Supporting Information for:

Characterization of CYP115 As a Gibberellin 3-Oxidase Indicates That Certain Rhizobia Can Produce Bioactive Gibberellin A_4

Ryan S. Nett[1], Tiffany Contreras[1], and Reuben J. Peters[1*]

[1]*Roy J. Carver Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames, IA 50011, USA*

**Corresponding author. E-mail: rjpeters@iastate.edu*

Table of Contents

SUPPORTING METHODS

Culture and growth conditions

Unless noted otherwise, chemicals and reagents were purchased from Thermo Fisher Scientific and Sigma-Aldrich®. Where indicated and unless stated otherwise, antibiotics were used at the following concentrations: kanamycin (Km) 50 μg mL^{-1}, chloramphenicol (Cm) 25 μg mL^{-1}, streptomycin (Sm) 500 μg mL^{-1}, and tetracycline (Tc) 20 μg mL^{-1}. *E. coli* strains were grown at 37˚ C in NZY media (5 g L^{-1} yeast extract, 10 g L^{-1} casein hydrolysate, 5 g L^{-1} NaCl, and 1 g L^{-1} $MgSO_4 \cdot 7H_2O$). *S. meliloti* 1021[1] was grown at 30 ˚C in LB-MC media (Luria-Bertani media supplemented with 2.5 mM $MgSO_4 \cdot 7H_2O$ and 2.5 mM $CaCl_2 \cdot 2H_2O$) using Sm as the selecting antibiotic. See **Table S4** for all strains used in this study.

Identification and genomic analysis of CYP115 and core GA operon in rhizobia

In plant pathogens containing the GA biosynthetic operon, CYP115 is always found tightly appended to the 5' end of the operon, and could thus be hypothesized to act in GA biosynthesis in these organisms. The presence of CYP115 in 20 rhizobial strains (**Table S1**) was found through protein BLAST searches on NCBI (https://blast.ncbi.nlm.nih.gov/blast/Blast.cgi) and the Joint Genome Institute (JGI; https://img.jgi.doe.gov/cgi-bin/mer/main.cgi?section=FindGenesBlast&page=gene SearchBlast) using the amino acid sequence of CYP115 from *Xanthomonas oryzae pv. oryzicola* BLS256, which contains the GA operon. The legumes from which these strains were isolated are listed as found at JGI and/or within the BioSample webpage for each strain on NCBI (**Table S1**). CYP115 and CYP112 pseudo-genes and fragments were further identified by using DNA BLAST searches with the NCBI/GenBank and JGI databases. CYP115 fragments were characterized as being ≥17 base-pairs (bp) in length with nearly 100% identity, in addition to being located within 50 bp of the start codon for CYP112 of the core GA biosynthetic operon. The average length of these fragments (with standard deviation) from 116 rhizobia is 158 ± 130 bp.

The total number of rhizobia with core GA biosynthetic operons (123) was estimated by comparing strains confirmed by BLAST searches to have CPS, KS, CYP117, CYP114 and CYP112 within the JGI database. This number is likely an underrepresentation as it excludes sequences from NCBI/GenBank (and other databases) that may not be available in JGI. Additionally, the confirmation of a full operon relies upon genomic scaffolds in which not only the presence, but also clustering, of operon genes can be observed. As many strains containing GA operon genes do not have well-assembled genomes, this results in many strains being omitted from this analysis.

The location of CYP115 in relation to the core GA operon (**Table S2**) was determined manually by using the graphic genome browser on NCBI. Briefly, CYP115 and the core operon were located via BLAST searches on NCBI, and DNA coordinates were used to calculate the distance between CYP115 genes and the closest gene of the GA operon. For many rhizobial strains with CYP115, this was not possible, as fully assembled genomes were not available, and CYP115 and the GA operon were sometimes located on separate contigs. Insertion sequence (IS) elements near CYP115 genes (**Table S3**) were located and analyzed manually with the graphic genome browser on NCBI by searching the genomic content in direct proximity to CYP115. Each identified IS element was further categorized and annotated based upon the conserved domains found by using the Conserved Domains tool on NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

CYP115 expression constructs

CYP115 expression constructs were created similarly to those used in previous study of the GA biosynthetic operon from *S. fredii* NGR234.[2] For expression of CYP115 from *S. meliloti* WSM4191 (SmCYP115), a clone was synthesized using GeneArt™ Gene Synthesis (Thermo Fisher Scientific). Included on the 3' of this clone was the intergenic region following CYP114 from *S. fredii* NGR234, along with the *S. fredii* NGR234 Fd_{GA}, which follows this intergenic region. Fd_{GA} has been shown to be necessary for full CYP114 functionality in GA biosynthesis,[2] and was included as it could be a potential redox partner for CYP115. The synthetic clone also contained a 5' leader sequence that contained a ribosome binding site, in addition to several upstream stop codons, as previously described.[2] This clone was amplified from the synthetic gene-containing plasmid (pMK-SmCYP115-Fd_{GA}) using primers specific to the 5' leader sequence and the 3'end of this dual gene fragment (see **Table S5** for a list of all primers used). Additionally, a construct without Fd_{GA} was amplified, using a reverse primer that anneals to the 3' end of SmCYP115. These primers contained restriction sites (BamHI for the forward primer, EcoRI for the reverse primer) allowing for subsequent digestion and ligation into the expression plasmid, pstb-LAFR5 (Tc^R), which has been previously shown to be useful for heterologous expression in *S. meliloti* 1021[3], and particularly for the genes found in the GA biosynthetic operon.[2] Notably, *S. meliloti* 1021 contains neither the GA biosynthetic operon, nor a homolog of CYP115.[4] This construct and the pstb-LAFR5 plasmid were digested using BamHI and EcoRI restriction enzymes (New England Biolabs® Inc.). The digested SmCYP115 and SmCYP115-Fd_{GA} clones were run on a 1% agarose gel and purified using the GenElute™ Gel Extraction Kit (Sigma-Aldrich®), then ligated into digested pstb-LAFR5 with T4 DNA ligase (New England Biolabs® Inc.).

CYP115 from *Mesorhizobium loti* MAFF303099 (MlCYP115) was cloned directly from genomic DNA, which was previously purified using Wizard® SV Genomic DNA Purification Kit (Promega). Analysis of CYP115 revealed several potential open reading frames (ORF) of similar length (~1200-1300 bp) and in the same reading frame. The longest predicted ORF of MlCYP115 (1278 bp; MlCYP115) was PCR amplified using forward and reverse primers specific to the 5' and 3' ends of this ORF (**Table S5**), and this construct was ligated into pCR™-Blunt II-TOPO® (Thermo Fisher Scientific). In a subsequent PCR reaction, the 5' leader sequence, along with a BamHI restriction site for cloning into pstb-LAFR5, was added to the 5' end via the forward primer and a 3' EcoRI restriction site was added via the reverse primer. This MlCYP115 construct was digested and ligated into pstb-LAFR5 in the same fashion as described for SmCYP115. Though functional activity was found with this clone, sequence alignment with other predicted CYP115 proteins (**Figure S6**) and the presence of a canonical Shine-Dalgarno sequence suggests that the native ORF is 1233 bp (410 amino acids). This shorter coding sequence was subsequently cloned and expressed, but exhibited indistinguishable activity from the longer clone.

Triparental mating to transfer constructs into *S. meliloti* 1021

The pstb-LAFR5-CYP115 constructs and an empty pstb-LAFR5 vector were transformed into chemically competent *E. coli* MM294A "donor" cells[5] and selected on LB agar plates containing Tc (20 μg mL^{-1}). Individual colonies were picked for growth in 5 mL NZY media cultures supplemented with Tc. Additionally, 5 mL cultures of *E. coli* MT616 "helper" cells[6] (Cm^R) were grown in NZY media, and 5 mL cultures of *S. meliloti* 1021 "recipient" cells (Sm^R) were grown in LB-MC media. All cultures were grown to late log phase, at which point 1 mL of each cultures was centrifuged at 5000 x *g* for 5 minutes to pellet the cells. The pellets were then washed with 1 mL of sterile 0.85% (w/v) NaCl in water, and centrifuged again. Pellets of the donor, helper, and recipient cells were resuspended in 50 μL of 0.85% NaCl, mixed in equal proportion, and 10-30 μL of this mixture was plated on non-selective LB-MC agar plates. Matings were incubated for 1-2 days, after which the spots were picked and resuspended in 1 mL of 0.85% NaCl.

