# Population genomics of Australian indigenous *Mesorhizobium* reveals diverse nonsymbiotic genospecies capable of nitrogen-fixing symbioses following horizontal gene transfer

Elena Colombi[1,2]†, Yvette Hill[3], Rose Lines[4], John T. Sullivan[5], MacLean G. Kohlmeier[3], Claus T. Christophersen[4,6,7], Clive W. Ronson[5], Jason J. Terpolilli[3] and Joshua P. Ramsay[1,2,*]

### Abstract

Mesorhizobia are soil bacteria that establish nitrogen-fixing symbioses with various legumes. Novel symbiotic mesorhizobia frequently evolve following horizontal transfer of symbiosis-gene-carrying integrative and conjugative elements (ICESyms) to indigenous mesorhizobia in soils. Evolved symbionts exhibit a wide range in symbiotic effectiveness, with some fixing nitrogen poorly or not at all. Little is known about the genetic diversity and symbiotic potential of indigenous soil mesorhizobia prior to ICESym acquisition. Here we sequenced genomes of 144 *Mesorhizobium* spp. strains cultured directly from cultivated and uncultivated Australian soils. Of these, 126 lacked symbiosis genes. The only isolated symbiotic strains were either exotic strains used previously as legume inoculants, or indigenous mesorhizobia that had acquired exotic ICESyms. No native symbiotic strains were identified. Indigenous nonsymbiotic strains formed 22 genospecies with phylogenomic diversity overlapping the diversity of internationally isolated symbiotic *Mesorhizobium* spp. The genomes of indigenous mesorhizobia exhibited no evidence of prior involvement in nitrogen-fixing symbiosis, yet their core genomes were similar to symbiotic strains and they generally lacked genes for synthesis of biotin, nicotinate and thiamine. Genomes of nonsymbiotic mesorhizobia harboured similar mobile elements to those of symbiotic mesorhizobia, including ICESym-like elements carrying aforementioned vitamin-synthesis genes but lacking symbiosis genes. Diverse indigenous isolates receiving ICESyms through horizontal gene transfer formed effective symbioses with *Lotus* and *Biserrula* legumes, indicating most nonsymbiotic mesorhizobia have an innate capacity for nitrogen-fixing symbiosis following ICESym acquisition. Non-fixing ICESym-harbouring strains were isolated sporadically within species alongside effective symbionts, indicating chromosomal lineage does not predict symbiotic potential. Our observations suggest previously observed genomic diversity amongst symbiotic *Mesorhizobium* spp. represents a fraction of the extant diversity of nonsymbiotic strains. The overlapping phylogeny of symbiotic and nonsymbiotic clades suggests major clades of *Mesorhizobium* diverged prior to introduction of symbiosis genes and therefore chromosomal genes involved in symbiosis have evolved largely independent of nitrogen-fixing symbiosis.

## DATA SUMMARY

All genome assemblies and sequence reads are available in NCBI BioProject PRJNA549135. Amplicon sequencing reads from soil DNA are available in PRJNA841043.

**Impact Statement**

Rhizobia are soil bacteria that can establish nitrogen-fixing symbioses with legumes. In rhizobia of the genus *Mesorhizobium*, genes for symbiosis are often located on mobile integrative and conjugative elements (ICESyms) that horizontally transfer to indigenous mesorhizobia in soils. However, symbiosis gene transfer does not always result in an effective nitrogen-fixing symbiosis, suggesting the genetic background impacts symbiotic performance. Since mesorhizobia are generally isolated from plant root nodules, very little is known about the diversity of indigenous soil *Mesorhizobium* prior to their acquisition of symbiosis genes. In this work we isolated 126 nonsymbiotic *Mesorhizobium* spp. from Australian soils and sequenced their genomes. Nonsymbiotic *Mesorhizobium* exhibited similar core and pangenome features to symbiotic isolates and overlapping phylogenetic diversity. Transfer of ICESyms to diverse nonsymbiotic *Mesorhizobium* spp. produced symbiotic strains with varying nitrogen-fixing efficiency, including some that were ineffective symbionts. Efficacy of symbiosis was not correlated with the lineage of the *Mesorhizobium* host chromosome, suggesting sporadic genetic incompatibilities in recipient genomes augment symbiotic effectiveness. These findings broaden our appreciation of *Mesorhizobium* diversity and reveal Australian soil *Mesorhizobium* populations are dominated by nonsymbiotic isolates that overlap the phylogenetic diversity of symbiotic isolates isolated internationally.

## INTRODUCTION

Rhizobia are a group of phylogenetically diverse soil bacteria capable of forming symbiosis with legumes through the formation of root nodules. Within root nodules, rhizobia fix atmospheric nitrogen ($N_2$) to produce $NH_3$, which is secreted to the plant and assimilated. In return, the host-plant supplies carbon, nutrients and a niche for the microsymbiont. Currently, at least 14 genera within the *Alphaproteobacteria* and *Betaproteobacteria*, are known to establish a legume symbiosis [1]. *Mesorhizobium* strains can establish nitrogen-fixing symbioses with legume species from temperate, tropical, sub-tropical and arctic areas [2]. These include cultivated legumes such as *Lotus* spp., *Biserrula pelecinus* [3, 4] and *Cicer* spp., [5] and a variety of uncultivated legumes [6–10] including *Robinia* [11], *Carmichaelia* [7], *Acacia* and *Sesbania* [12].

Bacterial genes that contribute to rhizobia-legume symbioses include nodulation (*nod*, *noe* and *nol*) and nitrogen fixation (*nif* and *fix*) genes. In *Mesorhizobium* spp., symbiosis genes are generally present on integrative and conjugative elements (ICEs) [13], which are chromosomally integrating mobile genetic elements (MGEs) that can excise from the chromosome and transfer using ICE-encoded conjugation machinery [14]. The first-described 'symbiosis ICE' (ICESym), ICE*Ml*Sym[R7A], is a 502 kb ICE enabling nitrogen-fixing symbiosis with *Lotus corniculatus*. Transfer of ICE*Ml*Sym[R7A] converts nonsymbiotic *Mesorhizobium* spp. into nitrogen-fixing symbionts of *L. corniculatus* [15]. Related ICESyms convert indigenous Australian *Mesorhizobium* spp. into microsymbionts of *B. pelecinus* [3, 4] and *Cicer arietinum* (chickpea) [16]. ICESyms form a subfamily within a much larger family of ICEs distributed throughout the proteobacteria, and are the largest of all characterized bacterial ICEs at ~350–850 kb [13]. ICESym-encoded site-specific recombinases target conserved chromosomal loci such as tRNA genes for integration. Tripartite ICESyms possess three distinct site-specific recombinases and following initial integration, two further inversions between the ICE and chromosome separate the tripartite ICE genome into three regions ($\alpha$, $\beta$ and $\gamma$) [17, 18]. Both monopartite and tripartite *Mesorhizobium* ICESyms have likely evolved from a common ancestor and carry related core genes for quorum sensing, conjugation, excision and regulation [13].

In Australia and New Zealand (NZ), where cultivated legumes have been introduced following European colonization [19, 20], soils often lack indigenous compatible microsymbionts [21–23], so inoculation with compatible rhizobia is a common agricultural practice [19, 24]. Several studies reveal rhizobia re-isolated from nodules of inoculated legumes are often genetically distinct to the inoculant strain [23, 25, 26] due to gene transfer from the inoculant to indigenous rhizobia. In some cases, ICESym-transfer to *Mesorhizobium* spp. in Australian soils produce exconjugants that fix nitrogen sub-optimally [27] or not at all [28], so the genetic background of indigenous *Mesorhizobium* spp. impacts the efficacy of legume inoculation. The genetic composition of the Australian indigenous *Mesorhizobium* spp. prior to ICESym acquisition is unknown. However, diverse nonsymbiotic mesorhizobia can be isolated from NZ soils and can acquire ICE*Ml*Sym[R7A] both *in situ* and *in vitro* [29].

Few studies have focussed directly on populations of nonsymbiotic rhizobia; however, most reveal diverse nonsymbiotic rhizobia coexist with symbiotic rhizobia in soils. Early studies demonstrated nonsymbiotic strains related to *Rhizobium leguminosarum* bv. *phaseoli* were converted into effective legume symbionts following transfer of symbiosis plasmids [30–32]. Diverse nonsymbiotic *Rhizobium* spp. coexist in rhizospheres with symbiotic rhizobia but symbiosis-plasmid transfer seems restricted between closely related species [33]. The diversity of *Bradyrhizobium* species in North American soils suggests populations are dominated by a few abundant lineages comprising both symbiotic and nonsymbiotic bacteria [34, 35]. In *Bradyrhizobium* spp., symbiosis genes are present on chromosomally integrated elements [36] and while many appear to have been independently acquired, many may have been subsequently lost from the same lineages [34]. Symbiotic and nonsymbiotic *Ensifer* (syn. *Sinorhizobium*) spp.

appear to form monophyletic symbiotic (*Sinorhizobium*) and nonsymbiotic (*Ensifer*) clades, with few exceptions, but these are now recognized as two different genera [37]. Nonsymbiotic isolates are also present in the *Sinorhizobium* genus and symbiosis plasmids have been acquired in several independent events [38]. Together, these observations highlight the complex interplay between MGEs carrying symbiosis genes and their bacterial hosts.

Members of the genus *Mesorhizobium* lacking genes for nitrogen-fixing legume symbiosis have been isolated from nodules [39], sea water [40], activated sludge used in wastewater treatment [41], photosynthetic amoeba [42], *Arabidopsis* root microbiota [43] and even human blood [44]. Australian soils appear devoid of *Mesorhizobium* capable of nodulating introduced legumes but indigenous *Mesorhizobium* spp. acquiring ICESyms from inoculant strains readily evolve. It is unclear if most indigenous *Mesorhizobium* can become nitrogen-fixing symbionts upon ICESym acquisition or if these exconjugants represent a subset of the population. In this work, we isolated 144 *Mesorhizobium* directly from cultivated and uncultivated Australian soils. Of these, 126 strains were nonsymbiotic *Mesorhizobium* spp. (NS-meso). Genome comparisons suggested NS-meso genomes are similar in composition and diversity to symbiotic isolates. We propose the NS-meso isolated in this study reflect the archetypal diversity present in the genus prior to the evolution of nitrogen-fixing symbiosis and that most NS-meso have an 'innate' capacity to evolve into nitrogen-fixing symbionts following ICESym acquisition.

