids PMERPH and r391. Mol Gen Genet 1991;228:294–299.
14. Rauch PJ, De Vas WM. Characterization of the novel nisin–sucrose conjugative transposon Tn5276 and its insertion in Lactococcus lactis. J Bacteriol 1992;174:1280–1287.
15. Ravatn R, Studer S, Springael D, Zehnder AJ, van der Meer JR. Chromosomal integration, tandem amplification, and deamplification in Pseudomonas putida F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from Pseudomonas sp. Strain B13. J Bacteriol 1998;180:4360–4369.
16. Beaber JW, Hochhut B, Waldor MK. Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from Vibrio cholerae. J Bacteriol 2002;184:4259–4269.
17. Drenkard E, Ausubel FM. Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation. Nature 2002;416:740–743.
18. Burrus V, Marrero J, Waldor MK. The current ICE age: Biology and evolution of sxt-related integrating conjugative elements. Plasmid 2006;55:173–183.
19. Sullivan JT, Ronsw CW. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-rfNA gene. Proc Natl Acad Sci U S A 1998;95:5145–5149.
20. Dimopoulou ID, Kartal SI, Harding RM, Peto TEA, Crook DW. Diversity of antibiotic resistance integrative and conjugative elements among haemophili. J Med Microbiol 2007;56:838–846.
21. Heather Z, Holden MTG, Steward KF, Parkhill J, Song L, et al. A novel streptococcal integrative conjugative element involved in iron acquisition. Mol Microbiol 2008;70:1274–1292.
22. Kung VL, Ozer EA, Hauser AR. The accessory genome of Pseudomonas aeruginosa. Microbiol Mol Biol Rev 2010;74:621–641.
23. Colombi E, Straub C, Künzel S, Templeton MD, Mc Cann HC, et al. Evolution of copper resistance in the kiwifruit pathogen Pseudomonas syringae pv. actinidiae through acquisition of integrative conjugative elements and plasmids. Environ Microbiol 2017;19:819–832.
24. Guédon G, Libante V, Coluzzi S, Payot S, Leblond-Bourget N. The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. Genes (Basel) 2017;8:337.
25. Larango M, Alexandre A, Oliveira S. Legume growth-promoting rhizobia: an overview on the Mesorhizobium genus. Microb Res 2014;169:2–17.
26. Hackett TL, Terpolilli JJ, Bekuma A, O’Hara GW, Sullivan JT, et al. Assembly and transfer of tripartite integrative and conjugative genetic elements. Proc Natl Acad Sci U S A 2016;113:12268–12273.
27. Greenlon A, Chang PL, Damelew ZM, Muleta A, Carrasquilla-Garcia N, et al. Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. Proc Natl Acad Sci U S A 2019;116:15200–15209.
28. Perry BJ, Sullivan JT, Colombi E, Murphy RJT, Ramsay JP, et al. Symbiosis islands of lotaeae-nodulating Mesorhizobium comprise three radiating lineages with concordant nod gene complements and nodulation host-range groupings. Microb Genom 2020;6.
29. Wang X, Luo Y, Liu D, Wang J, Wei S, et al. Complete genome sequence of the Robinia pseudacacia L. symbiont Mesorhizobium amyorphe CCNWS0123. Stand in Genomic Sci 2018;13:18.
30. Hill Y, Colombi E, Bonello E, Hackett T, Ramsay J, et al. Evolution of diverse effective N2-fixing microsymbionts of Cicer arietinum following horizontal transfer of the Mesorhizobium ciceri CC1192 symbiosis integrative and conjugative element. Appl Environ Microbiol 2020.
31. Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronsw CW. Nodulating strains of rhizobium loti arise through chromosomal symbiotic gene transfer in the environment. Proc Natl Acad Sci U S A 1995;92:8985–8989.
32. Ramsay JP, Sullivan JT, Stuart GS, Lamont IL, Ronsw CW. Excision and transfer of the Mesorhizobium loti R7A symbiosis island requires an integrase INTS, a novel recombination directionality factor RDFS, and a putative relaxase RLKS. Mol Microbiol 2004;52:723–734.
33. Ramsay JP, Sullivan JT, Jambari N, Ortori CA, Heeb S, et al. A LuxRF-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* 2009;73:1141–1155.

34. Ramsay JP, Major AS, Komarovsky VM, Sullivan JT, Dy RL, et al. A widely conserved molecular switch controls quorum sensing and symbiosis island transfer in *Mesorhizobium loti* through expression of a novel activator. *Mol Microbiol* 2013;87:1–13.

35. Ramsay JP, Tester LGL, Major AS, Sullivan JT, Edgar CD, et al. Ribosomal frameshifting and dual-target activation restrict quorum-sensing-activated transfer of a mobile genetic element. *Proc Natl Acad Sci USA* 2015;112:4104–4109.