Serial dilutions of this resuspension were then plated on LB-MC agar containing Sm (for selection of the *S. meliloti* 1021 recipient) and Tc (for selection of pstb-LAFR5 constructs). Colonies were screened by PCR using primers specific to the construct of interest, or in the case of the empty vector, primers specific to the region upstream of the multiple cloning site.

CYP115 heterologous expression in *S. meliloti* 1021

Incubations with wild-type (untransformed) strains of *S. meliloti* 1021 and heterologous strains expressing either an empty vector or pstb-LAFR5-CYP115 constructs were performed in a similar fashion to previous characterization of the GA biosynthetic operon.[2] Briefly, individual colonies were used to inoculate 5 mL LB-MC cultures containing appropriate antibiotics (Sm for *S. meliloti* 1021 wild-type and Sm + Tc for heterologous strains), and these cultures were shaken at 225 RPM at 30 ˚C until reaching late log phase (1-2 days). For initial experiments, 500 µL of each culture was used to inoculate 50 mL LB-MC cultures containing 5 mL phosphate buffer (pH 7) and the appropriate antibiotic(s). Cultures were grown at 30 ˚C with 225 RPM shaking until they reached early log phase (OD600 0.4-0.7), at which point substrate (5 µM), riboflavin (1 mM; for enhanced redox metabolism), δ-aminolevulinic acid (1 mM; to enhance heme synthesis for incorporation into the recombinant CYP), and $FeCl_3$ (0.1 mM; also for incorporation into heme of the expressed CYP) were added. GAs were purchased commercially from OlChemIm Ltd. (GA_9, GA_{20}, GA_4, and GA_1), or generously provided by Professor Peter Hedden (Rothamsted Research, Harpenden, U.K.; GA_9). Cultures were incubated at 30 ˚C with 225 RPM shaking for 3-5 days. Once the GA3ox functionality of CYP115 was observed, subsequent experiments were run with 10 mL cultures with the added supplements adjusted proportionally. To quantify relative levels of turnover of GA_9 to GA_4, incubations were run as described above in triplicate for heterologous cells expressing pstb-LAFR5 empty vector, pstb-LAFR5-SmCYP115, pstb-LAFR5-SmCYP115-Fd_{GA}, or pstb-LAFR5-MlCYP115.

Metabolite extraction, purification, and GC-MS analysis

After incubations, cultures were acidified to pH 3 with 5 M HCl to neutralize the free carboxylic acids found in many of the GA metabolites of interest here. Cultures were then extracted 3 times with an equivalent volume of ethyl acetate. These extracts were combined in a round bottom flask and dried with a rotary evaporator. The dried flask was extracted three times with 3 mL of ethyl acetate which was combined in glass tubes and dried under a gentle stream of N_2. The dried metabolites in each tube were resuspended in 1 mL hexanes for purification over a silica gel column (~1 mL). The 1 mL aliquots were loaded to the silica column and eluted with solutions of hexanes:ethyl acetate (v/v), starting with 100% hexanes and increasing polarity incrementally (by 10-15%) to 100% ethyl acetate with separate fractions collected for each elution solution. Purified fractions were dried under N_2 and resuspended in 500 µL ethyl acetate, after which diazomethane was added to produce the methyl esters of any free carboxylic acids, and this solution was incubated at room temperature for several hours. The methylated fractions were then dried under a gentle stream of N_2 and resuspended in BSTFA+TMCS [*N,O*-Bis(trimethylsilyl)trifluoroacetamide + Trimethylchlorosilane; 99:1 v/v] or MSTFA [*N*-Methyl-*N*-(trimethylsilyl)trifluoroacetamide] for 30 min at 80 ˚C to produce the trimethylsilyl ethers of any free hydroxyl groups. Samples were then dried down under a gentle stream of N_2 and resuspended in n-hexane for analysis with gas chromatography-mass spectrometry (GC-MS) using a 3900 Saturn GC with Saturn 2100T ion trap MS (Varian) equipped with an HP-5MS column (Agilent). For this analysis, 1 µL of each sample was injected in splitless mode with a column flow of 1.2 mL min^{-1} at an initial injector temperature of 250 ˚C and a column temperature of 50 ˚C, which was held for 3 minutes. The column temperature was increased to 300 ˚C at a rate of 15 ˚C min^{-1}, and this temperature (300 ˚C) was held for 3 minutes. Mass spectra acquisition using electron ionization with a collection range of 90 *m/z* to 650 *m/z* began at 13 minutes and continued until the end of the run. For optimal peak resolution, publication quality GC-MS runs were

performed with a column temperature increase of 10° C min^{-1} and a mass range of 90-550 *m/z*. GA_4 (**2**) and GA_1 products were identified via comparison to an authentic standard and to published mass spectral data.[7]

For relative quantification of samples run in triplicate, chromatograms were extracted with the 270 and 284 mass-to-charge ions (*m/z* 270 to distinguish GA_9 and *m/z* 284 to distinguish GA_4), with GA_9 and GA_4 peak areas being integrated manually within the Varian GC-MS software. Relative turnover was calculated by dividing the area of the GA_4 peak by the combined areas of the GA_9 and GA_4 peaks and multiplying this by 100 to give percent turnover. Student's t-test (two sample, assuming equal variance) was performed in Microsoft Excel to determine any statistically significant differences between the tested samples.

CYP115 and GA operon sequence analysis

The identity between CYP115 nucleotide and protein sequences were calculated using the tools available on the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) website. For nucleotide sequences, the MUSCLE program (default parameters) was used, while for proteins the Needle EMBOSS pairwise alignment tool (default parameters) was utilized with the *M. loti* MAFF303099 CYP115 compared to each rhizobial CYP115 (**Table S1**).

Phylogenetic trees were created in MEGA7.[8] Nucleotide alignments were performed using the MUSCLE algorithm with the GA operon orthologs from *X. oryzae* pv. *oryzicola* BLS256 as the outgroup sequences for CYP115 and core GA biosynthetic operon alignments, while the *Azotobacter vinelandii* DJ nifK sequence was used as the outgroup in the nifK alignment. The trees shown in **Figs. 3, 4 & Fig. S3** were created via the Maximum-Likelihood statistical method using the highest ranked substitution model (as determined within MEGA7 for each alignment) with complete deletion and 1000 bootstrap analyses. To verify the observed phylogeny of CYP115 from *Mesorhizobium* sp. AA22 as closest to the ancestral node among the rhizobia, trees were also calculated using the Neighbor-Joining and Minimum evolution methods (**Figure S4**). The resulting trees are essentially congruent with that shown for the Maximum-Likelihood tree in **Figure 3**, including the placement of CYP115 from *Mesorhizobium.* sp. AA22.

For the phylogenetic trees generated from alignments with CPS and CYP114 from the GA operon, initial trees that included all strains of rhizobia with the core GA biosynthetic operon were created via the Maximum-Likelihood statistical method using the highest ranked substitutional model with complete deletion and 1000 bootstrap analyses (**Figure S5**). Inspection of these trees showed redundancy and very high similarity among certain clusters of sequences (i.e. no observable branch length). Accordingly, these clusters were condensed to individual sequences and the analysis repeated with this reduced set of sequences to present the representative tree shown in **Figure 4**.