## METHODS

Detailed methods are provided in the supplementary material. Software commands used are available at https://github.com/EC-Rufina/NS-Meso and https://github.com/EC-Rufina/MesICE.

### Isolation of *Mesorhizobium* from soil and sequencing

Four-gram soil samples (Table S1 available in the online version of this paper) were washed with sterile water, and dilutions in 10 mM $MgCl_2$ were plated on G/RDM [45, 46] containing cycloheximide (100 µg ml$^{-1}$), quintozene (5 µg ml$^{-1}$), nystatin (50 µg ml$^{-1}$), fosfomycin (50 µg ml$^{-1}$), nicotinate (1 µg ml$^{-1}$), biotin (20 ng ml$^{-1}$) and thiamine (1 µg ml$^{-1}$), and incubated at 28 °C for 10 days. Single colonies were PCR-screened with *Mesorhizobium*-specific 16S-rRNA primers (Table S2). Genome extractions for Oxford Nanopore Technologies MinION and PacBio RSII sequencing were performed as previously described [47]. For assembly of complete genomes with MiniION or PacBio reads, long-reads alone were first assembled with Flye 2.7.1 [48] and then assemblies were polished with Illumina reads as previously described [49]. Genome extractions for Illumina NextSeq sequencing were performed with FavorPrep Tissue Genomic DNA Extraction Mini Kit (Favorgen). Illumina adapters were removed with Trimmomatic 0.39 [50], reads were corrected using Lighter (v1.1.1) [51]. Genomes sequenced only with Illumina reads were assembled with SPAdes 3.5.0 [52].

### Genome analyses

Phylogenies were constructed as described previously [13], except PRANK [53] was used for alignments. Trees were plotted with ggtree [54]. Pangenomes were evaluated with Proteinortho [55]. NodABC and NifHDK searches were performed as previously described [38]. *Mesorhizobium* and *Rhizobium leguminosarum* Rlv3841 genes involved in symbiosis were initially searched for in *M. japonicum* R7A with blastp, and Proteinortho [55] was used to identify orthologs of R7A proteins in other *Mesorhizobium* genomes. Identification of MGEs was performed as previously described [13]. Each MGE was queried against other genomes with blastn and considered present in other strains if >80% pairwise nucleotide identity and >80% coverage were shared.

### Microbiome analysis

Soil DNA was isolated using DNeasy PowerSoil Kit (QIAGEN). Barcoding qPCRs with 16S V4 and *atpD* primers (Table S3), and sequencing were carried out as in [56]. Reads were processed using mothur 1.32.1 software as per the MiSeq standard operating procedure [57]. PCR errors were accounted for by removing any sequences with ambiguous bases. For the 16S amplicon sequencing, reads longer than 275 bp were removed, for the *atpD* amplicon sequencing reads longer than 300 bp and reads shorter than 294 bp were discarded. Potential chimaeras were identified and removed using the mothur chimaera.vsearch command with default parameters.

### ICE*Ml*Sym$^{R7A}$ transfer experiments

ICE*Ml*Sym$^{R7A}$ was conjugated from strain R7A* as in [49]. Briefly, donor and recipient strains were mixed and spotted on a TY agar plate and incubated at 28 °C for 24 h. Exconjugants were selected on G/RDM containing tetracycline (50 µg ml$^{-1}$) or neomycin (200 µg ml$^{-1}$), without supplemented nicotinate, biotin or thiamine.

### Legume symbioses

Pots containing sterile sand mix or soil collected from Badgingarra were sown with germinated *B. pelecinus* seeds as previously described [58]. Pots were sown with four germinated seeds and four replicates for each treatment. Plants were grown for 69 days.
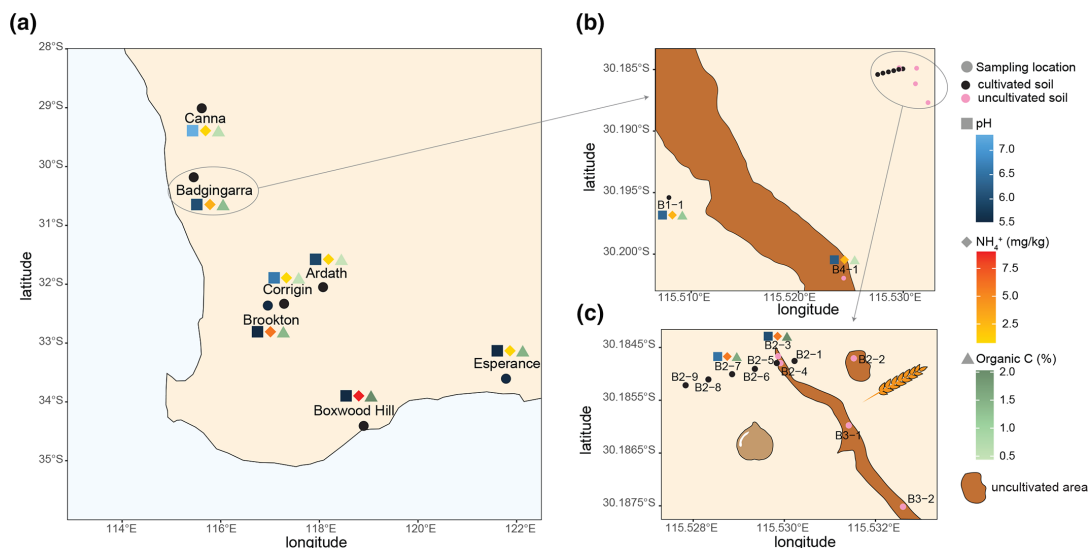
**Fig. 1.** Map of soil sites. (a) Sampling sites across western Australia. (b),(c) Location of the sampling sites in Badgingarra. Soils were analysed for pH, and content of $NH_4^+$ and organic C, results are indicated by squares, diamonds or triangles coloured as indicated in the key on the right. The chickpea seed and the ear of wheat cartoons in (c) indicate that the paddocks were cultivated with these crops at the time of sampling.

Nodule collection and treatments were conducted as previously described [59, 60]. Shoots were dried at 60 °C and desiccated prior to weighing. *Lotus japonicus* ecotype Gifu [61] was used to test the symbiotic capacity of R7A and exconjugants as previously described [62].

## RESULTS

### Diverse nonsymbiotic mesorhizobia can be isolated from Australian soils

Soil was collected from a 0–20 cm profile in areas of cultivated and uncultivated land in western Australia (Figure 1, Table S1). Thirteen soil samples were taken from a farm in Badgingarra (sites B1-1 to B4-1), including paddocks subjected to rotational cropping of grains and legumes, and neighbouring uncultivated areas containing native *Eucalyptus* spp., *Daviesia* spp., and various weed grass species (Fig. 1b, c). An additional six samples across southern-western Australia were taken from cultivated farmland with diverse soil characteristics (Fig. 1a).

To isolate *Mesorhizobium* from soil and reduce contaminating organisms, a modified semi-selective glucose rhizobium defined medium (G/RDM [45, 46]) was developed (see Methods for complete medium recipe). Antimicrobial susceptibility screening of *Mesorhizobium* strains in our collection revealed all were highly resistant to fosfomycin. We suspected this was partly due to strains carrying *fosX*. Deletion of *fosX* in *M. japonicum* R7A only reduced the fosfomycin resistance from >1 g ml$^{-1}$ to a minimum inhibitory concentration of ~300 mg ml$^{-1}$. Deletion of *fosX* in *M. ciceri* CC1192 reduced fosfomycin resistance from >900 mg ml$^{-1}$ to a minimum inhibitory concentration of ~100 mg ml$^{-1}$. This suggests that *Mesorhizobium* species are resistant to fosfomycin through various molecular mechanisms [45, 46]. A final concentration of 50 mg ml$^{-1}$ fosfomycin was used in the isolation medium along with vitamins nicotinate, biotin and thiamine and antifungals cycloheximide, quintozene and nystatin.

Soil washes were plated on solid semi-selective medium described above and incubated for 10 days. Approximately 900 individual colonies with mucoid and convex morphology were PCR-screened using *Mesorhizobium*-specific primers (Table S2), and 150 isolates were selected for short-read genome sequencing. 16S rRNA-gene sequences confirmed 144 isolates were *Mesorhizobium* spp. (Table S3). Of these, 126 were nonsymbiotic *Mesorhizobium* (NS-meso) and only 18 isolates carried nodulation (*nodABC*) and nitrogenase (*nifHDK*) genes. Of these, four strains shared 99.9% average nucleotide identity (ANI) across their entire genome with the commercial chickpea inoculant *M. ciceri* CC1192. The remaining 14 isolates were not identified as re-isolated commercial inocula, however, BLASTN searches using ICESyms (as previously described [13]) revealed they were NS-mesos that had acquired the ICESym of CC1192 (one isolate) or the ICESym of the *Biserrula* inoculant strain *M. ciceri* bv. *biserrulae* WSM1497 (Table S3).