36. Verdonk CJ, Sullivan JT, Williman KM, Nicholson L, Bastholm TR, et al. Delineation of the integrase-attachment and origin-of-transfer regions of the symbiosis island *ICEMI*SymR7A. *Plasmid* 2019;104:102416.

37. Haskett TL, Terpolilli JJ, Ramachandran VK, Verdonk CJ, Poole PS, et al. Sequential induction of three recombination directionality factors directs assembly of tripartite integrative and conjugative elements. *PLoS Genet* 2018;14:e1007292.

38. Toussaint A, Merlin C, Monchy S, Benotmane MA, Leplae R, et al. The bifhenyl- and 4-chlorobiphenyl-catabolic transposon Tn4371, a member of a new family of genomic islands related to IncP and Ti plasmids. *Appl Environ Microbiol* 2003;69:4837–4845.

39. Kolmogorov M, Yuan J, Lin Y, Pevzner PA, Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.

40. Li H. Minimap²: Pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.

41. Vasar R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 2017;27:737–746.

42. Walker BJ, Abeel T, Shea T, Priest M, Abouelhamed A, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;9:e112963.

43. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, et al. Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 2015;16:294.

44. Song L, Florea L, Langmead B. Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biol* 2014;15:509.

45. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. Spades: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.

46. Prokhorkin M, Pandey A, Moon H, Kim W, Jeon H, et al. Host adaptation and microbial competition drive *Ralstonia solanacearum* phyotype I evolution in the Republic of Korea. *Microb Genom* 2020;6:e000461.

47. Seemann T. PROKKA: Rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.

48. Lehner M, Findt T, Steiner L, Marz M, Stadler PF, et al. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* 2011;12:124.

49. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.

50. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.

51. Schliep KP. PHANGORN: Phylogenetic analysis in *R*. *Bioinformatics* 2011;27:592–593.

52. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90k prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:1–8.

53. de Lajudie PM, Andrews M, Ardley J, Eardly B, Jumas-Bilak E, et al. Minimal standards for the description of new genera and species of rhizobia and agrobacteria. *Int J Syst Evol Microbiol* 2019;69:1852–1863.

54. Eddy SR. Accelerated profile HMM searches. *PLOS Comp Biol* 2011;7:10.

55. Abby SS, Curj Guglielmini J, Néron B, Touchon M, et al. Identification of protein secretion systems in bacterial genomes. *Sci Rep* 2016;6:1–14.

56. Francia MV, Varsaki A, Garcillán-Barcia MP, Latorre A, Drainas C, et al. A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiol Rev* 2004;28:79–100.

57. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local Alignment Search Tool. *J Mol Biol* 1990;215:403–410.

58. Sullivan JT, Trzebiatowski JR, Cruickshank RW, Gouzy J, Brown SD, et al. Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. *J Bacteriol* 2002;184:3086–3095.

59. Reeve W, Sullivan J, Ronson C, Tian R, Bräu L, et al. Genome sequence of the *Lotus corniculatus* microsymbiont *Mesorhizobium loti* strain R88B. *Stand in Genomic Sci* 2014;9:3.

60. Arndt D, Grant JR, Marcus A, Saged T, Pon A, et al. PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16–21.

61. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, et al. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.

62. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012;28:1647–1649.

63. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of Phyml 3.0. *Syst Biol* 2010;59:307–321.

64. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 2006;23:254–267.

65. Didelot X, Wilson DJ. Collalframe: Efficient inference of recombination in whole bacterial genomes. *PLOS Comp Biol* 2015;11:e1004041.

66. Wheeler TJ, Eddy SR. NHMMER: DNA homology search with profile hmm. *Bioinformatics* 2013;29:2487–2489.

67. Sullivan JT, Brown SD, Ronson CW. The *NifA-RpoN* regulon of *Mesorhizobium loti* strain R7A and its symbiotic activation by a novel LacI/GalR-family regulator. *PLoS One* 2013;8:e53762.

68. Domazet-Lošo M, Haubold B. Alignment-free detection of local similarity among viral and bacterial genomes. *Bioinformatics* 2011;27:1466–1472.

69. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–2504.

70. Paradis E, Schliep K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in *r*. *Bioinformatics* 2019;35:526–528.

71. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, et al. EggnoG 5.0: A hierarchical, functional and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019;47:D309–D314.

72. Brynildsrud R, Bohlin J, Scheffer L, Erdholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with scopy. *Genome Biol* 2016;17:238.

73. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, et al. BAGEL4: A user-friendly web server to thoroughly mine riffs and bacteriocins. *Nucleic Acids Res* 2018;46:W278–W281.

74. Biswas A, Staals RHJ, Morales SE, Fineran PC, Brown CM. CRISP-PRDetect: A flexible algorithm to define CRISP PR arrays. *BMC Genomics* 2016;17:356.