Table S1. Rhizobial strains containing a presumed full-length CYP115. (a.a. = amino acids)

Rhizobia strain	Legume host (original isolation)	protein length (a.a.)	% identity to MlCYP115	GenBank Accession Version
Mesorhizobium loti MAFF303099	*Lotus pedunculatus*	410	-	WP_044549111.1
Mesorhizobium loti R7A	*Lotus corniculatus*	410	97.8	WP_027033340.1
Mesorhizobium loti CJ3sym	*Lotus corniculatus*	410	97.8	WP_027033340.1
Mesorhizobium loti R88b	*Lotus corniculatus*	410	97.8	WP_027033340.1
Mesorhizobium loti (*Mesorhizobium erdmanii*) USDA3471	*Lotus corniculatus*	410	97.6	WP_027056407.1
Mesorhizobium loti NZP2042	*Lotus sp.*	410	96.6	WP_064987301.1
Mesorhizobium loti NZP2014	*Lotus sp.*	410	96.6	WP_064987301.1
Mesorhizobium loti TONO	*Lotus japonicus*	410	96.6	BAV50319.1
Mesorhizobium sp. AA22	*Biserrula pelecinus L.*	410	91.5	WP_065010997.1
Mesorhizobium sp. SEMIA 3007	*Pisum sativum*	410	97.8	WP_069091147.1
Mesorhizobium sp. (*Mesorhizobium metallidurans*) STM4661	*Anthyllis vulneraria*	398	95.1	WP_036240890.1
Mesorhizobium sp. WSM2561	*Lessertia diffusa*	410	90.0	WP_027155432.1
Mesorhizobium sp. WSM3626	*Lessertia diffusa*	409	88.8	WP_027145827.1
Rhizobium mongolense USDA1844	*Medicago ruthenica*	410	91.0	WP_022718896.1
Rhizobium favelukesii OR191 (LPU83)	*Medicago sativa*	410	90.2	WP_024317976.1
Sinorhizobium arboris (*Ensifer arboris*) LMG 14919	*Prosopis chilensis*	410	91.2	WP_028002554.1
Sinorhizobium fredii CCBAU83666 AJQR	*Glycine max*	410	92.2	WP_037436960.1
Sinorhizobium medicae (*Ensifer medicae*) WSM1369	*Medicago sphaerocarpa*	425	87.1	WP_018009729.1
Sinorhizobium meliloti (*Ensifer medicae*) WSM4191	*Melilotus messanensis/Melilotus siculus*	425	86.8	WP_028055591.1
Sinorhizobium terangae (*Ensifer* sp.) WSM1721	*Indigofera sp.*	410	91.2	WP_026621921.1

Table S2. Analysis of CYP115 genomic location and synteny. (kb = kilobase, bp = base pairs)

Rhizobia strain	Location	Distance from GA operon	3' CYP112 fragment (present/absent); length
Mesorhizobium loti MAFF303099	Not in operon	~200 kb	Present; 203 bp
Mesorhizobium loti R7A	Not in operon	~105 kb	Present; 222 bp
Mesorhizobium loti CJ3sym	Not in operon	~105 kb	Present; 221 bp
Mesorhizobium loti R88b	Not in operon	~105 kb	Present; 221 bp
Mesorhizobium loti USDA3471	Not in operon	~98 kb	Present; 206 bp
Mesorhizobium loti NZP2042	Not in operon	~85 kb	Present; 20 bp
Mesorhizobium loti NZP2014	Not in operon	N.A.[a]	Present; 20 bp
Mesorhizobium loti TONO	Not in operon	~134 kb	Present; 223 bp
Mesorhizobium sp. AA22	5' end of operon	Precedes CYP112 by 17 bp	*b*
Mesorhizobium sp. SEMIA 3007	Not in operon	~126 kb	Present; 213 bp
Mesorhizobium sp. STM4661	Not in operon	~153 kb	Present; 145 bp
Mesorhizobium sp. WSM2561	Not in operon	~10 kb upstream of operon; opposite orientation	Present; 84 bp
Mesorhizobium sp. WSM3626	Not in operon	~3 kb upstream of operon; opposite orientation	Absent
Rhizobium mongolense USDA1844	Not in operon	N.A.[a]	Present; 252 bp
Rhizobium favelukesii OR191 (LPU83)	Not in operon	N.A.[a]	Present; 241 bp
Sinorhizobium arboris LMG 14919	5' end of operon	Precedes CYP112 by 739 bp	Present; 225 bp
Sinorhizobium fredii CCBAU83666 AJQR	5' end of operon	Precedes CYP112 by 736 bp	Present; 225 bp
Sinorhizobium medicae WSM1369	3' end of operon	Follows IDI by 78 bp	Absent
Sinorhizobium meliloti WSM4191	3' end of operon	Follows IDI by 138 bp	Absent
Sinorhizobium terangae WSM1721	Not in operon	N.A.[a]	Present; 176 bp

[a]N.A. = not available; this is due to CYP115 and the GA operon being found in different genomic scaffolds, and therefore relative location cannot be determined.
[b]CYP115 is adjacent to CYP112, and thus no 3' CYP112 is possible.

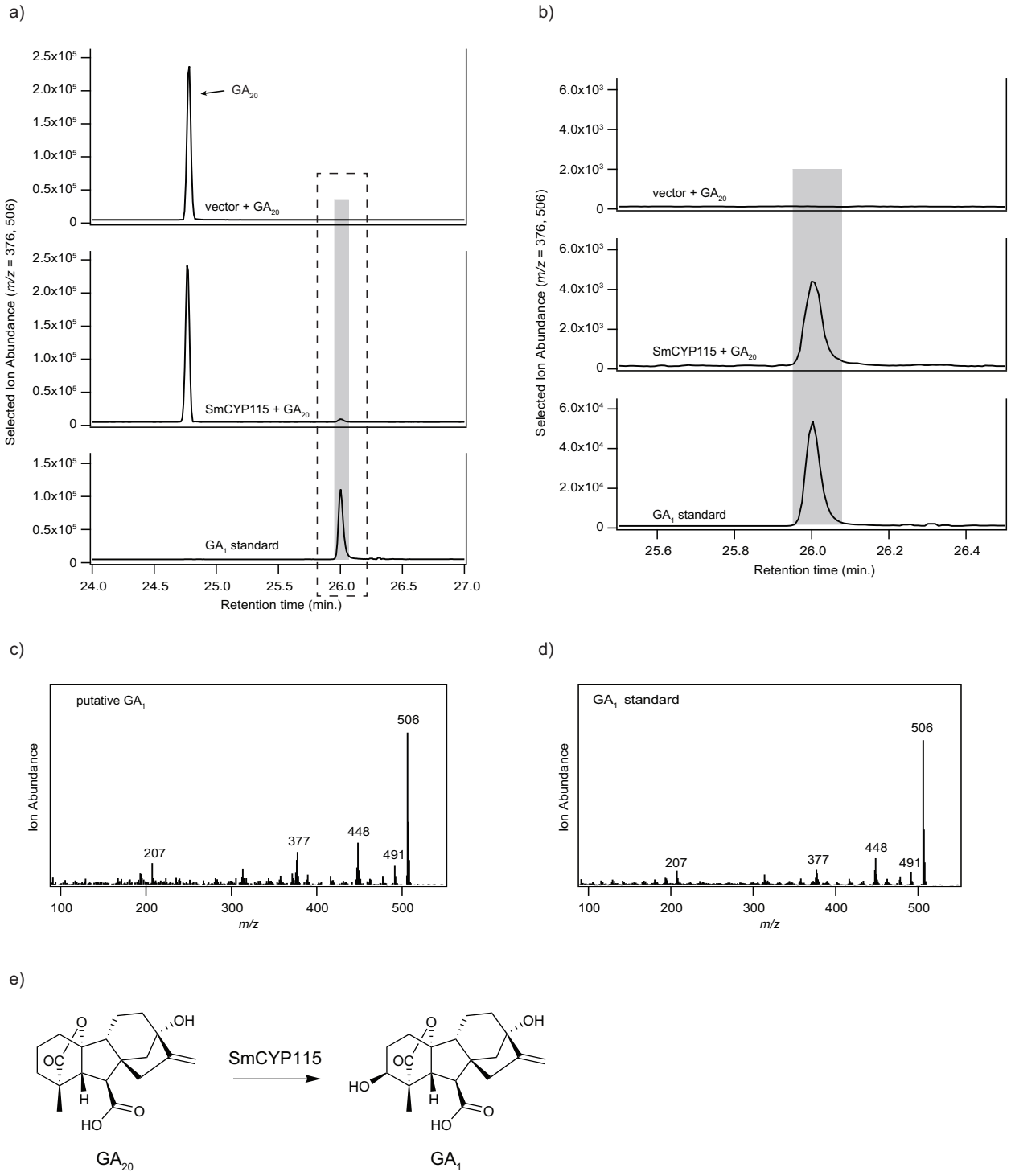


Figure S1. CYP115 converts GA_{20} into bioactive GA_1. a) When incubated in cells expressing SmCYP115, GA_{20} is converted into a compound corresponding to GA_1 as shown here through GC-MS chromatograms. b) Zoom-in on putative GA_1, as indicated by the dotted box from the first panel. c) Mass spectrum of the putative GA_1 product. d) Mass spectrum of an authentic GA_1 standard. e) Reaction catalyzed by SmCYP115, which corresponds to a C-3β hydroxylation of GA_{20} to form GA_1, a bioactive GA. Chromatograms and mass spectra are representative of compounds derivatized to methyl esters and/or trimethylsilyl ethers.