Pairwise whole-genome comparisons clustered the NS-meso into 22 genospecies with ANI>95%. To name novel genospecies, we extended a scheme used by Greenlon *et al.* [5]. Briefly, novel genospecies falling in clades (M-number) were designated an M-number and a new letter (e.g. *M.* sp. M2I), and new clades and genospecies were assigned a new M-number and letter (e.g. *M.* sp. M11A). A maximum-likelihood core-gene phylogeny (Fig. 2) was constructed using a representative from each NS-meso
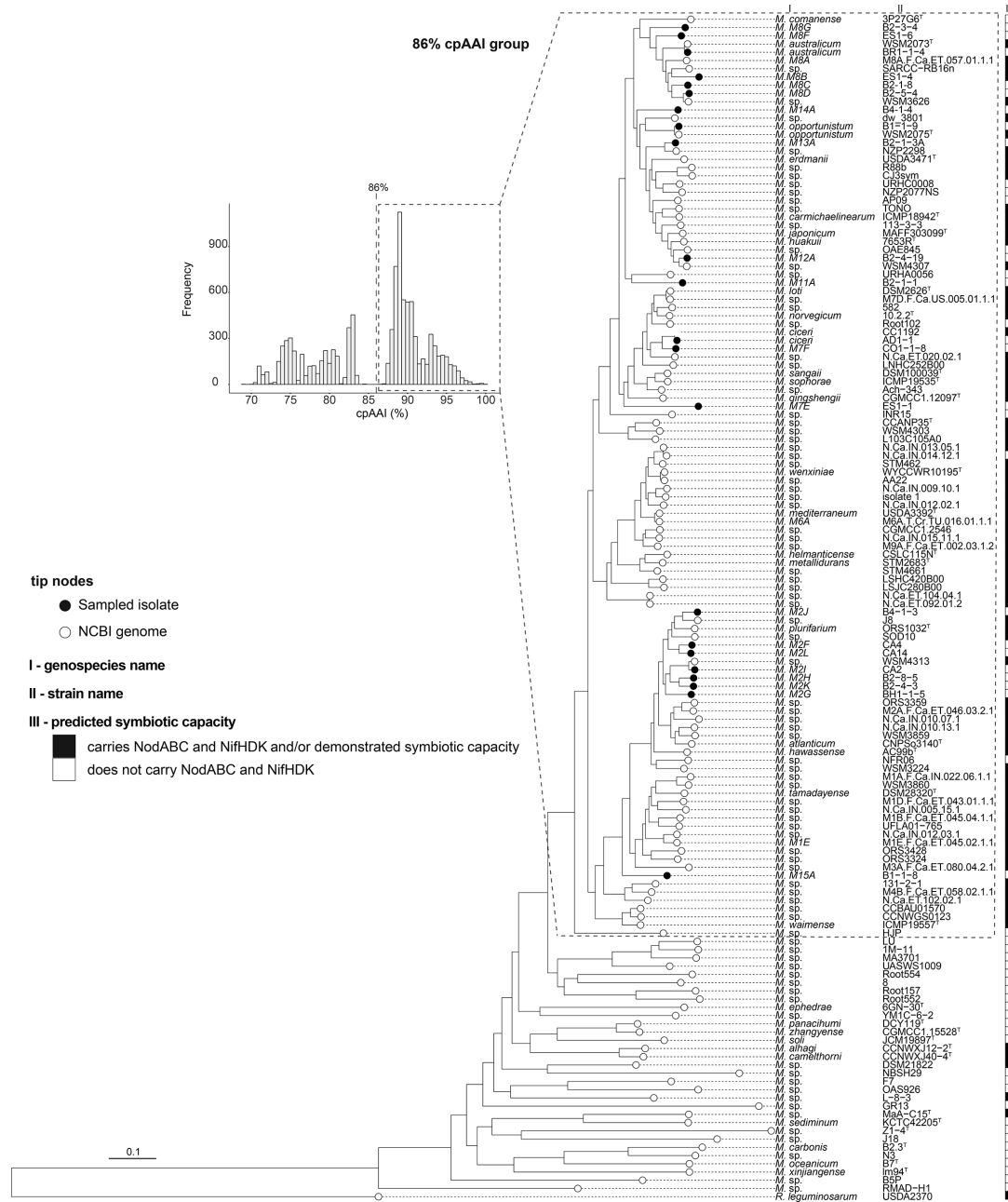
**Fig. 2.** Overview of the *Mesorhizobium* genus. Maximum-likelihood tree constructed on 436 single-copy core genes with RAxML. The tree was rooted with the type-strain *Rhizobium leguminosarum* biovar *viciae* USDA2370 (MRDL00000000). Strains were predicted to be symbiotic if they harboured NodABC and NifHDK genes and/or if they had published symbiotic capacity, and were assumed nonsymbiotic if they lacked these genes or demonstrated symbiotic capacity. Unfilled circles at tip nodes indicate NCBI-downloaded reference genomes (type strains are indicated by [T]), and black-filled circles at tip nodes indicate *Mesorhizobium* isolated in this study. Scale bar indicates substitutions per site. The histogram in the upper-left displays the distribution of core-proteome amino-acid identity (cpAAI) (bin width 0.5 %) calculated from the same 436 core genes used in tree generation, and the dotted line indicates species within the cpAAI>86%.

genospecies together with *Mesorhizobium* spp. genospecies representatives deposited in NCBI (Table S4 and Table S5). This tree segregated most symbiotic *Mesorhizobium* spp. (which included 22 type strains, all symbiotic) from previously isolated nonsymbiotic *Mesorhizobium* spp. (which included 11 nonsymbiotic type strains and two symbiotic type strains). The 'symbiotic cluster' shared a pairwise core-proteome average amino acid identity (cpAAI) of >85–86%, a measure previously deemed suitable to delineate *Rhizobiaceae* genera and other genera [37, 63, 64] (Fig. 2). The distribution of cpAAI scores revealed a clear demarcation between genospecies sharing >87% cpAAI and others sharing <84% cpAAI, suggesting those with less than ~86% cpAAI could
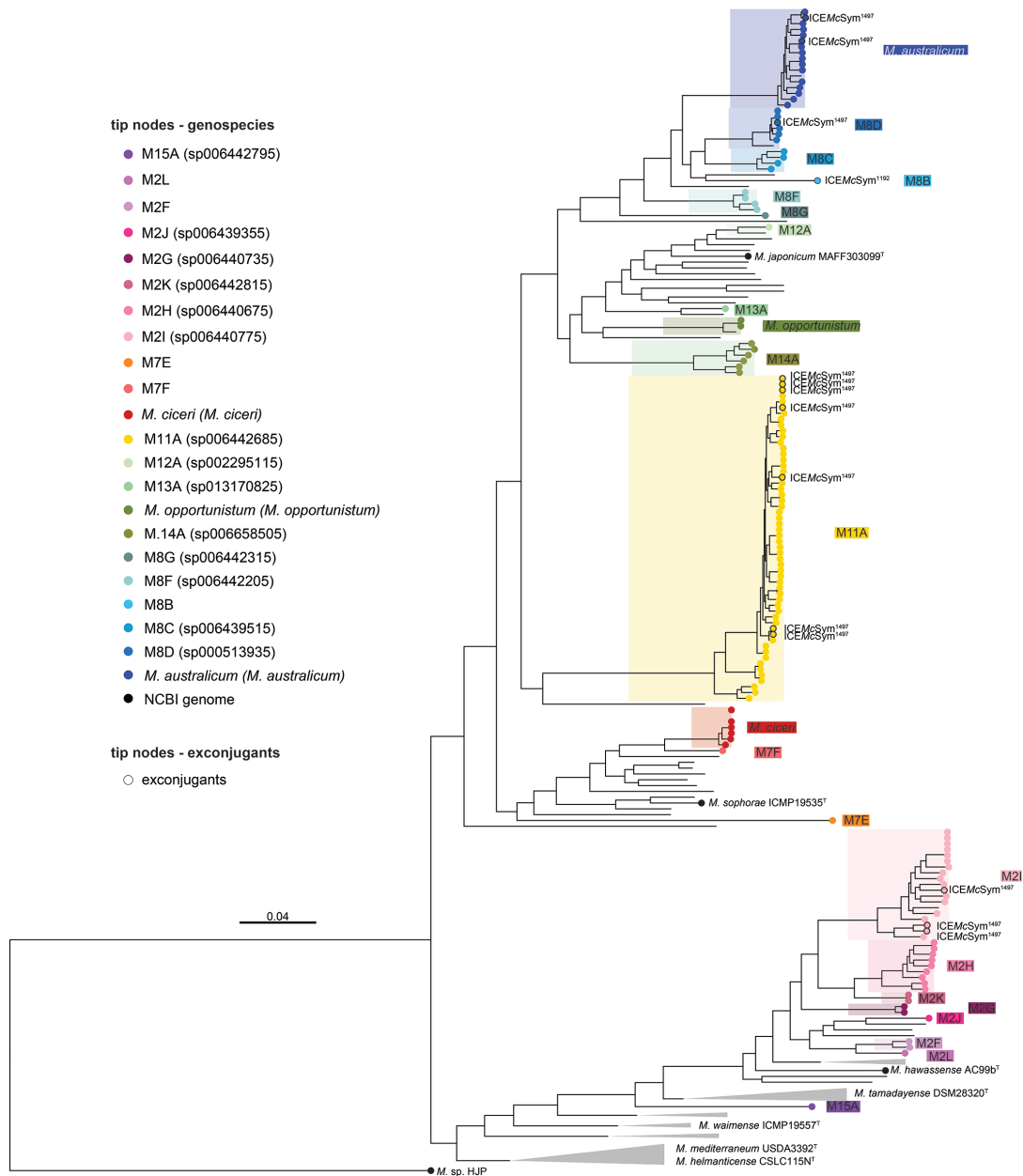
**Fig. 3.** Core-gene phylogeny of *Mesorhizobium* isolated from soil during this study. In total, 1,067 single-copy core-genes were used to build a maximum-likelihood tree using RAxML. The tree was rooted with the *M.* sp. HJP. Coloured tip nodes indicate the strains were isolated in this study, with colour of the tip nodes representing the species the strain belongs to. The genospecies name attributed by the Genome Taxonomy Database is reported in brackets. Coloured tip nodes with a black border indicate that the strain carries an ICESym of a commercial inoculum, the name of the ICESym is indicated adjacent to the tip node. Other genomes were downloaded from NCBI, grey triangles indicate a collapsed clade and black tip nodes highlight selected type strains. A complete tree with all strain labels is presented in Fig. S1. Scale bar indicates substitutions per site.

be considered distinct genera. All NS-meso isolated here grouped within the 'symbiotic' cpAAI>86% cluster. NS-meso were dispersed throughout the cpAAI>86% cluster, and of the 22 genospecies isolated, only seven genospecies had been previously sequenced (Table S6).