75. Bekuma AA, Terpolilli JJ, Tiwari RP, O’Hara G, Howieson J, et al. *Biserrula pelecinus*: A promising forage legume for the central Ethiopian highlands. *Grass Forage Sci* 2021;60:1–11.
Garcillán-Barcia MP, Francia MV, de la Cruz F. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev 2009;33:657–687.

Li PL, Hwang I, Miyagi H, True H, Farrand SK. Essential components of the Ti plasmid TRB system, a type IV macromolecular transporter. J Bacteriol 1999;181:5033–5041.

Li PL, Everhart DM, Farrand SK. Genetic and sequence analysis of the ptrc58 Trb locus, encoding a mating-pair formation system related to members of the type IV secretion family. J Bacteriol 1998;180:6164–6172.

Cho H, Winans SC, TraA, TraC and TraD autorepress two divergent quorum-regulated promoters near the transfer origin of the Ti plasmid of Agrobacterium tumefaciens. Mol Microbiol 2002;63:1769–1782.

Carlton TM, Sullivan JT, Stuart GS, Hutt K, Lamont IL, et al. Ferrichrome utilization in a mesorhizobial population: Microevolution of a three-locus system. Environ Microbiol 2007;9:2923–2932.

Haase J, Kalkum M, Lanka E. TrbK, a small cytoplasmic membrane lipoprotein, functions in entry exclusion of the INCP alpha plasmid rp4. J Bacteriol 1996;178:6720–6729.

Bourdès A, Rudder S, East AK, Poole PS. Mining the Sinorhizobium meliloti transportome to develop fret biosensors for sugars, dicarboxylates and cyclic polyols. PLoS One 2012;7:e43578.

Marchetti M, Capela D, Poincloux R, Benmerad N, Auric M-C, et al. Queuosine biosynthesis is required for Sinorhizobium meliloti-induced cytoskeletal modifications on HeLa cells and symbiosis with Medicago truncatula. PLoS One 2013;8:e56043.

Ronson CW, Lyttleton P, Robertson JG. C(4)-dicarboxylate transport mutants of Rhizobium trifolii form ineffective nodules on Medicago sativa. Proc Natl Acad Sci U S A 1981;78:4284–4288.

Boesten L, Batut J, Boistard P. DctBD-dependent and -independent expression of the Sinorhizobium (Rhizobium) meliloti C4-dicarboxylate transport gene (dctA) during symbiosis. MPMI 1998;11:878–886.

Zehner S, Schoder G, Wenzel M, Lang K, Göttert M. Expression of the Bradyrhizobium japonicum type III secretion system in legume nodules and analysis of the associated TTS box promoter. Mol Plant Microbe Interact 2008;21:1087–1093.

Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. BMC Microbiol 2005;5:1–13.

Sugawara M, Takahashi S, Umehara Y, Iwano H, Tsurumaru H, et al. Variation in bradyrhizobial NOPP effector determines symbiotic incompatibility with Rj2-soybeans via effector-triggered immunity. Nat Commun 2018;9:1–12.

Cianciootto NP, White RC. Expanding role of type II secretion in bacterial pathogenesis and beyond. Infect Immun 2017;85.

Burrous V, Waldor MK. Formation of SXT Tandem arrays and SXT-R391 hybrids. J Bacteriol 2004;186:2636–2645.

Wozniak RA, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, et al. Comparative ICE genomics: Insights into the evolution of the sxt/r391 family of ieces. PLoS Genet 2009;5:12.

Spagnoletti M, Ceccarelli D, Rieu X, Fondi M, Taviani E, et al. Acquisition and evolution of SXT-R391 integrative conjugative elements in the seventh-pandemic Vibrio cholerae lineage. mBio 2AD;5:e01356-01314.

Bryant D, Moutlon V. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 2004;21:255–265.

Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. Genetics 2006;172:2665–2681.

Kuykendall LD, Shao JY, Hartung JS. “Ca. Liberibacter asiaticus” proteins orthologous with pSymA-encoded proteins of Sinorhizobium meliloti: hypothetical roles in plant host interaction. PLoS One 2012;7:e38725.

Makarova KS, Wolf YI, Alkhnbash OS, Costa F, Shah SA, et al. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol 2015;13:722–736.

Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol 2017;37:67–78.

Biswas A, Gagnon JN, Brouns SJJ, Fineran PC, Brown CM. CRISPRTarget: Bioinformatic prediction and analysis of CRRNA targets. RNA Biol 2013;10:817–827.

Li X, Wang H, Tong W, Feng L, Wang L, et al. Exploring the evolutionary dynamics of Rhizobium plasmids through bipartite network analysis. Environ Microbiol 2020;22:934–951.

Cury J, Oliveira PH, de la Cruz F, Rocha EPC. Host range and genetic plasticity explain the coexistence of integrative and extrachromosomal mobile genetic elements. Mol Biol Evol 2018;35:2230–2239.

Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol 2005;3:711–721.