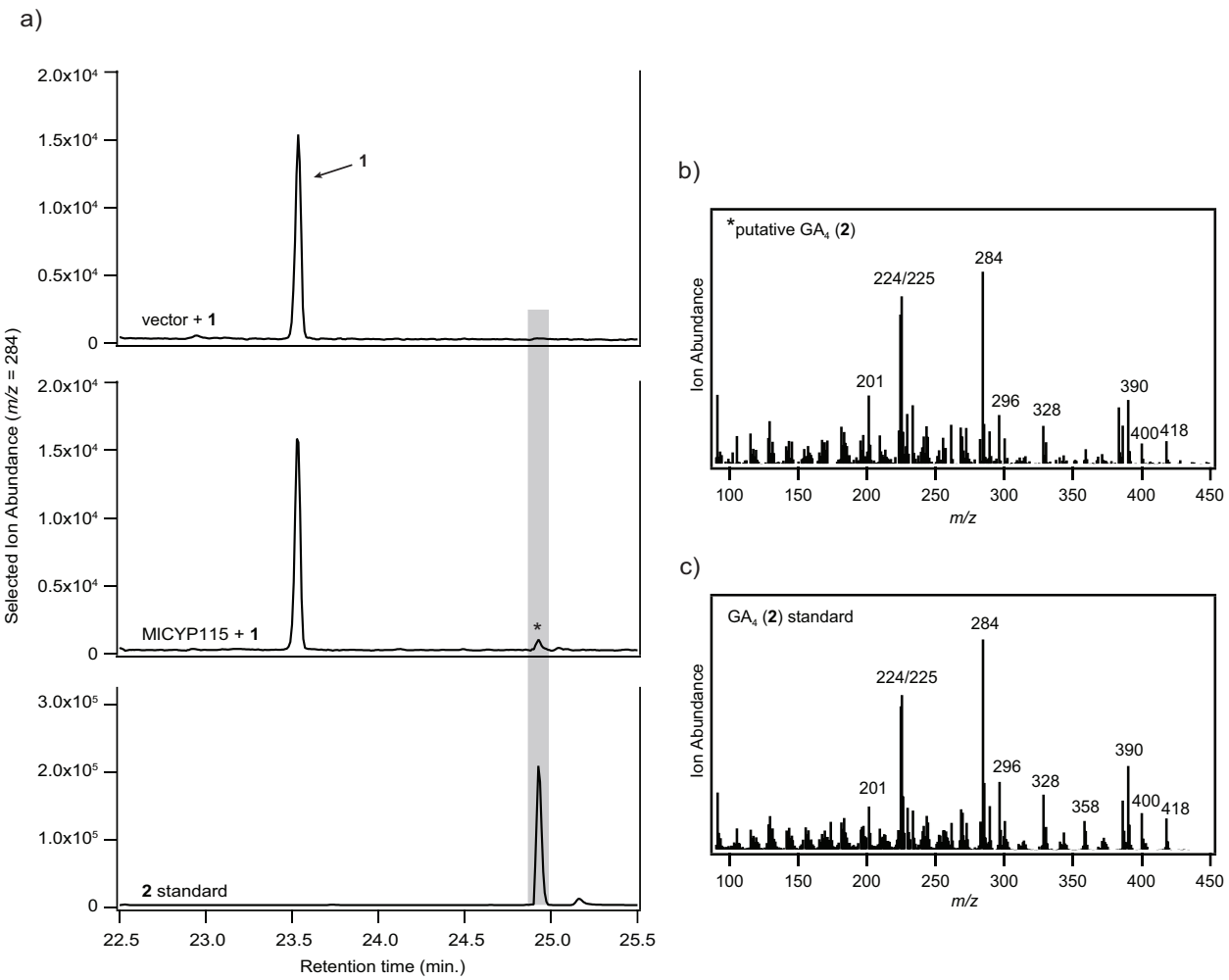


Figure S2. CYP115 function is conserved in *Mesorhizobium loti* MAFF303099. a) Although turnover is low (<1%), when GA_9 (**1**) is incubated with cells expressing CYP115 from *M. loti* MAFF303099, transformation to GA_4 (**2**) is observed, as shown here through GC-MS chromatograms. b) Mass spectrum of putative **2** (indicated by * in corresponding chromatogram). c) Mass spectrum of an authentic **2** standard. Chromatograms and mass spectra correspond to compounds derivatized as methyl esters and/or trimethylsilyl ethers.

Table S3. Insertion sequence (IS) elements flanking CYP115 genes.

Rhizobia strain	Annotated IS element present? (yes/no)	Location relative to CYP115	Type of element	GenBank accession version
Mesorhizobium loti MAFF303099	yes	5'	IS4/IS5 transposase	WP_010913789.1
Mesorhizobium loti R7A	no	-	-	-
Mesorhizobium loti CJ3sym	no	-	-	-
Mesorhizobium loti R88b	no	-	-	-
Mesorhizobium loti USDA3471	no	-	-	-
Mesorhizobium loti NZP2042	yes	5'	IS4/IS5 transposase	WP_032921559.1
Mesorhizobium loti NZP2014	no	-	-	-
Mesorhizobium loti TONO	no	-	-	-
Mesorhizobium sp. AA22	no	-	-	-
Mesorhizobium sp. SEMIA 3007	yes	5'	IS4/IS5 transposase	WP_069091149.1
Mesorhizobium sp. STM4661	no	-	-	-
Mesorhizobium sp. WSM2561	yes	5'	DDE transposase	WP_027155434.1
Mesorhizobium sp. WSM3626	yes	3'	integrase (partial)	M653_RS0125265[a]
Rhizobium mongolense USDA1844	yes (x2)	both 5'	1. DNA replication protein DnaC 2. transposase (COG3316)	1. A3C3_RS42030[b] 2. A3C3_RS42035[b]
Rhizobium favelukesii OR191 (LPU83)	yes (x2)	both 5'	1. transposase (DUF772 superfamily) 2. DDE transposase	1. WP_051166744.1 2. A3A1_RS40375[a]
Sinorhizobium arboris LMG 14919	no	-	-	-
Sinorhizobium fredii CCBAU83666 AJQR	yes	5'	integrase (rve superfamily)	WP_037436957.1
Sinorhizobium medicae WSM1369	yes (x2)	both 3'	1. transposase (pfam14319) 2. transposase (Y2_Tnp super family)	1. WP_018009730.1 2. WP_026160775.1
Sinorhizobium meliloti WSM4191	yes	3'	transposase (pfam00872)	SINMEL_RS0135775[b]
Sinorhizobium terangae WSM1721	yes	5'	transposase (COG3464)	WP_026621920.1

[a]Gene cut off at end of contig and no protein accession was provided. Shown instead is the gene name.
[b]Gene is annotated as a pseudo or partial gene, and no protein accession is provided. Shown instead is the gene name.