To further evaluate relationships between NS-meso and symbiotic strains, a maximum-likelihood core-genes phylogeny was constructed with NS-meso together with previously sequenced *Mesorhizobium* grouped within the cpAAI>86% threshold (Figs 3 and S1). The NS-meso included documented recipients of the tripartite ICESym ICE*Mc*Sym[1271], *M. opportunistum* and *M. australicum* [65], and recipients of ICE*Mc*Sym[1192] [16], genospecies M2I and M12A. Other previously isolated genospecies included those from legume nodules in the Mediterranean basin (*M. ciceri*), NZ, USA, Canada and Japan (genospecies M13A),

and in South Africa (genospecies M8D) (Table S6). Of the 15 novel genospecies, M2F to M2L clustered with members of the previously defined 'clade M2' isolated from *Cicer* spp. in Ethiopia, India and Morocco [5], and with symbiotic strains isolated from Brazil (*M. atlanticum* CNPSo3140), Japan (*M.* sp. J8), Eritrea (*M.* sp. WSM3859), South Africa (*M.* sp. WSM3224) and Senegal (*M. plurifarium* ORS1032). Seven novel genospecies M8B-M8G, and *M. australicum*, clustered with genospecies M8A, previously identified in India [5]. Several genospecies isolated in this study have also been assigned species names by Genome Taxonomy Database (GTDB) and these are indicated in Fig. 3. The most common NS-meso genospecies was M11A, isolated 56 times across five of the seven locations sampled. M11A was particularly abundant in the Badgingarra soil, representing 46.3% of strains isolated. The 14 ICESym exconjugants isolated from soil here were also distributed throughout the NS-meso phylogeny (Fig. 3).

A rarefaction curve generated from genospecies isolated here did not plateau (Fig. S2), and abundance-based coverage estimation suggested eight genospecies remain undiscovered [66]. To further estimate diversity of *Mesorhizobium* spp. in soil, amplicon sequencing of DNA from each of the Badgingarra soil samples was carried out using primers for the 16S rRNA gene and a 330 bp region of the *atpD* gene conserved across the genus (Table S2). Approximately 0.001% of 16S rRNA amplicons (Table S7) and ~2.45% of *atpD* amplicons were classified as *Mesorhizobium* using MOTHUR [67], which estimated 232 operational taxonomic units (OTU) of unclassified *Mesorhizobium* (Table S8). A tree constructed from *atpD* amplicon sequences together with *atpD* extracted from genome sequences (Fig. S3) placed the amplicon-derived *atpD* sequences throughout the tree, suggesting the additional sequences reflected unsampled diversity within the identified clades isolated here. The M11A OTU was present in 10 of the 13 sites, consistent with its frequent isolation. Overall, these observations suggest that while culture-based sampling underestimated species-level diversity, the phylogenetic distribution of isolated samples correlated well with amplicon sequences, suggesting the culture-based sampling strategy was relatively unbiased.

## Soil and nodule-isolated ICE*Mc*Sym[1497] exconjugants are capable of nitrogen-fixing symbioses with *Biserrula*

Previous examination of Australian indigenous *Mesorhizobium* spp. carrying ICE*Mc*Sym[1192] transferred from chickpea inoculant CC1192, or ICE*Mc*Sym[1271] from the *Biserrula* inoculant WSM1271, indicated ICE*Mc*Sym[1192] exconjugants can be as effective as CC1192 in nitrogen fixation [16, 68] but most ICE*Mc*Sym[1271] exconjugants perform poorly compared to WSM1271, or do not fix nitrogen at all (Fix[-]) [69]. With our sampling, we isolated 13 indigenous strains that had acquired the tripartite ICE*Mc*Sym[1497] of *M. ciceri* WSM1497, the present commercial inoculant for *B. pelecinus* in Australia. Seven were of the prevalent M11A lineage, two were *M. australicum*, three were genospecies M2I and one was M8D. We isolated eight further ICE*Mc*Sym[1497] exconjugants from *B. pelecinus* cv. Casbah plant nodules grown in glasshouse conditions with the Badgingarra soil samples (and additionally sequenced their genomes, Table S9). Of these, seven isolates were genospecies M11A, while one was a strain of *M. opportunistum*. Seventeen soil or nodule-isolated exconjugants were tested on *B. pelecinus* cv. Casbah. All induced nodule formation and increased plant dry weights to varying degrees compared to N- controls. Only one strain of genospecies M11A and one M2I strain induced nodule formation but exhibited a Fix[-] phenotype (Fig. 4a, Table S10). Thus, these experiments confirmed our previous observations that there is a wide range in symbiotic effectiveness amongst individual exconjugants carrying the same ICESym. These data also clarified that while the *Mesorhizobium* strain genetic background played an important role, the lack of symbiotic effectiveness sporadically observed for individual isolates was not associated with them belonging to a particular genospecies.

During isolation of *Mesorhizobium* from *B. pelecinus* cv. Casbah nodules described above, we isolated an additional *M. opportunistum* strain, B283B1A, which carried an ICESym distinct from those previously characterized (Table S9). Further interrogation of the farm-site history revealed it was inoculated in 1996 with WSM1558, a strain isolated in Italy for consideration as a commercial inoculant [70]. The genome of WSM1558 was sequenced to completion using long- and short-read sequencing (CP097252). B283B1A genome contigs were aligned with the completed WSM1558 genome, which revealed it was near-identical (99.99% coverage, 99.99% nucleotide identity), suggesting that it was a descendant of WSM1558 introduced 25 years ago. Unlike *M. opportunistum* WSM2075 (harbouring ICE*Mc*Sym[1271]), which forms Fix[-] *B. pelecinus* nodules [65], *M. opportunistum* B283B1A formed an effective symbiosis with *B. pelecinus* cv. Casbah (Fig. 4a).

## Phylogenetically diverse NS-meso can form effective symbiosis with *Lotus japonicus* following laboratory transfer of ICE*Ml*Sym[R7A]

To further explore the capacity of NS-meso to become legume symbionts, we transferred ICE*Ml*Sym[R7A] from strain *M. japonicum* R7A* [49] to NS-meso strains in laboratory conjugation experiments. ICE*Ml*Sym[R7A] was successfully introduced into six NS-meso belonging to genospecies *M. opportunistum*, *M. ciceri*, *M. australicum*, M11A and M2K (Table S11). The ICE*Ml*Sym[R7A] exconjugants were tested on *L. japonicus* Gifu (Fig. 4b and c). While none of the NS-meso strains lacking ICE*Ml*Sym[R7A] were able to induce nodule formation, the same strains carrying ICE*Ml*Sym[R7A] exhibited similar nodulation kinetics to R7A. Exconjugants of B1-9, B2-3-2 and BR1-1-3 produced plants with weights not statistically different from R7A. One of the two *M. australicum* exconjugants exhibited a Fix[-] phenotype (Fig. 4c). As observed in the above experiments with *Biserrula,* these data support the notion that most NS-meso genospecies have an innate capacity to form nitrogen-fixing symbioses following ICESym acquisition and that Fix[-] exconjugants arise sporadically in several genospecies.
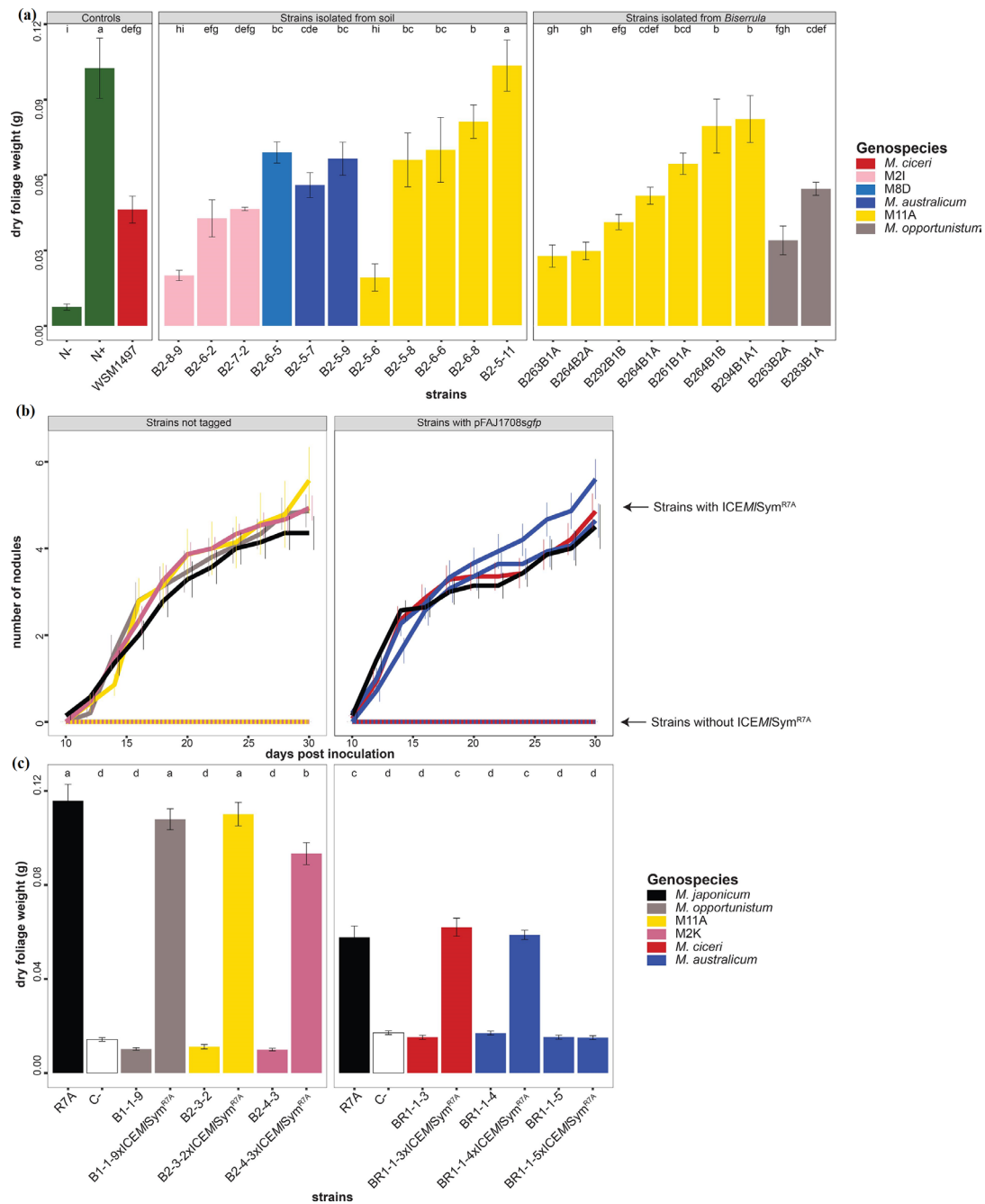
**Fig. 4.** Legume symbioses with NS-meso exconjugants carrying ICESyms. (a) *B. pelecinus* cv. Casbah dry foliage weight at 69 days post-inoculation. Uninoculated (N−) and N-fed (N+) (supplied as KNO$_3$) plants were included as negative and positive controls, respectively. Treatments are shown with standard errors of the means, and the treatments that share a letter are not significantly different according to the least significant difference test ($P \leq 0.05$). (b) Number of nodules developed by the *L. japonicus* Gifu plants grown in test tubes. (c) *L. japonicus* Gifu foliage weight at 30 days post-inoculation. The experiment was performed with 15 replicates and repeated three times. Treatments are shown with standard errors of the means, and treatments that share a letter are not significantly different according to the least significant difference test ($P \leq 0.01$).

## NS-meso and symbiotic *Mesorhizobium* have similar core genomes

The above experiments suggest most NS-meso have a capacity to form nitrogen-fixing symbioses with legumes following ICESym acquisition. This implies that genes not encoded by ICESyms, but nonetheless essential for symbiosis, are highly conserved on the *Mesorhizobium* chromosome. Pangenome analysis of the NS-meso genomes estimated a pangenome of 30,249 genes and a core genome (defined as genes present in all the strains) of 2,469 protein-coding genes. This is significantly higher than our previous

estimate of 1,670 core genes for 41 closed and completed genome sequences of the genus *Mesorhizobium* [13]. However, that analysis included distantly related nonsymbiotic members outside the >86% cpAAI threshold. For a more direct comparison, the 'symbiotic' pangenome was recalculated for 19 completely sequenced nitrogen-fixing symbiotic mesorhizobia within the 86% cpAAI threshold with their ICESym sequences removed, which produced a core genome estimate of 3,149. Core genomes of NS-meso and 19 symbiotic mesorhizobia overlapped by 74%, yet 813 genes considered core in symbiotic strains were not core in NS-meso genomes.

To gauge the likelihood that the average NS-meso strain had a genetic capacity to become a legume symbiont following ICESym acquisition, we searched genomes for genes involved in nodulation and nitrogen-fixation in *Mesorhizobium* spp. (Table S12), and for homologues of proteins involved in symbiosis and rhizosphere/host colonization identified from *R. leguminosarum* bv. *viciae* Rlv3841 insertion-sequencing experiments [71] (Table S13). This revealed that, aside from symbiosis-associated genes typically found on ICESyms, genes involved in symbiosis that were defined as core genes in the 19 symbiotic strains were also core genes in NS-meso genomes. Of ~600 symbiosis-associated genes, only 26 were classified as core genes in symbiotic strains but not in NS-meso (Tables S12 and S13). Seventeen of these were absent in symbiotically effective exconjugants identified here (Fig. 4), suggesting they were not essential for symbiosis. The remaining nine core genes absent in NS-meso were homologues of Rlv3841 proteins with putative functions in thiamine-binding, nitrate/sulfonate/bicarbonate transport, haem-export, transcriptional regulation, a maltose O-acetyltransferase, an amido-phosphoribosyl-transferase, a TetR/AcrR-family transcriptional-regulator and a N-acetyltransferase. Overall, these analyses indicated genes likely critical for nitrogen fixation and symbiosis that are not present on ICESyms are generallypresent in the genomes of NS-meso.

## Characterization of the NS-meso mobilome and identification of archetypal members of the ICESym family

We previously characterized the mobilomes of 41 completely sequenced *Mesorhizobium* strains, and revealed that ICEs and integrative and mobilizable elements (IME) are predominant over plasmids and bacteriophages. Here we identified MGEs in NS-meso genomes by searching for conjugative relaxase (MOB) proteins using hidden Markov models and inspecting nearby sequences to identify plasmids, ICEs and IMEs [13]. The genomes of B2-1-1 (M11A), B1-1-8 (M15A), B2-8-5 (M2H), B2-1-8 (M8C) and B4-1-4 (M14A) were completed with long-read PacBio and/or ONT MinION sequencing (Table S3) to facilitate MGE demarcation (Table S14). We identified 298 MOB-encoding genes in the NS-meso genomes (11 strains of five different genospecies did not harbour any MOB gene) (Table S15). Excluding ICESyms of commercial inoculants, 34 ICEs and 35 plasmids were identified. Ninety-four IMEs were related to IME*Ml*[R88b]-1, a common IME identified in numerous symbiotic strains [13]. Novel IMEs were identified in CO1-1-4, AD1-1-1 and B2-4-6 (Table S15). Overall, 135 contigs with a MOB-encoding gene could not be accurately categorized due to fragmentation of sequence contigs. Each recognized MGE and each contig carrying a MOB gene was compared with all other NS-meso genomes to evaluate their distribution. This revealed individual MGEs were largely restricted to a single genospecies, except for ICESyms and IMEMsp.[CA2] (Figure S4), which was present in M2I and M2L strains isolated from the same soil sample. IMEMsp.[CA2] carries putative heat-shock response and circadian oscillator KaiBC genes [72] and a *traACD* mobilisation locus related to that of IME*Ml*[R88b]-1 [73].

All ICEs identified here and identified previously in *Mesorhizobium* strains, carry related genes for conjugation and regulation, typified by the presence of unique genes flanking the type IV secretion system (*trb*-gene) cluster (*msi031* and *msi021*) and conserved genes for regulation and DNA transfer, *fseA*, *rdfS* and *rlxS* [13]. We identified 76 *msi031-trb-msi021*-gene clusters (Table S16). Trees were constructed from alignments of the *msi031-trb-msi021*-gene clusters together with clusters previously identified on ICEs and plasmids in *Mesorhizobium* [13]. The various NS-meso *msi031-trb-msi021*-gene clusters were interspersed throughout the tree amongst previously identified *msi031-trb-msi021* clusters (Fig. S5). Interestingly, three of the NS-meso ICE *msi031-trb-msi021*-gene clusters grouped with the ICESym clusters (Figs 5A and S5), and like all other ICESyms (but not other ICEs/plasmids in *Mesorhizobium*), lacked the *trbK* gene. This indicated these ICEs shared a recent common ancestor with the ICESyms despite them lacking symbiosis genes. One was located adjacent to the *phe*-tRNA gene in B4-1-3 (ICE*Msp*[B4-1-3]-1) and another adjacent to the *met*-tRNA gene in CA8 (ICE*Msp*[CA8]). Both sites are documented integration sites for ICESyms. The third ICE, ICE*Msp*[B4-1-4], was revealed to be the first example of a tripartite ICE lacking symbiosis genes.

## A tripartite ICE lacking symbiosis genes likely shares a common ancestor with the ICESym family

ICE*Msp*[B4-1-4] (Fig. 5) is integrated into the same three chromosomal attachment sites (*phe*-tRNA, *guaA* and *met*-tRNA) as all identified tripartite ICESyms. Despite lacking the many genes required for symbiosis, ICE*Msp*[B4-1-4] was larger than the median length for tripartite ICESyms (~560 kb) at 626 kb. ICE*Msp*[B4-1-4] carried two gene clusters likely involved in the broad-specificity catabolism of phosphonates [74] and a plethora of metabolic enzymes with predicted involvement in fatty acid degradation, metabolism of butanoate, aspartate, asparagine, glutamate and proline.

Like other ICEs in the genus *Mesorhizobium*, ICE*Msp*[B4-1-4] carries homologues of *rdfS* (required for ICE excision) and *fseA*, which encodes the transcriptional activator of *rdfS*, FseA. Like other ICESyms, ICE*Msp*[B4-1-4] encodes *luxRI*-family quorum-sensing genes directly upstream of the *fseA* gene. However, the ICE*Msp*[B4-1-4]*luxRI*-family genes are not orthologous with those on ICESyms (*traR1/traR2* and *traI1/traI2*). Comparisons revealed the ICE*Msp*[B4-1-4] LuxR homologue shared only 30% amino-acid identity with

**Fig. 5.** ICE*M*sp^B4-1-4, a tripartite ICE lacking symbiosis genes. Gene annotations are colour coded as follows: black, ICE transfer genes; purple, quorum-sensing genes; brown, *msi287*; green, vitamin biosynthesis; blue, mobile genes such as transposases, integrases, and recombinases; azure, genes likely involved in the broad-specificity catabolism of phosphonates. An asterisk (*) indicates frameshifting required for gene to be translated.