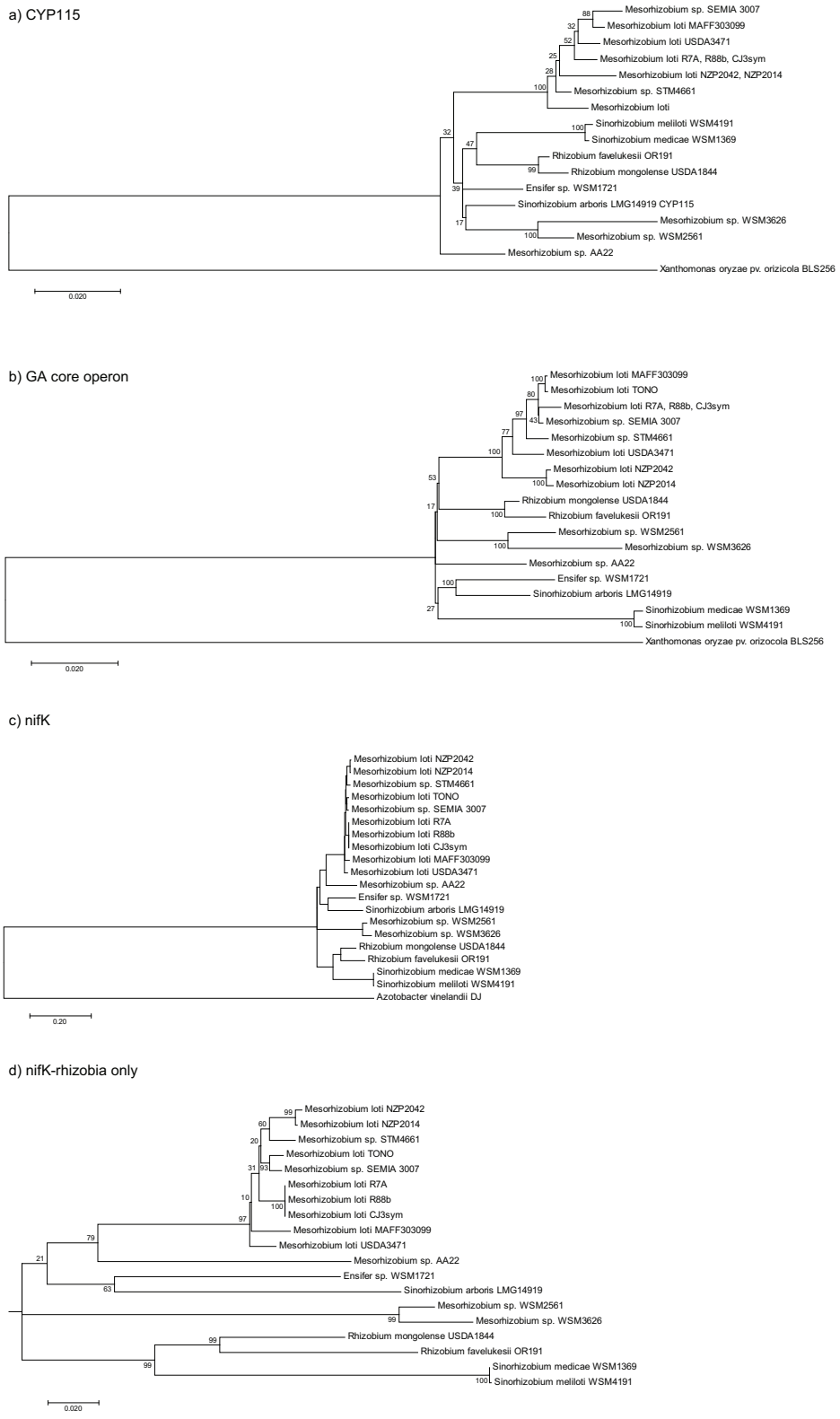


Figure S3. Phylogenetic analysis of CYP115. Shown are phylogenetic trees (Tamura 3-parameter model, gamma distribution, complete deletion of gaps, 1000 bootstraps) for a) CYP115, b) core GA operon, and c) nifK nucleotide sequences for rhizobia strains containing a full length CYP115. d) Zoom-in of rhizobia nifK subtree (excludes *Azotobacter vinlandii* DJ outgroup). All sequences were aligned with MUSCLE.

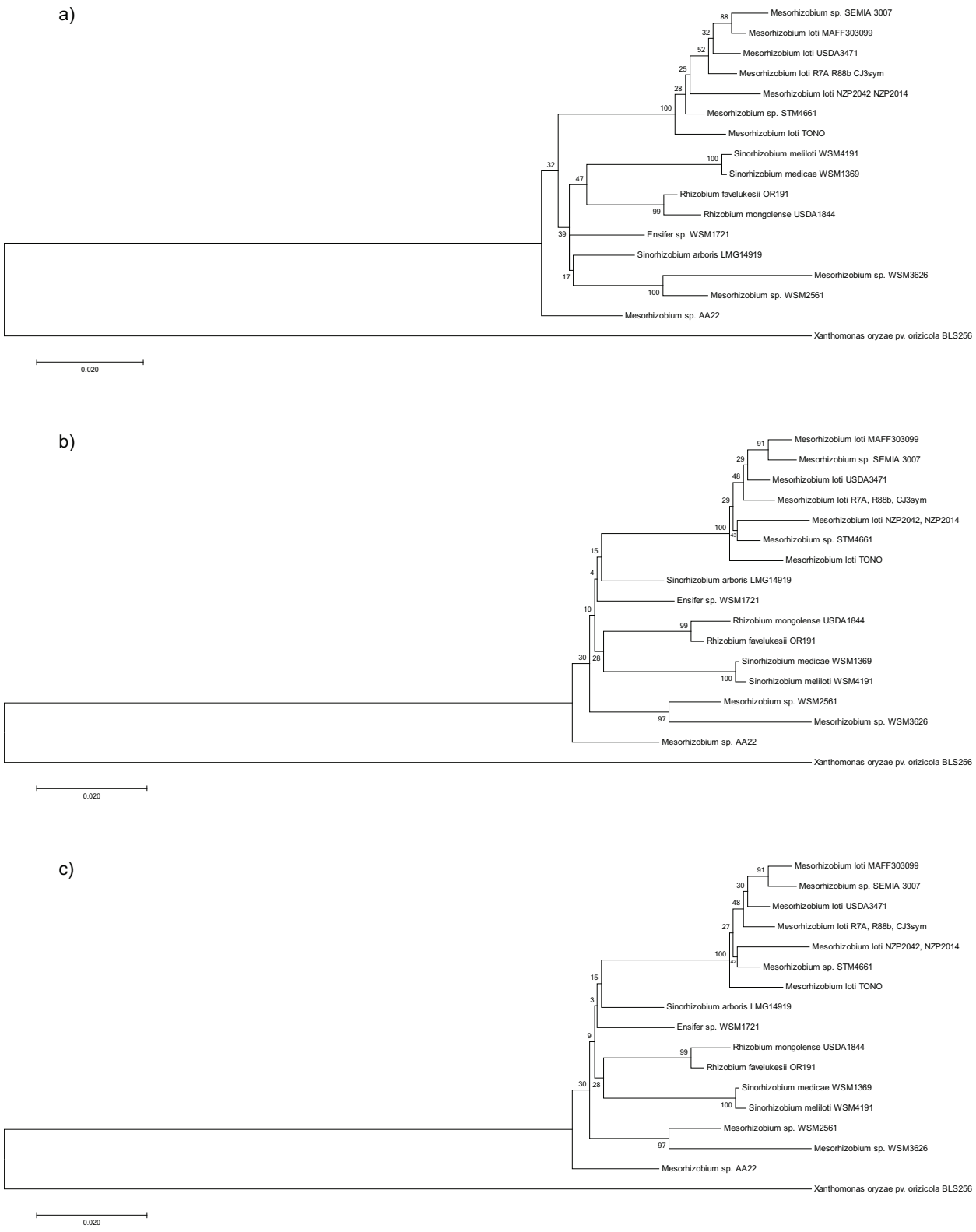
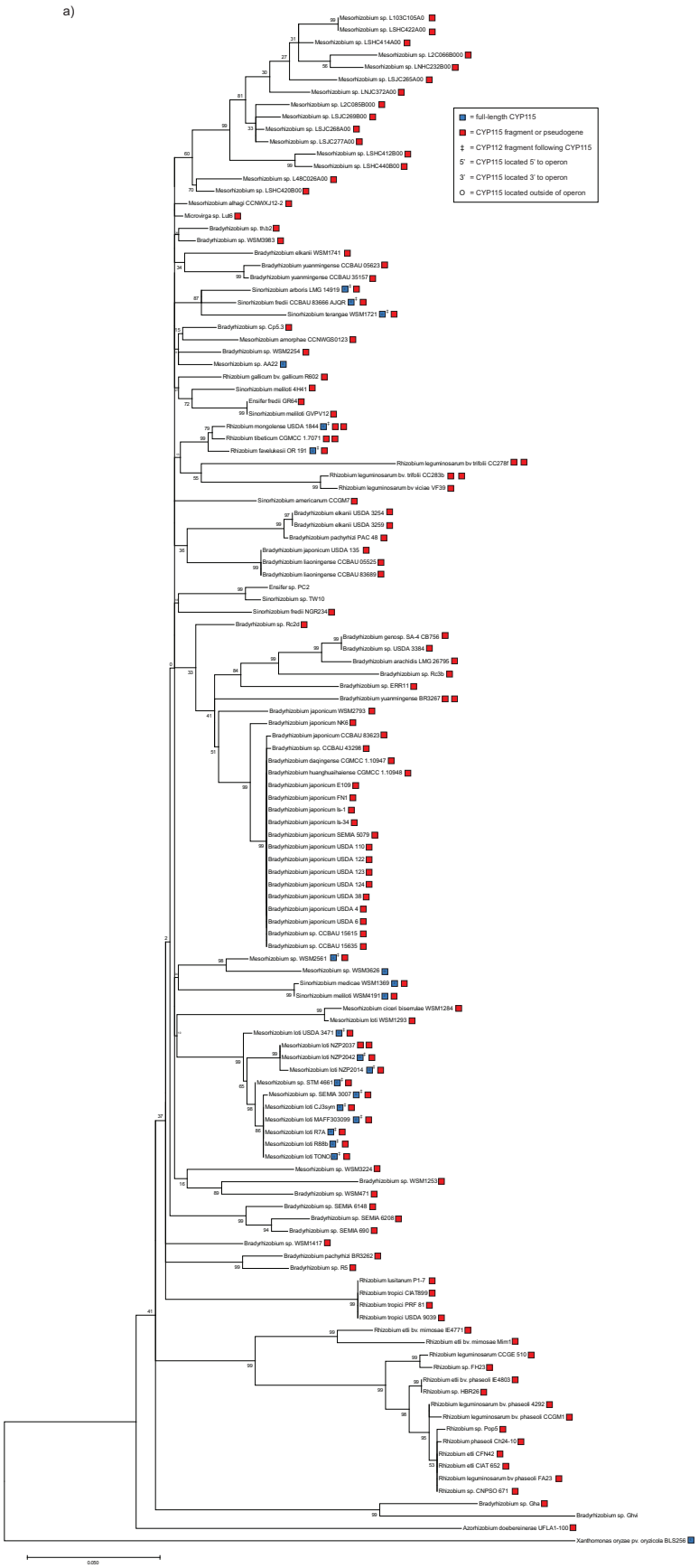