TraR of ICE*Ml*Sym^R7A. ICE*M*sp^B4-1-4 also lacked a conserved *tra*-box sequence upstream of its *luxI* homologue. Thus, ICE*M*sp^B4-1-4 carries distinct quorum-sensing genes that nonetheless presumably control expression of *fseA* and stimulate ICE excision and transfer. ICE*M*sp^B4-1-4 also lacks a gene for QseM, an antiactivator that binds and inhibits TraR and FseA [49]. In summary, while ICE*M*sp^B4-1-4 closely resembles ICESyms, it lacks key regulatory genes common to all other ICESyms.

In previous work, we defined a set of 66 genes conserved on ICESyms [13] but generally absent on all other *Mesorhizobium* ICEs. Of these genes, ICE*M*sp^B4-1-4 carries the *nadABC* for nicotinate biosynthesis, and *bioBFDAZ* for biotin biosynthesis, together present as a single cluster and sharing synteny with genes on ICE*Ml*Sym^R7A, suggesting these vitamin synthesis genes are an ancestral feature of these ICEs. Interestingly, 115 of the NS-meso genomes lacked both *nadABC* and *bioBFDAZ* clusters, and only three NS-meso carried both operons. The only other gene present on ICE*M*sp^B4-1-4 and conserved on ICESyms was *msi287*, which on ICE*Ml*Sym^R7A is present downstream of *nifH* and the nitrogenase genes. Msi287 domain composition and tertiary structure predictions suggest it is a PhoB/FixJ/MtrA-like DNA-binding response regulator. ICE*M*sp^B4-1-4 encodes three copies of Msi287. Each copy shares 50–61 % pairwise amino-acid identity with the ICE*Ml*Sym^R7A copy, suggesting Msi287 paralogues are common components of several distinct regulatory systems on ICE*M*sp^B4-1-4.

A tree constructed from the ICE*M*sp^B4-1-4 site-specific recombinase *intS*, *intM* and *intG,* together with related sequences on other ICEs and ICESyms [13], grouped the ICE*M*sp^B4-1-4 sequences with those present on tripartite ICESyms, rather than with genes present on other ICEs that target the same integration sites (Fig. 6b). In both this tree and the tree constructed from *msi031-trb-msi021*-gene clusters (Fig. 6), ICE*M*sp^B4-1-4 sequences were positioned at the base of the ICESym clade, suggesting ICE*M*sp^B4-1-4 is related to an early common ancestor of tripartite ICESyms. Altogether these observations suggest the ancestor of ICE*M*sp^B4-1-4 diverged from the ancestor of ICESyms early in evolution and perhaps prior to acquisition of symbiosis genes, rather than it being a relative of an extant tripartite ICESym that has subsequently lost symbiosis genes.

## DISCUSSION

The aim of this work was to assess the indigenous *Mesorhizobium* populations in Australian soils by directly isolating them from soil and comparing their genomes with other mesorhizobia. We discovered genetically diverse indigenous *Mesorhizobium* spp. are present in Australian soil and that they overwhelmingly lack genes for nitrogen-fixing symbiosis. Isolated NS-meso shared >86% cpAAI in pairwise comparisons with symbiotic *Mesorhizobium* spp. isolated worldwide. In core-gene trees, the NS-meso phylogeny overlapped with symbiotic strains and extended the genus by 15 genospecies. Despite lacking genes for symbiosis, NS-meso exhibited similar core genomes to symbiotic strains and harboured similar families of mobile genetic elements. These included the first discovered tripartite ICE lacking symbiosis genes, which may resemble the ancestor of extant ICESyms.

A wide diversity of nonsymbiotic *Mesorhizobium* spp. were isolated from soil, and a subset of these strains were capable of evolving into symbionts through ICESym acquisition. Although this study only focused on a limited number of sampling sites, the consistency with which NS-meso were isolated suggests that NS-meso are likely to be widespread in Australian soils. In contrast, there have been very few reports of indigenous symbiotic *Mesorhizobium* spp. isolated from Australian soils, with *Bradyrhizobium* most frequently identified in isolations from nodules of Australian native legumes [75–78]. In this study, the
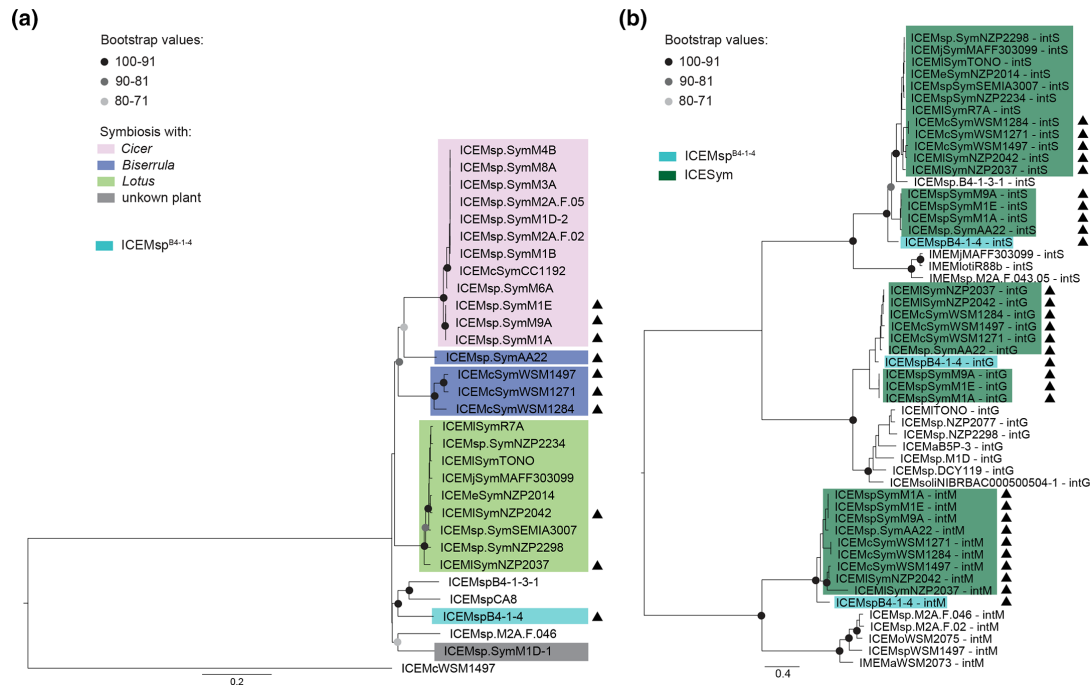
**Fig. 6.** ICE*M*sp^B4-1-4 shares a common ancestor with the archetypal ICESym. (a) Maximum-likelihood tree of ICESym clade ICEs. The tree was constructed with RAxML and it is based on the concatenation of alignments of 16 single-copy backbone genes, the tree was rooted with ICE*Mc*^WSM1497, scale bar indicates substitutions per site. Light green indicates the ICESym specifies symbiosis with *Lotus*, blue with *Biserrula*, pink with *Cicer*, grey with an unknown plant. The tripartite ICE*M*sp^B4-1-4 is highlighted in azure. Triangles indicate tripartite ICEs. (b) Maximum-likelihood tree of integrases specifying integration in *guaA* (*intG*), *met*-tRNA (*intM*) and *phe*-tRNA (*intS*). The tree was constructed with RAxML and rooted at midpoint, the scale bar indicates substitutions per site. Dark green highlight indicates integrases of ICESyms, and triangles indicate the integrase belongs to a tripartite ICE. The integrases of the tripartite ICE*M*sp^B4-1-4 are highlighted in azure.

only symbiotic strains isolated were previously introduced inoculants or were strains that acquired ICESyms from introduced inoculants. Thus, it seems likely that the frequent isolation of novel *Mesorhizobium* microsymbionts from nodules of inoculated agricultural legumes [16, 27, 68, 79] can largely be explained by nonsymbiotic *Mesorhizobium* soil saprophytes acquiring ICESyms from inoculant strains, and not from the evolution of existing indigenous symbiotic *Mesorhizobium* spp.

The phylogenetic diversity of NS-meso isolated here overlaps the diversity of symbiotic *Mesorhizobium* strains isolated internationally. It is possible that all *Mesorhizobium* evolved from a symbiotic ancestor and in Australian soils all the extant clades subsequently lost their symbiosis ICEs (or plasmids). However, a simpler explanation is that the extant diversity observed in both nonsymbiotic and symbiotic members of the genus *Mesorhizobium* evolved prior to the arrival of genes for nitrogen-fixing symbiosis. Further genomic analysis of indigenous rhizobia isolated both from Australian native legumes and soil may help to answer this question.

Transfer of symbiosis genes between distantly related bacteria does not frequently result in an effective nitrogen-fixing symbiosis. For example, *Agrobacterium tumefaciens* carrying the symbiosis-plasmid from *R. leguminosarum* bv. *viciae* VF39 nodulates *Pisum sativum* but does not fix nitrogen [80]. Transfer of the symbiosis-plasmid pRalta from the *Mimosa*-nodulating symbiont *Cupriavidus taiwanensis* to the plant pathogen *Ralstonia solanacearum* generates a non-nodulating *Ralstonia* that eventually forms nodules following spontaneous mutations but does not fix nitrogen [81]. Thus, there are clearly barriers preventing the evolution of symbiosis through gene acquisition. These barriers likely arise from differences in the genomic backgrounds of recipients and the specific adaptation of mobile elements to those backgrounds. ICESyms are probably specifically adapted to enabling a functional symbiosis in combination with the typical core-gene complement present in members of the genus *Mesorhizobium*. Analysis of the *Mesorhizobium* core genome revealed most non-ICESym encoded genes that are likely essential for symbiosis are core genes in NS-meso. Nevertheless, as discussed above, the *Mesorhizobium* core gene complement has likely evolved independent of any involvement in nitrogen-fixing symbiosis [82]. In this work, transfer of ICEsyms *in vivo* and *in vitro* to NS-meso produced both nitrogen-fixing and non-nitrogen-fixing symbionts, consistent with observations previously reported from laboratory and field-isolated strains [17, 27]. The wide variation in symbiotic effectiveness of ICESym exconjugants and the sporadic appearance of non-fixing exconjugants likely reflects the evolutionary independence of NS-meso prior to ICESym acquisition.

Previously it was unclear if the distinct regulatory and structural features of ICESyms and tripartite ICESyms were related to their involvement in symbiosis, or if perhaps the ICESym ancestor existed in its present form prior to acquiring symbiosis genes. The tripartite ICE*M*sp$^{B4-1-4}$ largely resembles other tripartite ICESyms but lacks genes for symbiosis. It is possible that ICE*M*sp$^{B4-1-4}$ once was an ICESym and subsequently lost its symbiosis genes, but other differences in ICE*M*sp$^{B4-1-4}$ suggest otherwise. The conjugation-gene sequences and site-specific recombinase sequences group basally with those of other ICESyms suggesting it branched off from them early on during its evolution. Overall, our analyses suggest ICE*M*sp$^{B4-1-4}$ is related to an ancestor of the ICESyms that probably lacked symbiosis genes. The only genes conserved on ICE*M*sp$^{B4-1-4}$ and all other ICESyms (aside from *msi287*) are genes for biosynthesis of nicotinic acid and biotin, which are not essential for symbiosis [83]. Interestingly, almost all of the NS-meso lacked these genes and those tested were auxotrophic for these vitamins, as were most nonsymbiotic isolates described by Sullivan *et al.* [29]. These observations suggest such vitamin auxotrophy could be advantageous for NS-meso. The lack of the *nad* genes would presumably reduce production of reactive oxygen species that contribute to bacterial death through ageing [84, 85], perhaps assisting their long-term survival in soil conditions not conducive for growth. Conversely, the vitamin prototrophy endowed by ICESyms and ICE*M*sp$^{B4-1-4}$ might make strains more competitive in the rhizosphere during more favourable conditions.

Overall, in this work we have exposed the unrecognized diversity of the nonsymbiotic members of the *Mesorhizobium* genus. While we only interrogated a relatively small number of sites and individual organisms, it is already clear the genome diversity in the NS-meso is expansive. We suspect that the presently characterized symbiotic members of the *Mesorhizobium* represent only a very small fraction of the underlying diversity in the genus. The observation that NS-meso isolated here can form an effective symbiosis following ICESym acquisition suggests there is much potential for generation of novel symbionts with improved symbiotic effectiveness or adaptations to local soil conditions. Finally, the large genome capacity of *Mesorhizobium* genomes and the facile transfer of 100's of kilobases of DNA via conjugation make the genus an attractive platform for rhizosphere engineering.

Conflicts of interest
The authors declare that there is no conflict of interest.

References

1. Remigi P, Zhu J, Young JPW, Masson-Boivin C. Symbiosis within symbiosis: evolving nitrogen-fixing legume symbionts. *Trends Microbiol* 2016;24:63–75.

2. Laranjo M, Alexandre A, Oliveira S. Legume growth-promoting rhizobia: an overview on the *Mesorhizobium* genus. *Microbiol Res* 2014;169:2–17.

3. Perry BJ, Sullivan JT, Colombi E, Murphy RJT, Ramsay JP, et al. Symbiosis islands of Loteae-nodulating *Mesorhizobium* comprise three radiating lineages with concordant *nod* gene complements and nodulation host-range groupings. *Microb Genom* 2020;6:mgen000426.

4. Nandasena KG, O'Hara GW, Tiwari RP, Yates R, Kishinevsky BD. Symbiotic relationships and root nodule ultrastructure of the pasture legume *Biserrula pelecinus* L.?a new legume in agriculture. *Soil Biology and Biochemistry* 2004;36:1309–1317.

5. Greenlon A, Chang PL, Damtew 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* 2019;116:15200–15209.

6. Weir BS, Turner SJ, Silvester WB, Park D-C, Young JM. Unexpectedly diverse *Mesorhizobium* strains and *Rhizobium leguminosarum* nodulate native legume genera of New Zealand, while introduced legume weeds are nodulated by *Bradyrhizobium* species. *Appl Environ Microbiol* 2004;70:5980–5987.

7. De Meyer SE, Andrews M, James EK, Willems A. *Mesorhizobium carmichaelinearum* sp. nov., isolated from *Carmichaelineae* spp. root nodules. *Int J Syst Evol Microbiol* 2019;69:146–152.

8. Lemaire B, Dlodlo O, Chimphango S, Stirton C, Schrire B, et al. Symbiotic diversity, specificity and distribution of rhizobia in native legumes of the Core Cape Subregion (South Africa). *FEMS Microbiol Ecol* 2015;91:1–17.

9. Lamin H, Alami S, Bouhnik O, Bennis M, Benkritly S, et al. Identification of the endosymbionts from *Sulla spinosissima* growing in a lead mine tailings in Eastern Morocco as *Mesorhizobium camelthorni* sv. *aridi*. *J Appl Microbiol* 2021;130:948–959.

10. Ampomah OY, Mousavi SA, Lindström K, Huss-Danell K. Diverse *Mesorhizobium* bacteria nodulate native *Astragalus* and *Oxytropis* in arctic and subarctic areas in Eurasia. *Syst Appl Microbiol* 2017;40:51–58.

11. Zhou PF, Chen WM, Wei GH. *Mesorhizobium robiniae* sp. nov., isolated from root nodules of *Robinia pseudoacacia*. *Int J Syst Evol Microbiol* 2010;60:2552–2556.

12. Degefu T, Wolde-meskel E, Frostegård A. Multilocus sequence analyses reveal several unnamed *Mesorhizobium* genospecies nodulating *Acacia* species and *Sesbania sesban* trees in Southern regions of Ethiopia. *Syst Appl Microbiol* 2011;34:216–226.

13. Colombi E, Perry BJ, Sullivan JT, Bekuma AA, Terpolilli JJ, et al. Comparative analysis of integrative and conjugative mobile genetic elements in the genus *Mesorhizobium Microb Genom* 2021;7:10.

14. Wozniak RAF, Waldor MK. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* 2010;8:552–563.

15. Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc Natl Acad Sci* 1995;92:8985–8989.

16. Hill Y, Colombi E, Bonello E, Haskett T, Ramsay J, et al. Evolution of diverse effective N$_2$-fixing microsymbionts of *Cicer arietinum* following horizontal transfer of the *Mesorhizobium ciceri* CC1192 symbiosis integrative and conjugative element. *Appl Environ Microbiol* 2020;87:e02558-20.

17. Haskett 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* 2016;113:12268–12273.

18. 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.

19. Davidson BR, Davidson HF. *Legumes, the Australian Experience: The Botany, Ecology and Agriculture of Indigenous and Immigrant Legumes*. Research Studies Press Ltd, 1993.

20. Nichols PGH, Revell CK, Humphries AW, Howie JH, Hall EJ, *et al*. Temperate pasture legumes in Australia—their history, current use, and future prospects. *Crop Pasture Sci* 2012;63:691.

21. Lange RT. Nodule bacteria associated with the indigenous *Leguminosae* of South-Western Australia. *J Gen Microbiol* 1961;26:351–359.

22. Corbin E, Brockwell J, Gault R. Nodulation studies on chickpea (Cicer arietinum). *Aust J Exp Agric* 1977;17:126.

23. Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc Natl Acad Sci* 1995;92:8985–8989.

24. Simonsen AK, Dinnage R, Barrett LG, Prober SM, Thrall PH. Symbiosis limits establishment of legumes outside their native range at a global scale. *Nat Commun* 2017;8:14790.

25. Hebb M, Richardson AE, Reid R, Brockwell J. PCR as an ecological tool to determine the establishment and persistence of *Rhizobium* strains introduced into the field as seed inoculant Diane. *Aust J Agric Res* 1998;49:923.

26. Ballard RA, Charman N, McInnes A, Davidson JA. Size, symbiotic effectiveness and genetic diversity of field pea rhizobia (*Rhizobium leguminosarum* bv. *viciae*) populations in South Australian soils. *Soil Biology and Biochemistry* 2004;36:1347–1355.

27. Nandasena KG, O'Hara GW, Tiwari RP, Sezmiş E, Howieson JG. 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* 2007;9:2496–2511.

28. Reeve W, Nandasena K, Yates R, Tiwari R, O'Hara G, *et al*. Complete genome sequence of *Mesorhizobium* opportunistum type strain WSM2075(T.). *Stand Genomic Sci* 2013;9:294–303.

29. Sullivan JT, Eardly BD, van Berkum P, Ronson CW. Four unnamed species of nonsymbiotic rhizobia isolated from the rhizosphere of *Lotus corniculatus*. *Appl Environ Microbiol* 1996;62:2818–2825.

30. Jarvis BD, Ward LJ, Slade EA. Expression by soil bacteria of nodulation genes from *Rhizobium leguminosarum* biovar trifolii. *Appl Environ Microbiol* 1989;55:1426–1434.

31. Soberón-Chávez G, Nájera R. Isolation from soil of *Rhizobium leguminosarum* lacking symbiotic information . *Can J Microbiol* 1989;35:464–468.

32. Segovia L, Piñero D, Palacios R, Martínez-Romero E. Genetic structure of a soil population of nonsymbiotic *Rhizobium leguminosarum*. *Appl Environ Microbiol* 1991;57:426–433.

33. Pérez Carrascal OM, VanInsberghe D, Juárez S, Polz MF, Vinuesa P, *et al*. Population genomics of the symbiotic plasmids of sympatric nitrogen-fixing *Rhizobium* species associated with *Phaseolus vulgaris*. *Environ Microbiol* 2016;18:2660–2676.