Figure S4. Confirmation of CYP115 phylogeny in rhizobia. Phylogenetic trees (Tamura 3-parameter model, gamma distribution, complete deletion of gaps, 1000 bootstraps) representing CYP115 DNA alignments (MUSCLE) are consistent when using the a) Maximum Likelihood, b) Neighbor Joining, and c) Maximum Parsimony methods for tree construction.

a)


= full-length CYP115
= CYP115 fragment or pseudogene
‡ = CYP112 fragment following CYP115
5′ = CYP115 located 5′ to operon
3′ = CYP115 located 3′ to operon
O = CYP115 located outside of operon
Mesorhizobium sp. L103C105A0
Mesorhizobium sp. LSHC422A00
Mesorhizobium sp. LSHC414A00
Mesorhizobium sp. L2C066B000
Mesorhizobium sp. LNHC232B00
Mesorhizobium sp. LSJC265A00
Mesorhizobium sp. LNJC372A00
Mesorhizobium sp. L2C085B000
Mesorhizobium sp. LSJC269B00
Mesorhizobium sp. LSJC268A00
Mesorhizobium sp. LSJC277A00
Mesorhizobium sp. LSHC412B00
Mesorhizobium sp. LSHC440B00
Mesorhizobium sp. L48C026A00
Mesorhizobium sp. LSHC420B00
Mesorhizobium alhagi CCNWXJ12-2
Microvirga sp. Lut5
Bradyrhizobium sp. th32
Bradyrhizobium sp. WSM3983
Bradyrhizobium elkanii WSM1741
Bradyrhizobium yuanmingense CCBAU 05623
Bradyrhizobium yuanmingense CCBAU 35157
Sinorhizobium arboris LMG 14919
Sinorhizobium fredii CCBAU 83666 AJQR
Sinorhizobium terangae WSM1721
Bradyrhizobium sp. Cp5.3
Mesorhizobium amorphae CCNWGS0123
Bradyrhizobium sp. WSM2254
Mesorhizobium sp. AA22
Rhizobium gallicum bv. gallicum R602
Sinorhizobium meliloti 4H41
Ensifer fredii GR64
Sinorhizobium meliloti GVPV12
Rhizobium mongolense USDA 1844
Rhizobium tibeticum CGMCC 1.7071
Rhizobium favelukesii OR 191
Rhizobium leguminosarum bv trifolii CC278f
Rhizobium leguminosarum bv. trifolii CC283b
Rhizobium leguminosarum bv viciae VF39
Sinorhizobium americanum CCGM7
Bradyrhizobium elkanii USDA 3254
Bradyrhizobium elkanii USDA 3259
Bradyrhizobium pachyrhizi PAC 48
Bradyrhizobium japonicum USDA 135
Bradyrhizobium liaoningense CCBAU 05525
Bradyrhizobium liaoningense CCBAU 83689
Ensifer sp. PC2
Sinorhizobium sp. TW10
Sinorhizobium fredii NGR234
Bradyrhizobium sp. Rc2d
Bradyrhizobium genosp. SA-4 CB756
Bradyrhizobium sp. USDA 3384
Bradyrhizobium arachidis LMG 26795
Bradyrhizobium sp. Rc3b
Bradyrhizobium sp. ERR11
Bradyrhizobium yuanmingense BR3267
Bradyrhizobium japonicum WSM2793
Bradyrhizobium japonicum NK6
Bradyrhizobium japonicum CCBAU 83623
Bradyrhizobium sp. CCBAU 43296
Bradyrhizobium daqingense CGMCC 1.10947
Bradyrhizobium huanghuaihaiense CGMCC 1.10948
Bradyrhizobium japonicum E109
Bradyrhizobium japonicum FN1
Bradyrhizobium japonicum Is-1
Bradyrhizobium japonicum Is-34
Bradyrhizobium japonicum SEMIA 5079
Bradyrhizobium japonicum USDA 110
Bradyrhizobium japonicum USDA 122
Bradyrhizobium japonicum USDA 123
Bradyrhizobium japonicum USDA 124
Bradyrhizobium japonicum USDA 36
Bradyrhizobium japonicum USDA 4
Bradyrhizobium japonicum USDA 6
Bradyrhizobium sp. CCBAU 15615
Bradyrhizobium sp. CCBAU 15635
Mesorhizobium sp. WSM2561
Mesorhizobium sp. WSM3626
Sinorhizobium medicae WSM1369
Sinorhizobium meliloti WSM4191
Mesorhizobium ciceri biserrulae WSM1284
Mesorhizobium loti WSM1293
Mesorhizobium loti USDA 3471
Mesorhizobium loti NZP2037
Mesorhizobium loti NZP2042
Mesorhizobium loti NZP2014
Mesorhizobium sp. STM 4661
Mesorhizobium sp. SEMIA 3007
Mesorhizobium loti CJ3sym
Mesorhizobium loti MAFF303099
Mesorhizobium loti R7A
Mesorhizobium loti R88b
Mesorhizobium loti TONO
Mesorhizobium sp. WSM3224
Bradyrhizobium sp. WSM1253
Bradyrhizobium sp. WSM471
Bradyrhizobium sp. SEMIA 6148
Bradyrhizobium sp. SEMIA 6208
Bradyrhizobium sp. SEMIA 690
Bradyrhizobium sp. WSM1417
Bradyrhizobium pachyrhizi BR3262
Bradyrhizobium sp. R5
Rhizobium lusitanum P1-7
Rhizobium tropici CIAT899
Rhizobium tropici PRF 81
Rhizobium tropici USDA 9039
Rhizobium etli bv. mimosae IE4771
Rhizobium etli bv. mimosae Mim1
Rhizobium leguminosarum CCGE 510
Rhizobium sp. FH23
Rhizobium etli bv. phaseoli IE4803
Rhizobium sp. HBR26
Rhizobium leguminosarum bv. phaseoli 4292
Rhizobium leguminosarum bv. phaseoli CCGM1
Rhizobium sp. Pop5
Rhizobium phaseoli Ch24-10
Rhizobium etli CFN42
Rhizobium etli CIAT 652
Rhizobium leguminosarum bv phaseoli FA23
Rhizobium sp. CNPSO 671
Bradyrhizobium sp. Ghv
Bradyrhizobium sp. Ghvi
Azorhizobium doebereinerae UFLA1-100
Xanthomonas oryzae pv. oryzicola BLS256
0.050

b)