34. Hollowell AC, Regus JU, Gano KA, Bantay R, Centeno D, *et al*. Epidemic spread of symbiotic and non-symbiotic *Bradyrhizobium* genotypes across California. *Microb Ecol* 2016;71:700–710.

35. VanInsberghe D, Maas KR, Cardenas E, Strachan CR, Hallam SJ, *et al*. Non-symbiotic *Bradyrhizobium* ecotypes dominate North American forest soils. *ISME J* 2015;9:2435–2441.

36. Weisberg AJ, Rahman A, Backus D, Tyavanagimatt P, Chang JH, *et al*. Pangenome evolution reconciles robustness and instability of rhizobial symbiosis. *mBio* 2022;13:e0007422.

37. Kuzmanović N, Fagorzi C, Mengoni A, Lassalle F, diCenzo GC. Taxonomy of *Rhizobiaceae* revisited: proposal of a new framework for genus delimitation. *Int J Syst Evol Microbiol* 2022;72:005243.

38. Fagorzi C, Ilie A, Decorosi F, Cangioli L, Viti C, *et al*. Symbiotic and nonsymbiotic members of the genus *Ensifer* (syn. *Sinorhizobium*) are separated into two clades based on comparative genomics and high-throughput phenotyping. *Genome Biol Evol* 2020;12:2521–2534.

39. Porter SS, Chang PL, Conow CA, Dunham JP, Friesen ML. Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic *Mesorhizobium*. *ISME J* 2017;11:248–262.

40. Krick A, Kehraus S, Eberl L, Riedel K, Anke H, *et al*. A marine *Mesorhizobium* sp. produces structurally novel long-chain N-acyl-L-homoserine lactones. *Appl Environ Microbiol* 2007;73:3587–3594.

41. Siddiqi MZ, Thao NTP, Choi G, Kim D-C, Lee Y-W, *et al*. *Mesorhizobium denitrificans* sp. nov., a novel denitrifying bacterium isolated from sludge. *J Microbiol* 2019;57:238–242.

42. Lhee D, Lee J, Cho CH, Ha J-S, Jeong SE. *Paulinella micropora* KR01 holobiont genome assembly for studying primary plastid evolution. *Evol Biol*. DOI: 10.1101/794941

43. Bai Y, Müller DB, Srinivas G, Garrido-Oter R, Potthoff E, *et al*. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* 2015;528:364–369.

44. Lo S-C, Li B, Hung G-C, Lei H, Li T, *et al*. Isolation and characterization of two novel bacteria *Afipia cberi* and *Mesorhizobium hominis* from blood of a patient afflicted with fatal pulmonary illness. *PLoS One* 2013;8:e82673.

45. Ronson CW, Primrose SB. Carbohydrate metabolism in *Rhizobium trifolii*: identification and symbiotic properties of mutants. *J General Microbiology* 1979;112:77–88.

46. Ronson CW, Nixon BT, Albright LM, Ausubel FM. *Rhizobium meliloti* ntrA (rpoN) gene is required for diverse metabolic functions. *J Bacteriol* 1987;169:2424–2431.

47. Meade HM, Long SR, Ruvkun GB, Brown SE, Ausubel FM. Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. *J Bacteriol* 1982;149:114–122.

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

49. Ramsay JP, Bastholm TR, Verdonk CJ, Tambalo DD, Sullivan JT, *et al*. An epigenetic switch activates bacterial quorum sensing and horizontal transfer of an integrative and conjugative element. *Nucleic Acids Res* 2022;50:975–988.

50. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.

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

52. 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.

53. Löytynoja A. Phylogeny-aware alignment with PRANK. In: *Multiple Sequence Alignment Methods*. Springer, 2014. pp. 155–170.

54. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;8:28–36.

55. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, *et al*. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* 2011;12:124.

56. West KM, Heydenrych M, Lines R, Tucker T, Fossette S, *et al*. Development of a 16S metabarcoding assay for the environmental DNA (eDNA) detection of aquatic reptiles across northern Australia. *Mar Freshwater Res* 2021.

57. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013;79:5112–5120.

58. Yates R, Howieson J, Hungria M, Bala A, O'Hara G, *et al.* Authentication of rhizobia and assessment of the legume symbiosis in controlled plant growth systems. *Australian Centre for International Agricultural Research* 2016.

59. Hungria M, O'Hara GW, Zilli JE, Araujo RS, Deaker R *et al.* Isolation and growth of rhizobia. In: Howieson JG, Dilworth MJ (editors). Working with rhizobia: ACIAR; 2016. pp. 39-60.

60. Gerding M, Howieson JG, O'Hara GW, Real D, Bräu L. Establishment and survival of the South African legume Lessertia spp. and rhizobia in Western Australian agricultural systems. *Plant Soil* 2013;370:235–249.

61. Handberg K, Stougaard J. Lotus japonicus, an autogamous, diploid legume species for classical and molecular genetics. *Plant J* 1992;2:487–496.

62. Rodpothong P, Sullivan JT, Songsrirote K, Sumpton D, Cheung KWJ-T, *et al.* Nodulation gene mutants of Mesorhizobium loti R7A-nodZ and nolL mutants have host-specific phenotypes on Lotus spp. *Mol Plant Microbe Interact* 2009;22:1546–1554.

63. Wirth JS, Whitman WB. Phylogenomic analyses of a clade within the roseobacter group suggest taxonomic reassignments of species of the genera Aestuariivita, Citreicella, Loktanella, Nautella, Pelagibaca, Ruegeria, Thalassobius, Thiobacimonas and Tropicibacter, and the proposal of six novel genera. *Int J Syst Evol Microbiol* 2018;68:2393–2411.

64. Luo C, Rodriguez-R LM, Konstantinidis KT. MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Res* 2014;42:e73.

65. Nandasena KG, O'Hara GW, Tiwari RP, Willems A, Howieson JG. Mesorhizobium australicum sp. nov. and Mesorhizobium opportunistum sp. nov., isolated from Biserrula pelecinus L. in Australia. *Int J Syst Evol Microbiol* 2009;59:2140–2147.

66. Chao A, Lee S-M. Estimating the Number of Classes via Sample Coverage. *Journal of the American Statistical Association* 1992;87:210–217.

67. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75:7537–7541.

68. Elias NV, Herridge DF. Naturalised populations of mesorhizobia in chickpea (*Cicer arietinum* L.) cropping soils: effects on nodule occupancy and productivity of commercial chickpea. *Plant Soil* 2015;387:233–249.

69. Haskett T, Wang P, Ramsay J, O'Hara G, Reeve W, *et al.* Complete genome sequence of *Mesorhizobium ciceri* bv. *biserrulae* strain WSM1284, an efficient nitrogen-fixing microsymbiont of the pasture legume *Biserrula pelecinus*. *Genome Announc* 2016;4:e00514-16.

70. Loi A, Howieson JG, Cocks PS, Carr SJ. Genetic variation in populations of two Mediterranean annual pasture legumes (*Biserrula pelecinus* L. and *Ornithopus compressus* L.) and associated rhizobia. *Aust J Agric Res* 1999;50:303.

71. Wheatley RM, Ford BL, Li L, Aroney STN, Knights HE, *et al.* Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. *Proc Natl Acad Sci* 2020;117:23823–23834.

72. Xu Y, Mori T, Johnson CH. Cyanobacterial circadian clockwork: roles of KaiA, KaiB and the kaiBC promoter in regulating KaiC. *EMBO J* 2003;22:2117–2126.

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

74. Lockwood S, Greening C, Baltar F, Morales SE. Global and seasonal variation of marine phosphonate metabolism. *ISME J* 2022;16:2198–2212.

75. Marsudi ND, Glenn AR, Dilworth MJ. Identification and characterization of fast- and slow-growing root nodule bacteria from South-Western Australian soils able to nodulate *Acacia saligna*. *Soil Biol Biochem* 1999;31:1229–1238.

76. Thrall PH, Burdon JJ, Woods MJ. Variation in the effectiveness of symbiotic associations between native rhizobia and temperate Australian legumes: interactions within and between genera. *J Appl Ecology* 2000;37:52–65.

77. Lafay B, Burdon JJ. Small-subunit rRNA genotyping of rhizobia nodulating Australian *Acacia* spp. *Appl Environ Microbiol* 2001;67:396–402.

78. Lafay B, Burdon JJ, Ward N. Molecular diversity of legume root-nodule bacteria in Kakadu National Park, Northern Territory, Australia. *PLoS ONE* 2007;2:e277.

79. Nandasena KG, O'Hara GW, Tiwari RP, Howieson JG. 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* 2006;72:7365–7367.

80. Quandt J, Clark RG, Venter AP, Clark SRD, Twelker S, *et al.* Modified RP4 and Tn5-Mob derivatives for facilitated manipulation of large plasmids in Gram-negative bacteria. *Plasmid* 2004;52:1–12.

81. Marchetti M, Capela D, Glew M, Cruveiller S, Chane-Woon-Ming B, *et al.* Experimental evolution of a plant pathogen into a legume symbiont. *PLoS Biol* 2010;8:e1000280.

82. Garrido-Oter R, Nakano RT, Dombrowski N, Ma K-W, McHardy AC, *et al.* Modular traits of the *Rhizobiales* root microbiota and their evolutionary relationship with symbiotic rhizobia. *Cell Host Microbe* 2018;24:155–167.

83. Sullivan JT, Brown SD, Yocum RR, Ronson CW. The bio operon on the acquired symbiosis island of *Mesorhizobium* sp. strain R7A includes a novel gene involved in pimeloyl-CoA synthesis. *Microbiology* 2001;147:1315–1322.

84. Nyström T. The free-radical hypothesis of aging goes prokaryotic. *Cell Mol Life Sci* 2003;60:1333–1341.

85. Fredriksson A, Nyström T. Conditional and replicative senescence in *Escherichia coli*. *Curr Opin Microbiol* 2006;9:612–618.