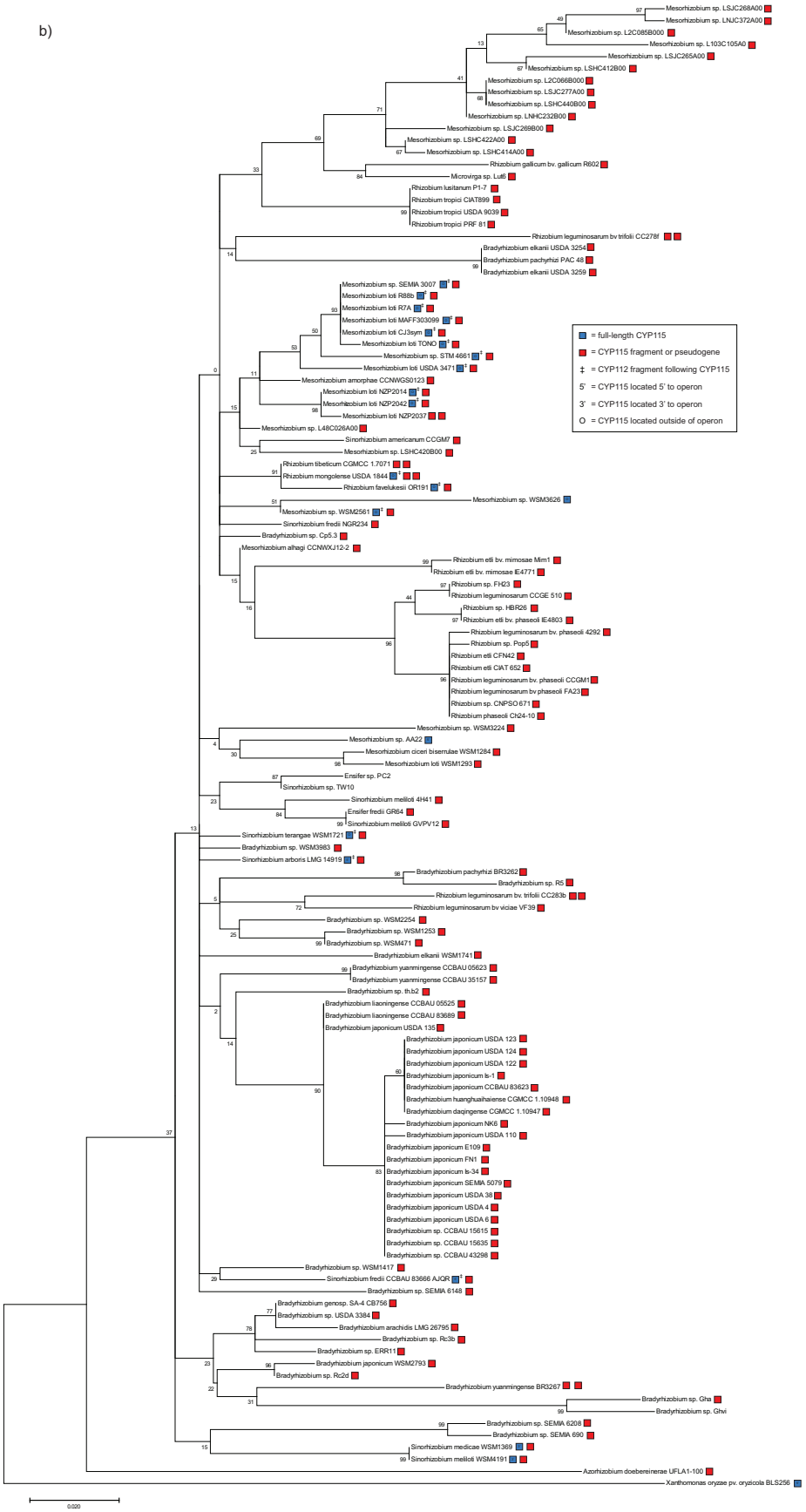

Mesorhizobium sp. LSJC268A00
Mesorhizobium sp. LNJC372A00
Mesorhizobium sp. L2C085B000
Mesorhizobium sp. L103C105A0
Mesorhizobium sp. LSJC265A00
Mesorhizobium sp. LSHC412B00
Mesorhizobium sp. L2C066B000
Mesorhizobium sp. LSJC277A00
Mesorhizobium sp. LSHC440B00
Mesorhizobium sp. LNHC232B00
Mesorhizobium sp. LSJC269B00
Mesorhizobium sp. LSHC422A00
Mesorhizobium sp. LSHC414A00
Rhizobium gallicum bv. gallicum R602
Microvirga sp. Lut6
Rhizobium lusitanum P1-7
Rhizobium tropici CIAT899
Rhizobium tropici USDA 9039
Rhizobium tropici PRF 81
Rhizobium leguminosarum bv trifolii CC278f
Bradyrhizobium elkanii USDA 3254
Bradyrhizobium pachyrhizi PAC 48
Bradyrhizobium elkanii USDA 3259
Mesorhizobium sp. SEMIA 3007
Mesorhizobium loti R88b
Mesorhizobium loti R7A
Mesorhizobium loti MAFF303099
Mesorhizobium loti CJ3sym
Mesorhizobium loti TONO
Mesorhizobium sp. STM 4661
Mesorhizobium loti USDA 3471
Mesorhizobium amorphae CCNWGS0123
Mesorhizobium loti NZP2014
Mesorhizobium loti NZP2042
Mesorhizobium loti NZP2037
Mesorhizobium sp. L48C026A00
Sinorhizobium americanum CCGM7
Mesorhizobium sp. LSHC420B00
Rhizobium tibeticum CGMCC 1.7071
Rhizobium mongolense USDA 1844
Rhizobium favelukesii OR191
Mesorhizobium sp. WSM3626
Mesorhizobium sp. WSM2561
Sinorhizobium fredii NGR234
Bradyrhizobium sp. Cp5.3
Mesorhizobium alhagi CCNWXJ12-2
Rhizobium etli bv. mimosae Mim1
Rhizobium etli bv. mimosae IE4771
Rhizobium sp. FH23
Rhizobium leguminosarum CCGE 510
Rhizobium sp. HBR26
Rhizobium etli bv. phaseoli IE4803
Rhizobium leguminosarum bv. phaseoli 4292
Rhizobium sp. Pop5
Rhizobium etli CFN42
Rhizobium etli CIAT 652
Rhizobium leguminosarum bv. phaseoli CCGM1
Rhizobium leguminosarum bv phaseoli FA23
Rhizobium sp. CNPSO 671
Rhizobium phaseoli Ch24-10
Mesorhizobium sp. WSM3224
Mesorhizobium sp. AA22
Mesorhizobium ciceri biserrulae WSM1284
Mesorhizobium loti WSM1293
Ensifer sp. PC2
Sinorhizobium sp. TW10
Sinorhizobium meliloti 4H41
Ensifer fredii GR64
Sinorhizobium meliloti GVPV12
Sinorhizobium terangae WSM1721
Bradyrhizobium sp. WSM3983
Sinorhizobium arboris LMG 14919
Bradyrhizobium pachyrhizi BR3262
Bradyrhizobium sp. R5
Rhizobium leguminosarum bv. trifolii CC283b
Rhizobium leguminosarum bv viciae VF39
Bradyrhizobium sp. WSM2254
Bradyrhizobium sp. WSM1253
Bradyrhizobium sp. WSM471
Bradyrhizobium elkanii WSM1741
Bradyrhizobium yuanmingense CCBAU 05623
Bradyrhizobium yuanmingense CCBAU 35157
Bradyrhizobium sp. th.b2
Bradyrhizobium liaoningense CCBAU 05525
Bradyrhizobium liaoningense CCBAU 83689
Bradyrhizobium japonicum USDA 135
Bradyrhizobium japonicum USDA 123
Bradyrhizobium japonicum USDA 124
Bradyrhizobium japonicum USDA 122
Bradyrhizobium japonicum Is-1
Bradyrhizobium japonicum CCBAU 83623
Bradyrhizobium huanghuaihaiense CGMCC 1.10948
Bradyrhizobium daqingense CGMCC 1.10947
Bradyrhizobium japonicum NK6
Bradyrhizobium japonicum USDA 110
Bradyrhizobium japonicum E109
Bradyrhizobium japonicum FN1
Bradyrhizobium japonicum Is-34
Bradyrhizobium japonicum SEMIA 5079
Bradyrhizobium japonicum USDA 38
Bradyrhizobium japonicum USDA 4
Bradyrhizobium japonicum USDA 6
Bradyrhizobium sp. CCBAU 15615
Bradyrhizobium sp. CCBAU 15635
Bradyrhizobium sp. CCBAU 43298
Bradyrhizobium sp. WSM1417
Sinorhizobium fredii CCBAU 83666 AJQR
Bradyrhizobium sp. SEMIA 6148
Bradyrhizobium genosp. SA-4 CB756
Bradyrhizobium sp. USDA 3384
Bradyrhizobium arachidis LMG 26795
Bradyrhizobium sp. Rc3b
Bradyrhizobium sp. ERR11
Bradyrhizobium japonicum WSM2793
Bradyrhizobium sp. Rc2d
Bradyrhizobium yuanmingense BR3267
Bradyrhizobium sp. Gha
Bradyrhizobium sp. Ghvi
Bradyrhizobium sp. SEMIA 6208
Bradyrhizobium sp. SEMIA 690
Sinorhizobium medicae WSM1369
Sinorhizobium meliloti WSM4191
Azorhizobium doebereinerae UFLA1-100
Xanthomonas oryzae pv. oryzicola BLS256
= full-length CYP115
= CYP115 fragment or pseudogene
‡ = CYP112 fragment following CYP115
5′ = CYP115 located 5′ to operon
3′ = CYP115 located 3′ to operon
O = CYP115 located outside of operon
0.020

Figure S5. Representative phylogenetic analysis of the core GA operon in rhizobia. Shown are the phylogenetic analyses (Tamura 3-parameter model, gamma distribution with invariant sites, complete deletion of gaps, 1000 bootstraps) for nucleotide alignments with the a) CPS and b) CYP114 from the GA operon (both aligned with MUSCLE). Only rhizobia confirmed to have each gene of the core GA operon are represented here.

Table S4. List of bacterial strains used in this study.

Strain	Description	Reference or source
One Shot® TOP10 Chemically Competent *E. coli*	F- *mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG*; for plasmid propagation and cloning	Thermo Fisher Scientific
OverExpress™ C41(DE3) Chemically Competent *E. coli*	F- *ompT hsdSB (rB- mB-) gal dcm* (DE3); for pCDF-BjSDR/Fd expression	Lucigen
E. coli MM294A	*pro-82 thi-1 endA hsdR17 supE44;* donor strains for triparental mating	[5]
E. coli MT616	MM294A *recA56* carrying plasmid pRK600, Cm^R; helper strains for triparental mating	[6]
Sinorhizobium meliloti 1021 (a.k.a. *Ensifer* meliloti 1021, *S. meliloti* Rm1021)	Rhizobial symbiont of *Medicago spp.;* Sm^R; heterologous host for expression of *S. fredii* GA operon genes	[1]
Mesorhizobium loti MAFF303099	Rhizobial symbiont of *Lotus japonicus*; genomic DNA was isolated from this strain for cloning	Genetic Resource Center, National Institute of Agrobiological Resources, Ministry of Agriculture, Forestry and Fisheries, Japan
*Sm*1021 pstb-LAFR5-SfCYP115	*S. meliloti* 1021 transformed for expression of *S. meliloti* WSM4191 CYP115; Sm^R, Tc^R	This study
*Sm*1021 pstb-LAFR5-SfCYP115-Fd	*S. meliloti* 1021 transformed for coexpression of *S. meliloti* WSM4191 CYP115 and *S. fredii* NGR234 Fd_{GA}; Sm^R, Tc^R	This study
*Sm*1021 pstb-LAFR5-MlCYP115	*S. meliloti* 1021 transformed for expression of *M. loti* MAFF303099 CYP115; Sm^R, Tc^R	This study

Table S5. List of primers used to create CYP115 expression constructs. Underlined type indicates restriction sites.

Name	Sequence (5' to 3')	Description
SmCYP115-F	CGCGGATCCTAACGTAACGT AAATGGAGTCACCTCT	Forward primer used to amplify synthetic SmCYP115 construct. Contains BamHI restriction site.
SmCYP115-R	CCGGAATTCCTATGCCGCAC CGCGGGTACGACCTAATG	Reverse primer used to amplify synthetic SmCYP115 construct *without* Fd_{GA} coexpression. Contains EcoRI restriction site.
SmCYP115-Fd_{GA}-R	CCGGAATTCTCAAACCGTTC CATCGTGTCCTCCTGG	Reverse primer used to amplify synthetic SmCYP115 construct *with* Fd_{GA} coexpression. Contains EcoRI restriction site.
MlCYP115-F	ATGCCGAGAAGTGGACCCG CGATGG	Forward primer used to amplify the longer MlCYP115 coding sequence (1278 nucleotides) from genomic DNA. While the genomic sequence contains a GTG start codon, this was changed to ATG for optimal expression.
MlCYP115-F'	ATGCGCGTAGAAAACGAT CATTGCG	Forward primer used to amplify the shorter predicted MlCYP115 coding sequence (1233 nucleotides) from the previously cloned longer sequence.
MlCYP115-R	CTAAGCGACGGAAGGGCGG AAGCTG	Reverse primer used to amplify MlCYP115 coding sequence from genomic DNA.
MlCYP115-F2	CGCGGATCCTAACGTAACGT AAATGGAGTCACCTCT ATGCCGAGAAGTGGACCCG CGATGG	Forward primer used to add the 5' leader sequence and BamHI restriction site to the long MlCYP115 coding sequence.
MlCYP115-F2'	CGCGGATCCTAACGTAACGT AAATGGAGTCACCTCTATGC GCGTAGAAAACGATCATTGC G	Forward primer used to add the 5' leader sequence and BamHI restriction site to the short MlCYP115 coding sequence.
MlCYP115-R2	CCGGAATTCCTAAGCGACGG AAGGGCGGAAGC	Reverse primer used to add EcoRI to the 3' end of the MlCYP115 coding sequence.

Figure S6. Sequence alignment of CYP115 proteins. The predicted CYP115 proteins of the two functionally characterized genes from *S. meliloti* WSM4191 (SmCYP115) and *M. loti* MAFF303099 (MlCYP115) are shown aligned with analogous proteins from two plant pathogens, *X. oryzae* pv. *oryzicola* BLS256 (XocCYP115) and *X. translucens* pv. *translucens* DSM 18974 (XttCYP115), showing that CYP115 length and sequence is highly similar between these distant lineages. Red highlights with white text indicates strict identity, red characters indicate similarity in a group, and a blue outline represents similarity across groups. The visualization of this alignment was created using ESPript 3.0.[9]

SUPPORTING REFERENCES

(1) Buikema, W. J., Long, S. R., Brown, S. E., van den Bos, R. C., Earl, C., and Ausubel, F. M. (1983) Physical and genetic characterization of *Rhizobium meliloti* symbiotic mutants. *J. Mol. Appl. Genet. 2*, 249–260.

(2) Nett, R. S., Montanares, M., Marcassa, A., Lu, X., Nagel, R., Charles, T. C., Hedden, P., Rojas, M. C., and Peters, R. J. (2017) Elucidation of gibberellin biosynthesis in symbiotic rhizobia reveals convergent evolution. *Nat. Chem. Biol. 13*, 69–74.

(3) Jones, K. M. (2012) Increased production of the exopolysaccharide succinoglycan enhances *Sinorhizobium meliloti* 1021 symbiosis with the host plant *Medicago truncatula*. *J. Bacteriol. 194*, 4322–4331.

(4) Galibert, F., Finan, T. M., Long, S. R., Puhler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M. J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R. W., Dreano, S., Federspiel, N. a, Fisher, R. F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., Hernandez-Lucas, I., Hong, A., Huizar, L., Hyman, R. W., Jones, T., Kahn, D., Kahn, M. L., Kalman, S., Keating, D. H., Kiss, E., Komp, C., Lelaure, V., Masuy, D., Palm, C., Peck, M. C., Pohl, T. M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thebault, P., Vandenbol, M., Vorholter, F. J., Weidner, S., Wells, D. H., Wong, K., Yeh, K. C., and Batut, J. (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science 293*, 668–672.

(5) Backman, K., and Boyer, H. W. (1983) Tetracycline resistance determined by pBR322 is mediated by one polypeptide. *Gene 26*, 197–203.

(6) Finan, T. M., Kunkel, B., De Vos, G. F., and Signer, E. R. (1986) Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. *J. Bacteriol. 167*, 66–72.

(7) Binks, R., MacMillan, J., and Pryce, R. J. (1969) Combined gas chromatography-mass spectrometry of the methyl esters of gibberellins A_1 to A_{24} and their trimethylsilyl ethers. *Phytochemistry 8*, 271–84.

(8) Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol. 33*, msw054.

(9) Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server *42*, 320–324.